SUPPLEMENTARY MATERIAL FOR *QUANTIFYING THE ACCURACY OF APPROXIMATE DIFFUSIONS AND MARKOV CHAINS*

JONATHAN H. HUGGINS AND JAMES ZOU

Appendix A. Exponential contractivity

A natural generalization of the strong concavity case is to assume that $\log \pi$ is strongly concave for x and x' far apart and that $\log \pi$ has "bounded convexity" when x and x' are close together. It turns out that in such cases Assumption 2.A still holds. More formally, the following assumption can be used even when the drift is not a gradient. For $f : \mathcal{X} \to \mathbb{R}^d$ and $r > 0$, let

$$\kappa(r) \triangleq \inf \left\{ -2\frac{(f(x) - f(x')) \cdot (x - x')}{r^2} : x, x' \in \mathcal{X}, \|x - x'\|_2 = r \right\}.$$

Define the constant $R_0 = \inf\{R \geq 0 : \kappa(r) \geq 0 \ \forall r \geq R\}$.

Assumption A.1 (Strongly log-concave tails)**.** *For the function $f \in C^1(\mathcal{X}, \mathbb{R}^d)$, there exist constants $R, \ell \in [0, \infty)$ and $k \in (0, \infty)$ such that*

$$\kappa(r) \geq -\ell \text{ for all } r \leq R \text{ and } \kappa(r) \geq k \text{ for all } r > R.$$

Furthermore, $\kappa(r)$ is continuous and $\int_0^1 r\kappa(r)^- \mathrm{d}r < \infty$.

Theorem A.1 (Eberle [7], Wang [12])**.** *If Assumption A.1 holds for $f = b$ then Assumption 2.A holds for*

$$C = \exp\left(\frac{1}{4}\int_0^{R_0} r\kappa(r)\mathrm{d}s\right)$$

$$\frac{1}{\log(1/\rho)} \leq \begin{cases} \frac{3e}{2}\max(R^2, 8k^{-1}) & \text{if } \ell R_0^2 \leq 8 \\ 8\sqrt{2\pi}R^{-1}\ell^{-1/2}(\ell^{-1} + k^{-1})e^{\ell R^2/8} + 32R^{-2}k^{-2} & \text{otherwise.} \end{cases}$$

For detailed calculations for the case of a mixture of Gaussians model, see Gorham et al. [9].

Appendix B. Proofs of the main results in Section 3

We state all our results in the more general case of a diffusion on a convex space $\mathcal{X} \subseteq \mathbb{R}^d$. We begin with some additional definitions. Any set $\mathcal{G} \subseteq C(\mathcal{X})$ defines an *integral probability metric* (IPM)

$$d_{\mathcal{G}}(\mu, \nu) = \sup_{\phi \in \mathcal{G}} |\mu(\phi) - \nu(\phi)|,$$

where μ and ν are measures on $\mathcal{X}$. The *Wasserstein metric* $d_{\mathcal{W}}$ corresponds to $\mathcal{W} \triangleq \{\phi \in C(\mathcal{X}) \,|\, \|\phi\|_L \leq 1\}$, The set $\mathcal{H} \triangleq \{\phi \in C^1(\mathcal{X}) \,|\, \|h\|_L \leq 1\}$ will be used to define an IPM $d_{\mathcal{H}}$. For a set $\mathcal{Z} \subseteq \mathbb{R}^n$, we use $\partial\mathcal{Z}$ to denote the boundary of $\mathcal{Z}$.

Suppose $\|b - \tilde{b}\|_2 \le \epsilon$. We first state several standard properties of the Wasserstein metric and invariant measures of diffusions. The proofs are included here for completeness.

Lemma B.1. *For any $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $d_{\mathcal{H}}(\mu,\nu) = d_{\mathcal{W}}(\mu,\nu)$.*

Proof sketch. The result follows since any Lipschitz function is continuous and a.e.-differentiable, and continuously differentiable functions are dense in the class of continuous and a.e.-differentiable functions. □

We use the notation $(X_t)_{t\ge0} \sim \mathsf{Diff}(b,\Sigma)$ if X_t is a diffusion defined by

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \Sigma \mathrm{d}W_t - n_t L(\mathrm{d}t).$$

A diffusion X_t is said to be *strong Feller* if its semigroup operator $(\pi_t\phi)(x) \triangleq \mathbb{E}[\phi(X_{x,t})]$, $\phi \in C(\mathcal{X})$, satisfies the property that for all bounded ϕ, $\pi_t\phi$ is bounded and continuous.

Proposition B.2. *Assume Assumption 2.B(1) holds and let $(X_t)_{t\ge0} \sim \mathsf{Diff}(b, I)$. Then for each $x \in \mathcal{X}$, $X_{x,t}$ has the invariant density π and is strong Feller.*

Proof. The existence of the diffusions follows from Tanaka [11, Theorem 4.1], the strong Feller property follows from Ethier and Kurtz [8, Ch. 8, Theorems 1.5 & 1.6], and the fact that π is the unique stationary measure follows since $\mathcal{A}_b^*\pi = 0$. □

By the same proof as Proposition B.2, we have

Proposition B.3 (Diffusion properties)**.** *For $f \in C^0(\mathcal{X}, \mathbb{R}^d)$ with $\|f\|_L < \infty$, the diffusion $(X_t)_{t\ge0} \sim \mathsf{Diff}(f, I)$ exists and has an invariant distribution π_f.*

Proposition B.4 (Expectation of the generator)**.** *For $f \in C^0(\mathcal{X}, \mathbb{R}^d)$, let the diffusion $(X_t)_{t\ge0} \sim \mathsf{Diff}(f, I)$ have invariant density π_f and assume that linear functions are π_f-integrable. Then for all $\phi \in C^2(\mathcal{X})$ such that $\|\phi\|_L < \infty$ and $\mathcal{A}_f\phi$ is π_f-integrable, $\pi_f(\mathcal{A}_f\phi) = 0$.*

Proof. Let P_t be the semigroup operator associated with $(X_t)_{t\ge0}$:

$$(P_t\phi)(x) = \mathbb{E}[\phi(X_{x,t})].$$

Since by hypothesis linear functions are π_f-integrable and ϕ is Lipschitz, ϕ is π_f-integrable. Thus, $P_t\phi$ is π_f-integrable and by the definition of an invariant measure (see [1, Definition 1.2.1] and subsequent discussion),

$$\pi_f(P_t\phi) = \pi_f\phi. \tag{B.1}$$

Using the fact that $\partial_t P_t = P_t\mathcal{A}_f$ [1, Eq. (1.4.1)], differentiating both size of Eq. (B.1), applying dominated convergence, and using the hypothesis that $\mathcal{A}_f\phi$ is π_f-integrable yields

$$0 = \partial_t \pi_f(P_t\phi) = \pi_f(\partial_t P_t\phi) = \pi_f(P_t\mathcal{A}_f\phi) = \pi_f(\mathcal{A}_f\phi).$$

□

We next show that the solution to Eq. (4.1) is Lipschitz continuous with a Lipschitz constant depending on the mixing properties of the diffusion associated with the generator.

Proposition B.5 (Differential equation solution properties)**.** *If Properties 2.A and Assumption 2.B(1) hold, then for any $h \in C^1(\mathcal{X})$ with $\|h\|_L < \infty$, the function*

$$u_h(x) \triangleq \int_0^\infty (\pi(h) - \mathbb{E}[h(X_{x,t})]) \,\mathrm{d}t$$

exists and satisfies

$$\|u_h\|_L \leq \frac{C}{\log(1/\rho)} \|h\|_L \tag{B.2}$$

$$(\mathcal{A}_b u_h)(x) = h(x) - \pi(h). \tag{B.3}$$

Proof. We follow the approach of Mackey and Gorham [10]. By Assumption 2.A and the definition of Wasserstein distance, we have that there is a coupling between $X_{x,t}$ and $X_{x',t}$ such that

$$\mathbb{E}[\|X_{x,t} - X_{x',t}\|_2] \leq C\|x - x'\|_2 \rho^t.$$

The function u_h is well-defined since for any $x \in \mathcal{X}$,

$$\begin{aligned}
\int_0^\infty |\pi(h) - \mathbb{E}[h(X_{x,t})]| \,\mathrm{d}t &= \int_0^\infty \left| \int_{\mathcal{X}} (\mathbb{E}[h(X_{x',t})] - \mathbb{E}[h(X_{x,t})]) \, \pi(x') \,\mathrm{d}x' \right| \mathrm{d}t \\
&\leq \sup_{z \in \mathcal{X}} \|\nabla h(z)\|_2 \int_0^\infty \int_{\mathcal{X}} \mathbb{E}[\|X_{x,t} - X_{x',t}\|_2] \pi(x') \,\mathrm{d}x' \,\mathrm{d}t \\
&= \sup_{z \in \mathcal{X}} \|\nabla h(z)\|_2 \int_0^\infty \int_{\mathcal{X}} \|x - x'\|_2 C \rho^t \pi(x') \,\mathrm{d}x' \,\mathrm{d}t \\
&\leq \|h\|_L \, \mathbb{E}_{X\sim\pi}[\|x - X\|_2] \int_0^\infty C\rho^t \,\mathrm{d}t \\
&< \infty,
\end{aligned}$$

where the first line uses the property that $\pi(h) = \int_{\mathcal{X}} \mathbb{E}[h(X_{x',t})]\pi(x')\mathrm{d}x'$ and the final inequality follows from Assumption 2.B(2) and the assumption that $0 < \rho < 1$. Furthermore, u_h has bounded Lipschitz constant since for any $x, x' \in \mathcal{X}$,

$$\begin{aligned}
|u_h(x) - u_h(x')| &= \left| \int_0^\infty \mathbb{E}[h(X_{x,t}) - h(X_{x',t})] \,\mathrm{d}t \right| \\
&\leq \sup_{z \in \mathcal{X}} \|\nabla h(z)\|_2 \int_0^\infty \mathbb{E}[\|X_{x,t} - X_{x',t}\|_2] \,\mathrm{d}t \\
&\leq \|h\|_L \|x - x'\|_2 \int_0^\infty C\rho^t \,\mathrm{d}t \\
&= \frac{C\|h\|_L}{\log(1/\rho)} \|x - x'\|_2.
\end{aligned}$$

Finally, we show that $(\mathcal{A}_b u_h)(x) = h(x) - \pi(h)$. Recall that for $h \in C(\mathcal{X})$, the semigroup operator is given by $(\pi_t h)(x) = \mathbb{E}[h(X_{x,t})]$. Since $X_{x,t}$ is strong Feller for all $x \in \mathcal{X}$ by Proposition B.2, for all $t \geq 0$, its generator satisfies [8, Ch. 1, Proposition 1.5]

$$h - \pi_t h = \mathcal{A}_b \int_0^t (\pi(h) - \pi_s h) \,\mathrm{d}s. \tag{B.4}$$

Hence,

$$\begin{aligned}
&|h(x) - \pi(h) - [h(x) - (\pi_t h)(x)]| \\
&= \left| \int_{\mathcal{X}} \mathbb{E}[h(X_{x,t})] - \mathbb{E}[h(X_{x',t})]\pi(x')\,\mathrm{d}x' \right| \\
&\le \sup_{z\in\mathcal{X}} \|\nabla h(z)\|_2 \int_{\mathcal{X}} \mathbb{E}[\|X_{x',t} - X_{x,t}\|_2]\pi(x')\,\mathrm{d}x' \\
&\le \|h\|_L\, \mathbb{E}_{X\sim\pi}[\|x - X\|_2] C\rho^t.
\end{aligned}$$

Thus, conclude that the left-hand side of Eq. (B.4) converges pointwise to $h(x) - \pi(h)$ as $t \to \infty$. Since $\mathcal{A}_b$ is closed [8, Ch. 1, Proposition 1.6], the right-hand side of Eq. (B.4) limits to $\mathcal{A}_b u_h$. Hence, u_h solves Eq. (B.3). □

We can now prove the main result bounding the Wasserstein distance between the invariant distributions of the original and perturbed diffusions.

Proof of Theorem 3.1. By Proposition B.3 and Assumption 2.B, the hypotheses of Proposition B.4 hold for $f = \tilde{b}$. Let $\mathcal{F} \triangleq \{u_h \mid h \in \mathcal{H}\}$. Then

$$\begin{aligned}
d_{\mathcal{W}}(\pi, \tilde{\pi}) &= \sup_{h\in\mathcal{H}} |\pi(h) - \tilde{\pi}(h)| \quad \text{by definition and Assumption 2.B} \\
&= \sup_{h\in\mathcal{H}} |\pi(\mathcal{A}_b u_h) - \tilde{\pi}(\mathcal{A}_b u_h)| \quad \text{by Eq. (B.3)} \\
&= \sup_{h\in\mathcal{H}} |\tilde{\pi}(\mathcal{A}_b u_h)| \quad \text{by Proposition B.4} \\
&= \sup_{u\in\mathcal{F}} |\tilde{\pi}(\mathcal{A}_b u)| \quad \text{by definition of } \mathcal{F} \\
&= \sup_{u\in\mathcal{F}} |\tilde{\pi}(\mathcal{A}_b u - \mathcal{A}_{\tilde{b}} u)| \quad \text{by Proposition B.4} \\
&= \sup_{u\in\mathcal{F}} |\tilde{\pi}(\nabla u \cdot b - \nabla u \cdot \tilde{b})| \qquad \text{by definition of } \mathcal{A}_b \\
&\le \sup_{u\in\mathcal{F}} |\tilde{\pi}(\|\nabla u\|_2 \|b - \tilde{b})\|_2)| \\
&\le \frac{C\epsilon}{\log(1/\rho)} \qquad \text{by Eq. (B.2) and } \|b - \tilde{b}\|_2 \le \epsilon.
\end{aligned}$$

□

A similar analysis can be used to bound the Wasserstein distance between π and $\tilde{\pi}$ when the approximate drift $\tilde{b}$ is itself stochastic.

Proof of Theorem 3.4. We will need to consider the joint diffusions $Z_t = (X_t, Y_t)$ and $\tilde{Z}_t = (\tilde{X}_t, \tilde{Y}_t)$ on $\mathcal{Z} \triangleq \mathcal{X} \times \mathbb{R}^d$, where

$$\begin{aligned}
\mathrm{d}Z_t &= (b(X_t), b_{aux}(Y_t))\,\mathrm{d}t + (\sqrt{2}\,\mathrm{d}W_t^X, \Sigma\,\mathrm{d}W_t^Y) - n_t L(\mathrm{d}t) \\
\mathrm{d}\tilde{Z}_t &= (\tilde{b}(\tilde{X}_t, \tilde{Y}_t), b_{aux}(\tilde{Y}_t))\,\mathrm{d}t + (\sqrt{2}\,\mathrm{d}\tilde{W}_t^X, \Sigma\,\mathrm{d}\tilde{W}_t^Y) - n_t \tilde{L}(\mathrm{d}t).
\end{aligned}$$

Notice that X_t and Y_t are independent and the invariant distribution of X_t is π. Let π_Z and $\tilde{\pi}_Z$ be the invariant distributions of Z_t and $\tilde{Z}_t$, respectively. Also note that the generators for Z_t and $\tilde{Z}_t$ are, respectively,

$$\begin{aligned}
\mathcal{A}_Z \phi(z) &= \nabla\phi \cdot (b(x), b_{aux}(y)) + \Delta\phi_x(z) + \Sigma^\top\Sigma : H\phi_y(z) \\
\mathcal{A}_{\tilde{Z}} \phi(z) &= \nabla\phi \cdot (\tilde{b}(x, y), b_{aux}(y)) + \Delta\phi_x(z) + \Sigma^\top\Sigma : H\phi_y(z).
\end{aligned}$$

where H is the Hessian operator.

By Proposition B.3 and 2.B, the hypotheses of Proposition B.4 hold for $f(x, y) = (\tilde{b}(x, y), b_{aux}(y))$. Let $\mathcal{H}_Z \triangleq \{h \in C^1(\mathcal{Z}) \,|\, \|h\|_L \le 1\}$ and $\mathcal{F}_Z \triangleq \{u_h \,|\, h \in \mathcal{H}_Z\}$. Also, for $z = (x, y) \in \mathcal{Z}$, let $\mathrm{id}_Y(z) = y$. Then, by reasoning analogous to that in the proof of Theorem 3.1,

$$\begin{aligned}
d_{\mathcal{W}}(\pi, \tilde{\pi}) &\le d_{\mathcal{W}}(\pi_Z, \tilde{\pi}_Z) \\
&= \sup_{h \in \mathcal{H}_Z} |\pi_Z(h) - \tilde{\pi}_Z(h)| \\
&= \sup_{u \in \mathcal{F}_Z} |\tilde{\pi}_Z(\mathcal{A}_Z u - \mathcal{A}_{\tilde{Z}} u)| \\
&= \sup_{u \in \mathcal{F}_Z} |\tilde{\pi}_Z(\nabla u \cdot b - \nabla u \cdot \tilde{b})| \\
&= \sup_{u \in \mathcal{F}_Z} |\mathbb{E}[\nabla u(\tilde{X}, \tilde{Y}) \cdot \mathbb{E}[b(\tilde{X}) - \tilde{b}(\tilde{X}, \tilde{Y}) \,|\, \tilde{X}]]| \\
&\le \sup_{u \in \mathcal{F}_Z} |\mathbb{E}[\|\nabla u(\tilde{X}, \tilde{Y})\|_2 \|\mathbb{E}[b(\tilde{X}) - \tilde{b}(\tilde{X}, \tilde{Y}) \,|\, \tilde{X}]\|_2]| \\
&\le \frac{C\, \tilde{\pi}(\epsilon)}{\log(1/\rho)}.
\end{aligned}$$

□

Proof of Theorem 3.5. The proof is very similar to that of Theorem 3.1, the only difference is in the Lipshitz coefficient of the differential equation solution $u_h(x)$ in B.5. Using polynomial contractivity, we have

$$\begin{aligned}
|u_h(x) - u_h(x')| &= \left| \int_0^\infty \mathbb{E}[h(X_{x,t}) - h(X_{x',t})] \,\mathrm{d}t \right| \\
&\le \sup_{z \in \mathcal{X}} \|\nabla h(z)\|_2 \int_0^\infty \mathbb{E}[\|X_{x,t} - X_{x',t}\|_2] \,\mathrm{d}t \\
&\le \|h\|_L \|x - x'\|_2 \int_0^\infty C(t + \beta)^{-\alpha} \,\mathrm{d}t \\
&= \frac{C\|h\|_L}{(\alpha - 1)\beta^{\alpha - 1}} \|x - x'\|_2.
\end{aligned}$$

Plugging in this Lipschitz constant, we have

$$d_{\mathcal{W}}(\pi, \tilde{\pi}) \le \frac{C\epsilon}{(\alpha - 1)\beta^{\alpha - 1}}.$$

□

Appendix C. Checking the Integrability Condition

The following result gives checkable conditions under which Assumption 2.B(3) holds. Let $\mathbb{B}_R \triangleq \{x \in \mathbb{R}^d \,|\, \|x\|_2 \le R\}$.

Proposition C.1 (Ensuring integrability)**.** *Assumption 2.B(3) is satisfied if* $b = \nabla \log \pi$*,* $\tilde{b} = \nabla \log \tilde{\pi}$*,* $\|b - \tilde{b}\|_2 \le \epsilon$*, and either*

(1) *there exist constants* $R > 0, B > 0, \delta > 0$ *such that for all* $x \in \mathcal{X} \setminus \mathbb{B}_R$*,* $\|b(x) - \tilde{b}(x)\|_2 \le B/\|x\|_2^{1+\delta}$*; or*
(2) *there exists a constant* $R > 0$ *such that for all* $x \in \mathcal{X} \setminus \mathbb{B}_R$ $x \cdot (b(x) - \tilde{b}(x)) \ge 0$.

Proof. For case (1), first we note that since $\int_{\mathcal{X}}(\pi(x)-\tilde{\pi}(x))\,\mathrm{d}x = 0$, by the (generalized) intermediate value theorem, there exists $x^* \in \mathcal{X}$ such that $\pi(x^*) - \tilde{\pi}(x^*) = 0$, and hence $\log \pi(x^*) - \log \tilde{\pi}(x^*) = 0$. Let $p[x^*, x]$ be any path from x^* to x. By the fundamental theorem of calculus for line integrals,

$$
\begin{aligned}
|\log \pi(x) - \log \tilde{\pi}(x)| &= \left|\log \tilde{\pi}(x^*) - \log \pi(x^*) + \int_{\gamma[x^*,x]} (b(r) - \tilde{b}(r)) \cdot \mathrm{d}r\right| \\
&= \left|\int_{\gamma[x^*,x]} (b(r) - \tilde{b}(r)) \cdot r'(t)\,\mathrm{d}t\right| \\
&\le \int_{\gamma[x^*,x]} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t.
\end{aligned}
$$

First consider $x \in \mathcal{X} \cap \mathbb{B}_R$. Choosing $p[x^*, x]$ to be the linear path $\gamma[x^*, x]$, we have

$$
\begin{aligned}
|\log \pi(x) - \log \tilde{\pi}(x)| &\le \epsilon \int_{\gamma[x^*,x]} \|r'(t)\|_2\,\mathrm{d}t \\
&= \epsilon \|x - x^*\|_2 \\
&\le (R + \ell^*)\epsilon, \qquad \text{(C.1)}
\end{aligned}
$$

where $\ell^* \triangleq \|x^*\|_2$.

Next consider $x \in \mathcal{X} \setminus \mathbb{B}_R$. Let $\ell \triangleq \|x\|_2$ and $x' = \frac{R}{\ell}x$. Choose $p[x^*, x]$ to consist of the concatenation of the linear paths $\gamma[x^*, 0]$, $\gamma[0, x']$, and $\gamma[x', 0]$, so

$$
\begin{aligned}
&\int_{p[x^*,x]} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t \\
&= \int_{\gamma[x^*,0]} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t + \int_{\gamma[0,x']} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t \\
&\quad + \int_{\gamma[x',x]} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t.
\end{aligned}
$$

Now, we bound each term:

$$
\begin{aligned}
\int_{\gamma[x^*,0]} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t &\le \ell^*\epsilon \\
\int_{\gamma[0,x']} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t &\le R\epsilon \\
\int_{\gamma[x',x]} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2\,\mathrm{d}t &\le (\ell - R)B \int_0^1 \frac{1}{(R + (\ell - R)t)^{1+\delta}} \\
&= (\ell - R)B \left[\frac{1}{(\ell - R)R^\delta} - \frac{1}{(\ell - R)\ell^\delta}\right] \\
&\le \frac{B}{R^\delta}.
\end{aligned}
$$

It follows that there exists a constant $\tilde{B} > 0$ such that for all $x \in \mathcal{X}$, $|\log \pi(x) - \log \tilde{\pi}(x)| < \tilde{B}$. Hence $\tilde{B}^{-1}\pi < \tilde{\pi} < \tilde{B}\pi$, so ϕ is π-integrable if and only if it is $\tilde{\pi}$-integrable.

Case (2) requires a slightly more delicate argument. Let x^* and ℓ^* be the same as in case (1). For $x \in \mathcal{X} \cap \mathbb{B}_R$, it follows from Eq. (C.1) that

$$
\log \pi(x) - \log \tilde{\pi}(x) \ge -(R + \ell^*)\epsilon.
$$

For $x \in \mathcal{X} \setminus \mathbb{B}_R$, arguing as above yields

$$\begin{aligned}\log \pi(x) - \log \tilde{\pi}(x) &= \int_{p[x^*,x]} (b(r) - \tilde{b}(r)) \cdot \mathrm{d}r \\ &\geq -\int_{p[x^*,r']} \|b(r) - \tilde{b}(r)\|_2 \|r'(t)\|_2 \,\mathrm{d}t \\ &\quad + \int_{\gamma[x',x]} (b(r) - \tilde{b}(r)) \cdot r'(t)\,\mathrm{d}t \\ &\geq -(R+\ell^*)\epsilon + \int_{\gamma[x',x]} (b(q(t)x) - \tilde{b}(q(t)x)) \cdot ax \,\mathrm{d}t \\ &\geq -(R+\ell^*)\epsilon,\end{aligned}$$

where we have used the fact that $r(t) = q(t)x$ for some linear function $q(t)$ with slope $a > 0$. Combining the previous two displays, conclude that for all $x \in \mathcal{X}$, $\tilde{\pi}(x) \leq e^{(R+\ell^*)\epsilon}\pi(x)$, hence Assumption 2.B(3) holds. □

We suspect Proposition C.1 continues to hold even when $b \neq \nabla \log \pi$ and $\tilde{b} \neq \nabla \log \tilde{\pi}$. Note that condition (1) always holds if $\mathcal{X}$ is compact, but also holds for unbounded $\mathcal{X}$ as long as the error in the gradients decays sufficiently quickly as $\|x\|_2$ grows large. Given an approximate distribution for which $\|b - \tilde{b}\|_2 \leq \epsilon/2$, it is easy to construct a new distribution that satisfies condition (2):

Proposition C.2. *Assume that $\tilde{\pi}$ satisfies $\|b - \tilde{b}\|_2 \leq \epsilon/2$ and let*

$$f_R(x) \triangleq -\frac{\epsilon x}{2\|x\|_2}\left\{(2\|x\|_2/R - 1)\mathbb{1}[R/2 \leq \|x\|_2 < R] + \mathbb{1}[\|x\|_2 \geq R]\right\}.$$

Then the distribution

$$\tilde{\pi}_R(x) \propto \tilde{\pi}(x)e^{f_R(x)}$$

satisfies condition (2) of Proposition C.1.

Proof. Let $\tilde{b}_R \triangleq \nabla \log \tilde{\pi}_R$. First we verify that $\|b - \tilde{b}_R\|_2 \leq \epsilon$. For $x \in \mathcal{X} \cap \mathbb{B}_{R/2}$, $\tilde{\pi}_R(x) = \tilde{\pi}(x)$, so $\|b(x) - \tilde{b}_R(x)\|_2 \leq \epsilon/2$. Otherwise $x \in \mathcal{X} \setminus \mathbb{B}_{R/2}$, in which case since $\|f_R(x)\| \leq \epsilon/2$ it follows that $\|b(x) - \tilde{b}_R(x)\|_2 \leq \epsilon$. To verify condition (2), calculate that for $x \in \mathcal{X} \setminus \mathbb{B}_R$,

$$x \cdot (b(x) - \tilde{b}_R(x)) = x \cdot \left(b(x) - \tilde{b}(x) - \frac{\epsilon x}{2\|x\|_2}\right) \geq \frac{\epsilon\|x\|_2}{2} - \frac{x \cdot \epsilon x}{2\|x\|_2} = 0.$$

□

By taking R very large in Proposition C.2, we can ensure the integrability condition holds without having any practical effect on the approximating drift since $\tilde{b}_R(x) = \tilde{b}(x)$ for all $x \in \mathbb{B}_{R/2}$. Thus, it is safe to view Assumption 2.B(3) as a mild regularity condition.

Appendix D. Approximation Results for Piecewise Deterministic Markov Processes

In the section we obtain results for a broader class of PDMPs which includes the ZZP a special case [2]. The class of PDMPs we consider are defined on the space $E \triangleq \mathbb{R}^d \times \mathcal{B}$, where $\mathcal{B}$ is a finite set. Let $A \in C^0(E, \mathbb{R}_+^{\mathcal{B}})$ and let $F \in C^0(E, \mathbb{R}^d)$ be such that for each $\theta \in \mathcal{B}$, $F(\cdot, \theta)$ is a smooth vector field for which the differential equation $\partial_t x_t = F(x_t, \theta)$ with initial condition $x_0 = x$ has a unique global solution. For $\phi \in C(E)$, the standard differential operator $\nabla_x \phi(x, \theta) \in \mathbb{R}^d$ is given by $(\nabla_x \phi(x, \theta))_i \triangleq \frac{\partial \phi}{\partial x_i}(x, \theta)$ for $i \in [d]$ and the discrete differential operator $\nabla_\theta \phi(x, \theta) \in \mathbb{R}^{\mathcal{B}}$ is given by $(\nabla_\theta \phi(x, \theta))_{\theta'} \triangleq \phi(x, \theta') - \phi(x, \theta)$. The PDMP $(X_t, \Theta_t)_{t \geq 0}$ determined by the pair (F, A) has infinitesimal generator

$$\mathcal{A}_{F,A}\phi = F \cdot \nabla_x \phi + A \cdot \nabla_\theta \phi.$$

We consider the cases when either or both of A and F are approximated (in the case of ZZP, only A is approximated while F is exact). The details of the polynomial contractivity condition depend on which parts of (F, A) are approximated. We use the same notation for the true and approximating PDMPs with, respectively, infinitesimal generators $\mathcal{A}_{F,A}$ and $\mathcal{A}_{\tilde{F},\tilde{A}}$, as we did for the ZZPs in Section 6.

Assumption D.2 (**PDMP error and polynomial contractivity**).

(1) *There exist $\epsilon_F, \epsilon_A \geq 0$ such that $\|F - \tilde{F}\|_2 \leq \epsilon_F$ and $\|A - \tilde{A}\|_1 \leq \epsilon_A$.*
(2) *For each $(x, \theta) \in E$, let $\mu_{x,\theta,t}$ denote the law of the PDMP $(X_{x,\theta,t}, \Theta_{x,\theta,t})$ with generator $\mathcal{A}_{F,A}$. There exist constants $\alpha > 1$ and $\beta > 0$ and a function $B \in C(E \times E, \mathbb{R}_+)$ such that for all $x, x' \in \mathbb{R}^d$ and $\theta, \theta' \in \mathcal{B}$,*

$$d_{\mathcal{W}}(\mu_{x,\theta,t}, \mu_{x',\theta',t}) \leq B(x, \theta, x', \theta')(t + \beta)^{-\alpha}.$$

Furthermore, if $\epsilon_F > 0$, then there exists $C_F > 0$ such that $B(x, \theta, x', \theta) \leq C_F\|x - x'\|_2$ and if $\epsilon_A > 0$, then there exists $C_A > 0$ such that $B(x, \theta, x, \theta') \leq C_A$. If $\epsilon_F = 0$ take $C_F = 0$ and if $\epsilon_A = 0$ take $C_A = 0$.

We also require some regularity conditions similar to those for diffusions.

Assumption D.3 (**PDMP regularity conditions**). *Let π and $\tilde{\pi}$ denote the stationary distributions of the PDMPs with, respectively, infinitesimal generators $\mathcal{A}_{F,A}$ and $\mathcal{A}_{\tilde{F},\tilde{A}}$.*

(1) *The stationary distributions π and $\tilde{\pi}$ exist.*
(2) *The target density satisfies $\int_E x^2 \pi(\mathrm{d}x, \mathrm{d}\theta) < \infty$.*
(3) *If a function $\phi \in C(E, \mathbb{R})$ is π-integrable then it is $\tilde{\pi}$-integrable.*

Theorem D.1 (PDMP error bounds). *If Assumptions D.2 and D.3 hold, then*

$$d_{\mathcal{W}}(\pi, \tilde{\pi}) \leq \frac{C_F \epsilon_F + C_A \epsilon_A}{(\alpha - 1)\beta^{\alpha - 1}}.$$

Proof sketch. For $h \in C_L(\mathbb{R}^d)$, we need to solve

$$h - \pi(h) = \mathcal{A}_{F,A} u.$$

Similarly to before, the solution is

$$u_h(x, \theta) \triangleq \int_0^\infty (\pi(h) - \mathbb{E}[h(X_{x,\theta,t})])\, \mathrm{d}t,$$

which can be verified as in the diffusion case using Assumptions D.2(2) and D.3. Furthermore, for $x, x' \in \mathbb{R}^d$ and $\theta, \theta' \in \mathcal{B}$, by Assumption D.2(2),

$$\begin{aligned} |u_h(x, \theta) - u_h(x', \theta)| &\leq \|h\|_L \int_0^\infty C_F \|x - x'\|_2 (t + \beta)^{-\alpha}\, \mathrm{d}t \\ &= \frac{C\|h\|_L}{(\alpha - 1)\beta^{\alpha - 1}} \|x - x'\|_2 \end{aligned}$$

and

$$|u_h(x,\theta) - u_h(x,\theta')| \le \|h\|_L \int_0^\infty C_A(t+\beta)^{-\alpha}\,\mathrm{d}t = \frac{C_A\|h\|_L}{(\alpha-1)\beta^{\alpha-1}}.$$

We bound $d_{\mathcal{W}}(\pi,\tilde{\pi})$ as in Theorem 3.4, but now using the fact that for $u = u_h$, $h \in C_L(\mathbb{R}^d)$, we have

$$\begin{aligned}\mathcal{A}_{F,A}u_h - \mathcal{A}_{\tilde{F},\tilde{A}}u_h &= (F-\tilde{F})\cdot\nabla_x u_h + (A-\tilde{A})\cdot\nabla_\theta u_h \\ &\le \|F-\tilde{F}\|_2\|\nabla_x u_h\|_2 + \|A-\tilde{A}\|_1\|\nabla_\theta u_h\|_\infty \\ &\le \frac{C_F\epsilon_F + C_A\epsilon_A}{(\alpha-1)\beta^{\alpha-1}}.\end{aligned}$$

□

D.1. **Hamiltonian Monte Carlo.** We can write an idealized form of Hamiltonian Monte Carlo (HMC) as a PDMP $(X_t,P_t)_{t\ge 0}$ by having the momentum vector $P_t \in \mathbb{R}^d$ refresh at a constant rate λ. Let R_t be a compound Poisson process with rate $\lambda > 0$ and jump size distribution $\mathcal{N}(0,M)$, where $M \in \mathbb{R}^{d\times d}$ is a positive-definite mass matrix. That is, if Γ_t is a homogenous Poisson (counting) process with rate λ and $J_i \sim \mathcal{N}(0,M)$, then

$$R_t \sim \sum_{i=1}^{\Gamma_t} J_i.$$

We can then write the HMC dynamics as

$$\begin{aligned}\mathrm{d}X_t &= M^{-1}P_t\,\mathrm{d}t \\ \mathrm{d}P_t &= \nabla\log\pi(X_T)\,\mathrm{d}t + \mathrm{d}R_t.\end{aligned}$$

The infinitesimal generator for $(X_t,P_t)_{t\ge 0}$ is

$$\begin{aligned}&\mathcal{A}_{\lambda,M,\pi}\phi(x,p) \\ &= M^{-1}p\cdot\nabla_x\phi(x,p) + \nabla\log\pi(x)\cdot\nabla_p\phi(x,p) + \lambda\left(\int\phi(x,p')\nu_M(\mathrm{d}p') - \phi(x,p)\right),\end{aligned}$$

where ν_M is the density of $\mathcal{N}(0,M)$. Let $\mu_{x,p,t}$ denote the law of $(X_{x,p,t},P_{x,p,t})$ with generator $\mathcal{A}_{\lambda,M,\pi}$. The proof of the following theorem is similar to that for Theorem D.1:

Theorem D.2 (HMC error bounds)**.** *Assume that:*

1. $\|\nabla\log\pi - \nabla\log\tilde{\pi}\|_2 \le \epsilon$.
2. *there exist constants* $C > 0$ *and* $0 < \rho < 1$ *such that*
$$d_{\mathcal{W}}(\mu_{x,p,t},\mu_{x,p',t}) \le C\|p-p'\|_2\rho^t.$$
3. *The stationary distributions of the PDMPs with, respectively, infinitesimal generators* $\mathcal{A}_{\lambda,M,\pi}$ *and* $\mathcal{A}_{\lambda,M,\tilde{\pi}}$*, exist (they are, respectively,* $\pi\times\mu_M$ *and* $\tilde{\pi}\times\mu_M$*).*
4. *The target density satisfies* $\int_E x^2\pi(\mathrm{d}x) < \infty$.
5. *If a function* $\phi \in C(\mathbb{R}^d,\mathbb{R})$ *is* π*-integrable then it is* $\tilde{\pi}$*-integrable.*

Then

$$d_{\mathcal{W}}(\pi,\tilde{\pi}) \le \frac{C\epsilon}{\log(1/\rho)}.$$

Appendix E. Analysis of computational–statistical trade-off

In this section we prove Theorem 5.1. In order to apply results on the approximation accuracy of ULA [3–5], we need the following property to hold for the exact and approximate drift functions.

Assumption E.4 (Strong log-concavity). *There exists a positive constant $k_f > 0$ such that for all $x, x' \in \mathcal{X}$,*

$$(f(x) - f(x')) \cdot (x - x') \le -k_f \|x - x'\|_2^2.$$

We restate the convexity smoothness requirements given in Assumption 5.D with some additional notations.

Assumption E.5.

(1) *The function $\log \pi_0 \in C^3(\mathbb{R}^d, \mathbb{R})$ is k_0-strongly concave, $L_0 \triangleq \|\nabla \log \pi_0\|_L < \infty$, and $\|H[\partial_j \log \pi_0]\|_* \le M_0 < \infty$ for $j = 1, \dots, d$.*
(2) *There exist constants k_ϕ, L_ϕ, and M_ϕ such that for $i = 1, \dots, N$, the function $\phi_i \in C^3(\mathbb{R}, \mathbb{R})$ is k_ϕ-strongly concave, $\|\phi_i'\|_L \le L_\phi < \infty$, and $\|\phi_i'''\|_\infty \le M_\phi < \infty$.*

Note that under Assumption E.4, there is a unique $x^\star \in \mathbb{R}^d$ such that $f(x^\star) = 0$. Our results in this section on based on the following bound on the Wasserstein distance between the law of ULA Markov chain and π_f:

Theorem E.1 ([5, Theorem 3], [6, Corollary 3]). *Assume that E.4 holds and the $L_f \triangleq \|f\|_L < \infty$. Let $\kappa_f \triangleq 2k_f L_f/(k_f + L_f)$ and let $\mu_{x,T}$ denote the law of $X'_{x,T}$. Take $\gamma_i = \gamma_1 i^{-\alpha}$ with $\alpha \in (0,1)$ and set*

$$\gamma_1 = 2(1-\alpha)\kappa_f^{-1}(2/T)^{1-\alpha} \log\left(\frac{\kappa_f T}{2(1-\alpha)}\right).$$

If $\gamma_1 < 1/(k_f + L_f)$, then

$$d_{\mathcal{W}}^2(\mu_{x,T}, \pi_f) \le 16(1-\alpha)L_f^2 \kappa_f^{-3} d T^{-1} \log\left(\frac{\kappa_f T}{2(1-\alpha)}\right).$$

For simplicity we fix $\alpha = 1/2$, though the same results hold for all $\alpha \in (0,1)$, just with different constants. Take $\{\gamma_i\}_{i=1}^\infty$ as defined in Theorem E.1. Let $x^\star = \arg\max_x \mathcal{L}(x)$, let $S_k \triangleq \sum_{i=1}^N \|y_i\|_2^k$, and let $A \triangleq \sum_{i=1}^N y_i y_i^\top$. The drift for this model is

$$b(x) \triangleq \nabla \mathcal{L}(x) = \nabla \log \pi_0(x) + \sum_{i=1}^N \phi_i'(x \cdot y_i) y_i.$$

By Taylor's theorem, the j-th component of $b(x)$ can be rewritten as

$$\begin{aligned} b_j(x) &= \partial_j \log \pi_0(x^\star) + \nabla \partial_j \log \pi_0(x^\star) \cdot (x - x^\star) + R(\partial_j \log \pi_0, x) \\ &\quad + \sum_{i=1}^N \phi_i'(x^\star \cdot y_i) y_{ij} + \phi_i''(x^\star \cdot y_i) y_{ij} y_i \cdot (x - x^\star) + R(\phi_i'(\cdot \cdot y_i) y_{ij}, x) \\ &= \nabla \partial_j \log \pi_0(x^\star) \cdot (x - x^\star) + R(\partial_j \log \pi_0, x) \\ &\quad + \sum_{i=1}^N \phi_i''(x^\star \cdot y_i) y_{ij} y_i \cdot (x - x^\star) + R(\phi_i'(\cdot \cdot y_i) y_{ij}, x), \end{aligned} \tag{E.1}$$

where

$$R(f,x) \triangleq \|x-x^\star\|_2^2 \int_0^1 (1-t) Hf(x^\star + t(x-x^\star))\,\mathrm{d}t.$$

Hence we can approximate the drift with a first-order Taylor expansion around $x^\star$:

$$\tilde{b}(x) \triangleq (H\log\pi_0)(x^\star)(x-x^\star) + \sum_{i=1}^N \phi_i''(x^\star\cdot y_i)y_i y_i^\top (x-x^\star).$$

Observe that Assumption E.4 is satisfied for $f = b$ and $f = \tilde{b}$ with $k_f = k_N \triangleq k_0 + k_\phi \|A\|_*$. Furthermore, Assumption 2.B is satisfied with $\|\tilde{b}\|_L \le L_N \triangleq L_0 + L_\phi S_2$ and $\|b\|_L \le L_N$ as well since

$$\begin{aligned}\|\phi_i'(x_1\cdot y_i)y_i - \phi_i'(x_2\cdot y_i)y_i\|_2 &\le |\phi_i'(x_1\cdot y_i) - \phi_i'(x_2\cdot y_i)|\|y_i\|_2 \\ &\le L_\phi |x_1\cdot y_i - x_2\cdot y_i|\|y_i\|_2 \\ &\le L_\phi \|y_i\|_2^2 \|x_1 - x_2\|_2.\end{aligned}$$

Thus, b and $\tilde{b}$ satisfy the same regularity conditions.

We next show that they cannot deviate too much from each other. Using Eq. (E.1) and regularity assumptions we have

$$\begin{aligned}\|b(x)-\tilde{b}(x)\|_2^2 &= \sum_{j=1}^d \left(R(\partial_j \log\pi_0, x) + \sum_{i=1}^N R(\phi_i'(\cdot\cdot y_i)y_{ij}, x)\right)^2 \\ &\le \|x-x^\star\|_2^4 \sum_{j=1}^d \left(M_0 + \sum_{i=1}^N M_\phi \|y_i\|_2^2 y_{ij}\right)^2 \\ &\le d\|x-x^\star\|_2^4 \left(M_0 + M_\phi \sum_{i=1}^N \|y_i\|_2^3\right)^2.\end{aligned}$$

It follows from [5, Theorem 1(ii)] that

$$\tilde{\pi}(\|b-\tilde{b}\|_2) \le d^{3/2} M_N k_N^{-1},$$

where $M_N \triangleq M_0 + M_\phi S_3$.

Putting these results together with Theorems 3.1 and E.1 and applying the triangle inequality, we conclude that

$$\begin{aligned} d_{\mathcal{W}}^2(\mu_T^\star, \pi) &\le \frac{(k_N+L_N)^3 d}{k_N^3 L_N} \frac{\log\left(\frac{2k_N L_N}{k_N+L_N} T\right)}{T} \\ d_{\mathcal{W}}^2(\tilde{\mu}_{\tilde{T}}^\star, \pi) &\le \frac{2(k_N+L_N)^3 d}{k_N^3 L_N} \frac{\log\left(\frac{2k_N L_N}{k_N+L_N} \tilde{T}\right)}{\tilde{T}} + \frac{2d^3 M_N^2}{k_N^4}.\end{aligned}$$

In order to compare the bounds we must make the computational budgets of the two algorithms equal. Recall that we measure computational cost by the number of d-dimensional inner products performed, so ULA with b costs TN and ULA with $\tilde{b}$ costs $(\tilde{T}+N)d$. Equating the two yields $\tilde{T} = N(T/d-1)$, so we must assume that $T > d$. For the purposes of asymptotic analysis, assume also that S_i/N is bounded from above and bounded away from zero. Under these assumptions, in the case of $k_\phi > 0$, we conclude that

$$d^2_{\mathcal{W}}(\mu^\star_T, \pi) = \tilde{O}\left(\frac{d}{TN}\right) \qquad \text{and} \qquad d^2_{\mathcal{W}}(\tilde{\mu}^\star_T, \pi) = \tilde{O}\left(\frac{d^2}{N^2 T} + \frac{d^3}{N^2}\right),$$

establishing the result of Theorem 5.1. For large N, the approximate ULA with $\tilde{b}$ is more accurate.

References

[1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2014.

[2] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *Electronic Communications in Probability*, 17(0), 2012.

[3] S. Bubeck, R. Eldan, and J. Lehec. Finite-Time Analysis of Projected Langevin Monte Carlo. In *Advances in Neural Information Processing Systems*, July 2015.

[4] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017.

[5] A. Durmus and E. Moulines. Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm. *HAL*, pages 1–25, Apr. 2016.

[6] A. Durmus and E. Moulines. Supplement to "Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm". *HAL*, pages 1–17, Apr. 2016.

[7] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, pages 1–36, Oct. 2015.

[8] S. N. Ethier and T. G. Kurtz. Markov processes: characterization and convergence, volume 282. John Wiley & Sons, 2009.

[9] J. Gorham, A. B. Duncan, S. J. Vollmer, and L. Mackey. Measuring Sample Quality with Diffusions. *arXiv.org*, Nov. 2016.

[10] L. Mackey and J. Gorham. Multivariate Stein Factors for Strongly Log-concave Distributions. *arXiv.org*, 2015.

[11] H. Tanaka. Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.*, 9:163–177, 1979.

[12] J. Wang. L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli*, 22(3):1598–1616, Aug. 2016.

Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology
URL: http://www.jhhuggins.org/
E-mail address: jhuggins@mit.edu

Stanford University
URL: http://sites.google.com/site/jamesyzou/
E-mail address: jamesyzou@gmail.com