Fractional Langevin Monte Carlo: Exploring Lévy Driven Stochastic Differential Equations for MCMC

SUPPLEMENTARY DOCUMENT

Umut Şimşekli [1]

1. Numerically Stable Computation

In this section, we focus on the computation of the following quantity:

$$\frac{\mathcal{D}^\gamma\{-\phi(x)\partial_x U(x)\}}{\phi(x)} \approx \frac{1}{h^\gamma}\sum_{k=-K}^{K} g_{\gamma,k}\frac{-\phi(x-kh)\partial_x U(x-kh)}{\phi(x)}. \tag{S1}$$

Since $\phi(x) = \exp(-U(x))$, for very large values of $U(x)$ and $U(x-kh)$ we might easily end up with $0/0$ errors if we directly implement (8).

We now present a numerically more stable algorithm for computing (8) . We rewrite the above equation as follows:

$$\frac{1}{h^\gamma}\sum_{k=-K}^{K} g_{\gamma,k}\frac{-\phi(x-kh)\partial_x U(x-kh)}{\phi(x)} = \frac{1}{h^\gamma}\sum_{k=-K}^{K} g_{\gamma,k}\Big[-\partial_x U(x-kh)\exp\Big(\underbrace{U(x)-U(x-kh)}_{\ell_k}\Big)\Big] \tag{S2}$$

$$= \frac{1}{h^\gamma}\sum_{k=-K}^{K} g_{\gamma,k}\Big[-\partial_x U(x-kh)\exp(\ell_k - \ell^* + \ell^*)\Big] \tag{S3}$$

$$= \frac{\exp \ell^*}{h^\gamma}\sum_{k=-K}^{K} g_{\gamma,k}\Big[-\partial_x U(x-kh)\exp(\ell_k - \ell^*)\Big] \tag{S4}$$

where $\ell^* = \max_{k\in[\![-K,K]\!]} \ell_k$. This numerical approach is similar to the well-known 'log-sum-exp' trick.

2. Proof of Theorem 1

Before proving Theorem 1, we present the following proposition that will be helpful for our analysis.

Proposition 1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function and assume that $\mathcal{D}^\gamma f(x)$ is well-defined for some $\gamma \in \mathbb{R}$. Then, the following equality holds:*

$$\partial_x \mathcal{D}^\gamma f(x) = \mathcal{D}^\gamma \partial_x f(x). \tag{S5}$$

Proof. By definition we have:

$$\mathcal{D}^\gamma f(x) = \mathcal{F}^{-1}\{|\omega|^\gamma \hat{f}(\omega)\}, \tag{S6}$$

$$\partial_x f(x) = \mathcal{F}^{-1}\{i\omega \hat{f}(\omega)\}, \tag{S7}$$

[1]LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France. Correspondence to: Umut Şimşekli <umut.simsekli@telecom-paristech.fr>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

where $\mathcal{F}$ denotes the Fourier transform, $\hat{f}(\omega) = \mathcal{F}\{f(x)\}$, and $i = \sqrt{-1}$. By using these definitions, we obtain:

$$\begin{aligned}
\partial_x \mathcal{D}^\gamma f(x) &= \mathcal{F}^{-1}\{\mathcal{F}\{\partial_x \mathcal{D}^\gamma f(x)\}\} && \text{(S8)}\\
&= \mathcal{F}^{-1}\{i\omega \mathcal{F}\{\mathcal{D}^\gamma f(x)\}\} && \text{(S9)}\\
&= \mathcal{F}^{-1}\{i\omega |\omega|^\gamma \hat{f}(\omega)\} && \text{(S10)}\\
&= \mathcal{F}^{-1}\{|\omega|^\gamma \mathcal{F}\{\mathcal{F}^{-1}\{i\omega \hat{f}(\omega)\}\}\} && \text{(S11)}\\
&= \mathcal{F}^{-1}\{|\omega|^\gamma \mathcal{F}\{\partial_x f(x)\}\} && \text{(S12)}\\
&= \mathcal{F}^{-1}\{\mathcal{F}\{\mathcal{D}^\gamma \partial_x f(x)\}\} && \text{(S13)}\\
&= \mathcal{D}^\gamma \partial_x f(x). && \text{(S14)}
\end{aligned}$$

This completes the proof. □

2.1. Proof of Theorem 1

Proof. Let us define $q(X,t)$ as the probability density function of the state X at time t. By Proposition 1 in (Schertzer et al., 2001), we obtain the fractional Fokker-Planck equation associated with the SDE given in (3) as follows:

$$\partial_t q(X,t) = -\partial_X[b(X,\alpha)q(X,t)] - \mathcal{D}^\alpha q(X,t). \tag{S15}$$

By using the definition of $b(X,\alpha)$ we obtain

$$\begin{aligned}
\partial_t q(X,t) &= -\partial_X[\frac{\mathcal{D}^{\alpha-2}\{-\phi(X)\partial_X U(X)\}}{\phi(X)} q(X,t)] - \mathcal{D}^\alpha q(X,t) && \text{(S16)}\\
&= -\partial_X[\frac{\mathcal{D}^{\alpha-2}\{-\pi(X)\partial_X U(X)\}}{\pi(X)} q(X,t)] - \mathcal{D}^\alpha q(X,t). && \text{(S17)}
\end{aligned}$$

Here, we used the fact that $\pi(X) = \phi(X)/Z$, where $Z = \int \phi(X)dX$. By using $-\partial_X U(X) = \partial_X \log \pi(X) = \frac{\partial_X \pi(X)}{\pi(X)}$, we obtain:

$$\partial_t q(X,t) = -\partial_X[\frac{\mathcal{D}^{\alpha-2}\{\partial_X \pi(X)\}}{\pi(X)} q(X,t)] - \mathcal{D}^\alpha q(X,t) \tag{S18}$$

We can verify that $\pi(X)$ is an invariant measure of the Markov process $(X_t)_{t\geq 0}$ by checking

$$\begin{aligned}
-\partial_X[\frac{\mathcal{D}^{\alpha-2}\{\partial_X \pi(X)\}}{\pi(X)}\pi(X)] - \mathcal{D}^\alpha \pi(X) &= -\partial_X[\mathcal{D}^{\alpha-2}\{\partial_X \pi(X)\}] - \mathcal{D}^\alpha \pi(X) && \text{(S19)}\\
&= -\partial_X^2[\mathcal{D}^{\alpha-2}\{\pi(X)\}] - \mathcal{D}^\alpha \pi(X) && \text{(S20)}\\
&= \mathcal{D}^2[\mathcal{D}^{\alpha-2}\{\pi(X)\}] - \mathcal{D}^\alpha \pi(X) && \text{(S21)}\\
&= \mathcal{D}^\alpha\{\pi(X)\} - \mathcal{D}^\alpha\{\pi(X)\} && \text{(S22)}\\
&= 0. && \text{(S23)}
\end{aligned}$$

Here, we used the semigroup property of the Riesz potentials $\mathcal{D}^a\mathcal{D}^b f(x) = \mathcal{D}^{a+b} f(x)$ in (S22) and Proposition 1 in (S20). If $b(X,\alpha)$ is Lipschitz continuous, by (Schertzer et al., 2001) we can conclude that $\pi(X)$ is the unique invariant measure of the Markov process $(X_t)_{t\geq 0}$.

□

3. Proof of Corollary 1

Proof. By Theorem 1, we know that $\pi(X)$ is the unique invariant distribution of the Markov process $(X_t)_t$. Then, the claim directly follows Theorem 2 of (Panloup, 2008), provided that there exists $p \in (0, 1/2]$ and $q \in [1/2, 1]$, such that the following conditions hold:

$$\int_{|x|>1} v(x)|x|^{2p} < \infty, \quad \text{and} \quad \int_{|x|\leq 1} v(x)|x|^{2q} < \infty, \tag{S24}$$

where $v(x)$ is the Lévy-measure of the symmetric α-stable Lévy process, defined as

$$v(x) = \frac{1}{|x|^{\alpha+1}}. \tag{S25}$$

It is easy to see that these conditions hold with $p \in (0, 1/2]$ and $q \in (\alpha/2, 1]$. Therefore, we can directly apply Theorem 2 of (Panloup, 2008) in order to obtain the desired result. □

4. Proof of Theorem 2

Before proving Theorem 2, we first bound $|\mathcal{D}^\gamma f_\pi(x) - \Delta_h^\gamma f_\pi(x)|$ and $|\Delta_h^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)|$, which will be useful in our analysis.

Çelik & Duman (2012) showed that $|\mathcal{D}^\gamma f(x) - \Delta_h^\gamma f(x)| = \mathcal{O}(h^2)$ for $1 < \gamma \leq 2$. However, we cannot directly use their result. For completeness, we adapt the proof of Lemma 2.2 in (Çelik & Duman, 2012), and prove that we obtain a bound of the same order for $-1 < \gamma < 0$.

Lemma 1. *Assume $f(x) \in \mathcal{C}^3(\mathbb{R})$ and all derivatives up to order three belong to $\mathcal{L}_1(\mathbb{R})$. Let Δ_h^γ be the operator defined in (7). Then, for $-1 < \gamma < 0$, the following bound holds:*

$$|\mathcal{D}^\gamma f(x) - \Delta_h^\gamma f(x)| = \mathcal{O}(h^2) \tag{S26}$$

as h goes to zero.

Proof. We follow the same proof technique given in (Çelik & Duman, 2012). We first make use of the generator of (7) given as follows: (Ortigueira, 2006)

$$|2\sin(z/2)|^\gamma = \sum_{k=-\infty}^{\infty} g_{\gamma,k} \exp(ikz). \tag{S27}$$

Now, consider the Fourier transform of $\Delta_h^\gamma f(x)$

$$\mathcal{F}\{\Delta_h^\gamma f(x)\} = \sum_{k=-\infty}^{\infty} g_{\gamma,k} \exp(ikh\omega)\hat{f}(\omega) \tag{S28}$$

where $\mathcal{F}\{f(x)\} = \int_{-\infty}^{\infty} f(x)\exp(i\omega x)dx$ and $\hat{f}(\omega) \triangleq \mathcal{F}\{f(x)\}$. Then, we have

$$\mathcal{F}\{\Delta_h^\gamma f(x)\} = |2\sin\frac{\omega h}{2}|^\gamma \hat{f}(\omega) \tag{S29}$$

Let us define $\hat{\varphi}(h,\omega) = |\omega|^\gamma(1 - \frac{|2\sin\frac{\omega h}{2}|^\gamma}{|\omega h|^\gamma})\hat{f}(\omega)$. Then we have

$$-\frac{1}{h^\gamma}\mathcal{F}\{\Delta_h^\gamma f(x)\} = -|\omega|^\gamma \hat{f}(\omega) + \hat{\varphi}(h,\omega). \tag{S30}$$

Let us define $z = \omega h$ and $v(z) = \frac{|2\sin(z/2)|^\gamma}{|z|^\gamma}$. Now, we will bound the function $v(z)$. By using a Taylor expansion, we obtain

$$v(z) = \left|\frac{2}{z}\right|^\gamma \left|\frac{z}{2} - \left(\frac{z}{2}\right)^3\frac{1}{3!} + \left(\frac{z}{2}\right)^5\frac{1}{5!} - \cdots\right|^\gamma \tag{S31}$$

$$= \left|1 - \left(\frac{z}{2}\right)^2\frac{1}{3!} + \left(\frac{z}{2}\right)^4\frac{1}{5!} - \cdots\right|^\gamma \tag{S32}$$

Since $\gamma < 0$, for small enough z, we have

$$v(z) \leq \left(1 - \left|\left(\frac{z}{2}\right)^2\frac{1}{3!} + \left(\frac{z}{2}\right)^4\frac{1}{5!} + \cdots\right|\right)^\gamma \tag{S33}$$

$$= 1 - \gamma\left|\left(\frac{z}{2}\right)^2\frac{1}{3!} + \left(\frac{z}{2}\right)^4\frac{1}{5!} + \cdots\right| + \gamma(\gamma-1)\left|\left(\frac{z}{2}\right)^2\frac{1}{3!} + \left(\frac{z}{2}\right)^4\frac{1}{5!} + \cdots\right|^2 - \cdots \tag{S34}$$

$$\leq 1 + C_0 z^2 \tag{S35}$$

$$= \mathcal{O}(1 + z^2). \tag{S36}$$

By our assumptions, we have

$$|\hat{f}(\omega)| \leq C_1(1+|\omega|)^{-3}. \tag{S37}$$

Therefore, we obtain

$$\begin{aligned}
|\hat{\varphi}(h,\omega)| &= |\omega|^{\gamma}|v(\omega h)-1||\hat{f}(\omega)| &\text{(S38)}\\
&\leq |\omega|^{\gamma}C_0|\omega h|^2 C_1(1+|\omega|)^{-3} &\text{(S39)}\\
&\leq C_2h^2(1+|\omega|)^{\gamma+2}(1+|\omega|)^{-3} &\text{(S40)}\\
&= C_2h^2(1+|\omega|)^{\gamma-1}. &\text{(S41)}
\end{aligned}$$

Since $-1<\gamma<0$, the inverse Fourier transform of $\hat{\varphi}(h,\omega)$ exists. Then we consider the inverse Fourier transform of (S30) and obtain

$$-\frac{1}{h^{\gamma}}\Delta_h^{\gamma} f_\pi(x) = \mathcal{D}^{\gamma} f(x) + \varphi(h,x) \tag{S42}$$

where

$$\varphi(h,x) \triangleq \mathcal{F}^{-1}\{\hat{\varphi}(h,\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\varphi}(h,\omega)\exp(-i\omega x)d\omega. \tag{S43}$$

By using the bound for $|\hat{\varphi}(h,\omega)|$, we obtain

$$\begin{aligned}
|\varphi(h,x)| &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\varphi}(h,\omega)\exp(-i\omega x)d\omega &\text{(S44)}\\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{\varphi}(h,\omega)|d\omega &\text{(S45)}\\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}C_2h^2(1+|\omega|)^{\gamma-1}d\omega &\text{(S46)}\\
&\leq C_3h^2. &\text{(S47)}
\end{aligned}$$

Finally, we conclude that

$$\begin{aligned}
|\mathcal{D}^{\gamma} f(x) - \frac{1}{h^{\gamma}}\Delta_h^{\gamma} f_\pi(x)| &= |\varphi(h,x)| &\text{(S48)}\\
&\leq C_3h^2. &\text{(S49)}
\end{aligned}$$

□

Now, we bound the term $|\Delta_h^{\gamma} f_\pi(x) - \Delta_{h,K}^{\gamma} f_\pi(x)|$.

Lemma 2. *Assume* $\left|f_\pi(x-kh)\right| \leq C\exp(-|k|h)$ *for some* $C>0$ *and* $|k|>K$ *for some* $K<\infty$*, where* $K\in\mathbb{N}_+$*. Then the following bound holds:*

$$\left|\Delta_h^{\gamma} f_\pi(x) - \Delta_{h,K}^{\gamma} f_\pi(x)\right| = \mathcal{O}\Big(\frac{1}{hK}\Big). \tag{S50}$$

Proof. By definition we have

$$\begin{aligned}
\left|\Delta_h^{\gamma} f_\pi(x) - \Delta_{h,K}^{\gamma} f_\pi(x)\right| &= \left|h^{-\gamma}\sum_{k\notin[\![-K,K]\!]} g_{\gamma,k} f_\pi(x-kh)\right| &\text{(S51)}\\
&\leq h^{-\gamma}\sum_{k\notin[\![-K,K]\!]} g_{\gamma,k}\left|f_\pi(x-kh)\right| &\text{(S52)}
\end{aligned}$$

By the hypothesis and the symmetry of the coefficients ($g_{\gamma,k} = g_{\gamma,-k}$), we have

$$\left|\Delta_h^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)\right| \leq Ch^{-\gamma} \sum_{k=K+1}^{\infty} g_{\gamma,k} \exp(-kh) \tag{S53}$$

From (Ortigueira, 2006; Çelik & Duman, 2012), we know that $g_{\gamma,k} = \mathcal{O}(\frac{1}{k^{\gamma+1}})$, then we obtain

$$\left|\Delta_h^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)\right| \leq Ch^{-\gamma} \sum_{k=K+1}^{\infty} \frac{1}{k^{\gamma+1}} \exp(-kh) \tag{S54}$$

$$= Ch \sum_{k=K+1}^{\infty} \frac{1}{(hk)^{\gamma+1}} \exp(-kh) \tag{S55}$$

$$\leq Ch \int_K^\infty (yh)^{-(\gamma+1)} \exp(-yh)\, dy \tag{S56}$$

By making a change of variables, we obtain

$$\left|\Delta_h^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)\right| \leq C \int_{hK}^\infty y^{-(\gamma+1)} \exp(-y)\, dy \tag{S57}$$

$$= C\Gamma(-\gamma, hK), \tag{S58}$$

where $\Gamma(\cdot,\cdot)$ denotes the incomplete gamma function (Borwein & Chan, 2009). Then by using Theorem 2.4 of (Borwein & Chan, 2009), we obtain the desired result as follows:

$$\left|\Delta_h^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)\right| \leq C\frac{1}{hK}. \tag{S59}$$

□

4.1. Proof of Theorem 2

Proof. We decompose the error as follows:

$$\left|\mathcal{D}^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)\right| \leq \left|\mathcal{D}^\gamma f_\pi(x) - \Delta_h^\gamma f_\pi(x)\right| + \left|\Delta_h^\gamma f_\pi(x) - \Delta_{h,K}^\gamma f_\pi(x)\right|. \tag{S60}$$

Then we obtain the desired result by applying Lemmas 1 and 2 to the first and the second terms of the left hand side of the above inequality. □

5. Proof of Theorem 3

Before presenting the proof of Theorem 3, let us define the following SDEs which will be useful in the analysis:

$$dX_t = b(X_{t-},\alpha)dt + dL_t^\alpha \tag{S61}$$

$$dY_t = \tilde{b}_{h,K}(Y_{t-},\alpha)dt + dL_t^\alpha \tag{S62}$$

where b and $\tilde{b}$ are defined in (5) and (8), respectively. Here, (S61) is our main SDE, (S62) is another SDE whose drift is $\tilde{b}$.

Let us first present the following lemma, which will be useful for proving Theorem 3.

Lemma 3. *Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be the solution processes of the SDEs (S61) and (S62). Assume that both $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are geometrically ergodic with their unique invariant measures and $|\partial_x g|$ is bounded. Further assume that the truncation parameter K is chosen in such a way that **H**4 holds for any x. Then the following bound on the weak error holds:*

$$\left|\mathbb{E}[g(X_t) - g(Y_t)]\right| \leq C\big(1 - \exp(-\lambda t)\big)\big(h^2 + \frac{1}{hK}\big), \tag{S63}$$

for some $C, \lambda > 0$.

Proof. We follow a standard approach for weak error analysis in SDEs. We make use of the semigroups associated with $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$, given as $\mathcal{P}_t^X g(x) \triangleq \mathbb{E}[g(X_t)]$ and $\mathcal{P}_t^Y g(x) \triangleq \mathbb{E}[g(Y_t)]$. Then, we rewrite the weak error by using the semigroups, given as follows (Kohatsu-Higa, 2015):

$$\mathbb{E}[g(X_t) - g(Y_t)] = \mathcal{P}_t^X g(x) - \mathcal{P}_t^Y g(x) \tag{S64}$$

$$= \int_0^t \partial_s \{\mathcal{P}_s^X \mathcal{P}_{t-s}^Y g(x)\} ds. \tag{S65}$$

We now investigate the integrand, as follows:

$$\partial_s \{\mathcal{P}_s^X \mathcal{P}_{t-s}^Y g(x)\} = (\partial_s \mathcal{P}_s^X)\mathcal{P}_{t-s}^Y g(x) + \mathcal{P}_s^X(\partial_s \mathcal{P}_{t-s}^Y) g(x) \tag{S66}$$

$$= \mathcal{P}_s^X \mathcal{A}^X \mathcal{P}_{t-s}^Y g(x) - \mathcal{P}_s^X \mathcal{A}^Y \mathcal{P}_{t-s}^Y g(x) \tag{S67}$$

$$= \mathcal{P}_s^X \{\mathcal{A}^X - \mathcal{A}^Y\} \mathcal{P}_{t-s}^Y g(x), \tag{S68}$$

where $\mathcal{A}^X$ and $\mathcal{A}^Y$ are the generators of the SDEs in (S61) and (S62), respectively, and they are defined as follows (Duan, 2015):

$$\mathcal{A}^X f(x) \triangleq b(x,\alpha)\partial_x f(x) + \int_{\mathbb{R}\setminus 0} [f(x+y) - f(x) - \mathbb{1}_{\{|y|<1\}} y \partial_x f(x)] v(dy), \tag{S69}$$

$$\mathcal{A}^Y f(x) \triangleq \tilde{b}_{h,K}(x,\alpha)\partial_x f(x) + \int_{\mathbb{R}\setminus 0} [f(x+y) - f(x) - \mathbb{1}_{\{|y|<1\}} y \partial_x f(x)] v(dy), \tag{S70}$$

for a differentiable function f, where $\mathbb{1}$ is the indicator function and $v(dy)$ is the Lévy-measure of the symmetric α-stable Lévy process defined in (S25). Since these SDEs have the same volatility, the difference $(\mathcal{A}^X - \mathcal{A}^Y)f(x)$ simplifies and it is equal to $(b(x,\alpha) - \tilde{b}_{h,K}(x,\alpha))\partial_x f(x)$. Accordingly, we obtain the following expression:

$$\partial_s \{\mathcal{P}_s^X \mathcal{P}_{t-s}^Y g(x)\} = \mathcal{P}_s^X (b(x,\alpha) - \tilde{b}_{h,K}(x,\alpha))\partial_x \mathcal{P}_{t-s}^Y g(x), \tag{S71}$$

$$= \mathcal{P}_s^X (b(x,\alpha) - \tilde{b}_{h,K}(x,\alpha))\mathcal{P}_{t-s}^Y \partial_x g(x) \tag{S72}$$

where we assumed the interchangeability of integration and differentiation. By the ergodicity assumptions, we have:

$$|\mathcal{P}_s^X f(x)| \leq C \exp\big(-\lambda_x s\big)\|f\|_\infty, \tag{S73}$$

$$|\mathcal{P}_{t-s}^Y f(x)| \leq C \exp\big(-\lambda_y (t-s)\big)\|f\|_\infty \tag{S74}$$

for some $C, \lambda_x, \lambda_y > 0$ and a bounded function f. By injecting (S72) into (S65) and then using the boundedness assumption on $\partial_x g$, (S73), (S74), and Theorem 2, we obtain the following inequality: (for some $C > 0$)

$$\Big|\mathbb{E}[g(X_t) - g(Y_t)]\Big| \leq C\Big(h^2 + \frac{1}{hK}\Big)\int_0^t \exp(-\lambda_x s) ds \tag{S75}$$

$$\leq C\big(1 - \exp(-\lambda_x t)\big)\Big(h^2 + \frac{1}{hK}\Big), \tag{S76}$$

as desired. This completes the proof.

□

5.1. Proof of Theorem 3

Proof. Let us first define the following quantities:

$$\nu(g) = \int g(X)\pi(dX) \tag{S77}$$

$$\tilde{\nu}(g) = \int g(Y)\tilde{\pi}(dY) \tag{S78}$$

where π, and $\tilde{\pi}$ are the unique invariant measures of (S61) and (S62), respectively. And let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be the solution processes of the SDEs (S61) and (S62). By the triangle inequality, we have

$$\left|\nu(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)\right| \leq \left|\nu(g) - \tilde{\nu}(g)\right| + \left|\tilde{\nu}(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)\right|. \tag{S79}$$

Due to the ergodicity assumptions, we can rewrite the right hand side of the above inequality as follows:

$$\begin{aligned}\left|\nu(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)\right| &\leq \left|\lim_{t\to\infty} \mathbb{E}\big[g(X_t) - g(Y_t)\big]\right| + \left|\tilde{\nu}(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)\right| && \text{(S80)}\\ &= \lim_{t\to\infty}\left|\mathbb{E}\big[g(X_t) - g(Y_t)\big]\right| + \left|\tilde{\nu}(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)\right| && \text{(S81)}\end{aligned}$$

where (S81) can be obtained by the reverse triangle inequality and the squeeze theorem. By (Panloup, 2008), we have almost surely

$$\left|\tilde{\nu}(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)\right| = 0. \tag{S82}$$

By Lemma 3, we have

$$\lim_{t\to\infty}\left|\mathbb{E}\big[g(X_t) - g(Y_t)\big]\right| \leq C(h^2 + \frac{1}{hK}), \tag{S83}$$

for some $C > 0$. Finally, by injecting (S82), and (S83) in (S81), we obtain the desired result:

$$|\nu(g) - \lim_{N\to\infty} \tilde{\nu}_N(g)| \leq C(h^2 + \frac{1}{hK}). \tag{S84}$$

This completes the proof.

□

Remark 1. *The assumption **H**5 is not very restrictive and for the SDE* (3) *it can be easily satisfied if the following conditions hold:*

A1) $xb(x,\alpha) <= -ax^2 + c$ *with* $a, c > 0$ *.*
A2) Let $S_K^i(x) = \sum_{k=-K}^{K} g_{k,\gamma} f_\pi^{(i)}(x - kh)$*, where* $f_\pi^{(i)}$ *is the* i*'th derivative of* f_π*. Then* $\{S_K\}_{K>0}$ *converges uniformly when* $K \to \infty$ *and is bounded, for any* x *belonging to a compact set and* $i >= 1$*.*

A1 is a standard growth condition. A2 is mild due to the nature of f_π*, and it ensures* X_t *to have a smooth density (see (Picard, 1996)). We note that, if the SDE* (3) *satisfies A1-2, then it is easy to show that so does the perturbed SDE defined in **H**5.*

6. Proof of Corollary 2

Proof. It is easy to check that (14) corresponds to using (8) with $h = r(x)$. Then we obtain the first part of the conclusion by directly applying Theorem 2. The second part of the conclusion can be proved by using the same proof technique presented in Theorem 3. □

Remark 2. *Corollary 2 implies that the weak error FLA depends heavily on the structure of the target density. If the high probability region of* π *is concentrated in a particular area,* K_x *would be small and vice versa. On the other hand, if* x *is near a mode of* π *or* $f_\pi(x)$ *is symmetric around* x*, or* f_π *varies very slowly with* x*,* $r(x)$ *can be arbitrarily small. Finally, Corollary 2 expresses the overall error in terms of* K_x *and* $r(x)$*, and illustrates the roles of these terms.*

7. A Note on the Experiments Conducted on SG-FLA

In the SG-FLA experiments, we monitored the training likelihood and we did not observe that SG-FLA is able to find a better mode in a systematic way. However, we did observe that SG-FLA is more robust to the size of the minibatches – therefore to the variance of the stochastic gradients – when compared to SGLD. We believe that this observation is caused by the fact that the jumps in SG-FLA provide robustness against stochastic gradients and the choice of the step sizes.

References

Borwein, J. M. and Chan, O. Uniform bounds for the complementary incomplete gamma function. *Mathematical Inequalities and Applications*, 12:115–121, 2009.

Çelik, C. and Duman, M. Crank–Nicolson method for the fractional diffusion equation with the Riesz fractional derivative. *Journal of Computational Physics*, 231(4):1743–1750, 2012.

Duan, J. *An Introduction to Stochastic Dynamics*. Cambridge University Press, New York, 2015.

Kohatsu-Higa, A. A short course on weak approximations for Lévy driven SDEs. In *Stochastic Analysis: A Series of Lectures*, pp. 247–270. Springer, 2015.

Ortigueira, M. D. Riesz potential operators and inverses via fractional centred derivatives. *International Journal of Mathematics and Mathematical Sciences*, 2006, 2006.

Panloup, F. Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *The Annals of Applied Probability*, 18(2):379–426, 2008.

Picard, J. On the existence of smooth densities for jump processes. *Probability Theory and Related Fields*, 105(4):481–511, 1996.

Schertzer, D, Larchevêque, M, Duan, J, Yanovsky, VV, and Lovejoy, S. Fractional Fokker–Planck equation for nonlinear stochastic differential equations driven by non-Gaussian Lévy stable noises. *Journal of Mathematical Physics*, 42(1): 200–212, 2001.