Non-Asymptotic Analysis of Fractional Langevin Monte Carlo for Non-Convex Optimization
SUPPLEMENTARY DOCUMENT

Thanh Huy Nguyen [1] **Umut Şimşekli** [1] **Gaël Richard** [1]

S1. Proof of Lemma 1

Proof. Let $q(X,t)$ be the probability density of $X(t)$. By Proposition 1 in (Schertzer et al., 2001) (see also Section 7 of the same study), the fractional Fokker-Planck equation associated with (9) is given as follows:

$$\partial_t q(X,t) = -\sum_{i=1}^{d} \frac{\partial[(b(X,\alpha))_i q(X,t)]}{\partial X_i} - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} q(X,t).$$

Using definition (10) of b, we have

$$\begin{aligned}
\partial_t q(X,t) &= -\sum_{i=1}^{d} \frac{\partial}{\partial_{X_i}} [\frac{\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(-\beta\phi(X)\frac{\partial f(X)}{\partial_{X_i}})}{\phi(X)} q(X,t)] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} q(X,t) \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial_{X_i}} [\frac{\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(-\beta\pi(X)\frac{\partial f(X)}{\partial_{X_i}})}{\pi(X)} q(X,t)] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} q(X,t) \\
&= -\sum_{i=1}^{d} \frac{\partial}{\partial_{X_i}} [\frac{\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(\frac{\partial \pi(X)}{\partial_{X_i}})}{\pi(X)} q(X,t)] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} q(X,t).
\end{aligned}$$

Here, we used $\pi(X) = \phi(X)/\int \phi(X)\mathrm{d}X$ in the second equality and $-\beta\frac{\partial}{\partial X_i} f(X) = \frac{\partial}{\partial X_i}\log\pi(X) = \frac{\partial\pi(X)/\partial X_i}{\pi(X)}$ in the third equality. Next, by replacing q by π on the right hand side of the above equality, we have:

$$\begin{aligned}
-\sum_{i=1}^{d} \frac{\partial}{\partial_{X_i}} [\frac{\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(\frac{\partial \pi(X)}{\partial_{X_i}})}{\pi(X)} \pi(X,t)] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} \pi(X,t) &= -\sum_{i=1}^{d} \frac{\partial}{\partial_{X_i}} [\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(\frac{\partial \pi(X)}{\partial_{X_i}})] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} \pi(X,t) \\
&= -\sum_{i=1}^{d} \frac{\partial^2}{\partial_{X_i^2}} [\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(\pi(X))] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} \pi(X,t) \\
&= \sum_{i=1}^{d} \mathcal{D}^{2}_{X_i} [\beta^{-1}\mathcal{D}^{\alpha-2}_{X_i}(\pi(X))] - \beta^{-1} \sum_{i=1}^{d} \mathcal{D}^{\alpha}_{X_i} \pi(X,t)
\end{aligned}$$

[1]LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France. Correspondence to: Thanh Huy Nguyen <thanh.nguyen@telecom-paristech.fr>.

Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

$$= \sum_{i=1}^{d} \mathcal{D}_{X_i}^{\alpha}[\beta^{-1}\pi(X)] - \beta^{-1}\sum_{i=1}^{d}\mathcal{D}_{X_i}^{\alpha}\pi(X,t) = 0.$$

Here, we used Proposition 1 in (Şimşekli, 2017), $\mathcal{D}^2 u(x) = -\frac{\partial}{\partial x^2}u(x)$, and the semi-group property of the Riesz derivation $\mathcal{D}^a\mathcal{D}^b u(x) = \mathcal{D}^{a+b}u(x)$. This proves that π is an invariant measure of the Markov process $(X(t))_{t\geq 0}$. □

S2. Proof of Proposition 1

Proof. By Corollary 1.3 in (Liang & Wang, 2018), the assumptions imply that there exist constants $\bar{C} > 0$ and $\bar{C}_1 > 0$ such that $\mathcal{W}_1(\mu_{3t}, \pi) \leq \bar{C}\beta e^{-\bar{C}_1 t}$.

Let P_{3t} be the coupling of μ_{3t} and π that such that $\mathcal{W}_1(\mu_{3t},\pi) = \int \|X_3(t) - \hat{W}\|\mathrm{d}P_{3t}$. For $0 < \lambda < 1$, by Hölder inequality,

$$\begin{aligned}\mathcal{W}_\lambda^\lambda(\mu_{3t},\pi) \leq& \int \|X_3(t) - \hat{W}\|^\lambda \mathrm{d}P_{3t}\\ \leq& \Big(\int \|X_3(t) - \hat{W}\|\mathrm{d}P_{3t}\Big)^\lambda\\ =&\mathcal{W}_1^\lambda(\mu_{3t},\pi)\end{aligned}$$

For $\alpha > \lambda > 1$,

$$\begin{aligned}\mathcal{W}_\lambda^\lambda(\mu_{3t},\pi) \leq& \int \|X_3(t) - \hat{W}\|^\lambda \mathrm{d}P_{3t}\\ \leq& \int \|X_3(t) - \hat{W}\|^\delta \|X_3(t) - \hat{W}\|^{\lambda-\delta}\mathrm{d}P_{3t}\\ \leq& \Big(\int \|X_3(t) - \hat{W}\|\mathrm{d}P_{3t}\Big)^\delta \Big(\int \|X_3(t) - \hat{W}\|^{(\lambda-\delta)/(1-\delta)}\mathrm{d}P_{3t}\Big)^{1-\delta}\\ =&\mathcal{W}_1^\delta(\mu_{3t},\pi)\Big(\int \|X_3(t) - \hat{W}\|^{(\lambda-\delta)/(1-\delta)}\mathrm{d}P_{3t}\Big)^{1-\delta},\end{aligned}$$

where we used Holder's inequality for $\delta < 1$ such that $(\lambda-\delta)/(1-\delta) < \alpha$, and $\int \|X_3(t) - \hat{W}\|^{(\lambda-\delta)/(1-\delta)}\mathrm{d}P_{3t}$ is bounded by a constant, by Assumption **H**5.

Finally, we have

$$\mathcal{W}_\lambda(\mu_{3t},\pi) \leq C\beta e^{-C_1 t},$$

for some constants $C, C_1 > 0$ and for $0 < \lambda < \alpha$. This completes the proof.

□

Remark 1. *Let us consider the case where the dimension* d *is equal to* 1 *(the extension for* $d > 1$ *is similar). The first part of Assumption* ***H****5 can be satisfied under the following (rather non-trivial) assumptions. Assume that there exist constants* $P, C_1, C_2, C_3, C_4, C_5, C_6 > 0$ *such that:*

$$f'(z) > 0 \text{ if } z > P, \tag{S1}$$

$$\int_{|z|\leq P} |\phi(z)f'(z)|\mathrm{d}z = C_1 > 0 \tag{S2}$$

$$\int_{z<-P} \phi(z)|f'(z)||z|^{1-\alpha}\mathrm{d}z = C_2 > 0 \tag{S3}$$

$$\int_{z>P} \phi(z)f'(z)|z|^{1-\alpha}\mathrm{d}z = C_3 > 0, \tag{S4}$$

$$\text{if } |z| \leq P : \Big|\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big| \leq C_4|x-y| \;\; \forall x,y\in\mathbb{R}, \tag{S5}$$

$$\textit{if } z < -P : \Big|\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big| \leq C_5|x-y||z|^{1-\alpha} \ \ \forall x, y \in \mathbb{R}, \tag{S6}$$

$$\textit{if } z > P : \Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big) \leq C_6|z|^{1-\alpha}(y-x) \ \ \forall x, y \in \mathbb{R} \textit{ s.t } x > y, \tag{S7}$$

$$C_1C_4 + C_2C_5 < C_3C_6. \tag{S8}$$

By definition of Riesz potential, we have:

$$\begin{aligned}
b(x) - b(y) &= \int_{\mathbb{R}} \frac{\phi(z)f'(z)}{\phi(x)|x-z|^{\alpha-1}}\mathrm{d}z - \int_{\mathbb{R}} \frac{\phi(z)f'(z)}{\phi(y)|y-z|^{\alpha-1}}\mathrm{d}z \\
&= \int_{\mathbb{R}} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z \\
&= \int_{|z|\leq P} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z \\
&\quad + \int_{z<-P} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z \\
&\quad + \int_{z>P} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z.
\end{aligned}$$

By these assumptions, we estimate the first term on the right hand side in the above expression of $b(x) - b(y)$*, for* $x > y$*, as follows:*

$$\begin{aligned}
\Big|\int_{|z|\leq P} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z\Big| &\leq \int_{|z|\leq P} |\phi(z)f'(z)|\Big|\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big|\mathrm{d}z \\
&\leq \int_{|z|\leq P} |\phi(z)f'(z)|C_4|x-y|\mathrm{d}z \\
&= C_1C_4|x-y| \\
&= C_1C_4(x-y).
\end{aligned}$$

For the remaining terms, we have:

$$\begin{aligned}
\Big|\int_{z<-P} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z\Big| &\leq \int_{z<-P} |\phi(z)f'(z)|\Big|\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big|\mathrm{d}z \\
&\leq \int_{z<-P} \phi(z)|f'(z)|C_5|z|^{1-\alpha}|x-y|\mathrm{d}z \\
&= C_2C_5|x-y| \\
&= C_2C_5(x-y),
\end{aligned}$$

and

$$\begin{aligned}
\int_{z>P} \phi(z)f'(z)\Big(\frac{1}{\phi(x)|x-z|^{\alpha-1}} - \frac{1}{\phi(y)|y-z|^{\alpha-1}}\Big)\mathrm{d}z &\leq \int_{z>P} \phi(z)f'(z)C_6|z|^{1-\alpha}(y-x)\mathrm{d}z \\
&= C_3C_6(y-x) \\
&= -C_3C_6(x-y).
\end{aligned}$$

By combining these estimates, we get, for $x > y$*:*

$$b(x) - b(y) \leq (C_1C_4 + C_2C_5 - C_3C_6)(x-y).$$

Thus, $(b(x) - b(y))(x-y) \leq (C_1C_4 + C_2C_5 - C_3C_6)(x-y)^2$*. Since* $C_1C_4 + C_2C_5 - C_3C_6 < 0$*, this inequality for drift* b *makes the first part of Assumption* ***H**5* *hold.*

S3. Proof of Lemma 2

In this section, we precise the statement of Lemma 2 and provide the proof.

Lemma S1. *Let V and W be two random variables on $\mathbb{R}^d$ which have μ and ν as the probability measures and let g be a function in $C^1(\mathbb{R}^d, \mathbb{R})$. Assume that for some $c_1 > 0, c_2 \geq 0$ and $0 \leq \gamma < 1$,*

$$\|\nabla g(w)\| \leq c_1 \|w\|^\gamma + c_2, \qquad \forall w \in \mathbb{R}^d$$

then the following bound holds:

$$\Big| \int g \mathrm{d}\mu - \int g \mathrm{d}\nu \Big| \leq \Big(c_1 \Big(\mathbb{E}_{\mathbf{P}} \|W\|^{\gamma p} \Big)^{\frac{1}{p}} + c_1 \Big(\mathbb{E}_{\mathbf{P}} \|V\|^{\gamma p} \Big)^{\frac{1}{p}} + c_2 \Big) \mathcal{W}_q(\mu, \nu).$$

Proof. We have

$$\begin{aligned} g(v) - g(w) &= \int_0^1 \langle w - v, \nabla g((1-t)v + tw) \rangle \mathrm{d}t \\ &\leq \int_0^1 \|w - v\| \|\nabla g((1-t)v + tw)\| \mathrm{d}t && \text{(by Cauchy-Schwarz)} \\ &\leq \int_0^1 \|w - v\| (c_1((1-t)\|v\| + t\|w\|)^\gamma + c_2) \mathrm{d}t && \text{(by the assumption on } \nabla g) \\ &\leq \|w - v\| \Big(c_1(\|v\| + \|w\|)^\gamma + c_2 \Big) \\ &\leq \|w - v\| (c_1 \|v\|^\gamma + c_1 \|w\|^\gamma + c_2). && \text{(by lemma S11)} \end{aligned}$$

Now let $\mathbf{P}$ be a joint probability distribution of μ and ν that achieves $\mathcal{W}_\lambda(\mu, \nu)$, that is, $\mathbf{P} = \mathcal{L}((W, V))$ with $\mu = \mathcal{L}(W)$ and $\nu = \mathcal{L}(V)$. We have

$$\begin{aligned} \int g \mathrm{d}\mu - \int g \mathrm{d}\nu &= \mathbb{E}_{\mathbf{P}}[g(W) - g(V)] \\ &\leq [\mathbb{E}_{\mathbf{P}}(c_1 \|W\|^\gamma + c_1 \|V\|^\gamma + c_2)^p]^{\frac{1}{p}} [\mathbb{E}_{\mathbf{P}} \|W - V\|^q]^{\frac{1}{q}} \\ &\leq \Big(c_1 \Big(\mathbb{E}_{\mathbf{P}} \|W\|^{\gamma p} \Big)^{\frac{1}{p}} + c_1 \Big(\mathbb{E}_{\mathbf{P}} \|V\|^{\gamma p} \Big)^{\frac{1}{p}} + c_2 \Big) \mathcal{W}_q(\mu, \nu), \end{aligned}$$

where we have used Holder's inequality and Minkowski's inequality. □

S4. Proof of Lemma 3

Proof. We define a real function F_λ as follows:

$$F_\lambda(y) \triangleq \|y\|^\lambda. \tag{S9}$$

It is clear that F_λ is a C^1 function. Let $Y(t) \triangleq X_1(t) - X_2(t)$. By the chain rule,

$$\begin{aligned} \mathrm{d}F_\lambda(Y(t)) &= \langle \nabla F_\lambda(Y(t)), b_1(X_1(t-), \alpha) - b_2(X_2(t-), \alpha) \rangle \mathrm{d}t \\ &= \lambda \|X_1(t) - X_2(t)\|^{\lambda-2} \langle X_1(t) - X_2(t), b_1(X_1(t-), \alpha) - b_2(X_2(t-), \alpha) \rangle \mathrm{d}t. \end{aligned} \tag{S10}$$

By integrating both sides of (S10) with respect to t, we arrive at

$$F_\lambda(Y(t)) = F_\lambda(Y(0)) + \int_0^t \lambda \|X_1(t) - X_2(t)\|^{\lambda-2} \langle X_1(t) - X_2(t), b_1(X_1(t-), \alpha) - b_2(X_2(t-), \alpha) \rangle \mathrm{d}s$$

$$= \int_0^t \lambda \|X_1(t) - X_2(t))\|^{\lambda-2} \langle X_1(t) - X_2(t), b_1(X_1(t-), \alpha) - b_2(X_2(t-), \alpha)\rangle \mathrm{d}s.$$

By definition of Wasserstein distance, we have

$$\mathcal{W}_\lambda(\mu_{1t}, \mu_{2t}) = \inf\{(\mathbb{E}[F_\lambda(Y(t))])^{1/\lambda}\},$$

which is the desired result. □

S5. Proof of Theorem 3

In this section, we first precise the statement of Theorem 3 and then provide the corresponding proof.

Theorem S1. *Let $\mathbb{E}\|L^\alpha(1)\|^\lambda \triangleq l_{\alpha,\lambda,d} < \infty$. We also define the following quantities:*

$$P_1(\eta) \triangleq \Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{p_1}} + (c\eta)^{\frac{1}{p_1}} + (2\eta(b+m))^{\frac{(q-1)}{2}} + 2^{\frac{(q-1)}{2}}(\eta B)^{(q-1)} + \Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)}{\alpha}} l_{\alpha,(q-1)p_1,d}^{\frac{1}{p_1}}$$
$$+\eta^{q-1}M^{q-1}\Big((2\eta(b+m))^{\frac{(q-1)\gamma}{2}} + 2^{\frac{(q-1)\gamma}{2}}(\eta B)^{(q-1)\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)\gamma}{\alpha}} l_{\alpha,(q-1)p_1\gamma,d}^{\frac{1}{p_1}}\Big),$$
$$P_2(\eta) \triangleq M\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{q_1}} + (c\eta)^{\frac{1}{q_1}} + (2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^{\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big),$$
$$Q_1(\eta) \triangleq c^{\frac{1}{p_1}} + (\mathbb{E}\|X_2(0)\|^{(q-1)p_1})^{\frac{1}{p_1}} + \eta^{q-1}\Big(M^{q-1}(\mathbb{E}\|X_2(0)\|^{(q-1)p_1\gamma})^{\frac{1}{p_1}} + B^{(q-1)}\Big) + \Big(\frac{\eta}{\beta}\Big)^{\frac{q-1}{\alpha}} l_{\alpha,(q-1)p_1,d}^{\frac{1}{p_1}},$$
$$Q_2 \triangleq M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}} + Mc^{\frac{1}{q_1}}.$$

Under additional assumption on the step-size: $0 < \eta \leq \frac{m}{M^2}$, we have

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{2t}) \leq q\eta\Big(k^2 P_1(\eta)P_2(\eta) + k^{1+1/p_1}P_1(\eta)Q_2 + k^{1+1/q_1}P_2(\eta)Q_1(\eta) + kQ_1(\eta)Q_2\Big).$$

Proof. From Lemma 3, we have

$$\begin{aligned}
\mathcal{W}_q^q(\mu_{1t}, \mu_{2t}) =& \mathbb{E}\Big[\int_0^t q\,\|X_1(s) - X_2(s))\|^{q-2}\langle X_1(s) - X_2(s), b_1(X_1(s-), \alpha) - b_2(X_2(s-), \alpha)\rangle \mathrm{d}s\Big]\\
=& \sum_{j=0}^{k-1} \mathbb{E}\Big[\int_{j\eta}^{(j+1)\eta} q\,\|X_1(s) - X_2(s))\|^{q-2}\langle X_1(s) - X_2(s), b_1(X_1(s-), \alpha) - b_2(X_2(s-), \alpha)\rangle \mathrm{d}s\Big]\\
\leq& \sum_{j=0}^{k-1} \mathbb{E}\Big[\int_{j\eta}^{(j+1)\eta} q\,\|X_1(s) - X_2(s)\|^{q-1} c_\alpha \|\nabla f(X_1(s)) - \nabla f(X_2(j\eta))\| \mathrm{d}s\Big]\\
=& q\sum_{j=0}^{k-1} \int_{j\eta}^{(j+1)\eta} \mathbb{E}\Big[\|X_1(s) - X_2(s)\|^{q-1} c_\alpha \|\nabla f(X_1(s)) - \nabla f(X_2(j\eta))\|\Big] \mathrm{d}s\\
\leq& q\sum_{j=0}^{k-1} \int_{j\eta}^{(j+1)\eta} \Big[\mathbb{E}\|X_1(s) - X_2(s)\|^{(q-1)p_1}\Big]^{\frac{1}{p_1}} \Big[\mathbb{E}\|c_\alpha(\nabla f(X_1(s)) - \nabla f(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}} \mathrm{d}s,
\end{aligned}$$

where we have used Cauchy-Schwarz inequality in the third line and Holder's inequality in the last line.

Since $(q-1)p_1 < 1$ by Assumption **H**4, using Lemma S11 twice, we have:

$$\begin{aligned}
\Big(\mathbb{E}\|X_1(s) - X_2(s)\|^{(q-1)p_1}\Big)^{\frac{1}{p_1}} \leq& \Big(\mathbb{E}\|X_1(s)\|^{(q-1)p_1} + \mathbb{E}\|X_2(s)\|^{(q-1)p_1}\Big)^{\frac{1}{p_1}}\\
\leq& \Big[\mathbb{E}\Big(\|X_1(s)\|^{(q-1)p_1}\Big)\Big]^{\frac{1}{p_1}} + \Big[\mathbb{E}\Big(\|X_2(s)\|^{(q-1)p_1}\Big)\Big]^{\frac{1}{p_1}}
\end{aligned}$$

Then, by applying Lemma S4 and Lemma S7 for $s \in [j\eta, (j+1)\eta)$, we obtain:

$$\begin{aligned}&\Big(\mathbb{E}\|X_1(s)-X_2(s)\|^{(q-1)p_1}\Big)^{\frac{1}{p_1}}\\&\quad\leq\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{q-1}+\Big[\mathbb{E}\|X_2(0)\|^{(q-1)p_1}+j\Big((2\eta(b+m))^{\frac{(q-1)p_1}{2}}+2^{\frac{(q-1)p_1}{2}}(\eta B)^{(q-1)p_1}\\&\quad+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)p_1}{\alpha}}l_{\alpha,(q-1)p_1,d}\Big)+(s-j\eta)^{(q-1)p_1}\Big(M^{(q-1)p_1}\Big(\mathbb{E}\|X_2(0)\|^{(q-1)p_1\gamma}+j\Big((2\eta(b+m))^{\frac{(q-1)p_1\gamma}{2}}\\&\quad+2^{\frac{(q-1)p_1\gamma}{2}}(\eta B)^{(q-1)p_1\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)p_1\gamma}{\alpha}}l_{\alpha,(q-1)p_1\gamma,d}\Big)\Big)+B^{(q-1)p_1}\Big)+\Big(\frac{s-j\eta}{\beta}\Big)^{\frac{(q-1)p_1}{\alpha}}l_{\alpha,(q-1)p_1,d}\Big]^{\frac{1}{p_1}}.\end{aligned}$$

Next, using Lemma S11, the inequalities $j < j+1$ and $s - j\eta \leq \eta$ for $s \in [j\eta, (j+1)\eta)$, we get

$$\begin{aligned}\Big(\mathbb{E}\|X_1(s)-X_2(s)\|^{(q-1)p_1}\Big)^{\frac{1}{p_1}}\leq&\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{q-1}+(\mathbb{E}\|X_2(0)\|^{(q-1)p_1})^{\frac{1}{p_1}}+(j+1)^{\frac{1}{p_1}}\Big((2\eta(b+m))^{\frac{(q-1)}{2}}\\&+2^{\frac{(q-1)}{2}}(\eta B)^{(q-1)}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1,d}\Big)+\eta^{q-1}\Big(M^{q-1}\Big((\mathbb{E}\|X_2(0)\|^{(q-1)p_1\gamma})^{\frac{1}{p_1}}\\&+(j+1)^{\frac{1}{p_1}}\Big((2\eta(b+m))^{\frac{(q-1)\gamma}{2}}+2^{\frac{(q-1)\gamma}{2}}(\eta B)^{(q-1)\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)\gamma}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1\gamma,d}\Big)\Big)\\&+B^{(q-1)}\Big)+\Big(\frac{\eta}{\beta}\Big)^{\frac{q-1}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1,d}.\end{aligned}$$

We note that $s < (j+1)\eta$ and $q-1 < \frac{1}{p_1}$ (from the assumptions). Hence,

$$\begin{aligned}\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{q-1}\leq&\Big(c\Big((j+1)\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{\frac{1}{p_1}}\\\leq&(j+1)^{\frac{1}{p_1}}\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)\Big)^{\frac{1}{p_1}}+c^{\frac{1}{p_1}},\end{aligned}$$

where the last inequality is an application of Lemma S11. By replacing this inequality into the previous one and rearranging the terms, we have

$$\begin{aligned}&\Big(\mathbb{E}\|X_1(s)-X_2(s)\|^{(q-1)p_1}\Big)^{\frac{1}{p_1}}\\&\leq c^{\frac{1}{p_1}}+(\mathbb{E}\|X_2(0)\|^{(q-1)p_1})^{\frac{1}{p_1}}+\eta^{q-1}\Big(M^{q-1}(\mathbb{E}\|X_2(0)\|^{(q-1)p_1\gamma})^{\frac{1}{p_1}}+B^{(q-1)}\Big)+\Big(\frac{\eta}{\beta}\Big)^{\frac{q-1}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1,d}\\&\quad+(j+1)^{\frac{1}{p_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)\Big)^{\frac{1}{p_1}}+(2\eta(b+m))^{\frac{(q-1)}{2}}+2^{\frac{(q-1)}{2}}(\eta B)^{(q-1)}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1,d}\\&\quad+\eta^{q-1}M^{q-1}\Big((2\eta(b+m))^{\frac{(q-1)\gamma}{2}}+2^{\frac{(q-1)\gamma}{2}}(\eta B)^{(q-1)\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)\gamma}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1\gamma,d}\Big)\Big)\\&\leq c^{\frac{1}{p_1}}+(\mathbb{E}\|X_2(0)\|^{(q-1)p_1})^{\frac{1}{p_1}}+\eta^{q-1}\Big(M^{q-1}(\mathbb{E}\|X_2(0)\|^{(q-1)p_1\gamma})^{\frac{1}{p_1}}+B^{(q-1)}\Big)+\Big(\frac{\eta}{\beta}\Big)^{\frac{q-1}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1,d}\\&\quad+(j+1)^{\frac{1}{p_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{p_1}}+(c\eta)^{\frac{1}{p_1}}+(2\eta(b+m))^{\frac{(q-1)}{2}}+2^{\frac{(q-1)}{2}}(\eta B)^{(q-1)}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1,d}\\&\quad+\eta^{q-1}M^{q-1}\Big((2\eta(b+m))^{\frac{(q-1)\gamma}{2}}+2^{\frac{(q-1)\gamma}{2}}(\eta B)^{(q-1)\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)\gamma}{\alpha}}l^{\frac{1}{p_1}}_{\alpha,(q-1)p_1\gamma,d}\Big)\Big)\\&=Q_1(\eta)+(j+1)^{\frac{1}{p_1}}P_1(\eta),\end{aligned}$$

Here, we have used Lemma S11 in the last inequality. Now, consider the following quantity

$$\Big[\mathbb{E}\|c_\alpha(\nabla f(X_1(s))-\nabla f(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}}\leq\Big[\mathbb{E}\Big(M\|X_1(s)-X_2(j\eta)\|^{\gamma}\Big)^{q_1}\Big]^{\frac{1}{q_1}}$$

$$\le\Big[\mathbb{E}\Big(M\|X_1(s)\|^\gamma+M\|X_2(j\eta)\|^\gamma\Big)^{q_1}\Big]^{\frac{1}{q_1}}$$
$$\le\Big[\mathbb{E}\Big(M^{q_1}\|X_1(s)\|^{\gamma q_1}\Big)\Big]^{\frac{1}{q_1}}+\Big[\mathbb{E}\Big(M^{q_1}\|X_2(j\eta)\|^{\gamma q_1}\Big)\Big]^{\frac{1}{q_1}},$$

where we have used Assumption **H**2, Lemma S11 and Minkowski's inequality. By Lemma S4 and Lemma S7, we have

$$\Big[\mathbb{E}\|c_\alpha\nabla f(X_1(s))-c_\alpha\nabla f(X_2(j\eta))\|^{q_1}\Big]^{\frac{1}{q_1}}\le M\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma+\Big[M^{q_1}(\mathbb{E}\|X_2(0)\|^{\gamma q_1})$$
$$+M^{q_1}j\Big((2\eta(b+m))^{\frac{\gamma q_1}{2}}+2^{\frac{\gamma q_1}{2}}(\eta B)^{\gamma q_1}+\Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma q_1}{\alpha}}l_{\alpha,\gamma q_1,d}\Big)\Big]^{\frac{1}{q_1}}.$$

By using Lemma S11 and the inequality $j<j+1$, we have

$$\Big[\mathbb{E}\|c_\alpha\nabla f(X_1(s))-c_\alpha\nabla f(X_2(j\eta))\|^{q_1}\Big]^{\frac{1}{q_1}}\le M\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma+M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}}$$
$$+M(j+1)^{\frac{1}{q_1}}\Big((2\eta(b+m))^{\frac{\gamma}{2}}+2^{\frac{\gamma}{2}}(\eta B)^{\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}}l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big).$$

We note that $s<(j+1)\eta$ and $\gamma<\frac{1}{q_1}$ (from the assumptions). Hence,

$$\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma\le\Big(c\Big((j+1)\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{\frac{1}{q_1}}$$
$$\le(j+1)^{\frac{1}{q_1}}\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)\Big)^{\frac{1}{q_1}}+c^{\frac{1}{q_1}},$$

where the last inequality is an application of Lemma S11. By replacing this inequality into the previous one and rearranging the terms, we have

$$\Big[\mathbb{E}\|c_\alpha\nabla f(X_1(s))-c_\alpha\nabla f(X_2(j\eta))\|^{q_1}\Big]^{\frac{1}{q_1}}\le M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}}+Mc^{\frac{1}{q_1}}+M(j+1)^{\frac{1}{q_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)\Big)^{\frac{1}{q_1}}$$
$$+(2\eta(b+m))^{\frac{\gamma}{2}}+2^{\frac{\gamma}{2}}(\eta B)^{\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}}l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big)$$
$$\le M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}}+Mc^{\frac{1}{q_1}}+M(j+1)^{\frac{1}{q_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{q_1}}$$
$$+(c\eta)^{\frac{1}{q_1}}+(2\eta(b+m))^{\frac{\gamma}{2}}+2^{\frac{\gamma}{2}}(\eta B)^{\gamma}+\Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}}l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big)$$
$$=Q_2+(j+1)^{\frac{1}{q_1}}P_2(\eta).$$

Here, we have used Lemma S11 in the last inequality. By combining the above inequalities, we get

$$\mathbb{E}\Big[\int_0^t q\,\|X_1(s)-X_2(s))\|^{q-2}\langle X_1(s)-X_2(s),b_1(X_1(s-),\alpha)-b_2(X_2(s-),\alpha)\rangle\mathrm{d}s\Big]$$
$$\le\sum_{j=0}^{k-1}q\eta\Big((j+1)P_1(\eta)P_2(\eta)+(j+1)^{\frac{1}{p_1}}P_1(\eta)Q_2+(j+1)^{\frac{1}{q_1}}P_2(\eta)Q_1(\eta)+Q_1(\eta)Q_2\Big)$$
$$\le q\eta\Big(k^2P_1(\eta)P_2(\eta)+k^{1+1/p_1}P_1(\eta)Q_2+k^{1+1/q_1}P_2(\eta)Q_1(\eta)+kQ_1(\eta)Q_2\Big).$$

The final conclusion follows from this inequality. □

S5.1. Proof of Corollary 1

Proof. In order to get the results from the bound obtained by Theorem S1, we take the max power of k and the min power of η among the terms containing k and η but not containing β. For the terms containing β, we take the max power of k, min power of η, min power of $1/\beta$ and max power of d. We get

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{2t}) \leq C(k^2\eta + k^2\eta^{1+\min\{\gamma, q-1\}/\alpha}\beta^{-(q-1)\gamma/\alpha}d).$$

Since $\gamma < 1/p = (q-1)/q < q-1$, we finally obtain

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{2t}) \leq C(k^2\eta + k^2\eta^{1+\gamma/\alpha}\beta^{-(q-1)\gamma/\alpha}d).$$

□

S5.2. Proof of Corollary 2

Proof. The proof starts from the bound established in Corollary S2 then, follows the same lines of the proof of Corollary 1. □

S6. Proof of Theorem 2

Proof. We have the decomposition:

$$\begin{aligned}\mathbb{E}[f(W^k)] - f^* =& \mathbb{E}[f(X_2(k\eta))] - f^* \\ =& (\mathbb{E}[f(X_2(k\eta))] - \mathbb{E}[f(X_1(k\eta))]) + (\mathbb{E}[f(X_1(k\eta))] - \mathbb{E}[f(X_3(k\eta))]) + (\mathbb{E}[f(X_3(k\eta))] - \mathbb{E}[f(\hat{W}))]) \\ &+ (\mathbb{E}[f(\hat{W}))] - f^*).\end{aligned}$$

By Corollary 2, Corollary 4, Lemma 4 and Lemma 5, there exists a constant C' independent of k, η and β such that

$$\begin{aligned}\mathbb{E}[f(W^k)] - f^* \leq& C'\Big(k^{1+\frac{1}{q}}\eta^{\frac{1}{q}} + k^{1+\frac{1}{q}}\eta^{\frac{1}{q}+\frac{\gamma}{\alpha q}}\beta^{-\frac{(q-1)\gamma}{\alpha q}}d + k^{\gamma+\frac{\gamma+q}{q}}\eta^{\gamma+\frac{1}{q}}\beta^{-\frac{\gamma}{\alpha}}d + k^{\gamma+\frac{\gamma+q}{q}}\eta^{\frac{1}{q}} \\ &+ \beta\frac{b+d/\beta}{m}\exp(-\lambda_*\beta^{-1}t)\Big) + \frac{\beta^{-\gamma-1}Mc_\alpha^{-1}}{1+\gamma} + \beta^{-1}\log\Big(\frac{(2e(b+d/\beta))^{d/2}\Gamma(d/2+1)\beta^d}{(dm)^{d/2}}\Big).\end{aligned}$$

Here, we note that $k\eta = t$. then by taking the largest power of k, smallest powers of η and β^{-1} among the terms containing all of three parameters k, η and β, there exist a constant C satisfying the following inequality:

$$\begin{aligned}\mathbb{E}[f(W^k)] - f^* \leq& C\Big(k^{1+\max\{\frac{1}{q}, \gamma+\frac{\gamma}{q}\}}\eta^{\frac{1}{q}} + k^{1+\max\{\frac{1}{q}, \gamma+\frac{\gamma}{q}\}}\eta^{\frac{1}{q}+\frac{\gamma}{\alpha q}}\beta^{-\frac{(q-1)\gamma}{\alpha q}}d + \beta\frac{b+d/\beta}{m}\exp(-\lambda_*\beta^{-1}k\eta)\Big) \\ &+ \frac{\beta^{-\gamma-1}Mc_\alpha^{-1}}{1+\gamma} + \beta^{-1}\log\Big(\frac{(2e(b+d/\beta))^{d/2}\Gamma(d/2+1)\beta^d}{(dm)^{d/2}}\Big).\end{aligned}$$

□

S6.1. Discussion on smoothness assumptions

Let us recall the four constraints given in **H**4:

$$\begin{gathered}(1/p + 1/q) = (1/p_1 + 1/q_1) = 1 \\ \gamma p < 1, \quad \gamma q_1 < 1, \quad (q-1)p_1 < 1.\end{gathered}$$

We will refer to these conditions as the *first*, *second*, *third*, and *fourth* conditions, respectively. Our aim is to find a condition on γ (more precisely, the maximum value of γ) such that there exist $p, q, p_1, q_1 > 0$ satisfying these four conditions.

First, suppose that $p > q_1$. Then, the maximum value of γ is decided by the second constraint. Since we want γ to be as large as possible, it is natural to choose a smaller p. We can observe that, as we decrease p, due to the first and the fourth constraints, the value of q_1 needs to be increased. If we continue decreasing p, then q_1 continues to be increased and soon becomes strictly greater than p. At this moment, the maximum value of γ is decided by the third constraint, not by the second constraint anymore, and from this point on, it is more plausible to decrease q_1.

By this intuition, it is reasonable to choose p to be equal to q_1, which implies that $p_1 = q$. Accordingly, the fourth constraint becomes: $(q-1)q < 1$. By noting that $q > 1$, solving this constraint gives $1 < q < (1+\sqrt{5})/2$. Then by the first constraint, we have $p > (3+\sqrt{5})/2$, and the second constraint gives $\gamma < 1/p < (3-\sqrt{5})/2$.

This upper bound for γ is a number between 0.38 and 0.39 and tells us that there exist p, q, p_1, q_1 satisfying the four constraints if and only if $0 \leq \gamma < (3-\sqrt{5})/2$.

Let us take a closer look at Theorem 2. Since $\gamma(q+1) < (3-\sqrt{5})(3+\sqrt{5})/4 = 1$, we have $\gamma + \gamma/q = \gamma(q+1)/q < 1/q$ Hence,

$$1 + \max\{1/q, \gamma + \gamma/q\} = 1 + 1/q.$$

Let ε_1 and ε_2 be positive numbers such that

$$\begin{aligned} &1/q - \varepsilon_1 = 2/(1+\sqrt{5}) = (\sqrt{5}-1)/2, \\ &\gamma + \varepsilon_2 = (3-\sqrt{5})/2. \end{aligned}$$

then, if $q = p_1$ is approximately equal to $(1+\sqrt{5})/2$ and γ is approximately equal to $(3-\sqrt{5})/2$, we imply that ε_1 and ε_2 become very small and

$$\begin{aligned} &1/q \approx (\sqrt{5}-1)/2, \\ &1/q + \gamma/(\alpha q) \approx (\sqrt{5}-1)/2 + (\sqrt{5}-2)/\alpha, \\ &(q-1)\gamma/(q\alpha) \approx (7-3\sqrt{5})/(2\alpha). \end{aligned}$$

For example, the values $\alpha = 1.65, \gamma = 0.38, p = q_1 = 2.63, q = p_1 = q/(q-1) \approx 1.613$ satisfy Assumption **H**4. Hence, the bound in Theorem 2 can be expressed as follows:

Corollary S1. *Under conditions **H**1-**H**7, for* $\alpha = 1.65, \gamma = 0.38, p = q_1 = 2.63, q = p_1 = q/(q-1) \approx 1.613$ *and for* $0 < \eta < \frac{m}{M^2}$*, there exists a positive constant* C *independent of* k *and* η *such that the following bound holds:*

$$\begin{aligned} \mathbb{E}[f(W^k)] - f^* \leq & C\left\{k^{1.62}\eta^{0.61} + \frac{k^{1.62}\eta^{0.75}d}{\beta^{0.0875}} + \frac{\beta b + d}{m}\exp(-\frac{\lambda_* k\eta}{\beta})\right\} + \frac{Mc_\alpha^{-1}}{1.38\beta^{1.38}} \\ &+ \frac{1}{\beta}\log\frac{(2e(b+\frac{d}{\beta}))^{\frac{d}{2}}\Gamma(\frac{d}{2}+1)\beta^d}{(dm)^{\frac{d}{2}}}. \end{aligned}$$

Proof. The result is a direct consequence of Theorem 2. □

S7. Proof of Theorem 4

In this section, we precise the statement of Theorem 4 and provide the full proof.

Theorem S2. *We have the following estimate:*

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{3t}) \leq qt\Big(M(c^{q-1} + c_b^{q-1})(c^\gamma + c_b^\gamma)\Big(t\Big(\frac{d}{\beta^{1/\alpha}} + 1\Big) + 1\Big)^{q-1+\gamma} + L(c^{q-1} + c_b^{q-1})\Big(t\Big(\frac{d}{\beta^{1/\alpha}} + 1\Big) + 1\Big)^{q-1}\Big),$$

where c *and* c_b *are constants defined in Lemma S4 and Lemma S5.*

Proof. From Lemma 3, we have

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{3t}) = \mathbb{E}\Big[\int_0^t q\,\|X_1(s) - X_3(s))\|^{q-2}\langle X_1(s) - X_3(s), b_1(X_1(s-), \alpha) - b(X_3(s-), \alpha)\rangle \mathrm{d}s\Big]$$

$$
\begin{aligned}
&= \int_0^t q \, \|X_1(s) - X_3(s))\|^{q-2} \langle X_1(s) - X_3(s), b_1(X_1(s-), \alpha) - b(X_3(s-), \alpha)\rangle \mathrm{d}s \\
&\leq \mathbb{E}\Big[\int_0^t q \, \|X_1(s) - X_3(s)\|^{q-1} \|c_\alpha \nabla f(X_1(s)) + b(X_3(s), \alpha)\| \mathrm{d}s \Big] \\
&= q \int_0^t \mathbb{E}\Big[\|X_1(s) - X_3(s)\|^{q-1} \|c_\alpha \nabla f(X_1(s)) + b(X_3(s), \alpha)\| \Big] \mathrm{d}s \\
&\leq q \int_0^t \Big[\mathbb{E}\|X_1(s) - X_3(s)\|^{(q-1)p_1} \Big]^{\frac{1}{p_1}} \Big[\mathbb{E}\|c_\alpha \nabla f(X_1(s)) + b(X_3(s), \alpha)\|^{q_1} \Big]^{\frac{1}{q_1}} \mathrm{d}s,
\end{aligned}
$$

where we have used Cauchy-Schwarz inequality in the third line and Holder's inequality in the last line.

Since $(q-1)p_1 < 1$ by Assumption **H**4, using Lemma S11 twice, we have:

$$
\begin{aligned}
\Big(\mathbb{E}\|X_1(s) - X_3(s)\|^{(q-1)p_1} \Big)^{\frac{1}{p_1}} &\leq \Big(\mathbb{E}\|X_1(s)|^{(q-1)p_1} + \mathbb{E}\|X_3(s)\|^{(q-1)p_1} \Big)^{\frac{1}{p_1}} \\
&\leq \Big[\mathbb{E}\Big(\|X_1(s)\|^{(q-1)p_1} \Big) \Big]^{\frac{1}{p_1}} + \Big[\mathbb{E}\Big(\|X_3(s)\|^{(q-1)p_1} \Big) \Big]^{\frac{1}{p_1}}
\end{aligned}
$$

Then, by applying Lemma S4 and Lemma S5 we obtain:

$$
\Big(\mathbb{E}\|X_1(s) - X_3(s)\|^{(q-1)p_1} \Big)^{\frac{1}{p_1}} \leq \Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{q-1} + \Big(c_b\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{q-1}.
$$

Now, consider the following quantity

$$
\begin{aligned}
\Big[\mathbb{E}\|c_\alpha \nabla f(X_1(s)) + b(X_3(s), \alpha)\|^{q_1} \Big]^{\frac{1}{q_1}} &\leq \Big[\mathbb{E}\big(\|c_\alpha \nabla f(X_1(s)) - c_\alpha \nabla f(X_3(s))\| + \|c_\alpha \nabla f(X_3(s)) + b(X_3(s), \alpha)\| \big)^{q_1} \Big]^{\frac{1}{q_1}} \\
&\leq \Big[\mathbb{E}\big(M\|X_1(s) - X_3(s)\|^\gamma + L \big)^{q_1} \Big]^{\frac{1}{q_1}} \\
&\leq \Big[\mathbb{E}\Big(M\|X_1(s)\|^\gamma + M\|X_3(s)\|^\gamma + L \Big)^{q_1} \Big]^{\frac{1}{q_1}} \\
&\leq \Big[\mathbb{E}\Big(M^{q_1}\|X_1(s)\|^{\gamma q_1} \Big) \Big]^{\frac{1}{q_1}} + \Big[\mathbb{E}\Big(M^{q_1}\|X_3(s)\|^{\gamma q_1} \Big) \Big]^{\frac{1}{q_1}} + L,
\end{aligned}
$$

where we have used Assumption **H**2, Assumption **H**6, Lemma S11 and Minkowski's inequality. By Lemma S4 and Lemma S5, we have

$$
\Big[\mathbb{E}\|c_\alpha \nabla f(X_1(s)) + b(X_3(s), \alpha)\|^{q_1} \Big]^{\frac{1}{q_1}} \leq M\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{\gamma} + M\Big(c_b\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{\gamma} + L.
$$

By combining the above inequalities, we get

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^t q \, \|X_1(s) - X_3(s))\|^{q-2} \langle X_1(s) - X_3(s), b_1(X_1(s-), \alpha) - b(X_3(s-), \alpha)\rangle \mathrm{d}s \Big] \\
&\quad \leq q \int_0^t \Big(\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{q-1} + \Big(c_b\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{q-1} \Big) \Big(M\Big(c\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{\gamma} \\
&\qquad + M\Big(c_b\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big) \Big)^{\gamma} + L \Big) \mathrm{d}s \\
&\quad = q \int_0^t \Big(M(c^{q-1} + c_b^{q-1})(c^\gamma + c_b^\gamma)\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big)^{q-1+\gamma} + L(c^{q-1} + c_b^{q-1})\Big(s\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big)^{q-1} \Big) \mathrm{d}s \\
&\quad \leq qt\Big(M(c^{q-1} + c_b^{q-1})(c^\gamma + c_b^\gamma)\Big(t\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big)^{q-1+\gamma} + L(c^{q-1} + c_b^{q-1})\Big(t\Big(\frac{d}{\beta^{1/\alpha}} + 1 \Big) + 1 \Big)^{q-1} \Big).
\end{aligned}
$$

The final conclusion follows from this inequality. □

S7.1. Proof of Corollary 3

Proof. First, we replace t by $k\eta$. Then, by following the same lines of the proof of Corollary 1, we get

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{3t}) \leq C(k^{q+\gamma}\eta + k^{q+\gamma}\eta^q\beta^{-\frac{q-1}{\alpha}}d^{q-1+\gamma}).$$

By assumption **H**4, $q-1 < 1/p_1$ and $\gamma < 1/q_1$. It implies that $d^{q-1+\gamma} < d^{1/p_1+1/q_1} = d$. Hence, we have

$$\mathcal{W}_q^q(\mu_{1t}, \mu_{3t}) \leq C(k^{q+\gamma}\eta + k^{q+\gamma}\eta^q\beta^{-\frac{q-1}{\alpha}}d).$$

□

S7.2. Proof of Corollary 4

Proof. By Lemma 2, Lemma S4 and Lemma S5, we have

$$\begin{aligned} c_\alpha|\mathbb{E}[f(X_1(t))] - \mathbb{E}[f(X_3(t))]| \leq& \Big(M\big(\mathbb{E}\|X_1(t)\|^{\gamma p}\big)^{\frac{1}{p}} + M\big(\mathbb{E}\|X_3(t)\|^{\gamma p}\big)^{\frac{1}{p}} + B\Big)\mathcal{W}_q(\mu_{1t}, \mu_{3t}) \\ \leq& \Big(M\Big(c\Big(t\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma + M\Big(c_b\Big(t\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma + B\Big)\mathcal{W}_q(\mu_{1t}, \mu_{3t}). \end{aligned}$$

Then by Theorem 4, we have

$$\begin{aligned} &c_\alpha|\mathbb{E}[f(X_1(t))] - \mathbb{E}[f(X_3(t))]| \\ &\leq\Big(M\Big(c\Big(t\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma + M\Big(c_b\Big(t\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^\gamma + B\Big)\Bigg(qt\Big(M(c^{q-1}+c_b^{q-1})(c^\gamma+c_b^\gamma) \\ &\quad\Big(t\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)^{q-1+\gamma} + L(c^{q-1}+c_b^{q-1})\Big(t\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)^{q-1}\Big)\Bigg)^{\frac{1}{q}}. \end{aligned}$$

Applying Lemma S11 twice, we get

$$\begin{aligned} &c_\alpha|\mathbb{E}[f(X_1(t))]-\mathbb{E}[f(X_3(t))]| \\ &\leq\Big(M(c^\gamma+c_b^\gamma)\Big(\frac{t^\gamma d^\gamma}{\beta^{\gamma/\alpha}}+t^\gamma+1\Big)+B\Big)\Bigg((qt)^{1/q}\Big(M^{1/q}(c^{q-1}+c_b^{q-1})^{1/q}(c^\gamma+c_b^\gamma)^{1/q} \\ &\quad\Big(\frac{td}{\beta^{1/\alpha}}+t+1\Big)^{(q-1+\gamma)/q} + L^{1/q}(c^{q-1}+c_b^{q-1})^{1/q}\Big(\frac{td}{\beta^{1/\alpha}}+t+1\Big)^{(q-1)/q}\Big)\Bigg) \\ &\leq\Big(M(c^\gamma+c_b^\gamma)\Big(\frac{t^\gamma d^\gamma}{\beta^{\gamma/\alpha}}+t^\gamma+1\Big)+B\Big)\Bigg((qt)^{1/q}\Big(M^{1/q}(c^{q-1}+c_b^{q-1})^{1/q}(c^\gamma+c_b^\gamma)^{1/q}\Big(\frac{(td)^{(q-1+\gamma)/q}}{\beta^{(q-1+\gamma)/(q\alpha)}} \\ &\quad+t^{(q-1+\gamma)/q}+1\Big) + L^{1/q}(c^{q-1}+c_b^{q-1})^{1/q}\Big(\frac{(td)^{(q-1)/q}}{\beta^{(q-1)/(q\alpha)}}+t^{(q-1)/q}+1\Big)\Big)\Bigg). \end{aligned}$$

Now, by replacing $t = k\eta$ we find that, among the terms containing β, the largest power of d, the largest power of k and the smallest power of η are $\gamma + \frac{q-1+\gamma}{q}$, $\gamma + \frac{\gamma+q}{q}$ and $\gamma + \frac{1}{q}$, respectively. For the smallest power of β^{-1}, we need to compare the following quantities: γ/α, $(q-1+\gamma)/(q\alpha)$ and $(q-1)/(q\alpha)$.

It is obvious that $(q-1+\gamma)/(q\alpha) > (q-1)/(q\alpha)$. Next, from the relation $\gamma < 1/p = (q-1)/q$, we have $\gamma/\alpha < (q-1)/(q\alpha)$. Thus, the smallest power of β^{-1} is γ/α. Hence, we have the following bound:

$$c_\alpha|\mathbb{E}[f(X_1(t))] - \mathbb{E}[f(X_3(t))]| \leq C\Big(k^{\gamma+\frac{\gamma+q}{q}}\eta^{\gamma+\frac{1}{q}}\beta^{-\frac{\gamma}{\alpha}}d^{\gamma+\frac{q-1+\gamma}{q}} + k^{\gamma+\frac{\gamma+q}{q}}\eta^{\frac{1}{q}}\Big),$$

for some constant $C > 0$. For the power of d, using that $\gamma < 1/p$, $q - 1 < 1/p_1$ and $\gamma < 1/q_1$ we have

$$\begin{aligned}\gamma + \frac{q-1+\gamma}{q} \leq& 1/p + \frac{1/p_1 + 1/q_1}{q} \\ =& 1/p + 1/q \\ =& 1.\end{aligned}$$

Finally, we have

$$c_\alpha |\mathbb{E}[f(X_1(t))] - \mathbb{E}[f(X_3(t))]| \leq C\Big(k^{\gamma+\frac{\gamma+q}{q}}\eta^{\gamma+\frac{1}{q}}\beta^{-\frac{\gamma}{\alpha}}d + k^{\gamma+\frac{\gamma+q}{q}}\eta^{\frac{1}{q}}\Big).$$

□

S8. Proof of Lemma 4

Proof. By Lemma 2, we have

$$c_\alpha |\mathbb{E}[f(X_3(t))] - \mathbb{E}[f(\hat{W})]| \leq \Big(M\big(\mathbb{E}\|X_3(t)\|^{\gamma p}\big)^{\frac{1}{p}} + M\big(\mathbb{E}\|\hat{W}\|^{\gamma p}\big)^{\frac{1}{p}} + B\Big)\mathcal{W}_q(\mu_{3t}, \pi).$$

Assumption **H**7 says that $\mathbb{E}\|\hat{W}\|^{\gamma p}$ is bounded by a constant depending on b, m and β. In addition, by Proposition 1, $\lim_{t\to\infty} \mathcal{W}_{\gamma p}(\mu_{3t}, \pi) = 0$, and by Theorem 7.12 in (Villani, 2003), it follows that

$$\lim_{t\to\infty} \mathbb{E}\|X_3(t)\|^{\gamma p} = \mathbb{E}\|\hat{W}\|^{\gamma p}.$$

Thus, $\mathbb{E}\|X_3(t)\|^{\gamma p}$ is bounded by a constant independent of t. Finally, since $q < \alpha$, by Proposition 1 again, $\mathcal{W}_q(\mu_{3t}, \pi) \leq C\beta e^{-\lambda_* t/\beta}$. Hence, using the bound in Assumption **H**7, there exists constant C such that

$$|\mathbb{E}[f(X_3(t))] - \mathbb{E}[f(\hat{W})]| \leq C\beta\frac{b + d/\beta}{m}\exp(-\lambda_*\beta^{-1}t).$$

□

S9. Proof of Lemma 5

Proof. The proof is adapted from (Raginsky et al., 2017), Section 3.5. First, we have the decomposition:

$$\begin{aligned}\mathbb{E}[f(\hat{W})] =& \int_{\mathbb{R}^d} f(w)\frac{\exp(-\beta f(w))}{\int_{\mathbb{R}^d}\exp(-\beta f(v))\mathrm{d}v}\mathrm{d}w \\ =& \frac{1}{\beta}\Big(-\int_{\mathbb{R}^d}\frac{\exp(-\beta f(w))}{\int_{\mathbb{R}^d}\exp(-\beta f(v))\mathrm{d}v}\log\frac{\exp(-\beta f(w))}{\int_{\mathbb{R}^d}\exp(-\beta f(v))\mathrm{d}v}\mathrm{d}w - \log\int_{\mathbb{R}^d}\exp(-\beta f(v))\mathrm{d}v\Big).\end{aligned}$$

The first term in the parentheses is the differential entropy of the probability density of $\hat{W}$, which has a finite second moment (due to Assumption **H**7). Hence, it is upper-bounded by the differential entropy of a Gaussian density with the same second moment:

$$-\int_{\mathbb{R}^d}\frac{\exp(-\beta f(w))}{\int_{\mathbb{R}^d}\exp(-\beta f(v))\mathrm{d}v}\log\frac{\exp(-\beta f(w))}{\int_{\mathbb{R}^d}\exp(-\beta f(v))\mathrm{d}v}\mathrm{d}w \leq \frac{d}{2}\log\Big(\frac{2\boldsymbol{\pi} e(b + d/\beta)}{dm}\Big).$$

By Lemma S3, we have

$$-\log\int_{\mathbb{R}^d}\exp(-\beta f(w))\mathrm{d}w \leq \beta f(w^*) + \frac{\beta^{-\gamma}Mc_\alpha^{-1}}{1+\gamma} - \log\Big(\frac{\boldsymbol{\pi}^{d/2}\beta^{-d}}{\Gamma(d/2+1)}\Big).$$

Then, it implies that

$$\begin{aligned}\mathbb{E}[f(\hat{W})] \leq & \frac{d\beta^{-1}}{2}\log\Big(\frac{2\boldsymbol{\pi} e(b+d/\beta)}{dm}\Big) + f(w^*) + \frac{\beta^{-\gamma-1}Mc_\alpha^{-1}}{1+\gamma} - \beta^{-1}\log\Big(\frac{\boldsymbol{\pi}^{d/2}\beta^{-d}}{\Gamma(d/2+1)}\Big) \\ = & f(w^*) + \frac{\beta^{-\gamma-1}Mc_\alpha^{-1}}{1+\gamma} + \beta^{-1}\log\Big(\frac{(2e(b+d/\beta))^{d/2}\Gamma(d/2+1)\beta^d}{(dm)^{d/2}}\Big),\end{aligned}$$

which leads to desired result. □

S10. Proof of Corollary 5

Proof. By triangular inequality, we have

$$\mathcal{W}_q(\mu_{2t},\pi) \leq \mathcal{W}_q(\mu_{2t},\mu_{1t}) + \mathcal{W}_q(\mu_{1t},\mu_{3t}) + \mathcal{W}_q(\mu_{3t},\pi).$$

Then, using Corollary 1, Corollary 3 and Proposition 1, we get

$$\begin{aligned}\mathcal{W}_q(\mu_{2t},\pi) \leq & C\Big((k^2\eta + k^2\eta^{1+\gamma/\alpha}\beta^{-\gamma(q-1)/\alpha}d)^{1/q} + (k^{q+\gamma}\eta + k^{q+\gamma}\eta^q\beta^{-(q-1)/\alpha}d)^{1/q} + \beta e^{-\lambda_* k\eta/\beta}\Big) \\ \leq & C\Big(k^{2/q}\eta^{1/q} + k^{2/q}\eta^{1/q+\gamma/(q\alpha)}\beta^{-\gamma(q-1)/(q\alpha)}d^{1/q} + k^{1+\gamma/q}\eta^{1/q} + k^{1+\gamma/q}\eta\beta^{-(q-1)/(q\alpha)}d^{1/q} \\ & + \beta e^{-\lambda_* k\eta/\beta}\Big),\end{aligned}$$

where, we have used Lemma S11 for the second inequality. Then, similar to the proof of Corollary 1, we obtain

$$\mathcal{W}_q(\mu_{2t},\pi) \leq C\Big(k^{\max\{2,q+\gamma\}/q}\eta^{1/q} + k^{\max\{2,q+\gamma\}/q}\eta^{1/q+\gamma/(q\alpha)}\beta^{-\gamma(q-1)/(q\alpha)}d^{1/q} + \beta e^{-\lambda_* k\eta/\beta}\Big).$$

□

S11. Proof of Theorem 5

Proof. Since each function $x \mapsto f^{(i)}(x)$ satisfies assumptions **H**1-**H**7, it is easy to check that f_k also satisfies these assumptions (with the same constants and the same parameters) for all k. Then by repeating exactly the same lines as in the proof of Lemma S7, we obtain the same estimates for the moments of X_2. Now by following the same steps as in the proof of Theorem S1, we first have

$$\mathcal{W}_q^q(\mu_{1t},\mu_{2t}) \leq q\sum_{j=0}^{k-1}\int_{j\eta}^{(j+1)\eta}\Big[\mathbb{E}\|X_1(s)-X_2(s)\|^{(q-1)p_1}\Big]^{\frac{1}{p_1}}\Big[\mathbb{E}\|c_\alpha(\nabla f(X_1(s)) - \nabla f_k(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}}\mathrm{d}s,$$

then

$$\begin{aligned}&\Big(\mathbb{E}\|X_1(s)-X_2(s)\|^{(q-1)p_1}\Big)^{\frac{1}{p_1}} \\ &\leq c^{\frac{1}{p_1}} + (\mathbb{E}\|X_2(0)\|^{(q-1)p_1})^{\frac{1}{p_1}} + \eta^{q-1}\Big(M^{q-1}(\mathbb{E}\|X_2(0)\|^{(q-1)p_1\gamma})^{\frac{1}{p_1}} + B^{(q-1)}\Big) + \Big(\frac{\eta}{\beta}\Big)^{\frac{q-1}{\alpha}} l_{\alpha,(q-1)p_1,d}^{\frac{1}{p_1}} \\ &\quad + (j+1)^{\frac{1}{p_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{p_1}} + (c\eta)^{\frac{1}{p_1}} + (2\eta(b+m))^{\frac{(q-1)}{2}} + 2^{\frac{(q-1)}{2}}(\eta B)^{(q-1)} + \Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)}{\alpha}} l_{\alpha,(q-1)p_1,d}^{\frac{1}{p_1}} \\ &\quad + \eta^{q-1}M^{q-1}\Big((2\eta(b+m))^{\frac{(q-1)\gamma}{2}} + 2^{\frac{(q-1)\gamma}{2}}(\eta B)^{(q-1)\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{(q-1)\gamma}{\alpha}} l_{\alpha,(q-1)p_1\gamma,d}^{\frac{1}{p_1}}\Big)\Big) \\ &= Q_1(\eta) + (j+1)^{\frac{1}{p_1}}P_1(\eta),\end{aligned}$$

where $P_1(\eta)$ and $Q_1(\eta)$ are defined in Theorem S1. Now, by Minkowski's inequality, we have

$$\Big[\mathbb{E}\|c_\alpha(\nabla f(X_1(s)) - \nabla f_k(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}} = \Big[\mathbb{E}\|c_\alpha(\nabla f(X_1(s)) - \nabla f(X_2(j\eta)) + \nabla f(X_2(j\eta))$$

$$
\begin{aligned}
&- \nabla f_k(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}} \\
\le&\Big[\mathbb{E}\|c_\alpha(\nabla f(X_1(s)) - \nabla f(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}} + \Big[\mathbb{E}\|c_\alpha(\nabla f(X_2(j\eta)) \\
&- \nabla f_k(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}}.
\end{aligned}
$$

As in the proof of Theorem S1, the following inequality holds:

$$
\begin{aligned}
\Big[\mathbb{E}\|c_\alpha \nabla f(X_1(s)) - c_\alpha \nabla f(X_2(j\eta))\|^{q_1}\Big]^{\frac{1}{q_1}} \le& M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}} + Mc^{\frac{1}{q_1}} + M(j+1)^{\frac{1}{q_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{q_1}} \\
&+ (c\eta)^{\frac{1}{q_1}} + (2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^\gamma + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big) \\
=& Q_2 + (j+1)^{\frac{1}{q_1}} P_2(\eta),
\end{aligned}
$$

where $P_2(\eta)$ and Q_2 are defined in Theorem S1. Using the additional assumption, Lemma S7, and Lemma S11, we get

$$
\begin{aligned}
\Big[\mathbb{E}\|c_\alpha(\nabla f(X_2(j\eta)) - \nabla f_k(X_2(j\eta)))\|^{q_1}\Big]^{\frac{1}{q_1}} \le& \delta\Big[\mathbb{E}\Big(M^{q_1}\|X_2(j\eta)\|^{\gamma q_1}\Big)\Big]^{\frac{1}{q_1}} \\
\le& \delta\Big[M^{q_1}(\mathbb{E}\|X_2(0)\|^{\gamma q_1}) + M^{q_1} j\Big((2\eta(b+m))^{\frac{\gamma q_1}{2}} + 2^{\frac{\gamma q_1}{2}}(\eta B)^{\gamma q_1} \\
&+ \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma q_1}{\alpha}} l_{\alpha,\gamma q_1,d}\Big)\Big]^{\frac{1}{q_1}} \\
\le& \delta M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}} + \delta M(j+1)^{\frac{1}{q_1}}\Big((2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^\gamma \\
&+ \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big).
\end{aligned}
$$

By combining the two above inequalities, we obtain

$$
\begin{aligned}
\Big[\mathbb{E}\|c_\alpha \nabla f(X_1(s)) - c_\alpha \nabla f(X_2(j\eta))\|^{q_1}\Big]^{\frac{1}{q_1}} \le& (1+\delta)M(\mathbb{E}\|X_2(0)\|^{\gamma q_1})^{\frac{1}{q_1}} + Mc^{\frac{1}{q_1}} + M(j+1)^{\frac{1}{q_1}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{q_1}} \\
&+ (c\eta)^{\frac{1}{q_1}} + (1+\delta)(2\eta(b+m))^{\frac{\gamma}{2}} + (1+\delta)2^{\frac{\gamma}{2}}(\eta B)^\gamma \\
&+ (1+\delta)\Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma q_1,d}^{\frac{1}{q_1}}\Big) \\
=& Q_2' + (j+1)^{\frac{1}{q_1}} P_2'(\eta).
\end{aligned}
$$

Finally, we have

$$
\mathcal{W}_q^q(\mu_{1t}, \mu_{2t}) \le q\eta\Big(k^2 P_1(\eta)P_2'(\eta) + k^{1+1/p_1} P_1(\eta)Q_2' + k^{1+1/q_1} P_2'(\eta)Q_1(\eta) + kQ_1(\eta)Q_2'\Big).
$$

By considering the additional term δ, we arrive at the following bound:

$$
\mathcal{W}_q^q(\mu_{1t}, \mu_{2t}) \le C(1+\delta)(k^2\eta + k^2\eta^{1+\gamma/\alpha}\beta^{-\gamma(q-1)/\alpha} d).
$$

□

S12. Proof of Corollary 6

Proof. By Lemma 2,

$$
c_\alpha\Big|\mathbb{E}[f(X_1(k\eta))] - \mathbb{E}[f(X_2(k\eta))]\Big| \le \Big(M\Big(\mathbb{E}_{\mathbf{P}}\|X_1(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + M\Big(\mathbb{E}_{\mathbf{P}}\|X_2(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + B\Big)\mathcal{W}_q(\mu_{1t}, \mu_{2t}).
$$

Then, by following the same proof as in Corollary S2, Corollary 1 and using Theorem 5, we get

$$
c_\alpha\Big|\mathbb{E}[f(X_1(k\eta))] - \mathbb{E}[f(X_2(k\eta))]\Big| \le C(1+\delta)\Big(k^{1+\frac{1}{q}}\eta^{\frac{1}{q}} + k^{1+\frac{1}{q}}\eta^{\frac{1}{q}+\frac{\gamma}{\alpha q}}\beta^{-\frac{(q-1)\gamma}{\alpha q}} d\Big).
$$

□

S13. Technical Results

Corollary S2. *Along with $P_1(\eta), P_2(\eta), Q_1(\eta), Q_2$ in Lemma S1, we define, in addition, the following quantities:*

$$P_3(\eta) \triangleq M\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{p}} + (c\eta)^{\frac{1}{p}} + (2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^{\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma p,d}^{\frac{1}{p}}\Big)$$
$$Q_3 \triangleq M(\mathbb{E}\|X_2(0)\|^{\gamma p})^{\frac{1}{p}} + Mc^{\frac{1}{p}} + B.$$

For $0 < \eta < \frac{m}{M^2}$, we have the following bound:

$$\begin{aligned}
c_\alpha\big|\mathbb{E}[f(X_1(k\eta))]-\mathbb{E}[f(X_2(k\eta))]\big| \\
\le&(q\eta)^{\frac{1}{q}}\Big(k^{1+\frac{1}{q}}(P_1(\eta)P_2(\eta))^{\frac{1}{q}}P_3(\eta) + k^{1+\frac{1}{qp_1}}(P_1(\eta)Q_2)^{\frac{1}{q}}P_3(\eta) + k^{1+\frac{1}{qq_1}}(P_2(\eta)Q_1(\eta))^{\frac{1}{q}}P_3(\eta) \\
&+ k(Q_1(\eta)Q_2)^{\frac{1}{q}}P_3(\eta) + k^{\frac{2}{q}}(P_1(\eta)P_2(\eta))^{\frac{1}{q}}Q_3 + k^{\frac{1}{q}+\frac{1}{qp_1}}(P_1(\eta)Q_2)^{\frac{1}{q}}Q_3 \\
&+ k^{\frac{1}{q}+\frac{1}{qq_1}}(P_2(\eta)Q_1(\eta))^{\frac{1}{q}}Q_3 + k^{\frac{1}{q}}(Q_1(\eta)Q_2)^{\frac{1}{q}}Q_3\Big).
\end{aligned}$$

Proof. By Lemma 2,

$$c_\alpha\big|\mathbb{E}[f(X_1(k\eta))] - \mathbb{E}[f(X_2(k\eta))]\big| \le \Big(M\Big(\mathbb{E}_{\mathbf{P}}\|X_1(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + M\Big(\mathbb{E}_{\mathbf{P}}\|X_2(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + B\Big)\mathcal{W}_q(\mu_{1t}, \mu_{2t}).$$

Using Lemma S4 and Lemma S8, we have

$$\begin{aligned}
\Big(M\Big(\mathbb{E}_{\mathbf{P}}\|X_1(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + M\Big(\mathbb{E}_{\mathbf{P}}\|X_2(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + B\Big) \le& M\Big(c\Big(k\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{\gamma} + M\Big[(\mathbb{E}\|X_2(0)\|^{\gamma p}) \\
&+ k\Big((2\eta(b+m))^{\frac{\gamma p}{2}} + 2^{\frac{\gamma p}{2}}(\eta B)^{\gamma p} + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma p}{\alpha}} l_{\alpha,\gamma p,d}\Big)\Big]^{\frac{1}{p}} \\
&+ B.
\end{aligned}$$

By using Lemma S11, we obtain

$$\begin{aligned}
\Big(M\Big(\mathbb{E}_{\mathbf{P}}\|X_1(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + M\Big(\mathbb{E}_{\mathbf{P}}\|X_2(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + B\Big) \le& M\Big(c\Big(k\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{\gamma} + M(\mathbb{E}\|X_2(0)\|^{\gamma p})^{\frac{1}{p}} \\
&+ Mk^{\frac{1}{p}}\Big((2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^{\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma p,d}^{\frac{1}{p}}\Big) + B.
\end{aligned}$$

We note that $\gamma < \frac{1}{p}$. Hence,

$$\begin{aligned}
\Big(c\Big(k\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{\gamma} \le& \Big(c\Big(k\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)+1\Big)\Big)^{\frac{1}{p}} \\
\le& k^{\frac{1}{p}}\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)\Big)^{\frac{1}{p}} + c^{\frac{1}{p}},
\end{aligned}$$

where the last inequality is an application of Lemma S11. By replacing this inequality into the previous one and rearranging the terms, we have

$$\begin{aligned}
\Big(M\Big(\mathbb{E}_{\mathbf{P}}\|X_1(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + M\Big(\mathbb{E}_{\mathbf{P}}\|X_2(k\eta)\|^{\gamma p}\Big)^{\frac{1}{p}} + B\Big) \le& M(\mathbb{E}\|X_2(0)\|^{\gamma p})^{\frac{1}{p}} + Mc^{\frac{1}{p}} + B + Mk^{\frac{1}{p}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}+1\Big)\Big)^{\frac{1}{p}} \\
&+ (2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^{\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma p,d}^{\frac{1}{p}}\Big)
\end{aligned}$$

$$\begin{aligned}&\leq M(\mathbb{E}\|X_2(0)\|^{\gamma p})^{\frac{1}{p}} + Mc^{\frac{1}{p}} + B + Mk^{\frac{1}{p}}\Big(\Big(c\eta\Big(\frac{d}{\beta^{1/\alpha}}\Big)\Big)^{\frac{1}{p}} \\ &\quad + (c\eta)^{\frac{1}{p}} + (2\eta(b+m))^{\frac{\gamma}{2}} + 2^{\frac{\gamma}{2}}(\eta B)^{\gamma} + \Big(\frac{\eta}{\beta}\Big)^{\frac{\gamma}{\alpha}} l_{\alpha,\gamma p,d}^{\frac{1}{p}}\Big) \\ &= Q_3 + k^{\frac{1}{p}} P_3(\eta).\end{aligned}$$

Here, we have used Lemma S11 in the last inequality. Next, by Lemma S1 and Lemma S11,

$$\begin{aligned}\mathcal{W}_q(\mu_{1t},\mu_{2t}) &\leq (q\eta)^{\frac{1}{q}}\Big(k^2 P_1(\eta)P_2(\eta) + k^{1+1/p_1}P_1(\eta)Q_2 + k^{1+1/q_1}P_2(\eta)Q_1(\eta) + kQ_1(\eta)Q_2\Big)^{\frac{1}{q}} \\ &\leq (q\eta)^{\frac{1}{q}}\Big(k^{\frac{2}{q}}(P_1(\eta)P_2(\eta))^{\frac{1}{q}} + k^{\frac{1}{q}+\frac{1}{qp_1}}(P_1(\eta)Q_2)^{\frac{1}{q}} + k^{\frac{1}{q}+\frac{1}{qq_1}}(P_2(\eta)Q_1(\eta))^{\frac{1}{q}} + k^{\frac{1}{q}}(Q_1(\eta)Q_2)^{\frac{1}{q}}\Big).\end{aligned}$$

By combining the above two inequalities, we get

$$\begin{aligned}&c_\alpha\big|\mathbb{E}[f(X_1(k\eta))] - \mathbb{E}[f(X_2(k\eta))]\big| \\ &\leq (q\eta)^{\frac{1}{q}}\Big(Q_3 + k^{\frac{1}{p}}P_3(\eta)\Big)\Big(k^{\frac{2}{q}}(P_1(\eta)P_2(\eta))^{\frac{1}{q}} + k^{\frac{1}{q}+\frac{1}{qp_1}}(P_1(\eta)Q_2)^{\frac{1}{q}} + k^{\frac{1}{q}+\frac{1}{qq_1}}(P_2(\eta)Q_1(\eta))^{\frac{1}{q}} + k^{\frac{1}{q}}(Q_1(\eta)Q_2)^{\frac{1}{q}}\Big) \\ &= (q\eta)^{\frac{1}{q}}\Big(k^{1+\frac{1}{q}}(P_1(\eta)P_2(\eta))^{\frac{1}{q}}P_3(\eta) + k^{1+\frac{1}{qp_1}}(P_1(\eta)Q_2)^{\frac{1}{q}}P_3(\eta) + k^{1+\frac{1}{qq_1}}(P_2(\eta)Q_1(\eta))^{\frac{1}{q}}P_3(\eta) \\ &\quad + k(Q_1(\eta)Q_2)^{\frac{1}{q}}P_3(\eta) + k^{\frac{2}{q}}(P_1(\eta)P_2(\eta))^{\frac{1}{q}}Q_3 + k^{\frac{1}{q}+\frac{1}{qp_1}}(P_1(\eta)Q_2)^{\frac{1}{q}}Q_3 + k^{\frac{1}{q}+\frac{1}{qq_1}}(P_2(\eta)Q_1(\eta))^{\frac{1}{q}}Q_3 \\ &\quad + k^{\frac{1}{q}}(Q_1(\eta)Q_2)^{\frac{1}{q}}Q_3\Big).\end{aligned}$$

□

The following lemma is an extension of Lemma 1.2.3 in (Nesterov, 2013) to functions with Hölder continuous gradients.

Lemma S2. *Under Assumption **H**2, the following inequality holds for any $x, y \in \mathbb{R}^d$:*

$$c_\alpha|f(x) - f(y) - \langle\nabla f(y), x-y\rangle| \leq \frac{M}{1+\gamma}\|x-y\|^{1+\gamma}.$$

Proof. Let $g(t) \triangleq c_\alpha f(y + t(x-y))$. Then, $g'(t) = c_\alpha\langle\nabla f(y+t(x-y)), x-y\rangle$ and $\int_0^1 g'(t)\mathrm{d}t = g(1) - g(0) = c_\alpha(f(x) - f(y))$. We have

$$\begin{aligned}c_\alpha|f(x) - f(y) - \langle\nabla f(y), x-y\rangle| &= \Big|\int_0^1 g'(t)\mathrm{d}t - c_\alpha\langle\nabla f(y), x-y\rangle\Big| \\ &= \Big|\int_0^1 c_\alpha\langle\nabla f(y+t(x-y)), x-y\rangle\mathrm{d}t - c_\alpha\langle\nabla f(y), x-y\rangle\Big| \\ &= \Big|\int_0^1 c_\alpha\langle\nabla f(y+t(x-y)) - \nabla f(y), x-y\rangle\mathrm{d}t\Big|.\end{aligned}$$

By Cauchy-Schwarz inequality and Assumption **H**2, we have

$$\begin{aligned}c_\alpha|f(x) - f(y) - \langle\nabla f(y), x-y\rangle| &\leq \int_0^1 c_\alpha\|\nabla f(y+t(x-y)) - \nabla f(y)\|\|x-y\|\mathrm{d}t \\ &\leq \int_0^1 Mt^\gamma\|x-y\|^\gamma\|x-y\|\mathrm{d}t \\ &= \frac{M}{1+\gamma}\|x-y\|^{1+\gamma}.\end{aligned}$$

□

Lemma S3. *The normalized factor of π is bounded below, i. e.,*

$$\log \int_{\mathbb{R}^d} \exp(-\beta f(w)) \mathrm{d}w \geq -\beta f(w^*) - \frac{\beta^{-\gamma} M c_\alpha^{-1}}{1+\gamma} + \log\Big(\frac{\boldsymbol{\pi}^{d/2}\beta^{-d}}{\Gamma(d/2+1)}\Big).$$

Proof. We start by writing:

$$\begin{aligned}\log \int_{\mathbb{R}^d} \exp(-\beta f(w)) \mathrm{d}w =& -\beta f(w^*) + \log \int_{\mathbb{R}^d} \exp\big(-\beta(f(w) - f(w^*))\big)\mathrm{d}w \\ \geq& -\beta f(w^*) + \log \int_{\mathbb{R}^d} \exp\Big(-\frac{\beta M c_\alpha^{-1}}{1+\gamma}\|w - w^*\|^{1+\gamma}\Big)\mathrm{d}w.\end{aligned}$$

Here, we used Lemma S2, with $\nabla f(w^*) = 0$. For the second term on the right hand side, we have

$$\begin{aligned}\int_{\mathbb{R}^d} \exp\Big(-\frac{\beta M c_\alpha^{-1}}{1+\gamma}\|w - w^*\|^{1+\gamma}\Big)\mathrm{d}w =& \int_{\|w\|\leq\beta^{-1}} \exp\Big(-\frac{\beta M c_\alpha^{-1}}{1+\gamma}\|w\|^{1+\gamma}\Big)\mathrm{d}w \\ &+ \int_{\|w\|\geq\beta^{-1}} \exp\Big(-\frac{\beta M c_\alpha^{-1}}{1+\gamma}\|w\|^{1+\gamma}\Big)\mathrm{d}w \\ \geq& \int_{\|w\|\leq\beta^{-1}} \exp\Big(-\frac{\beta M c_\alpha^{-1}}{1+\gamma}\beta^{-1-\gamma}\Big)\mathrm{d}w + 0 \\ =& \exp\Big(-\frac{\beta^{-\gamma} M c_\alpha^{-1}}{1+\gamma}\Big)\int_{\|w\|\leq\beta^{-1}} 1\mathrm{d}w \\ =& \exp\Big(-\frac{\beta^{-\gamma} M c_\alpha^{-1}}{1+\gamma}\Big)\frac{\boldsymbol{\pi}^{d/2}\beta^{-d}}{\Gamma(d/2+1)},\end{aligned}$$

where, Γ denotes the Gamma function and $\boldsymbol{\pi}$ denotes Archimedes' constant (here, it is not the invariant distribution). Hence,

$$\log \int_{\mathbb{R}^d} \exp\Big(-\frac{\beta M c_\alpha^{-1}}{1+\gamma}\|w - w^*\|^{1+\gamma}\Big)\mathrm{d}w \geq -\frac{\beta^{-\gamma} M c_\alpha^{-1}}{1+\gamma} + \log\Big(\frac{\boldsymbol{\pi}^{d/2}\beta^{-d}}{\Gamma(d/2+1)}\Big).$$

By combining the above inequalities, we have the desired result. □

Lemma S4. *For $\lambda \in (0,1)$, there exists a constant c depending on m, b, α, such that*

$$\mathbb{E}\Big(\|X_1(t)\|^\lambda\Big)^{\frac{1}{\lambda}} \leq c\Big(t(d\beta^{-1/\alpha}+1)+1\Big), \qquad \forall t > 0, \beta \geq 1, 1 < \alpha < 2.$$

Proof. We follow exactly the same proof as Lemma 7.1 in (Xie & Zhang, 2017), with some modifications. Let $h(x) \triangleq (1+\|x\|^2)^{1/2}$. By Itô's formula, we have $\mathrm{d}h(X_1(t)) =$

$$\Bigg(\langle b_1(X_1(t)), \nabla h(X_1(t))\rangle + \int_{\mathbb{R}^d}\Big(h(X_1(t) + \beta^{-1/\alpha}x) - h(X_1(t)) - \mathbb{I}_{\|x\|<1}\langle\beta^{-1/\alpha}x, \nabla h(X_1(t))\rangle\Big)\nu(\mathrm{d}x)\Bigg)\mathrm{d}t + \mathrm{d}M(t), \tag{S11}$$

where $M(t)$ is a local martingale. Noticing that $\partial_i h(x) = x_i(1+\|x\|^2)^{-1/2}/2$ and using Assumption **H**3, we have

$$\begin{aligned}\langle b_1(x), \nabla h(x)\rangle =& \langle b_1(x), x\rangle(1+\|x\|^2)^{-1/2}/2 \\ \leq& (-m\|x\|^{1+\gamma} + b)(1+\|x\|^2)^{-1/2}/2 \\ =& (-m(\|x\|^{1+\gamma}+1) + m + b)(1+\|x\|^2)^{-1/2}/2.\end{aligned}$$

Since $(\|x\|^2+1)^{(1+\gamma)/2} \leq (\|x\|^{1+\gamma}+1)$ by Lemma S11, it follows that

$$\langle b_1(x), \nabla h(x)\rangle \leq (-m(\|x\|^2+1)^{(1+\gamma)/2} + m + b)(1+\|x\|^2)^{-1/2}/2$$

$$
\begin{aligned}
&=(-m(\|x\|^2+1)^{\gamma/2}+(m+b)(1+\|x\|^2)^{-1/2})/2\\
&\le(-m(\|x\|^2+1)^{\gamma/2}+m+b)/2\\
&=(-mh(x)^{\gamma}+m+b)/2.
\end{aligned}
$$

On the other hand, observing that

$$
|h(x+y)-h(x)|\le\|y\|\int_0^1\|\nabla h(x+sy)\|\mathrm{d}s\le\|y\|/2,
$$

and

$$
h(x+y)-h(x)-\langle y,\nabla h(x)\rangle\le\|y\|^2/2,
$$

we have

$$
\begin{aligned}
\int_{\mathbb{R}^d}\Big(h(X_1(t)+x)-h(X_1(t))-\mathbb{I}_{\|x\|<1}\langle x,\nabla h(X_1(t))\rangle\Big)\nu(\mathrm{d}x)\le&\frac{1}{2\beta^{2/\alpha}}\int_{\|x\|<1}\|x\|^2\nu(\mathrm{d}x)+\frac{1}{2\beta^{1/\alpha}}\int_{\|x\|\ge1}\|x\|\nu(\mathrm{d}x)\\
\le&C\frac{d}{\beta^{1/\alpha}},
\end{aligned}
$$

where the last inequality is due to Lemma S10. By integrating (S11) and combining the above inequalities, we have

$$
\begin{aligned}
h(X_1(t))-h(X_1(0))\le&\int_0^t\Big((-mh(X_1(s))^{\gamma}+m+b)/2+C\frac{d}{\beta^{1/\alpha}}\Big)\mathrm{d}s+M(t)\\
\le&\int_0^t\Big((m+b)/2+C\frac{d}{\beta^{1/\alpha}}\Big)\mathrm{d}s+M(t).
\end{aligned}
$$

By Lemma 3.8 in (Xie & Zhang, 2017), for $\lambda\in(0,1)$,

$$
\mathbb{E}\Big(\sup_{s\in[0,t]}h(X_1(s))^{\lambda}\Big)\le c_\lambda\Big(\mathbb{E}h(X_1(0))+((m+b)/2+C\frac{d}{\beta^{1/\alpha}})t\Big)^{\lambda}.
$$

This leads to the conclusion since $h(x)\ge\|x\|$. □

Lemma S5. *For $\lambda\in(0,1)$, there exists a constant c_b depending on L,m,b,α, such that*

$$
\mathbb{E}\Big(\|X_3(t)\|^{\lambda}\Big)^{\frac{1}{\lambda}}\le c_b\Big(t(d\beta^{-1/\alpha}+1)+1\Big),\qquad\forall t>0,\beta\ge1,1<\alpha<2.
$$

Proof. The proof is similar to the proof of Lemma S4. □

Lemma S6. *Let X be a scalar symmetric α-stable distribution with $\alpha<2$, i. e. $X\sim\mathcal{S}\alpha\mathcal{S}(1)$ (see Definition 2), then, for $-1<\lambda<\alpha$,*

$$
\mathbb{E}(|X|^{\lambda})=\frac{2^{\lambda}\Gamma((1+\lambda)/2)\Gamma(1-\lambda/\alpha)}{\Gamma(1/2)\Gamma(1-\lambda/2)}.
$$

Proof. The proof follows from Theorem 3 in (Shanbhag & Sreehari, 1977) (see also equation (13) in (Matsui et al., 2016)). □

Corollary S3. *The quantity $l_{\alpha,\lambda,d}\triangleq\mathbb{E}\|L^{\alpha}(1)\|^{\lambda}$ is finite for $0\le\lambda<\alpha$. For details, we have*

(a) If $1<\lambda<\alpha$, then

$$
\mathbb{E}\|L^{\alpha}(1)\|^{\lambda}\le d^{\lambda}\Big(\frac{2^{\lambda}\Gamma((1+\lambda)/2)\Gamma(1-\lambda/\alpha)}{\Gamma(1/2)\Gamma(1-\lambda/2)}\Big).
$$

(b) If $0\le\lambda\le1$, then

$$
\mathbb{E}\|L^{\alpha}(1)\|^{\lambda}\le d\Big(\frac{2^{\lambda}\Gamma((1+\lambda)/2)\Gamma(1-\lambda/\alpha)}{\Gamma(1/2)\Gamma(1-\lambda/2)}\Big).
$$

Proof. Since $L^\alpha(1)$, by definition, is a d-dimensional vector whose components are i.i.d symmetric α-stable distributions $L_i^\alpha(1)$ for $i \in \{1, \dots, d\}$, we have

$$\|L^\alpha(1)\| \leq \sum_{i=1}^{d} |L_i^\alpha(1)|$$

(a) $1 < \lambda < \alpha$. By using Minkowski's inequality and Lemma S6,

$$\begin{aligned}
(\mathbb{E}\|L^\alpha(1)\|^\lambda)^{1/\lambda} \leq & \Big(\mathbb{E}\Big[\Big(\sum_{i=1}^{d} |L_i^\alpha(1)|\Big)^\lambda\Big]\Big)^{1/\lambda} \\
\leq & \sum_{i=1}^{d} (\mathbb{E}|L_i^\alpha(1)|^\lambda)^{1/\lambda} \\
= & d\Big(\frac{2^\lambda \Gamma((1+\lambda)/2)\Gamma(1-\lambda/\alpha)}{\Gamma(1/2)\Gamma(1-\lambda/2)}\Big)^{1/\lambda}.
\end{aligned}$$

Thus, we have

$$\mathbb{E}\|L^\alpha(1)\|^\lambda \leq d^\lambda \Big(\frac{2^\lambda \Gamma((1+\lambda)/2)\Gamma(1-\lambda/\alpha)}{\Gamma(1/2)\Gamma(1-\lambda/2)}\Big).$$

(b) $0 \leq \lambda \leq 1$. By using Lemma S11 and Lemma S6 ,

$$\begin{aligned}
\mathbb{E}\|L^\alpha(1)\|^\lambda \leq & \mathbb{E}\Big[\Big(\sum_{i=1}^{d} |L_i^\alpha(1)|\Big)^\lambda\Big] \\
\leq & \sum_{i=1}^{d} \mathbb{E}|L_i^\alpha(1)|^\lambda \\
= & d\Big(\frac{2^\lambda \Gamma((1+\lambda)/2)\Gamma(1-\lambda/\alpha)}{\Gamma(1/2)\Gamma(1-\lambda/2)}\Big).
\end{aligned}$$

□

Lemma S7. *Let us denote the value $\mathbb{E}\|L^\alpha(1)\|^\lambda$ by $l_{\alpha,\lambda,d} < \infty$. For $0 < \eta \leq \frac{m}{M^2}$ and $s \in [j\eta, (j+1)\eta)$, we have the following estimates:*

(a) If $1 < \lambda < \alpha$ and $1 < \gamma\lambda < \alpha$ then

$$\begin{aligned}
\mathbb{E}\|X_2(j\eta)\|^\lambda &\leq B_{j,\lambda} \triangleq \Big(\Big(\mathbb{E}\|X_2(0)\|^\lambda\Big)^{\frac{1}{\lambda}} + j\Big((2\eta(b+m))^{\frac{1}{2}} + 2^{\frac{1}{2}}\eta B + \Big(\frac{\eta}{\beta}\Big)^{\frac{1}{\alpha}} l_{\alpha,\lambda,d}^{\frac{1}{\lambda}}\Big)\Big)^\lambda, \\
\mathbb{E}\|X_2(s)\|^\lambda &\leq \Big(B_{j,\lambda}^{\frac{1}{\lambda}} + (s-j\eta)\Big(MB_{j,\gamma\lambda}^{\frac{1}{\lambda}} + B\Big) + \Big(\frac{s-j\eta}{\beta}\Big)^{\frac{1}{\alpha}} l_{\alpha,\lambda,d}^{\frac{1}{\lambda}}\Big)^\lambda.
\end{aligned}$$

(b) If $0 \leq \lambda \leq 1$ then

$$\begin{aligned}
\mathbb{E}\|X_2(j\eta)\|^\lambda &\leq \bar{B}_{j,\lambda} \triangleq \mathbb{E}\|X_2(0)\|^\lambda + j\Big((2\eta(b+m))^{\frac{\lambda}{2}} + 2^{\frac{\lambda}{2}}(\eta B)^\lambda + \Big(\frac{\eta}{\beta}\Big)^{\frac{\lambda}{\alpha}} l_{\alpha,\lambda,d}\Big), \\
\mathbb{E}\|X_2(s)\|^\lambda &\leq \bar{B}_{j,\lambda} + (s-j\eta)^\lambda\Big(M^\lambda \bar{B}_{j,\gamma\lambda} + B^\lambda\Big) + \Big(\frac{s-j\eta}{\beta}\Big)^{\frac{\lambda}{\alpha}} l_{\alpha,\lambda,d}.
\end{aligned}$$

(c) If $1 < \lambda < \alpha$ and $0 \leq \gamma\lambda \leq 1$ then

$$\begin{aligned}
\mathbb{E}\|X_2(j\eta)\|^\lambda &\leq B_{j,\lambda}, \\
\mathbb{E}\|X_2(s)\|^\lambda &\leq \Big(B_{j,\lambda}^{\frac{1}{\lambda}} + (s-j\eta)\Big(M\bar{B}_{j,\gamma\lambda}^{\frac{1}{\lambda}} + B\Big) + \Big(\frac{s-j\eta}{\beta}\Big)^{\frac{1}{\alpha}} l_{\alpha,\lambda,d}^{\frac{1}{\lambda}}\Big)^\lambda.
\end{aligned}$$

Proof. Starting from

$$X_2((j+1)\eta) = X_2(j\eta) - \eta c_\alpha \nabla f(X_2(j\eta)) + \left(\frac{\eta}{\beta}\right)^{\frac{1}{\alpha}} L^\alpha(1),$$

we have either (by Minkowski, if $\lambda > 1$)

$$\left(\mathbb{E}\|X_2((j+1)\eta)\|^\lambda\right)^{\frac{1}{\lambda}} \le \left(\mathbb{E}\|X_2(j\eta) - \eta c_\alpha \nabla f(X_2(j\eta))\|^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{\eta}{\beta}\right)^{\frac{1}{\alpha}} \left(\mathbb{E}\|L^\alpha(1)\|^\lambda\right)^{\frac{1}{\lambda}}, \tag{S12}$$

or (by Lemma S11, if $0 \le \lambda \le 1$)

$$\mathbb{E}\|X_2((j+1)\eta)\|^\lambda \le \mathbb{E}\|X_2(j\eta) - \eta c_\alpha \nabla f(X_2(j\eta))\|^\lambda + \left(\frac{\eta}{\beta}\right)^{\frac{\lambda}{\alpha}} \mathbb{E}\|L^\alpha(1)\|^\lambda. \tag{S13}$$

We have

$$\begin{aligned}
\|X_2(j\eta) - \eta c_\alpha \nabla f(X_2(j\eta))\|^\lambda &= \|X_2(j\eta) - \eta c_\alpha \nabla f(X_2(j\eta))\|^{2\times\frac{\lambda}{2}} \\
&= \left(\|X_2(j\eta)\|^2 - 2\eta c_\alpha \langle X_2(j\eta), \nabla f(X_2(j\eta))\rangle + \eta^2 \|c_\alpha \nabla f(X_2(j\eta))\|^2\right)^{\frac{\lambda}{2}} \\
&\le \left(\|X_2(j\eta)\|^2 - 2\eta(m\|X_2(j\eta)\|^{1+\gamma} - b) + \eta^2(2M^2\|X_2(j\eta)\|^{2\gamma} + 2B^2)\right)^{\frac{\lambda}{2}},
\end{aligned} \tag{S14}$$

where we have used assumption **H**3 and Lemma S8. For $0 < \eta \le \frac{m}{M^2}$,

$$2\eta m(\|X_2(j\eta)\|^{1+\gamma} + 1) \ge 2\eta^2 M^2 \|X_2(j\eta)\|^{2\gamma}. \qquad \text{(since } 1+\gamma > 2\gamma \text{ and } \eta m > \eta^2 M^2\text{)}$$

Using this inequality we have

$$\begin{aligned}
\|X_2(j\eta) - \eta c_\alpha \nabla f(X_2(j\eta))\|^\lambda &\le \left(\|X_2(j\eta)\|^2 + 2\eta(b+m) + 2\eta^2 B^2\right)^{\frac{\lambda}{2}} \\
&\le \|X_2(j\eta)\|^\lambda + (2\eta(b+m))^{\frac{\lambda}{2}} + 2^{\frac{\lambda}{2}}(\eta B)^\lambda. \qquad \text{(by Lemma S11)}
\end{aligned} \tag{S15}$$

Consider the case where $\lambda > 1$. By (S12) and (S15),

$$\begin{aligned}
\left(\mathbb{E}\|X_2((j+1)\eta)\|^\lambda\right)^{\frac{1}{\lambda}} &\le \left(\mathbb{E}\|X_2(j\eta)\|^\lambda + (2\eta(b+m))^{\frac{\lambda}{2}} + 2^{\frac{\lambda}{2}}(\eta B)^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{\eta}{\beta}\right)^{\frac{1}{\alpha}} \left(\mathbb{E}\|L^\alpha(1)\|^\lambda\right)^{\frac{1}{\lambda}} \\
&\le \left(\mathbb{E}\|X_2(j\eta)\|^\lambda\right)^{\frac{1}{\lambda}} + (2\eta(b+m))^{\frac{1}{2}} + 2^{\frac{1}{2}}\eta B + \left(\frac{\eta}{\beta}\right)^{\frac{1}{\alpha}} l_{\alpha,\lambda,d}^{\frac{1}{\lambda}} \quad \text{(by Lemma S11)} \\
&\le \left(\mathbb{E}\|X_2(0)\|^\lambda\right)^{\frac{1}{\lambda}} + (j+1)\left((2\eta(b+m))^{\frac{1}{2}} + 2^{\frac{1}{2}}\eta B + \left(\frac{\eta}{\beta}\right)^{\frac{1}{\alpha}} l_{\alpha,\lambda,d}^{\frac{1}{\lambda}}\right).
\end{aligned}$$

For the case where $0 \le \lambda \le 1$, by (S13) and (S15),

$$\begin{aligned}
\mathbb{E}\|X_2((j+1)\eta)\|^\lambda &\le \mathbb{E}\|X_2(j\eta)\|^\lambda + (2\eta(b+m))^{\frac{\lambda}{2}} + 2^{\frac{\lambda}{2}}(\eta B)^\lambda + \left(\frac{\eta}{\beta}\right)^{\frac{\lambda}{\alpha}} l_{\alpha,\lambda,d} \\
&\le \mathbb{E}\|X_2(0)\|^\lambda + (j+1)\left((2\eta(b+m))^{\frac{\lambda}{2}} + 2^{\frac{\lambda}{2}}(\eta B)^\lambda + \left(\frac{\eta}{\beta}\right)^{\frac{\lambda}{\alpha}} l_{\alpha,\lambda,d}\right).
\end{aligned}$$

Now, from the identification, for $s \in [j\eta, (j+1)\eta)$,

$$X_2(s) = X_2(j\eta) + (s - j\eta)c_\alpha \nabla f(X_2(j\eta)) + \left(\frac{s - j\eta}{\beta}\right)^{\frac{1}{\alpha}} L^\alpha(1),$$

we have

$$\begin{aligned}\|X_2(s)\| &\leq \|X_2(j\eta)\| + (s-j\eta)c_\alpha\|\nabla f(X_2(j\eta))\| + \left(\frac{s-j\eta}{\beta}\right)^{\frac{1}{\alpha}}\|L^\alpha(1)\| \\ &\leq \|X_2(j\eta)\| + (s-j\eta)(M\|X_2(j\eta)\|^\gamma + B) + \left(\frac{s-j\eta}{\beta}\right)^{\frac{1}{\alpha}}\|L^\alpha(1)\|.\end{aligned}$$

For $\lambda > 1$,

$$\left(\mathbb{E}\|X_2(s)\|^\lambda\right)^{\frac{1}{\lambda}} \leq \left(\mathbb{E}\|X_2(j\eta)\|^\lambda\right)^{\frac{1}{\lambda}} + (s-j\eta)\left(M\left(\mathbb{E}\|X_2(j\eta)\|^{\gamma\lambda}\right)^{\frac{1}{\lambda}} + B\right) + \left(\frac{s-j\eta}{\beta}\right)^{\frac{1}{\alpha}} l_{\alpha,\lambda,d}^{\frac{1}{\lambda}}.$$

For $\lambda \leq 1$,

$$\mathbb{E}\|X_2(s)\|^\lambda \leq \mathbb{E}\|X_2(j\eta)\|^\lambda + (s-j\eta)^\lambda\left(M^\lambda\mathbb{E}\|X_2(j\eta)\|^{\gamma\lambda} + B^\lambda\right) + \left(\frac{s-j\eta}{\beta}\right)^{\frac{\lambda}{\alpha}} l_{\alpha,\lambda,d}.$$

By replacing the estimate of $\mathbb{E}\|X_2(j\eta)\|^\lambda$, we obtain the desired result.

□

Lemma S8. *Under assumptions **H**1 and **H**2 we have*

$$c_\alpha\|\nabla f(w)\| \leq M\|w\|^\gamma + B, \qquad \forall w \in \mathbb{R}^d.$$

Proof. By assumption **H**2 we have

$$c_\alpha\|\nabla f(w) - \nabla f(0)\| \leq M\|w-0\|^\gamma.$$

Since $c_\alpha\|\nabla f(0)\| \leq B$ by assumption **H**1, the conclusion follows. □

Lemma S9. *For the function* b *defined in Lemma 1, we have, for* $w \in \mathbb{R}^d$*,*

$$\begin{aligned}&\|b(w)\| \leq M\|w\|^\gamma + (B+L), \\ &\langle w, b(w)\rangle \leq (L-m)\|w\|^{1+\gamma} + (b+L).\end{aligned}$$

Proof. From assumption **H**6, it implies that

$$\|b(w)\| \leq c_\alpha\|\nabla f(w)\| + L.$$

Then, by Lemma S8,

$$\|b(w)\| \leq M\|w\|^\gamma + (B+L).$$

Next, by Cauchy-Schwarz inequality and assumption **H**6, we have

$$\langle w, b(w) + c_\alpha\nabla f(w)\rangle \leq \|w\|L.$$

Then, by assumption **H**3,

$$\begin{aligned}\langle w, b(w)\rangle &\leq -c_\alpha\langle w, \nabla f(w)\rangle + \|w\|L \\ &\leq -m\|w\|^{1+\gamma} + b + \|w\|L \\ &\leq -m\|w\|^{1+\gamma} + b + (\|w\|^{1+\gamma} + 1)L \\ &= (L-m)\|w\|^{1+\gamma} + (b+L).\end{aligned}$$

Here, we have used the inequality $\|w\| \leq \|w\|^{1+\gamma} + 1$. □

Lemma S10. *Let ν be the Lévy measure of a d-dimensional Lévy process L^α whose components are independent scalar symmetric α-stable Lévy processes $L_1^\alpha, \dots, L_d^\alpha$. Then there exists a constant $C > 0$ such that the following inequality holds with $\beta \geq 1$ and $2 > \alpha > 1$:*

$$\frac{1}{\beta^{2/\alpha}} \int_{\|x\|<1} \|x\|^2 \nu(\mathrm{d}x) + \frac{1}{\beta^{1/\alpha}} \int_{\|x\|\geq 1} \|x\| \nu(\mathrm{d}x) \leq C \frac{d}{\beta^{1/\alpha}}.$$

Proof. Using Lemma 4.1 in (Kallsen & Tankov, 2006), we have

$$\begin{aligned}\int_{\|x\|<1} \|x\|^2 \nu(\mathrm{d}x) &= \sum_{i=1}^d \int_{|x_i|<1} |x_i|^2 \frac{1}{|x_i|^{1+\alpha}} \mathrm{d}x_i \\ &= \sum_{i=1}^d \frac{2}{2-\alpha} \\ &= \frac{2d}{2-\alpha}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\int_{\|x\|\geq 1} \|x\| \nu(\mathrm{d}x) &= \sum_{i=1}^d \int_{|x_i|\geq 1} |x_i| \frac{1}{|x_i|^{1+\alpha}} \mathrm{d}x_i \\ &= \sum_{i=1}^d \frac{2}{\alpha-1} \\ &= \frac{2d}{\alpha-1}.\end{aligned}$$

Combining these two equalities, we have the desired conclusion. □

Lemma S11. *For $a, b \geq 0$ and $0 \leq \gamma \leq 1$, we have the following inequality:*

$$(a+b)^\gamma \leq a^\gamma + b^\gamma.$$

Proof. If $a = b = 0$, the inequality is trivial. Hence, let us assume that $a > b \geq 0$. We have

$$\begin{aligned}\left(1 + \frac{b}{a}\right)^\gamma &\leq 1 + \gamma \frac{b}{a} && \text{(by Bernoulli's inequality)} \\ &\leq 1 + \frac{b}{a} && \text{(since } 0 \leq \gamma \leq 1 \text{ and } \frac{b}{a} \geq 0\text{)} \\ &\leq 1 + \left(\frac{b}{a}\right)^\gamma. && \text{(since } 0 \leq \gamma \leq 1 \text{ and } 0 \leq \frac{b}{a} \leq 1\text{)}\end{aligned}$$

By multiplying both sides by $a^\gamma > 0$, we have the conclusion. □

S14. A Remark on the Global Hölder Condition

We note that the assumption **H**2 can be weakened to local Hölder continuity by using the localization techniques given in the proof of Proposition 4.2.2 of (Kunze, 2012). This approach requires rewriting all the expressions which use **H**2 in our proofs by using stopping-times in such a way that we can use the local Hölder continuity in the same way of (Kunze, 2012).

References

Kallsen, J. and Tankov, P. Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, 97(7):1551–1572, 2006.

Kunze, M. Stochastic differential equations. Lecture notes, University of Ulm, 2012.

Liang, M. and Wang, J. Gradient Estimates and Ergodicity for SDEs Driven by Multiplicative Lévy Noises via Coupling. *arXiv preprint arXiv:1801.05936*, 2018.

Matsui, M., Pawlas, Z., et al. Fractional absolute moments of heavy tailed distributions. *Brazilian Journal of Probability and Statistics*, 30(2):272–298, 2016.

Nesterov, Y. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

Raginsky, M., Rakhlin, A., and Telgarsky, M. Non-convex learning via stochastic gradient Langevin dynamics: a nonasymptotic analysis. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65, pp. 1674–1703, 2017.

Schertzer, D., Larchevêque, M., Duan, J., Yanovsky, V., 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.

Shanbhag, D. N. and Sreehari, M. On certain self-decomposable distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 38(3):217–222, 1977.

Şimşekli, U. Fractional Langevin Monte Carlo: Exploring Lévy Driven Stochastic Differential Equations for Markov Chain Monte Carlo. In *International Conference on Machine Learning*, 2017.

Villani, C. Topics in optimal transportation, volume 58 ofGraduate. *Studies in Mathematics*, 2003.

Xie, L. and Zhang, X. Ergodicity of stochastic differential equations with jumps and singular coefficients. *arXiv preprint arXiv:1705.07402*, 2017.