CONTENTS

A THE PROOF OF THE UPPER BOUND ON THE ERROR OF RLMC

This section is devoted to the proof of the upper bound on the error of sampling, measured in W_2-distance, of the randomized mid-point method for the vanilla Langevin Langevin diffusion. Since no other sampling method is considered in this section, without any risk of confusion, we will use the notation $\boldsymbol{\vartheta}_k$ instead of $\boldsymbol{\vartheta}_k^{\mathsf{RLMC}}$ to refer to the k-th iterate of the RLMC. We will also use the shorthand notation

$$f_k = f(\boldsymbol{\vartheta}_k), \qquad \nabla f_k := \nabla f(\boldsymbol{\vartheta}_k), \qquad \text{and} \qquad \nabla f_{k+U} := \nabla f(\boldsymbol{\vartheta}_{k+U}).$$

A.1 PROOF OF THEOREM 1

Let $\boldsymbol{\vartheta}_0 \sim \nu_0$ and $\boldsymbol{L}_0 \sim \pi$ be two random vectors in $\mathbb{R}^p$ defined on the same probability space. At this stage, the joint distribution of these vectors is arbitrary; we will take an infimum over all possible joint distributions with given marginals at the end of the proof. Note right away that the condition $Mh + \sqrt{\kappa}(Mh)^{3/2} \leqslant 1/4$ implies that $Mh + (Mh)^{3/2} \leqslant 1/4$, which also yields $Mh \leqslant 0.18$.

Assume that on the same probability space, we can define a Brownian motion $\boldsymbol{W}$, independent of $(\boldsymbol{\vartheta}_0, \boldsymbol{L}_0)$, and an infinite sequence of iid random variables, uniformly distributed in $[0,1]$, $U_0, U_1, \ldots$, independent of $(\boldsymbol{\vartheta}_0, \boldsymbol{L}_0, \boldsymbol{W})$. We define the Langevin diffusion

$$\boldsymbol{L}_t = \boldsymbol{L}_0 - \int_0^t \nabla f(\boldsymbol{L}_s)\mathrm{d}s + \sqrt{2}\,\boldsymbol{W}_t. \tag{17}$$

We also set

$$\begin{aligned}\boldsymbol{\vartheta}_{k+U} &= \boldsymbol{\vartheta}_k - hU_k\nabla f_k + \sqrt{2}\left(\boldsymbol{W}_{(k+U_k)h} - \boldsymbol{W}_{kh}\right)\\ \boldsymbol{\vartheta}_{k+1} &= \boldsymbol{\vartheta}_k - h\nabla f_{k+U} + \sqrt{2}\,(\boldsymbol{W}_{(k+1)h} - \boldsymbol{W}_{kh}).\end{aligned}$$

One can check that this sequence $\{\boldsymbol{\vartheta}_k\}$ has exactly the same distribution as the sequence defined in equation 6 and equation 7. Therefore,

$$\mathsf{W}_2^2(\nu_{k+1}, \pi) \leqslant \mathbb{E}[\|\boldsymbol{\vartheta}_{k+1} - \boldsymbol{L}_{(k+1)h}\|_2^2] := \|\boldsymbol{\vartheta}_{k+1} - \boldsymbol{L}_{(k+1)h}\|_{\mathbb{L}_2}^2 := x_{k+1}^2.$$

We will also consider the Langevin process on the time interval $[0, h]$ given by

$$\boldsymbol{L}'_t = \boldsymbol{L}'_0 - \int_0^t \nabla f(\boldsymbol{L}'_s)\,\mathrm{d}s + \sqrt{2}\,(\boldsymbol{W}_{kh+t} - \boldsymbol{W}_{kh}), \qquad \boldsymbol{L}'_0 = \boldsymbol{\vartheta}_k.$$

Note that the Brownian motion is the same as in equation 17.

Let us introduce one additional notation, the average of $\boldsymbol{\vartheta}_{k+1}$ with respect to U_k,

$$\bar{\boldsymbol{\vartheta}}_{k+1} = \mathbb{E}[\boldsymbol{\vartheta}_{k+1}|\boldsymbol{\vartheta}_k, \boldsymbol{W}, \boldsymbol{L}_0].$$

Since $\boldsymbol{L}_{(k+1)h}$ is independent of U_k, it is clear that

$$x_{k+1}^2 = \|\boldsymbol{\vartheta}_{k+1} - \bar{\boldsymbol{\vartheta}}_{k+1}\|_{\mathbb{L}_2}^2 + \|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}_{(k+1)h}\|_{\mathbb{L}_2}^2.$$

Furthermore, the triangle inequality yields

$$\|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}_{(k+1)h}\|_{\mathbb{L}_2} \leqslant \|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} + \|\boldsymbol{L}'_h - \boldsymbol{L}_{(k+1)h}\|_{\mathbb{L}_2}.$$

From the exponential ergodicity of the Langevin diffusion (Bhattacharya, 1978), we get

$$\|\boldsymbol{L}'_h - \boldsymbol{L}_{(k+1)h}\|_{\mathbb{L}_2} \leqslant e^{-mh}\|\boldsymbol{L}'_0 - \boldsymbol{L}_{kh}\|_{\mathbb{L}_2} = e^{-mh}\|\boldsymbol{\vartheta}_k - \boldsymbol{L}_{kh}\|_{\mathbb{L}_2} = e^{-mh}x_k.$$

Therefore, we get

$$\begin{aligned}x_{k+1}^2 &\leqslant \|\boldsymbol{\vartheta}_{k+1} - \bar{\boldsymbol{\vartheta}}_{k+1}\|_{\mathbb{L}_2}^2 + \big(\|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} + e^{-2mh}x_k\big)^2\\ &= \big(e^{-mh}x_k + \|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2}\big)^2 + \|\boldsymbol{\vartheta}_{k+1} - \bar{\boldsymbol{\vartheta}}_{k+1}\|_{\mathbb{L}_2}^2.\end{aligned} \tag{18}$$

The last term of the right-hand side can be bounded as follows

$$\begin{aligned}\|\boldsymbol{\vartheta}_{k+1} - \bar{\boldsymbol{\vartheta}}_{k+1}\|_{\mathbb{L}_2} &= h\|\nabla f_{k+U} - \mathbb{E}_U[\nabla f_{k+U}]\|_{\mathbb{L}_2}\\ &\leqslant h\|\nabla f_{k+U} - \nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2}.\end{aligned}$$

Using the definition of $\boldsymbol{\vartheta}_{k+U}$, we get

$$\|\boldsymbol{\vartheta}_{k+1} - \bar{\boldsymbol{\vartheta}}_{k+1}\|_{\mathbb{L}_2}^2 \leqslant (Mh)^2\Big((1/3)h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + hp\Big). \tag{19}$$

We will also need the following lemma, the proof of which is postponed.

Lemma 1. *If $Mh \leqslant 0.18$, then $\|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} \leqslant (Mh)^2\big\{0.7h\|\nabla f_k\|_{\mathbb{L}_2} + 1.2\sqrt{hp}\big\}$.*

One can check by induction that if for some $A \in [0,1]$ and for two positive sequences $\{B_k\}$ and $\{C_k\}$ the inequality $x_{k+1}^2 \leqslant \big\{(1-A)x_k + C_k\big\}^2 + B_k^2$ holds for every integer $k \geqslant 0$, then[5]

$$x_n \leqslant (1-A)^n x_0 + \sum_{k=0}^{n}(1-A)^{n-k}C_k + \bigg\{\sum_{k=0}^{n}(1-A)^{2(n-k)}B_k^2\bigg\}^{1/2} \tag{20}$$

In view of equation 20, equation 18, equation 19 and Lemma 1, for $\rho = e^{-mh}$, we get

$$\begin{aligned}
x_n &\leqslant \rho^n x_0 + (Mh)^2\sum_{k=0}^{n}\rho^{n-k}\big(0.7h\|\nabla f_k\|_{\mathbb{L}_2} + 1.2\sqrt{hp}\big)\\
&\qquad + Mh\bigg\{\sum_{k=0}^{n}\rho^{2(n-k)}\big((1/3)h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + hp\big)\bigg\}^{1/2}\\
&\leqslant \rho^n x_0 + 0.7(Mh)^2 h\sum_{k=0}^{n}\rho^{n-k}\|\nabla f_k\|_{\mathbb{L}_2} + 1.32\frac{M^2h\sqrt{hp}}{m}\\
&\qquad + \frac{Mh^2}{\sqrt{3}}\bigg\{\sum_{k=0}^{n}\rho^{2(n-k)}\|\nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2}^2\bigg\}^{1/2} + 0.92Mh\sqrt{p/m}.
\end{aligned} \tag{21}$$

We need a last lemma for finding a suitable upper bound on the right-hand side of the last display.

Lemma 2. *If $Mh \leqslant 0.18$ and $k \geqslant 1$, then the following inequalities hold*

$$h^2\sum_{k=0}^{n}\rho^{n-k}\|\nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2}^2 \leqslant 1.7Mh\rho^n\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + 4.4Mh(p/m) \leqslant 0.31\rho^n\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + 0.8(p/m).$$

The claim of this lemma and together with equation 21 entail that

$$\begin{aligned}
x_n &\leqslant \rho^n x_0 + 0.7(Mh)^2 h\sum_{k=0}^{n}\rho^{n-k}\|\nabla f_k\|_{\mathbb{L}_2} + \frac{Mh^2}{\sqrt{3}}\bigg\{\sum_{k=0}^{n}\rho^{2(n-k)}\|\nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2}^2\bigg\}^{1/2}\\
&\qquad + (1.32\sqrt{\kappa Mh} + 0.92)Mh\sqrt{p/m}\\
&\leqslant \rho^n x_0 + \big(0.74\sqrt{\kappa Mh} + 0.58\big)Mh\bigg\{h^2\sum_{k=0}^{n}\rho^{n-k}\|\nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2}^2\bigg\}^{1/2}\\
&\qquad + \big(1.32\sqrt{\kappa Mh} + 0.92\big)Mh\sqrt{p/m}\\
&\leqslant \rho^n x_0 + (0.42\sqrt{\kappa Mh} + 0.33)Mh\rho^{n/2}\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2} + \big(1.98\sqrt{\kappa Mh} + 1.44\big)Mh\sqrt{p/m}.
\end{aligned}$$

Assuming that h is such that $(\sqrt{\kappa Mh}+1)Mh \leqslant 1/4$ and noting that $\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2} \leqslant \mathsf{W}_2(\nu_0,\pi) + \sqrt{p/m}$, we arrive at the desired inequality.

A.2 Proof of technical lemmas

In this section, we present the proofs of two technical lemmas that have been used in the proof of the main theorem. The first lemma provides an upper bound on the error of the averaged iterate $\bar{\boldsymbol{\vartheta}}_{k+1}$ and the continuous time diffusion $\boldsymbol{L}'$ that starts from $\boldsymbol{\vartheta}_k$ and runs until the time h. This upper bound involves the norm of the gradient of the potential f evaluated at $\boldsymbol{\vartheta}_k$. The second lemma aims at bounding the discounted sums of the squared norms of these gradients.

[5] This is an extension of (Dalalyan & Karagulyan, 2019, Lemma 7). It essentially relies on the elementary $\sqrt{(a+b)^2+c^2} \leqslant a + \sqrt{b^2+c^2}$, which should be used to prove the induction step.

A.2.1 Proof of Lemma 1 (one-step mean discretisation error)

We have

$$\begin{aligned}
\|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} &= \left\|\boldsymbol{\vartheta}_k - h\mathbb{E}_U[\nabla f_{k+U}] - \boldsymbol{L}'_0 + \int_0^h \nabla f(\boldsymbol{L}'_s)\,\mathrm{d}s\right\|_{\mathbb{L}_2} \\
&= \left\|h\mathbb{E}_U[\nabla f_{k+U} - \nabla f(\boldsymbol{L}'_{Uh})]\right\|_{\mathbb{L}_2} \\
&\leqslant Mh\big\|\boldsymbol{\vartheta}_{k+U} - \boldsymbol{L}'_{Uh}\big\|_{\mathbb{L}_2} \\
&= Mh\big\|\boldsymbol{\vartheta}_k - U_k h\nabla f_k - \boldsymbol{L}'_0 + \int_0^{U_k h} \nabla f(\boldsymbol{L}'_s)\mathrm{d}s\big\|_{\mathbb{L}_2} \\
&= Mh\left\|\int_0^{U_k h} (\nabla f(\boldsymbol{L}'_s) - \nabla f_k)\,\mathrm{d}s\right\|_{\mathbb{L}_2} \\
&\leqslant Mh\int_0^h \|\nabla f(\boldsymbol{L}'_s) - \nabla f_k\|_{\mathbb{L}_2}\,\mathrm{d}s \\
&= Mh\int_0^h \big\|\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{L}'_0)\big\|_{\mathbb{L}_2}\,\mathrm{d}s. \qquad (22)
\end{aligned}$$

Let us define $\varphi(t) = \|\nabla f(\boldsymbol{L}'_t) - \nabla f(\boldsymbol{L}'_0)\|_{\mathbb{L}_2}$. Using the Lipschitz continuity of ∇f and the definition of $\boldsymbol{L}'$, we arrive at

$$\begin{aligned}
\varphi(t)^2 &\leqslant M^2\bigg\{\bigg\|\int_0^t \nabla f(\boldsymbol{L}'_s)\,\mathrm{d}s\bigg\|^2_{\mathbb{L}_2} + 2tp\bigg\} \\
&\leqslant M^2\bigg\{\bigg(t\|\nabla f_k\|_{\mathbb{L}_2} + \int_0^t \big\|\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{L}'_0)\big\|_{\mathbb{L}_2}\,\mathrm{d}s\bigg)^2 + 2tp\bigg\} \\
&\leqslant M^2\bigg\{\int_0^t \big\|\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{L}'_0)\big\|_{\mathbb{L}_2}\,\mathrm{d}s + \sqrt{t^2\|\nabla f_k\|^2_{\mathbb{L}_2} + 2tp}\bigg\}^2
\end{aligned}$$

or, equivalently,

$$\varphi(t) \leqslant M\int_0^t \varphi(s)\,\mathrm{d}s + M\sqrt{t^2\|\nabla f_k\|^2_{\mathbb{L}_2} + 2tp}.$$

Using the Grönwall inequality, we get

$$\varphi(t) \leqslant Me^{Mt}\sqrt{t^2\|\nabla f_k\|^2_{\mathbb{L}_2} + 2tp}.$$

Combining this inequality with the bound obtained in equation 22, and using the inequality $e^{Mh} \leqslant 1.2$, we arrive at

$$\begin{aligned}
\|\bar{\boldsymbol{\vartheta}}_{k+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} &\leqslant 1.2M^2h\int_0^h \sqrt{s^2\|\nabla f_k\|^2_{\mathbb{L}_2} + 2sp}\,\mathrm{d}s \\
&\leqslant 1.2M^2h\sqrt{h}\bigg\{\int_0^h \big(s^2\|\nabla f_k\|^2_{\mathbb{L}_2} + 2sp\big)\,\mathrm{d}s\bigg\}^{1/2} \\
&\leqslant 1.2M^2h\sqrt{h}\big\{(h^3/3)\|\nabla f_k\|^2_{\mathbb{L}_2} + h^2p\big\}^{1/2}.
\end{aligned}$$

This completes the proof.

A.2.2 Proof of Lemma 2 (Discounted sum of squared gradients)

We have

$$\begin{aligned}
f_{k+1} &\leqslant f_k + \nabla f_k^\top(\boldsymbol{\vartheta}_{k+1} - \boldsymbol{\vartheta}_k) + \frac{M}{2}\|\boldsymbol{\vartheta}_{k+1} - \boldsymbol{\vartheta}_k\|_2^2 \\
&\leqslant f_k - h\nabla f_k^\top\nabla f_{k+U} + \sqrt{2}\nabla f_k^\top\boldsymbol{\xi}_h + \frac{M}{2}\|h\nabla f_{k+U} - \sqrt{2}\,\boldsymbol{\xi}_k\|_2^2 \\
&\leqslant f_k - h\|\nabla f_k\|_2^2 + Mh\|\nabla f_k\|_2\|\boldsymbol{\vartheta}_{k+U} - \boldsymbol{\vartheta}_k\|_2 + \sqrt{2}\nabla f_k^\top\boldsymbol{\xi}_h + \frac{M}{2}\|h\nabla f_{k+U} - \sqrt{2}\,\boldsymbol{\xi}_k\|_2^2. \qquad (23)
\end{aligned}$$

One checks that

$$\|\boldsymbol{\vartheta}_{k+U} - \boldsymbol{\vartheta}_k\|_{\mathbb{L}_2}^2 = h^2\|U\nabla f_k\|_{\mathbb{L}_2}^2 + 2hp\mathbb{E}[U] = (h^2/3)\|\nabla f_k\|_{\mathbb{L}_2}^2 + hp \leqslant 0.011\frac{\|\nabla f_k\|_{\mathbb{L}_2}^2}{M^2} + hp$$

and, therefore,

$$\begin{aligned}M\|\nabla f_k\|_2\|\boldsymbol{\vartheta}_{k+U} - \boldsymbol{\vartheta}_k\|_2 &\leqslant \big(0.011\|\nabla f_k\|_{\mathbb{L}_2}^4 + M^2hp\|\nabla f_k\|_2^2\big)^{1/2}\\ &\leqslant 0.105\|\nabla f_k\|_{\mathbb{L}_2}^2 + 4.55M^2hp\\ &\leqslant 0.105\|\nabla f_k\|_{\mathbb{L}_2}^2 + 0.82Mp.\end{aligned}$$

Furthermore,

$$\begin{aligned}\|h\nabla f_{k+U} - \sqrt{2}\,\boldsymbol{\xi}_k\|_{\mathbb{L}_2} &\leqslant \|h\nabla f_k - \sqrt{2}\,\boldsymbol{\xi}_k\|_{\mathbb{L}_2} + h\|\nabla f_{k+U} - \nabla f_k\|_{\mathbb{L}_2}\\ &\leqslant \sqrt{h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + hp} + Mh\|\boldsymbol{\vartheta}_{k+U} - \boldsymbol{\vartheta}_k\|_{\mathbb{L}_2}\\ &\leqslant \sqrt{h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + hp} + h\sqrt{0.011\|\nabla f_k\|_{\mathbb{L}_2}^2 + M^2hp}\\ &\leqslant \sqrt{h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + hp} + \sqrt{0.011h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + 0.18^2hp}\end{aligned}$$

implying that

$$\begin{aligned}\frac{M}{2}\|h\nabla f_{k+U} - \sqrt{2}\,\boldsymbol{\xi}_k\|_{\mathbb{L}_2}^2 &\leqslant \frac{M}{2}\big(1.37h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + 1.4hp\big)\\ &\leqslant 0.124h^2\|\nabla f_k\|_{\mathbb{L}_2}^2 + 0.7Mhp.\end{aligned}$$

Combining these inequalities with equation 23, we get

$$\mathbb{E}[f_{k+1}] \leqslant \mathbb{E}[f_k] - 0.771h\|\nabla f_k\|_{\mathbb{L}_2}^2 + 1.52Mhp. \tag{24}$$

Set $S_n(f) = \sum_{k=0}^n \rho^{n-k} f_k$ and $S_n(\nabla f^2) = \sum_{k=0}^n \rho^{n-k}\|\nabla f_k\|_{\mathbb{L}_2}^2$. Using Lemma 3, we get

$$\mathbb{E}[f_{n+1}] - \rho^n\mathbb{E}[f_0] + \rho S_n(f) \leqslant S_n(f) - 0.771hS_n(\nabla f^2) + \frac{1.52Mhp}{1-\rho}$$

Since $mh \geqslant 1-\rho \geqslant 0.915mh$, we get

$$\begin{aligned}0.771hS_n(\nabla f^2) &\leqslant \rho^n\mathbb{E}[f_0] + (1-\rho)S_n(f) + 1.67\kappa p\\ &\leqslant \rho^n\mathbb{E}[f_0] + mhS_n(f) + 1.67\kappa p\\ &\leqslant \rho^n\mathbb{E}[f_0] + 0.5hS_n(\nabla f^2) + 1.67\kappa p\end{aligned}$$

where the last line follows from the Polyak-Lojasiewicz inequality. Rearranging the terms, we get

$$hS_n(\nabla f^2) \leqslant 3.7\rho^n\mathbb{E}[f_0] + 6.2\kappa p \tag{25}$$

Note that equation 25 is obtained under the Polyak-Lojasiewicz condition, without explicitly using the strong convexity of f. However, using the latter property, we can obtain a similar inequality with slightly better constants.

Indeed, equation 24 yields

$$h\mathbb{E}[\|\nabla f_k\|_2^2] \leqslant 1.3\big(\mathbb{E}[f_k] - \mathbb{E}[f_{k+1}]\big) + 1.98Mhp. \tag{26}$$

In what follows, without loss of generality, we assume that $f(\boldsymbol{\theta}^*) = \min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = 0$. In view of equation 26, we have

$$\begin{aligned}h\sum_{j=0}^k \rho^{k-j}\|\nabla f(\boldsymbol{\vartheta}_j)\|_{\mathbb{L}_2}^2 &\leqslant 1.3\sum_{j=0}^k \rho^{k-j}(\mathbb{E}[f(\boldsymbol{\vartheta}_j) - f(\boldsymbol{\vartheta}_{j+1})]) + \frac{1.98Mhp}{1-e^{-mh}}\\ &\leqslant 1.3\big(\rho^k\mathbb{E}[f(\boldsymbol{\vartheta}_0)] - \mathbb{E}[f_{k+1}]\big) + 1.3\sum_{j=1}^k \rho^{k-j}(1-\rho)\mathbb{E}[f(\boldsymbol{\vartheta}_j)] + 2.1\kappa p\\ &\leqslant 1.3\rho^{k+1}\mathbb{E}[f(\boldsymbol{\vartheta}_0)] + 1.3(1-\rho)\sum_{j=0}^k \rho^{k-j}\mathbb{E}[f(\boldsymbol{\vartheta}_j)] + 2.1\kappa p\\ &\leqslant \frac{1.3M}{2}\rho^{k+1}\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + \frac{1.3}{2}M(1-\rho)\sum_{j=0}^k \rho^{k-j}\|\boldsymbol{\vartheta}_j\|_{\mathbb{L}_2}^2 + 2.1\kappa p.\end{aligned}$$

We have, in addition

$$\|\boldsymbol{\vartheta}_{k+1}\|_{\mathbb{L}_2}^2 = \|\boldsymbol{\vartheta}_k - h\nabla f_k\|_{\mathbb{L}_2}^2 + 2hp \leqslant (1-mh)^2\|\boldsymbol{\vartheta}_k\|_{\mathbb{L}_2}^2 + 2hp.$$

Therefore,

$$\|\boldsymbol{\vartheta}_k\|_{\mathbb{L}_2}^2 \leqslant (1-mh)^{2k}\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + \frac{2hp}{2mh-(mh)^2} \leqslant (1-mh)^{2k}\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + \frac{1.1p}{m}.$$

Using this inequality in conjunction with the fact that $1 - mh \leqslant \rho$, we arrive at

$$h\sum_{j=0}^{k}\rho^{k-j}\|\nabla f(\boldsymbol{\vartheta}_j)\|_{\mathbb{L}_2}^2 \leqslant \frac{1.3M}{2}\rho^{k+1}\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + \frac{1.3}{2}M\rho^{2k}\|\boldsymbol{\vartheta}_0\|_{\mathbb{L}_2}^2 + 2.9\kappa p.$$

This completes the proof of the lemma.

B THE PROOF OF THE UPPER BOUND ON THE ERROR OF KLMC

The goal of this section is to present the proof of the bound on the error of sampling of the "standard" discretization of the kinetic Langevin diffusion. With a slight abuse of language, we will call it Euler-Maruyama discretized kinetic Langevin diffusion, or kinetic Langevin Monte Carlo (KLMC). To avoid complicated notation, and since there is no risk of confusion, throughout this section $\boldsymbol{\vartheta}_k$ and $\boldsymbol{v}_k$ will refer to $\boldsymbol{\vartheta}_k^{\mathsf{KLMC}}$ and $\boldsymbol{v}_k^{\mathsf{KLMC}}$, respectively. We will also use the following shorthand notation:

$$f_n = f(\boldsymbol{\vartheta}_n), \quad \boldsymbol{g}_n = \nabla f_n = \nabla f(\boldsymbol{\vartheta}_n), \quad \eta = \gamma h, \quad M_\gamma = M/\gamma.$$

The advantage of dealing with η instead of h is that the former is scale-free.

Note that the iterates of KLMC satisfy

$$\boldsymbol{v}_{n+1} = (1-\alpha\eta)\boldsymbol{v}_n - \alpha\eta\, \boldsymbol{g}_n + \sqrt{2\gamma\eta}\,\sigma\boldsymbol{\xi}_n \tag{27}$$

$$\boldsymbol{\vartheta}_{n+1} = \boldsymbol{\vartheta}_n + \gamma^{-1}\eta\big(\alpha\boldsymbol{v}_n - \beta\eta\boldsymbol{g}_n + \sqrt{2\gamma\eta}\,\tilde\sigma\bar{\boldsymbol{\xi}}_n\big), \tag{28}$$

where

$$\alpha = \frac{1-e^{-\eta}}{\eta} \in (0,1), \qquad \beta = \frac{e^{-\eta}-1+\eta}{\eta^2} \in (0,1/2),$$

$$\sigma^2 = \frac{1-e^{-2\eta}}{2\eta} \in (0,1), \qquad \tilde\sigma^2 = \frac{2(1-2\eta+2\eta^2-e^{-2\eta})}{(2\eta)^3} \in (0,1/3)$$

and $\boldsymbol{\xi}_n, \bar{\boldsymbol{\xi}}_n$ are two $\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$-distributed random vectors independent of $(\boldsymbol{\vartheta}_n, \boldsymbol{v}_n)$.

Since we assume throughout this section that $2Mh \leqslant 0.1$, $\gamma \geqslant 2M$ and $\kappa \geqslant 10$, we have

$$\alpha = \frac{1-\exp(-\eta)}{\eta} \geq 0.95, \quad \text{and} \quad mh = \frac{Mh}{\kappa} \leqslant \frac{Mh}{10} \leqslant \frac{1}{200}.$$

The latter, in particular, implies the following bound for ϱ:

$$1 - mh \leqslant \varrho = e^{-mh} \leqslant 1 - 0.99mh = 1 - 0.99m\eta/\gamma. \tag{29}$$

For any sequence $\omega = (\omega_n)_{n\in\mathbb{N}}$ of real numbers, we denote by $S_n(\omega)$ the ρ-discounted sum $\sum_{k=0}^n \rho^{n-k}\omega_k$. Below we present a simple lemma for the function $S_n(\cdot)$ that we will use repeatedly in this proof.

Lemma 3 (Summation by parts)**.** *Suppose $\omega = (\omega_n)_{n\in\mathbb{N}}$ is a sequence of real numbers and define $S_n^{+1}(\omega) := \sum_{k=0}^n \varrho^{n-k}\omega_{k+1}$. Then, the following identity is true*

$$S_n^{+1}(\omega) = \omega_{n+1} - \varrho^{n+1}\omega_0 + \varrho S_n(\omega).$$

Proof. The proof is based on simple algebra:

$$S_n^{+1}(\omega) = \omega_{n+1} + \sum_{j=1}^n \varrho^{n-j+1}\omega_j = \omega_{n+1} + \varrho\left(S_n(\omega) - \varrho^n\omega_0\right). \qquad \square$$

B.1 EXPONENTIAL MIXING OF CONTINUOUS-TIME KINETIC LANGEVIN DIFFUSION

Consider the kinetic Langevin diffusions

$$d\boldsymbol{L}_t = \boldsymbol{V}_t\, dt \qquad d\boldsymbol{V}_t = -\gamma\boldsymbol{V}_t\, dt - \gamma\nabla f(\boldsymbol{L}_t)dt + \sqrt{2}\gamma\, d\boldsymbol{W}_t \tag{30}$$

Proposition 1. *Let $\boldsymbol{V}_0, \boldsymbol{L}_0$ and $\boldsymbol{L}_0'$ be random vectors in $\mathbb{R}^p$. Let $(\boldsymbol{V}_t, \boldsymbol{L}_t)$ and $(\boldsymbol{V}_t', \boldsymbol{L}_t')$ be kinetic Langevin diffusions defined in equation 30 driven by the same Brownian motion and starting from $(\boldsymbol{V}_0, \boldsymbol{L}_0)$ and $(\boldsymbol{V}_0', \boldsymbol{L}_0')$ respectively. It holds for any $t \geqslant 0$ that*

$$\left\|\mathbf{C}\begin{bmatrix}\boldsymbol{V}_t - \boldsymbol{V}_t'\\ \boldsymbol{L}_t - \boldsymbol{L}_t'\end{bmatrix}\right\| \leqslant e^{-\{m\wedge(\gamma-M)\}t}\left\|\mathbf{C}\begin{bmatrix}\boldsymbol{V}_0 - \boldsymbol{V}_0'\\ \boldsymbol{L}_0 - \boldsymbol{L}_0'\end{bmatrix}\right\|, \quad \textit{with} \quad \mathbf{C} = \begin{bmatrix}\mathbf{I}_p & \mathbf{0}_p\\ \mathbf{I}_p & \gamma\mathbf{I}_p\end{bmatrix}.$$

Proof of Proposition 1. Set $\boldsymbol{Y}_t := \boldsymbol{V}_t - \boldsymbol{V}_t' + \gamma(\boldsymbol{L}_t - \boldsymbol{L}_t')$, $\boldsymbol{Z}_t := \boldsymbol{V}_t - \boldsymbol{V}_t'$, that is

$$\begin{bmatrix}\boldsymbol{Z}_t\\ \boldsymbol{Y}_t\end{bmatrix} = \mathbf{C}\begin{bmatrix}\boldsymbol{V}_t - \boldsymbol{V}_t'\\ \boldsymbol{L}_t - \boldsymbol{L}_t'\end{bmatrix}.$$

We note that by the Taylor expansion, we have

$$\nabla f(\boldsymbol{L}_t) - \nabla f(\boldsymbol{L}_t') = \mathbf{H}_t(\boldsymbol{L}_t - \boldsymbol{L}_t'),$$

where $\mathbf{H}_t := \int_0^1 \nabla^2 f(\boldsymbol{L}_t - x(\boldsymbol{L}_t - \boldsymbol{L}_t'))\,\mathrm{d}x$. By the definition of $(\boldsymbol{V}_t, \boldsymbol{L}_t)$ and $(\boldsymbol{V}_t', \boldsymbol{L}_t')$, we find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{V}_t - \boldsymbol{V}_t' + \gamma(\boldsymbol{L}_t - \boldsymbol{L}_t')) &= -\gamma\mathbf{H}_t(\boldsymbol{L}_t - \boldsymbol{L}_t')\\ &= -\mathbf{H}_t(\boldsymbol{Y}_t - \boldsymbol{Z}_t).\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{V}_t - \boldsymbol{V}_t') &= -\gamma(\boldsymbol{V}_t - \boldsymbol{V}_t') - \gamma\mathbf{H}_t(\boldsymbol{L}_t - \boldsymbol{L}_t')\\ &= -\gamma\boldsymbol{Z}_t - \mathbf{H}_t(\boldsymbol{Y}_t - \boldsymbol{Z}_t).\end{aligned}$$

This implies

$$\begin{aligned}\frac{d}{dt}\big[\,\|\boldsymbol{Y}_t\|^2 + \|\boldsymbol{Z}_t\|^2\,\big] &= 2\boldsymbol{Y}_t^\top(-\mathbf{H}_t\boldsymbol{Y}_t + \mathbf{H}_t\boldsymbol{Z}_t) + 2\boldsymbol{Z}_t^\top(-\gamma\boldsymbol{Z}_t - \mathbf{H}_t\boldsymbol{Y}_t + \mathbf{H}_t\boldsymbol{Z}_t)\\ &\leqslant 2\big(-m\,\|\boldsymbol{Y}_t\|^2 - \gamma\,\big\|\boldsymbol{Z}_t^2\big\| + M\,\|\boldsymbol{Z}_t\|^2\big)\\ &\leqslant -2(m\wedge(\gamma - M))\left\|\begin{bmatrix}\boldsymbol{Z}_t\\ \boldsymbol{Y}_t\end{bmatrix}\right\|^2.\end{aligned}$$

Invoking Gronwall's inequality, we get

$$\left\|\begin{bmatrix}\boldsymbol{Z}_t\\ \boldsymbol{Y}_t\end{bmatrix}\right\| \leqslant \exp\big(-\{m\wedge(\gamma - M)\}t\big)\left\|\begin{bmatrix}\boldsymbol{Z}_0\\ \boldsymbol{Y}_0\end{bmatrix}\right\|$$

as desired. □

B.2 Proof of Theorem 3

Let $\boldsymbol{\vartheta}_n$, $\boldsymbol{v}_n$ be the iterates of the KLMC algorithm. Let $(\boldsymbol{L}_t, \boldsymbol{V}_t)$ be the kinetic Langevin diffusion, coupled with $(\boldsymbol{\vartheta}_n, \boldsymbol{v}_n)$ through the same Brownian motion $(\boldsymbol{W}_t; t \geqslant 0)$ and starting from a random point $(\boldsymbol{L}_0, \boldsymbol{V}_0) \propto \exp(-f(\boldsymbol{y}) + \frac{1}{2\gamma}\|\boldsymbol{v}\|_2^2)$ such that $\boldsymbol{V}_0 = \boldsymbol{v}_0$. This means that

$$\begin{aligned}\boldsymbol{v}_{n+1} &= \boldsymbol{v}_n e^{-\eta} - \gamma\int_0^h e^{-\gamma(h-s)}\,\mathrm{d}s\nabla f(\boldsymbol{\vartheta}_n) + \sqrt{2}\,\gamma\int_0^h e^{-\gamma(h-s)}\,\mathrm{d}\boldsymbol{W}_s\\ \boldsymbol{\vartheta}_{n+1} &= \boldsymbol{\vartheta}_n + \int_0^h\left(\boldsymbol{v}_n e^{-\gamma u} - \gamma\int_0^u e^{-\gamma(u-s)}\,\mathrm{d}s\nabla f(\boldsymbol{\vartheta}_n) + \sqrt{2}\,\gamma\int_0^u e^{-\gamma(u-s)}\,\mathrm{d}\boldsymbol{W}_s\right)\mathrm{d}u.\end{aligned}$$

We also consider the kinetic Langevin diffusion, $(\boldsymbol{L}', \boldsymbol{V}')$, defined on $[0, h]$ with the starting point $(\boldsymbol{\vartheta}_n, \boldsymbol{v}_n)$ and driven by the Brownian motion $(\boldsymbol{W}_{nh+t} - \boldsymbol{W}_{nh}; t \in [0, h])$. It satisfies

$$\begin{aligned}\boldsymbol{V}_t' &= \boldsymbol{v}_n e^{-\gamma t} - \gamma\int_0^t e^{-\gamma(t-s)}\nabla f(\boldsymbol{L}_s')\,\mathrm{d}s + \sqrt{2}\,\gamma\int_0^t e^{-\gamma(t-s)}\,\mathrm{d}\boldsymbol{W}_s\\ \boldsymbol{L}_t' &= \boldsymbol{\vartheta}_n + \int_0^t \boldsymbol{V}_s'\,\mathrm{d}s.\end{aligned}$$

Our goal will be to bound the term x_n defined by

$$x_n = \left\|\mathbf{C}\begin{bmatrix}\boldsymbol{v}_n - \boldsymbol{V}_{nh}\\ \boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh}\end{bmatrix}\right\|_{\mathbb{L}_2} \quad \text{with} \quad \mathbf{C} = \begin{bmatrix}\mathbf{I}_p & \mathbf{0}_p\\ \mathbf{I}_p & \gamma\mathbf{I}_p\end{bmatrix}. \tag{31}$$

The triangle inequality yields

$$
\begin{aligned}
x_{n+1} &\leqslant \left\| \mathbf{C} \begin{bmatrix} \boldsymbol{v}_n - \boldsymbol{V}_h' \\ \boldsymbol{\vartheta}_n - \boldsymbol{L}_h' \end{bmatrix} \right\|_{\mathbb{L}_2} + \left\| \mathbf{C} \begin{bmatrix} \boldsymbol{V}_h' - \boldsymbol{V}_{nh} \\ \boldsymbol{L}_h' - \boldsymbol{L}_{nh} \end{bmatrix} \right\|_{\mathbb{L}_2} \\
&\leqslant \left\| \mathbf{C} \begin{bmatrix} \boldsymbol{v}_n - \boldsymbol{V}_h' \\ \boldsymbol{\vartheta}_n - \boldsymbol{L}_h' \end{bmatrix} \right\|_{\mathbb{L}_2} + \varrho x_n \\
&\leqslant \varrho x_n + \sqrt{2}\, \|\boldsymbol{v}_{n+1} - \boldsymbol{V}_h'\|_{\mathbb{L}_2} + \gamma \|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{L}_h'\|_{\mathbb{L}_2}, \qquad (32)
\end{aligned}
$$

where the second inequality follows from Proposition 1 (see also the proof of (Dalalyan & Riou-Durand, 2020, Prop. 1)), while the third inequality is a consequence of the elementary inequality $\sqrt{a^2 + (a+b)^2} \leqslant \sqrt{2}\, a + b$ for $a, b \geqslant 0$.

The next lemma gives an upper bound on the terms appearing in the right-hand side of equation 32.

Lemma 4. *If ∇f is M-Lipschitz continuous, then for every step-size $\eta = \gamma h \geqslant 0$ and every $\gamma \geqslant 0$, the following holds*

$$
\begin{aligned}
\|\boldsymbol{v}_{n+1} - \boldsymbol{V}_h'\|_{\mathbb{L}_2} &\leqslant \tfrac{1}{6}\big\{2\sqrt{\gamma p \eta} + 3\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big\} M_\gamma \eta^2 e^{M_\gamma \eta^2/2} \\
\gamma \|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{L}_h'\|_{\mathbb{L}_2} &\leqslant \tfrac{1}{6}\big(0.6\sqrt{\gamma p \eta} + \|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.25\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \big) M_\gamma \eta^3 e^{M_\gamma \eta^2/2},
\end{aligned}
$$

where $M_\gamma = M/\gamma$.

For $\eta \leqslant 0.2$ and $\gamma \geqslant 2M$, Lemma 4 implies

$$
\begin{aligned}
\|\boldsymbol{v}_{n+1} - \boldsymbol{V}_h'\|_{\mathbb{L}_2} &\leqslant M_\gamma \eta^2 (0.15\sqrt{\gamma p} + 0.51\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.17\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}) \\
\gamma \|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{L}_h'\|_{\mathbb{L}_2} &\leqslant M_\gamma \eta^3 (0.046\sqrt{\gamma p} + 0.17\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.043\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}).
\end{aligned}
$$

Therefore,

$$
\sqrt{2}\, \|\boldsymbol{v}_{n+1} - \boldsymbol{V}_h'\|_{\mathbb{L}_2} + \gamma \|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{L}_h'\|_{\mathbb{L}_2} \leqslant M_\gamma \eta^2 \big(0.23\sqrt{\gamma p} + 0.74\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.25\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big). \qquad (33)
$$

Combining equation 32 and equation 33, we get

$$
x_{n+1} \leqslant \varrho x_n + M_\gamma \eta^2 \big(0.23\sqrt{\gamma p} + 0.74\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.25\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big).
$$

From the last display, we infer that

$$
x_n \leqslant \varrho^n x_0 + M_\gamma \eta^2 \sum_{k=0}^{n-1} \varrho^{n-1-k} \big(0.23\sqrt{\gamma p} + 0.74\|\boldsymbol{v}_k\|_{\mathbb{L}_2} + 0.25\eta \|\boldsymbol{g}_k\|_{\mathbb{L}_2}\big).
$$

This implies that

$$
\begin{aligned}
x_n &\leqslant \varrho^n x_0 + \frac{0.23 M_\gamma \eta^2 \sqrt{\gamma p}}{1-\varrho} + 0.74 M_\gamma \eta^2 \sum_{k=1}^{n} \varrho^{n-k} \big(\|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2} + 0.33\eta \|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2}\big) \\
&\leqslant \varrho^n x_0 + \frac{0.23 M_\gamma \eta^2 \sqrt{\gamma p}}{1-\varrho} + \frac{0.74 M_\gamma \eta^2}{\sqrt{1-\varrho}} \bigg\{ \sum_{k=1}^{n} \varrho^{n-k} \big(\|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2}^2 + 0.33\eta^2 \|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2}^2\big) \bigg\}^{1/2}.
\end{aligned}
$$

In view of equation 29, $\varrho \leqslant 1 - 0.99 m\eta/\gamma$ and

$$
x_n \leqslant \varrho^n x_0 + 0.233\kappa\eta\sqrt{\gamma p} + 0.74 M_\gamma \eta \bigg\{ \frac{\gamma\eta}{m\varrho} \sum_{k=0}^{n} \varrho^{n-k} \big(\|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 + 0.33\eta^2 \|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2\big) \bigg\}^{1/2}.
$$

Proposition 2. *Assume that $\boldsymbol{v}_0 \sim \mathcal{N}(0, \gamma \mathbf{I}_p)$ is independent of $\boldsymbol{\vartheta}_0$. If $\gamma \geqslant 5M$, $\kappa \geqslant 10$ and $\eta \leqslant 1/10$ then*

$$
\begin{aligned}
\eta \sum_{k=0}^{n} \varrho^{n-k} \|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 &\leqslant 4.42 \varrho^n \gamma\, \mathbb{E}[f_0] + \frac{1.11\gamma^2 p}{m} + 4.98\big(x_n + 0.96\sqrt{\gamma p}\big)^2 \\
\eta \sum_{k=0}^{n} \varrho^{n-k} \|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 &\leqslant 3.93 \varrho^n \gamma \mathbb{E}[f_0] + \frac{1.87\gamma^2 p}{m} + 3.2\big(x_n + 0.96\sqrt{\gamma p}\big)^2.
\end{aligned}
$$

We can apply Proposition 2 and $\varrho \geqslant 0.998$ to infer that

$$\begin{aligned}
x_n &\leqslant \varrho^n x_0 + 0.233\kappa\eta\sqrt{\gamma p} + 0.74M_\gamma\eta\bigg\{\frac{\gamma}{m\varrho}\Big(3.98\varrho^n\gamma\,\mathbb{E}[f_0] + \frac{1.87\gamma^2 p}{m} + 3.25\big(x_n + 0.96\sqrt{\gamma p}\big)^2\Big)\bigg\}^{1/2}\\
&\leqslant \varrho^n x_0 + 0.233\kappa\eta\sqrt{\gamma p} + 0.74M_\gamma\eta\bigg\{\frac{\gamma}{m}\Big(3.99\varrho^n\gamma\,\mathbb{E}[f_0] + \frac{1.88\gamma^2 p}{m} + 3.26\big(x_n + 0.96\sqrt{\gamma p}\big)^2\Big)\bigg\}^{1/2}\\
&\leqslant \varrho^n x_0 + +0.233\kappa\eta\sqrt{\gamma p} + \frac{1.54M\eta}{\sqrt{m}}\sqrt{\varrho^n\mathbb{E}[f_0]} + 0.62\sqrt{\kappa}\,\eta x_n + 0.74M_\gamma\eta\Big(\frac{1.88\gamma^3 p}{m^2} + \frac{3\gamma^2 p}{m}\Big)^{1/2}\\
&\leqslant \varrho^n x_0 + 0.62\sqrt{\kappa}\,\eta x_n + \frac{1.54M\eta}{\sqrt{m}}\sqrt{\varrho^n\mathbb{E}[f_0]} + \frac{M\eta\sqrt{\gamma p}}{m}\big(0.233 + 0.74\sqrt{1.94}\big).
\end{aligned}$$

Therefore, under the condition $\sqrt{\kappa}\,\eta \leqslant 0.1$,

$$x_n \leqslant 1.07\varrho^n x_0 + \frac{1.65M\eta}{\sqrt{m}}\sqrt{\varrho^n\mathbb{E}[f_0]} + \frac{1.35M\eta\sqrt{\gamma p}}{m}.$$

Finally, one can check that $2x_n^2 \geqslant \gamma^2\|\boldsymbol{\vartheta}_n - \bar{\boldsymbol{\vartheta}}_n\|_{\mathbb{L}_2}^2 \geqslant \gamma^2\mathsf{W}_2^2(\nu_n^{\mathsf{KLMC}}, \pi)$ and $x_0 = \gamma\|\boldsymbol{\vartheta}_0 - \boldsymbol{L}_0\|_{\mathbb{L}_2} = \gamma\mathsf{W}_2(\nu_0, \pi)$. This completes the proof of the theorem.

B.3 Proof of Proposition 2 (discounted sums of squared gradients and velocities)

To ease the notation, we set $z_n := \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n]$ and define

$$S_n(z) := \sum_{k=0}^n \varrho^{n-k} z_k, \qquad S_n(g^2) := \sum_{k=0}^n \varrho^{n-k}\,\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2\,,$$
$$S_n(f) := \sum_{k=0}^n \varrho^{n-k}\mathbb{E}[f_k], \qquad S_n(v^2) := \sum_{k=0}^n \varrho^{n-k}\|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2\,.$$

Throughout the proof, we will need some technical results that will be stated as lemmas and their proof will be postponed to Appendix B.5.

Lemma 5. *If for some $M \geqslant 0$, the gradient ∇f is M-Lipschitz continuous, then for every step-size $h > 0$ and every $\gamma > 0$ it holds for the KLMC iterates defined in equation 28 that*

$$\big|z_{n+1} - (1-\alpha\eta)z_n + \alpha\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2\big| \leqslant \eta M_\gamma\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{5}{8}\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \tfrac{4}{3}\eta\gamma p - \tilde{\alpha}\eta z_n\big)$$

for some positive number $\tilde{\alpha}\eta \leqslant 0.14$.

Since $M_\gamma \leqslant 1/5$ and $\eta \leqslant 0.1$, we have

$$\alpha - \tfrac{5}{8}M_\gamma\eta^2 \geqslant \tfrac{1}{\eta}(1 - e^{-\eta}) - \tfrac{1}{8}\eta^2 \geqslant 10(1 - e^{-0.1}) - \tfrac{1}{8}0.1^2 \geqslant 0.94.$$

Therefore, we can rewrite the claim of Lemma 5 with the notation $\tilde{\beta} = \eta M_\gamma\tilde{\alpha}$ as follows:

$$z_{n+1} \leqslant (1 - \alpha\eta - \tilde{\beta}\eta)z_n + 0.2\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.67\gamma p\eta^2 - 0.94\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2.$$

Lemma 6. *Let $\tilde{\beta} \leqslant 0.014$ and $\eta \in [0, 0.1]$. If $z_0 = 0$ and the sequences $\{z_n\} \subset \mathbb{R}$, $\{\boldsymbol{v}_n\} \subset \mathbb{R}^p$ and $\{\boldsymbol{g}_n\} \subset \mathbb{R}^p$ satisfy the inequality*

$$z_{n+1} \leqslant (e^{-\eta} - \tilde{\beta}\eta)z_n + 0.2\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.67\gamma p\eta^2 - 0.94\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 \tag{34}$$

then for every $\varrho \in [0, 1]$ such that $\varrho \geqslant e^{-\eta}$, it holds that

$$S_n(g^2) \leqslant 1.09\alpha(S_n(z))_- + 0.213S_n(v^2) + \frac{0.73\eta\gamma p}{1-\varrho} - \frac{1.07z_{n+1}}{\eta}, \tag{35}$$

where $(S_n(z))_- = \max(0, S_n(z))$ is the negative part of $S_n(z)$ and $\alpha = (1 - e^{-\eta})/\eta$.

In order to get rid of the last term in equation 35, we need a bound on $(S_n(z))_-$. To this end, we use the smoothness of the function f, in conjunction with equation 28, to infer that

$$\begin{aligned}2\gamma\mathbb{E}[f_{n+1}-f_n] &\leqslant 2\gamma\mathbb{E}[\boldsymbol{g}_n^\top(\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n)] + M_\gamma\|\gamma(\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2\\ &\leqslant 2\alpha\eta(1-\beta M_\gamma\eta^2)z_n - \beta\eta^2(2-\beta M_\gamma\eta^2)\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + M_\gamma\alpha^2\eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{2}{3}M_\gamma\eta^3\gamma p\\ &\leqslant 2\alpha_0\eta z_n - 0.96\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.0182\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{2}{15}\eta^3\gamma p\end{aligned}$$

with $\alpha_0 = \alpha(1-\beta M_\gamma\eta^2) \geqslant \alpha(1-0.1\times 0.1^2) \geqslant 0.999\alpha$.

Lemma 7. *Let $\alpha_0,\gamma,\eta>0$. If the sequences $\{F_n\}\subset\mathbb{R}$, $\{\boldsymbol{g}_n\}\subset\mathbb{R}^p$ and $\{\boldsymbol{v}_n\}\subset\mathbb{R}^p$ satisfy $F_n\geqslant 0$ and*

$$2(F_{n+1}-F_n) \leqslant 2\alpha_0\eta z_n + 0.0182\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{2}{15}\eta^3\gamma p \tag{36}$$

then, for every $\varrho\in(0,1)$, it holds that

$$\alpha_0(\eta S_n(z))_- \leqslant \varrho^n F_0 + (1-\varrho)\gamma S_n(F) + 0.0182\eta S_n(v^2) + \frac{2\eta^3\gamma p}{15(1-\varrho)}.$$

In view of the strong convexity of the potential function and the assumption that $f(\boldsymbol{\theta}_*)=0$, the Polyak-Lojasiewicz inequality

$$f_n \leqslant \tfrac{1}{2m}\|\boldsymbol{g}_n\|^2$$

holds true. This implies that $(1-\varrho)S_n(f)\leqslant(\eta/\gamma)mS_n(f)\leqslant\frac{1}{2}(\eta/\gamma)S_n(g^2)$. Combining this inequality with the claim of Lemma 7, applied to $F_n=\gamma\mathbb{E}[f_n]$, we get

$$\boxed{0.999\alpha(S_n(z))_- \leqslant \frac{\varrho^n\gamma}{\eta}\mathbb{E}[f_0] + 0.5S_n(g^2) + 0.0182S_n(v^2) + \frac{0.14\eta\gamma^2 p}{m}.} \tag{37}$$

Let us now combine equation 35 and equation 37:

$$\begin{aligned}S_n(g^2) &\leqslant \frac{1.1\varrho^n\gamma}{\eta}\mathbb{E}[f_0] + 0.55S_n(g^2) + 0.02S_n(v^2) + \frac{0.16\eta\gamma^2 p}{m}\\ &\quad + 0.213S_n(v^2) + \frac{0.73\gamma^2 p}{m} + \frac{1.07|z_{n+1}|}{\eta}\\ &\leqslant \frac{1.1\varrho^n\gamma}{\eta}\mathbb{E}[f_0] + 0.55S_n(g^2) + 0.223S_n(v^2) + \frac{0.75\gamma^2 p}{m} + \frac{1.07|z_{n+1}|}{\eta}.\end{aligned}$$

Subtracting $0.55S_n(g^2)$ from both sides and dividing by 0.45, we obtain

$$\boxed{S_n(g^2) \leqslant \frac{2.45\varrho^n\gamma\mathbb{E}[f_0]}{\eta} + 0.5S_n(v^2) + \frac{1.7\gamma^2 p}{m} + \frac{2.38|z_{n+1}|}{\eta}.} \tag{38}$$

Let us now derive a bound for $S_n(v^2)$. We start with the following property, which is a direct consequence of the definition of $\boldsymbol{v}_{n+1}$:

$$\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}^2 - \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 \leqslant -\alpha\eta(2-\alpha\eta)\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 - 2\alpha\eta(1-\alpha\eta)z_n + \alpha^2\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2\eta\gamma p.$$

Using the same technique as before and applying Lemma 3, we deduce the following:

$$\begin{aligned}(\varrho-1)S_n(v^2) - \varrho^n\|\boldsymbol{v}_0\|_{\mathbb{L}_2}^2 &= (\varrho-1)S_n(v^2) - \varrho^n\gamma p\\ &\leqslant -\alpha\eta(2-\alpha\eta)S_n(v^2) - 2\alpha\eta(1-\alpha\eta)S_n(z) + \alpha^2\eta^2 S_n(g^2) + \frac{2\eta\gamma p}{1-\varrho}.\end{aligned}$$

Therefore, since $\varrho\geqslant 1-\frac{m}{\gamma}\eta$,

$$(2\alpha-\alpha^2\eta-\tfrac{m}{\gamma})\eta S_n(v^2) \leqslant -2\alpha\eta(1-\alpha\eta)S_n(z) + (\alpha\eta)^2 S_n(g^2) + \frac{2.021\gamma^2 p}{m}.$$

Since $\alpha=(1-e^{-\eta})/\eta$ with $\eta\leqslant 0.1$, from the last display, we infer that

$$S_n(v^2) \leqslant 1.02\alpha\big(S_n(z)\big)_- + 0.51\eta S_n(g^2) + \frac{1.13\gamma^2 p}{m\eta}.$$

Combining this inequality with equation 37, implies

$$\begin{aligned} S_n(v^2) &\leqslant \frac{1.03\varrho^n\gamma}{\eta}[f_0] + 0.562 S_n(g^2) + 0.019 S_n(v^2) + \frac{0.2\eta\gamma^2 p}{m} + 0.51\eta S_n(g^2) + \frac{1.13\gamma^2 p}{m\eta} \\ &\leqslant \frac{1.03\varrho^n\gamma}{\eta} f_0 + 0.62 S_n(g^2) + 0.019 S_n(v^2) + \frac{1.13\gamma^2 p}{m\eta} \end{aligned}$$

Therefore, subtracting $0.019 S_n(v^2)$ and dividing by $(1 - 0.019)$, we get

$$\boxed{S_n(v^2) \leqslant \frac{1.1\varrho^n\gamma}{\eta} f_0 + 0.64 S_n(g^2) + \frac{1.16\gamma^2 p}{m\eta}} \tag{39}$$

Combining equation 38 and equation 39, we arrive at

$$\begin{aligned} S_n(v^2) &\leqslant \frac{1.1\varrho^n\gamma}{\eta}\mathbb{E}[f_0] + \frac{1.16\gamma^2 p}{m\eta} + 0.64\Big(\frac{2.45\varrho^n\gamma}{\eta}\mathbb{E}[f_0] + 0.5 S_n(v^2) + \frac{1.7\gamma^2 p}{m} + \frac{2.38|z_{n+1}|}{\eta}\Big) \\ &\leqslant \frac{2.67\varrho^n\gamma}{\eta} f_0 + \frac{1.27\gamma^2 p}{m\eta} + \frac{1.53|z_{n+1}|}{\eta} + 0.32 S_n(v^2). \end{aligned}$$

Therefore, subtracting $0.32 S_n(v^2)$ and dividing by $(1 - 0.32)$, we get

$$\boxed{S_n(v^2) \leqslant \frac{3.93\varrho^n\gamma}{\eta} f_0 + \frac{1.87\gamma^2 p}{m\eta} + \frac{2.25|z_{n+1}|}{\eta}.}$$

Once again, combining with equation 38, we get

$$S_n(g^2) \leqslant \frac{2.45\varrho^n\gamma}{\eta}\mathbb{E}[f_0] + 0.5\Big(\frac{3.93\varrho^n\gamma}{\eta} f_0 + \frac{1.87\gamma^2 p}{m\eta} + \frac{2.25|z_{n+1}|}{\eta}\Big) + \frac{1.7\gamma^2 p}{m} + \frac{2.38|z_{n+1}|}{\eta}$$

that leads to

$$\boxed{S_n(g^2) \leqslant \frac{4.42\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0] + \frac{1.11\gamma^2 p}{m\eta} + \frac{3.51|z_{n+1}|}{\eta}.} \tag{40}$$

The last lemma we need is the one providing an upper bound on $|z_{n+1}|$.

Lemma 8. *For every $\eta \leqslant 0.1$ and $\gamma \geqslant 5M$, we have*

$$|z_{n+1}| \leqslant \big(1.19 x_n + 1.14\sqrt{\gamma p}\big)^2,$$

where x_n is given by equation 31.

Using Lemma 8 in conjunction with equation 39 and equation 40, we arrive at the inequalities stated in Proposition 2.

B.4 Proof of Lemma 4 (one-step discretization error)

We use the notation

$$\psi_0(t) = e^{-\gamma t}, \qquad \psi_1(t) = \frac{1 - e^{-\gamma t}}{\gamma}, \qquad \psi_2(t) = \frac{e^{-\gamma t} - 1 + \gamma t}{\gamma}$$

and note that

$$\psi_1(t) \leqslant t, \qquad \psi_2(t) \leqslant 0.5\gamma t^2.$$

Furthermore,

$$\begin{aligned} \|\boldsymbol{v}_{n+1} - \boldsymbol{V}'_h\|_{\mathbb{L}_2} &= \gamma\Big\|\int_0^h e^{-\gamma(h-s)}\big(\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{\vartheta}_n)\big)\,\mathrm{d}s\Big\|_{\mathbb{L}_2} \\ &\leqslant \gamma\int_0^h e^{-\gamma(h-s)}\big\|\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{\vartheta}_n)\big\|_{\mathbb{L}_2}\,\mathrm{d}s \\ &\leqslant M\gamma\int_0^h \big\|\boldsymbol{L}'_s - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2}\,\mathrm{d}s, \end{aligned} \tag{41}$$

where the last implication is due to the M-smoothness of the potential function f. On the one hand, for every $s \in [0, h]$, we have

$$
\begin{aligned}
\boldsymbol{L}'_s - \boldsymbol{\vartheta}_n &= \int_0^s \boldsymbol{V}'_u \, \mathrm{d}u \\
&= \psi_1(s)\, \boldsymbol{v}_n - \gamma \int_0^s \int_0^u e^{-\gamma(u-t)} \big(\nabla f(\boldsymbol{L}'_t) - \nabla f(\boldsymbol{\vartheta}_n)\big)\, \mathrm{d}t \mathrm{d}u - \psi_2(s)\, \boldsymbol{g}_n \\
&\quad + \sqrt{2}\,\gamma \int_0^s \int_0^u e^{-\gamma(u-t)}\, \mathrm{d}\boldsymbol{W}_t \, \mathrm{d}u \\
&= \psi_1(s)\, \boldsymbol{v}_n - \psi_2(s)\, \boldsymbol{g}_n + \sqrt{2}\,\gamma \int_0^s \psi_1(t)\, \mathrm{d}\boldsymbol{W}_t \\
&\quad - \gamma \int_0^s \psi_1(s-t) \big(\nabla f(\boldsymbol{L}'_t) - \nabla f(\boldsymbol{\vartheta}_n)\big)\, \mathrm{d}t.
\end{aligned}
$$

Therefore,

$$
\big\|\boldsymbol{L}'_s - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2} \leqslant s\, \|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.5\gamma s^2\, \|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \sqrt{2ps/3}\,\gamma s + M\gamma \int_0^s (s-t) \big\|\boldsymbol{L}'_t - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2}\, \mathrm{d}t.
$$

The last inequality combined with $s - t \leqslant h - t$ allows us to use the Grönwall lemma, which implies that

$$
\begin{aligned}
\big\|\boldsymbol{L}'_s - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2} &\leqslant \big(s\, \|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.5\gamma s^2\, \|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \sqrt{(2/3)ps}\,\gamma s\big) e^{M\gamma s(h-0.5s)} \\
&\leqslant \big(s\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.5\gamma s^2\, \|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \sqrt{(2/3)ps}\,\gamma s\big) e^{0.5M\gamma h^2}. \qquad (42)
\end{aligned}
$$

Combining the last display with equation 41, we get

$$
\begin{aligned}
\big\|\boldsymbol{v}_{n+1} - \boldsymbol{V}'_h\big\|_{\mathbb{L}_2} &\leqslant \big\{\tfrac{1}{2}\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \tfrac{1}{6}\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.33\sqrt{\gamma p\eta}\big\} M h^2 \gamma e^{M\gamma h^2/2} \\
&\leqslant \big\{\tfrac{1}{2}\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \tfrac{1}{6}\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.33\sqrt{\gamma p\eta}\big\} M_\gamma \eta^2 e^{M_\gamma \eta^2/2}.
\end{aligned}
$$

This completes the proof of the first inequality. To prove the second one, we again use the update rules of $\boldsymbol{\theta}_{n+1}$ and $\boldsymbol{L}'$:

$$
\begin{aligned}
\big\|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{L}'_h\big\|_{\mathbb{L}_2} &= \gamma \bigg\| \int_0^h \int_0^t e^{-\gamma^2(t-s)} \big(\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{\vartheta}_n)\big)\, \mathrm{d}s\, \mathrm{d}t \bigg\|_{\mathbb{L}_2} \\
&\leqslant \gamma \int_0^h \int_0^t \big\|\nabla f(\boldsymbol{L}'_s) - \nabla f(\boldsymbol{\vartheta}_n)\big\|_{\mathbb{L}_2}\, \mathrm{d}s\, \mathrm{d}t \\
&\leqslant M\gamma \int_0^h \int_0^t \big\|\boldsymbol{L}'_s - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2}\, \mathrm{d}s\, \mathrm{d}t.
\end{aligned}
$$

The last term can be bounded using equation 42. This yields

$$
\begin{aligned}
\gamma\big\|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{L}'_h\big\|_{\mathbb{L}_2} &\leqslant M_\gamma \gamma^3 e^{M_\gamma \eta^2/2} \int_0^h \int_0^t \big(s\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.5\gamma s^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \sqrt{(2/3)ps^3}\,\gamma\big)\, \mathrm{d}s\, \mathrm{d}t \\
&\leqslant M_\gamma \eta^3 e^{M_\gamma \eta^2/2} \big(\tfrac{1}{6}\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \tfrac{1}{24}\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.1\sqrt{p\eta\gamma}\big) \\
&\leqslant (1/6) M_\gamma \eta^3 e^{M_\gamma \eta^2/2} \big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.25\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.6\sqrt{p\eta\gamma}\big)
\end{aligned}
$$

as desired.

B.5 Proofs of the technical lemmas used in Proposition 2

B.5.1 Proof of Lemma 5

Since $z_n = \mathbb{E}[\boldsymbol{g}_n^\top \boldsymbol{v}_n]$, we have

$$
\big|z_{n+1} - z_n - \mathbb{E}[\boldsymbol{g}_n^\top (\boldsymbol{v}_{n+1} - \boldsymbol{v}_n)]\big| = \big|\mathbb{E}\big[(\boldsymbol{g}_{n+1} - \boldsymbol{g}_n)^\top \boldsymbol{v}_{n+1}\big]\big|. \qquad (43)
$$

On the one hand, definition equation 27 of $\boldsymbol{v}_{n+1}$ yields

$$\mathbb{E}[\boldsymbol{g}_n^\top(\boldsymbol{v}_{n+1}-\boldsymbol{v}_n)] = -\alpha\eta\mathbb{E}[\boldsymbol{g}_n^\top\boldsymbol{v}_n] - \alpha\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2. \tag{44}$$

On the other hand, the Cauchy-Schwartz inequality implies

$$\begin{aligned}\big|\mathbb{E}\big[(\boldsymbol{g}_{n+1}-\boldsymbol{g}_n)^\top\boldsymbol{v}_{n+1}\big]\big| &\leqslant \big\|\boldsymbol{g}_{n+1}-\boldsymbol{g}_n\big\|_{\mathbb{L}_2}\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}\\ &\leqslant M\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}.\end{aligned}$$

Similarly, using update rules equation 27 and equation 28 of the KLMC, and the triangle inequality we get

$$\begin{aligned}\gamma^2\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2 &\leqslant \eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{1}{4}\eta^4\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 - 2\alpha\beta\eta^2 z_n + \tfrac{2}{3}\eta^3\gamma p\\ \|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}^2 &\leqslant \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 - 2\alpha\eta(1-\alpha\eta)z_n + 2\eta\gamma p.\end{aligned}$$

Hence,

$$\begin{aligned}\tfrac{\gamma}{\eta}\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2} &\leqslant \frac{(\gamma/\eta)^2\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}^2}{2}\\ &\leqslant \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{5}{8}\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 - \alpha\eta(\beta+1-\alpha\eta)z_n + \tfrac{4}{3}\eta\gamma p. \end{aligned}\tag{45}$$

Therefore, combining equation 43, equation 44 and equation 45, we get

$$\begin{aligned}\big|z_{n+1}-(1-\alpha\eta)z_n + \alpha\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2\big| &\leqslant \eta M_\gamma\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{5}{8}\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \tfrac{4}{3}\eta\gamma p\big)\\ &\qquad - \eta^2 M_\gamma \underbrace{\alpha(\beta+1-\alpha\eta)}_{:=\tilde{\alpha}} z_n\end{aligned}$$

with $\tilde{\alpha}\eta \leqslant \alpha\eta(1.5-\alpha\eta) \leqslant 0.14$, as desired.

B.5.2 Proof of Lemma 6

We apply inequality equation 34 for every index $k \leqslant n$ multiply each side by ϱ^{n-k}:

$$\varrho^{n-k}z_{k+1} \leqslant (e^{-\eta}-\tilde{\beta}\eta)\varrho^{n-k}z_k + 0.2\varrho^{n-k}\eta\|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 + 0.67\eta^2\gamma p\varrho^{n-k} - 0.94\eta\varrho^{n-k}\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2.$$

Summing over k, applying Lemma 3 and taking into account that $z_0 = 0$, we get

$$\begin{aligned}z_{n+1} + \varrho S_n(z) &\leqslant (e^{-\eta}-\tilde{\beta}\eta)S_n(z) + 0.2\eta S_n(v^2) + \frac{0.67\eta^2\gamma p}{1-\varrho} - 0.94\eta S_n(g^2)\\ &\leqslant (e^{-\eta}-\tilde{\beta}\eta)S_n(z) + 0.2\eta S_n(v^2) + \frac{0.68\eta^2\gamma p}{1-\varrho} - 0.94\eta S_n(g^2).\end{aligned}$$

This implies that

$$0.94\eta S_n(g^2) \leqslant (\varrho - e^{-\eta} + \tilde{\beta}\eta)(-S_n(z)) + 0.2\eta S_n(v^2) + \frac{0.68\eta^2\gamma p}{1-\varrho} - z_{n+1}.$$

Note that $\rho - e^{-\eta} \geqslant 0$ and $\varrho - e^{-\eta} + \tilde{\beta}\eta \leqslant 1 - e^{-\eta} + 0.014\eta \leqslant 1.02(e^{-\eta}-1) = 1.02\alpha\eta$. Therefore,

$$0.94\eta S_n(g^2) \leqslant 1.02\alpha\eta(S_n(z))_- + 0.2\eta S_n(v^2) + \frac{0.68\eta^2\gamma p}{1-\varrho} - z_{n+1}.$$

Dividing both sides of the last display by 0.94η, we get

$$S_n(g^2) \leqslant 1.09\alpha(S_n(z))_- + 0.213 S_n(v^2) + \frac{0.73\eta\gamma p}{1-\varrho} - \frac{1.07 z_{n+1}}{\eta}.$$

This completes the proof of the lemma.

B.5.3 PROOF OF LEMMA 7

We write inequality equation 36 for all indices k and multiply both sides of it by ϱ^{n-k}. Summing the obtained inequalities and applying Lemma 3, we obtain the following:

$$2(\varrho-1)S_n(F) - 2\varrho^n F_0 \leqslant 2\alpha_0\eta S_n(z) + 0.0.182\eta S_n(v^2) + \frac{2\eta^3\gamma p}{15(1-\varrho)},$$

where the left-hand side is obtained using Lemma 3 and the fact that $F_n \geqslant 0$. Rearranging the terms and dividing by 2, we obtain

$$-\alpha_0\eta S_n(z) \leqslant \varrho^n F_0 + (1-\varrho)S_n(F) + 0.0.182\eta S_n(v^2) + \frac{2\eta^3\gamma p}{15(1-\varrho)}.$$

Since the right-hand side of the last display is nonnegative, we infer that

$$\alpha_0(\eta S_n(z))_- \leqslant \varrho^n F_0 + (1-\varrho)S_n(F) + 0.0.182\eta S_n(v^2) + \frac{2\eta^3\gamma p}{15(1-\varrho)},$$

which coincides with the claim of the lemma.

B.5.4 PROOF OF LEMMA 8

In view of Lemma 5, we have

$$\begin{aligned}
|z_{n+1}| &\leqslant (e^{-\eta} + 0.3\eta^2)|z_n| + \eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \eta\big(0.2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.2\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.027\eta\gamma p\big) \\
&\leqslant 0.56(\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \|\boldsymbol{v}_n\|_{\mathbb{L}_2})^2 + 0.0027\gamma p \\
&\leqslant 0.56\big(\|\boldsymbol{g}_n - \nabla f(\boldsymbol{L}_{nh})\|_{\mathbb{L}_2} + \|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2} + \sqrt{Mp} + \sqrt{\gamma p}\big)^2 + 0.0027\gamma p,
\end{aligned}$$

where we have used the facts $\|\nabla f(\boldsymbol{L}_{nh})\|_{\mathbb{L}_2} = \int \|\nabla f\|^2\, d\pi \leqslant Mp$ (Dalalyan & Karagulyan, 2019, Lemma 3) and $\mathbb{E}[\|\boldsymbol{V}_{nh}\|^2] = \gamma p$. Finally, one can note that

$$\begin{aligned}
\|\boldsymbol{g}_n - \nabla f(\boldsymbol{L}_{nh})\|_{\mathbb{L}_2} + \|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2} &\leqslant 0.5\gamma\|\boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh}\|_{\mathbb{L}_2} + \|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2} \\
&\leqslant 0.5\|\boldsymbol{v}_n - \boldsymbol{V}_{nh} + \gamma(\boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh})\|_{\mathbb{L}_2} + 1.5\|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2} \\
&\leqslant \sqrt{5/2}\, x_n.
\end{aligned}$$

Therefore,

$$|z_{n+1}| \leqslant \big(1.19x_n + 1.09\sqrt{\gamma p}\big)^2 + 0.0027\gamma p \leqslant \big(1.19x_n + 1.14\sqrt{\gamma p}\big)^2.$$

This completes the proof of the lemma.

C THE PROOF OF THE UPPER BOUND ON THE ERROR OF RKLMC

Consider the underdamped Langevin diffusion

$$\mathrm{d}\boldsymbol{L}_t = \boldsymbol{V}_t \,\mathrm{d}t, \qquad \text{where} \qquad \mathrm{d}\boldsymbol{V}_t = -\gamma \boldsymbol{V}_t \,\mathrm{d}t - \gamma \nabla f(\boldsymbol{L}_t)\,\mathrm{d}t + \sqrt{2}\gamma \,\mathrm{d}\boldsymbol{W}_t \tag{46}$$

for every $t \geqslant 0$, with given initial conditions $\boldsymbol{L}_0$ and $\boldsymbol{V}_0$. Throughout this section, we assume that $\boldsymbol{V}_0 \sim \mathcal{N}_p(0, \gamma \mathbf{I}_p)$ is independent of $\boldsymbol{L}_0$, and the couple $(\boldsymbol{V}_0, \boldsymbol{L}_0)$ is independent of the Brownian motion $\boldsymbol{W}$. We also assume that $\boldsymbol{L}_0$ is drawn from the target distribution π; this implies that the process $(\boldsymbol{L}_t, \boldsymbol{V}_t)$ is stationary.

In the sequel, we use the following shorthand notation

$$\eta = \gamma h, \quad g = \nabla f, \qquad f_n = f(\boldsymbol{\vartheta}_n), \qquad \boldsymbol{g}_n = g(\boldsymbol{\vartheta}_n), \qquad \boldsymbol{g}_{n+U} = g(\boldsymbol{\vartheta}_{n+U}), \qquad M_\gamma = M/\gamma.$$

The randomized midpoint discretization—proposed and studied in (Shen & Lee, 2019)—of the kinetic Langevin process equation 51, can be written as

$$\begin{aligned}
\boldsymbol{\vartheta}_{n+U} &= \boldsymbol{\vartheta}_n + \frac{1-e^{-U\eta}}{\gamma}\boldsymbol{v}_n - \int_0^{Uh} (1-e^{-\gamma(Uh-s)})\,\mathrm{d}s\,\nabla f_n + \sqrt{2}\int_0^{Uh} (1-e^{-\gamma(Uh-s)})\,d\bar{\boldsymbol{W}}_s\\
\boldsymbol{\vartheta}_{n+1} &= \boldsymbol{\vartheta}_n + \frac{1-e^{-\eta}}{\gamma}\boldsymbol{v}_n - \eta\frac{1-e^{-\eta(1-U)}}{\gamma}\nabla f_{n+U} + \sqrt{2}\int_0^{h} (1-e^{-\gamma(h-s)})\,\mathrm{d}\bar{\boldsymbol{W}}_s\\
\boldsymbol{v}_{n+1} &= \boldsymbol{v}_n e^{-\eta} - \eta e^{-\gamma(h-Uh)}\nabla f_{n+U} + \sqrt{2}\,\gamma\int_0^{h} e^{-\gamma(h-s)}\,\mathrm{d}\bar{\boldsymbol{W}}_s
\end{aligned} \tag{47}$$

where $\bar{\boldsymbol{W}}_s = \boldsymbol{W}_{nh+s} - \boldsymbol{W}_{nh}$. We rewrite these relations in the shorter form

$$\boldsymbol{\vartheta}_{n+U} = \boldsymbol{\vartheta}_n + \gamma^{-1}\eta\big(U\bar{\alpha}_1\boldsymbol{v}_n - U^2\eta\bar{\beta}_1\,\boldsymbol{g}_n + U\sqrt{2U\gamma\eta}\,\bar{\sigma}_1\boldsymbol{\xi}_1\big) \tag{48}$$

$$\boldsymbol{\vartheta}_{n+1} = \boldsymbol{\vartheta}_n + \gamma^{-1}\eta\big(\bar{\alpha}_2\boldsymbol{v}_n - \eta\bar{\beta}_2\,\boldsymbol{g}_{n+U} + \sqrt{2\gamma\eta}\,\bar{\sigma}_2\boldsymbol{\xi}_2\big) \tag{49}$$

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n - \eta\bar{\alpha}_2\boldsymbol{v}_n - 2\eta\bar{\beta}_3\boldsymbol{g}_{n+U} + \sqrt{2\gamma\eta}\,\bar{\sigma}_3\boldsymbol{\xi}_3 \tag{50}$$

where $\bar{\alpha}_1$, $\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\beta}_3$ and $\bar{\sigma}_1$ are positive random variables (with randomness inherited from U only) satisfying

$$\bar{\alpha}_1 \leqslant 1, \qquad \bar{\beta}_1 \leqslant 1/2, \qquad \bar{\beta}_2 \leqslant 1-U \leqslant 1, \qquad \bar{\beta}_3 \leqslant 1/2, \qquad \bar{\sigma}_1^2 \leqslant 1/3$$

and $\mathbb{E}[\bar{\beta}_2] \in [0.468, 0.5]$. Similarly, $\bar{\alpha}_2$, $\bar{\sigma}_2$ and $\bar{\sigma}_3$ are positive real numbers depending on γ and h such that

$$\bar{\alpha}_2 \leqslant 1, \qquad \bar{\sigma}_2^2 \leqslant 1/3, \qquad \bar{\sigma}_3^2 \leqslant 1.$$

We define

$$\bar{\boldsymbol{v}}_{n+1} := \mathbb{E}_U[\boldsymbol{v}_{n+1}], \qquad \bar{\boldsymbol{\vartheta}}_{n+1} := \mathbb{E}_U[\boldsymbol{\vartheta}_{n+1}].$$

The solution to SDE equation 46 starting from $(\boldsymbol{v}_n, \boldsymbol{\vartheta}_n)$ at the n-th iteration at time h admits the following integral formulation

$$\begin{aligned}
\boldsymbol{L}'_t &= \boldsymbol{\vartheta}_n + \int_0^t \boldsymbol{V}'_s \mathrm{d}s\\
\boldsymbol{V}'_t &= \boldsymbol{v}_n e^{-\gamma t} - \gamma\int_0^t e^{-\gamma(t-s)}\nabla f(\boldsymbol{L}'_s)\,\mathrm{d}s + \sqrt{2}\,\gamma\int_0^t e^{-\gamma(t-s)}\mathrm{d}\boldsymbol{W}_{nh+s}\,.
\end{aligned} \tag{51}$$

These expressions will be used in the proofs provided in the present section. Furthermore, without loss of generality, we assume that the $f(\boldsymbol{\theta}_*) = \min_{\boldsymbol{\theta}\in\mathbb{R}^p} f(\boldsymbol{\theta}) = 0$.

C.1 SOME PRELIMINARY RESULTS

We start with some technical results required to prove Theorem 2. They mainly assess the discretisation error as well as discounted sums of squared gradients and velocities.

Lemma 9 (Precision of the mid-point). *For every $h > 0$, it holds that*

$$\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|_{\mathbb{L}_2} \leqslant \gamma^{-1} M_\gamma \eta^3 e^{M_\gamma \eta^2/2}\Big(0.065\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + (1/6)\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \sqrt{\eta\gamma p/54}\Big).$$

Lemma 10 (Discretization error). *Let $(\boldsymbol{L}'_t, \boldsymbol{V}'_t)$ be the exact solution of the kinetic Langevin diffusion starting from $(\boldsymbol{\vartheta}_n, \boldsymbol{v}_n)$. If $\gamma \geqslant M$ and $h > 0$, it holds that*

$$\gamma\|\bar{\boldsymbol{\vartheta}}_{n+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} \leqslant \frac{M_\gamma^2 \eta^5 e^{M_\gamma \eta^2/2}}{\sqrt{3}}\Big(0.065\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + (1/6)\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \sqrt{\eta\gamma p/54}\Big)$$

$$\gamma\|\boldsymbol{\vartheta}_{n+1} - \bar{\boldsymbol{\vartheta}}_{n+1}\|_{\mathbb{L}_2} \leqslant M_\gamma \eta^3\big(0.26\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.106\sqrt{\eta\gamma p}\big) + \frac{\eta^2}{\sqrt{3}}\big(0.12 M_\gamma \eta^2 + 1\big)\|\boldsymbol{g}_n\|_{\mathbb{L}_2}$$

$$\|\bar{\boldsymbol{v}}_{n+1} - \boldsymbol{V}'_h\|_{\mathbb{L}_2} \leqslant M_\gamma^2 \eta^4 e^{M_\gamma \eta^2/2}\Big(0.065\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + (1/6)\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \sqrt{\eta\gamma p/54}\Big)$$

$$\|\boldsymbol{v}_{n+1} - \bar{\boldsymbol{v}}_{n+1}\|_{\mathbb{L}_2} \leqslant M_\gamma \eta^2\big(0.82\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.41\sqrt{\eta\gamma p}\big) + \frac{\eta^2}{\sqrt{3}}\big(0.55 M_\gamma \eta + 1\big)\big\|\boldsymbol{g}_n\big\|_{\mathbb{L}_2}.$$

Corollary 4. *If $\gamma \geqslant 2M$ and $\eta \leqslant 1/5$, it holds that*

$$\begin{aligned}
\gamma\|\bar{\boldsymbol{\vartheta}}_{n+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} &\leqslant M_\gamma^2\eta^5\big(\,0.038\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.098\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.084\sqrt{\eta\gamma p}\big),\\
\gamma\|\boldsymbol{\vartheta}_{n+1} - \bar{\boldsymbol{\vartheta}}_{n+1}\|_{\mathbb{L}_2} &\leqslant \eta^2\big(\,0.578\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.02\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.005\sqrt{\eta\gamma p}\big),\\
\|\bar{\boldsymbol{v}}_{n+1} - \boldsymbol{V}'_h\|_{\mathbb{L}_2} &\leqslant M_\gamma^2\eta^4\big(\,0.066\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.168\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.137\sqrt{\eta\gamma p}\big),\\
\|\boldsymbol{v}_{n+1} - \bar{\boldsymbol{v}}_{n+1}\|_{\mathbb{L}_2} &\leqslant \eta^2\big(\,0.591\big\|\boldsymbol{g}_n\big\|_{\mathbb{L}_2} + 0.164\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.082\sqrt{\eta\gamma p}\big).
\end{aligned}$$

Proposition 3. *If $\gamma \geqslant 5M$ and $\eta \leqslant 1/5$, then, for any $n \in \mathbb{N}$, the iterates of the RKLMC satisfy*

$$\eta\sum_{k=0}^{n}\varrho^{n-k}\|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 \leqslant 18.8\varrho^n\gamma\mathbb{E}[f_0] + 3.92(x_n + 1.5\sqrt{\gamma p})^2 + \frac{10.6\gamma^2 p}{m},$$

$$\eta\sum_{k=0}^{n}\varrho^{n-k}\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 \leqslant 21.7\varrho^n\gamma\mathbb{E}[f_0] + 4.88(x_n + 1.5\sqrt{\gamma p})^2 + \frac{11.2\gamma^2 p}{m},$$

where $\varrho = \exp(-mh)$ and $x_n = \big(\,\|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_n - \boldsymbol{V}_{nh} + \gamma(\boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh})\|_{\mathbb{L}_2}^2\,\big)^{1/2}$.

Proof of Proposition 3. We use the same shorthand notation as in the previous proofs and assume without loss of generality that $\boldsymbol{\theta}_* = 0$. Let us define $z_k = \mathbb{E}[\boldsymbol{v}_k^\top \boldsymbol{g}_k]$, and

$$S_n(z) := \sum_{k=0}^{n}\varrho^{n-k} z_k, \qquad S_n(g^2) := \sum_{k=0}^{n}\varrho^{n-k}\,\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2\,,$$

$$S_n(f) := \sum_{k=0}^{n}\varrho^{n-k}\mathbb{E}[f_k], \qquad S_n(v^2) := \sum_{k=0}^{n}\varrho^{n-k}\|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2\,.$$

We will need the following lemma, the proof of which is postponed.

Lemma 11. *For any $\gamma > 0$ and $h > 0$ satisfying $\gamma \geqslant 5M$ and any $\eta \leqslant 1/5$, the iterates of the randomized midpoint discretization of the kinetic Langevin diffusion satisfy*

$$\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}^2 \leqslant (1 - 1.47\eta)\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 - 2\bar{\alpha}_2\eta\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n] + 2\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2.12\gamma\eta p \tag{52}$$

$$\mathbb{E}[\boldsymbol{v}_{n+1}^\top\boldsymbol{g}_{n+1}] \leqslant 0.51\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + (1 - \bar{\alpha}_2\eta)\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n] - 0.97\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.9\eta^2\gamma p \tag{53}$$

$$\gamma\mathbb{E}[f_{n+1} - f_n] \leqslant 0.28\eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \bar{\alpha}_2\eta\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n] - 0.46\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.09\eta^3\gamma p.$$

From the first inequality equation 52 in Lemma 11, we infer that

$$S_n^{+1}(v^2) \leqslant (1 - 1.47\eta)S_n(v^2) - 2\bar{\alpha}_2\eta S_n(z) + 2\eta^2 S_n(g^2) + 2.12\gamma^2 p/m.$$

In view of Lemma 3 and the fact that $\|\boldsymbol{v}_0\|_{\mathbb{L}_2}^2 = \gamma p$, this implies that

$$\begin{aligned}
1.47\eta S_n(v^2) + 2\bar{\alpha}_2\eta S_n(z) &\leqslant S_n(v^2) - S_n^{+1}(v^2) + 2\eta^2 S_n(g^2) + 2.12\gamma^2 p/m\\
&\leqslant (1 - \varrho)S_n(v^2) + 2\eta^2 S_n(g^2) + 2.12\gamma^2 p/m + \gamma p.
\end{aligned}$$

Note that $1-\varrho \leqslant \frac{m\eta}{\gamma} \leqslant 0.02\eta$. Therefore, we obtain

$$(1.47-0.02)S_n(v^2)+2\bar{\alpha}_2 S_n(z) \leqslant 2\eta S_n(g^2)+\frac{2.14\gamma^2 p}{m\eta},$$

that is equivalent to

$$\boxed{S_n(v^2) \leqslant 1.38\bar{\alpha}_2 S_n(z)_- + 1.38\eta S_n(g^2)+\frac{1.48\gamma^2 p}{m\eta}.} \tag{54}$$

The second step is to use the second inequality equation 53 of Lemma 11. Note that $m\eta/\gamma \leqslant 1/500$ implies $1-\varrho \geqslant 0.998 m\eta/\gamma$. It then follows that

$$S_n^{+1}(z)=(1-\bar{\alpha}_2\eta)S_n(z)-0.97\eta S_n(g^2)+0.51\eta S_n(v^2)+0.9\eta\gamma^2 p/m.$$

This inequality, combined with Lemma 3, yields

$$\begin{aligned}0.97\eta S_n(g^2) &\leqslant -(\bar{\alpha}_2\eta+\varrho-1)S_n(z)+0.51\eta S_n(v^2)+|z_{n+1}|+0.9\eta\gamma^2 p/m\\ &\leqslant \bar{\alpha}_2\eta S_n(z)_- + 0.51\eta S_n(v^2)+|z_{n+1}|+0.9\eta\gamma^2 p/m.\end{aligned}$$

This can be rewritten as

$$\boxed{S_n(g^2) \leqslant 1.03\bar{\alpha}_2 S_n(z)_- + 0.53 S_n(v^2)+\frac{1.03|z_{n+1}|}{\eta}+\frac{0.93\gamma^2 p}{m}.} \tag{55}$$

Let us now proceed with a similar treatment for the last inequality of Lemma 11. Applying Lemma 3, we get $S_n^{+1}(f) \geqslant \varrho S_n(f)-\varrho^{n+1}\mathbb{E}[f_0] \geqslant (1-m\eta/\gamma)S_n(f)-\varrho^{n+1}\mathbb{E}[f_0]$, which leads to

$$-m\eta S_n(f) \leqslant \varrho^{n+1}\gamma\mathbb{E}[f_0]+0.28\eta^2 S_n(v^2)+\bar{\alpha}_2\eta S_n(z)-0.46\eta^2 S_n(g^2)+0.09\frac{\eta^2\gamma^2 p}{m}.$$

From this inequality, and the Polyak-Lojasievicz condition, one can infer that

$$\boxed{\bar{\alpha}_2 S_n(z)_- \leqslant \varrho^{n+1}\gamma\mathbb{E}[f_0]/\eta+0.28\eta S_n(v^2)+(0.5-0.46\eta)S_n(g^2)+0.09\frac{\eta\gamma^2 p}{m}.} \tag{56}$$

Combining equation 56 with equation 54, we get

$$\begin{aligned}S_n(v^2) \leqslant 1.38\Big(&\varrho^{n+1}\gamma\mathbb{E}[f_0]/\eta+0.28\eta S_n(v^2)+(0.5-0.46\eta)S_n(g^2)+0.09\frac{\eta\gamma^2 p}{m}\Big)\\ &+1.38\eta S_n(g^2)+\frac{1.48\gamma^2 p}{m\eta}.\end{aligned}$$

Since $\eta \leqslant 0.2$, it follows then

$$\boxed{S_n(v^2) \leqslant 0.8\Big(S_n(g^2)+\frac{1.8\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0]+\frac{2\gamma^2 p}{m\eta}\Big)\,.} \tag{57}$$

Similarly, combining equation 56 and equation 55, we get

$$\begin{aligned}S_n(g^2) \leqslant 1.03\Big(&\varrho^{n}\gamma\mathbb{E}[f_0]/\eta+0.28\eta S_n(v^2)+(0.5-0.46\eta)S_n(g^2)+0.09\frac{\eta\gamma^2 p}{m}\Big)\\ &+0.53 S_n(v^2)+\frac{1.03|z_{n+1}|}{\eta}+\frac{0.93\gamma^2 p}{m}.\end{aligned}$$

Since $\eta \leqslant 0.2$, it follows then

$$\boxed{S_n(g^2) \leqslant 1.05 S_n(v^2)+\frac{1.94\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0]+\frac{1.94|z_{n+1}|}{\eta}+\frac{0.94\gamma^2 p}{m}.} \tag{58}$$

Equations equation 57 and equation 58 together yield

$$\begin{aligned}S_n(g^2) &\leqslant 0.84\Big(S_n(g^2)+\frac{1.8\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0]+\frac{2\gamma^2 p}{m\eta}\Big)+\frac{1.94\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0]+\frac{1.94|z_{n+1}|}{\eta}+\frac{0.94\gamma^2 p}{m}\\ &\leqslant 0.84 S_n(g^2)+\frac{3.46\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0]+\frac{1.94|z_{n+1}|}{\eta}+\frac{1.78\gamma^2 p}{m\eta}.\end{aligned}$$

Hence, we get

$$\boxed{S_n(g^2) \leqslant \frac{21.7\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0] + \frac{12.2|z_{n+1}|}{\eta} + \frac{11.2\gamma^2 p}{m\eta}.}$$

Using once again equation equation 57, we arrive at

$$S_n(v^2) \leqslant 0.8\Big(\frac{21.7\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0] + \frac{12.2|z_{n+1}|}{\eta} + \frac{11.2\gamma^2 p}{m\eta} + \frac{1.8\varrho^{n+1}}{\eta}\,\mathbb{E}[f_0] + \frac{2\gamma^2 p}{m\eta}\Big),$$

which is equivalent to

$$\boxed{S_n(v^2) \leqslant \frac{18.8\varrho^n\gamma}{\eta}\,\mathbb{E}[f_0] + \frac{9.8|z_{n+1}|}{\eta} + \frac{10.6\gamma^2 p}{m\eta}.}$$

To complete the proof of the proposition, it remains to establish the suitable upper bound on $|z_{n+1}|$. To this end, we note that

$$\begin{aligned}
\|\boldsymbol{g}_n\|_{\mathbb{L}_2} &\leqslant M\,\|\boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh}\|_{\mathbb{L}_2} + \sqrt{Mp}\\
&\leqslant 0.2\,\|\gamma(\boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh})\|_{\mathbb{L}_2} + \sqrt{0.2\gamma p}\\
&\leqslant 0.3(x_n + 1.5\sqrt{\gamma p})\\
\|\boldsymbol{v}_n\|_{\mathbb{L}_2} &\leqslant \|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2} + \sqrt{\gamma p}\\
&\leqslant \|\boldsymbol{v}_n - \boldsymbol{V}_{nh}\|_{\mathbb{L}_2} + \sqrt{\gamma p}\\
&\leqslant x_n + \sqrt{\gamma p}.
\end{aligned}$$

Then, following the same steps as those used in the proof of the second inequality of Lemma 11, one can infer that

$$\begin{aligned}
|z_{n+1}| &\leqslant |z_n| + 0.97\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.51\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.09\eta^2\gamma p\\
&\leqslant \|\boldsymbol{g}_n\|_{\mathbb{L}_2}\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.1\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.051\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.001\gamma p\\
&\leqslant 1.1\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.301\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.001\gamma p\\
&\leqslant 0.099(x_n + 1.5\sqrt{\gamma p})^2 + 0.301(x_n + 1.1\sqrt{p})^2\\
&\leqslant 0.4\big(x_n + 1.5\sqrt{p}\big)^2.
\end{aligned}$$

This completes the proof of the proposition.

□

C.2 Proof of Theorem 2

Let $\boldsymbol{\vartheta}_{n+U}, \boldsymbol{\vartheta}_{n+1}, \boldsymbol{v}_{n+1}$ be the iterates of Algorithm. Let $(\boldsymbol{L}_t, \boldsymbol{V}_t)$ be the kinetic Langevin diffusion, coupled with $(\boldsymbol{\vartheta}_n, \boldsymbol{v}_n)$ through the same Brownian motion and starting from a random point $(\boldsymbol{L}_0, \boldsymbol{V}_0) \propto \exp(-f(\boldsymbol{\theta}) - \frac{1}{2}\|\boldsymbol{v}\|^2)$ such that $\boldsymbol{V}_0 = \boldsymbol{v}_0$. Let $(\boldsymbol{L}'_t, \boldsymbol{V}'_t)$ be the kinetic Langevin diffusion defined on $[0, h]$ using the same Brownian motion and starting from $(\boldsymbol{\vartheta}_n, \boldsymbol{v}_n)$.

Our goal will be to bound the term x_n defined by

$$x_n = \left\|\mathbf{C}\begin{bmatrix}\boldsymbol{v}_n - \boldsymbol{V}_{nh}\\ \boldsymbol{\vartheta}_n - \boldsymbol{L}_{nh}\end{bmatrix}\right\|_{\mathbb{L}_2} \quad \text{with} \quad \mathbf{C} = \begin{bmatrix}\mathbf{I}_p & \mathbf{0}_p\\ \mathbf{I}_p & \gamma\mathbf{I}_p\end{bmatrix}.$$

To this end, define

$$\bar{\boldsymbol{v}}_{n+1} = \mathbb{E}_U[\boldsymbol{v}_{n+1}], \qquad \bar{\boldsymbol{\vartheta}}_{n+1} = \mathbb{E}_U[\boldsymbol{\vartheta}_{n+1}].$$

Since $(\boldsymbol{V}_{(n+1)h}, \boldsymbol{L}_{(n+1)h})$ are independent of U, we have

$$x_{n+1}^2 = \left\|\mathbf{C}\begin{bmatrix}\boldsymbol{v}_{n+1} - \bar{\boldsymbol{v}}_{n+1}\\ \boldsymbol{\vartheta}_{n+1} - \bar{\boldsymbol{\vartheta}}_{n+1}\end{bmatrix}\right\|_{\mathbb{L}_2}^2 + \left\|\mathbf{C}\begin{bmatrix}\bar{\boldsymbol{v}}_{n+1} - \boldsymbol{V}_{(n+1)h}\\ \bar{\boldsymbol{\vartheta}}_{n+1} - \boldsymbol{L}_{(n+1)h}\end{bmatrix}\right\|_{\mathbb{L}_2}^2.$$

Using the triangle inequality and Proposition 1 (See also Proposition 1 from (Dalalyan & Riou-Durand, 2020)), we get

$$\left\|\mathbf{C}\begin{bmatrix}\bar{\boldsymbol{v}}_{n+1}-\boldsymbol{V}_{(n+1)h}\\ \bar{\boldsymbol{\vartheta}}_{n+1}-\boldsymbol{L}_{(n+1)h}\end{bmatrix}\right\|_{\mathbb{L}_2} \leqslant \left\|\mathbf{C}\begin{bmatrix}\bar{\boldsymbol{v}}_{n+1}-\boldsymbol{V}'_{h}\\ \bar{\boldsymbol{\vartheta}}_{n+1}-\boldsymbol{L}'_{h}\end{bmatrix}\right\|_{\mathbb{L}_2} + \left\|\mathbf{C}\begin{bmatrix}\bar{\boldsymbol{V}}'_{h}-\boldsymbol{V}_{(n+1)h}\\ \bar{\boldsymbol{L}}'_{h}-\boldsymbol{L}_{(n+1)h}\end{bmatrix}\right\|_{\mathbb{L}_2}$$
$$\leqslant \left\|\mathbf{C}\begin{bmatrix}\bar{\boldsymbol{v}}_{n+1}-\boldsymbol{V}'_{h}\\ \bar{\boldsymbol{\vartheta}}_{n+1}-\boldsymbol{L}'_{h}\end{bmatrix}\right\|_{\mathbb{L}_2} + \varrho x_n$$

where $\varrho = e^{-mh}$. Combining these inequalities, we get

$$x_{n+1}^2 \leqslant \big(\varrho x_n + y_{n+1}\big)^2 + z_{n+1}^2$$

where

$$y_{n+1} = \left\|\mathbf{C}\begin{bmatrix}\bar{\boldsymbol{v}}_{n+1}-\boldsymbol{V}'_{h}\\ \bar{\boldsymbol{\vartheta}}_{n+1}-\boldsymbol{L}'_{h}\end{bmatrix}\right\|_{\mathbb{L}_2}, \qquad z_{n+1} = \left\|\mathbf{C}\begin{bmatrix}\boldsymbol{v}_{n+1}-\bar{\boldsymbol{v}}_{n+1}\\ \boldsymbol{\vartheta}_{n+1}-\bar{\boldsymbol{\vartheta}}_{n+1}\end{bmatrix}\right\|_{\mathbb{L}_2}^2.$$

This yields[6]

$$x_n \leqslant \varrho^n x_0 + \sum_{k=1}^{n}\varrho^{n-k}y_k + \bigg(\sum_{k=1}^{n}\varrho^{2(n-k)}z_k^2\bigg)^{1/2}$$
$$\leqslant \varrho^n x_0 + \bigg(\frac{1}{1-\varrho}\sum_{k=1}^{n}\varrho^{n-k}y_k^2\bigg)^{1/2} + \bigg(\sum_{k=1}^{n}\varrho^{2(n-k)}z_k^2\bigg)^{1/2},$$

where the second inequality follows from the Cauchy-Schwarz inequality and the formula of the sum of a geometric progression. Using the fact that $\|\mathbf{C}[a,b]^\top\|^2 = \|a\|^2+\|a+\gamma b\|^2 \leqslant 3\|a\|^2+2\gamma^2\|b\|^2$, we arrive at

$$y_{n+1}^2 \leqslant 3\|\bar{\boldsymbol{v}}_{n+1}-\boldsymbol{V}'_h\|_{\mathbb{L}_2}^2 + 2\gamma^2\|\bar{\boldsymbol{\vartheta}}_{n+1}-\boldsymbol{L}'_h\|_{\mathbb{L}_2}^2,$$
$$z_{n+1}^2 \leqslant 3\|\boldsymbol{v}_{n+1}-\bar{\boldsymbol{v}}_{n+1}\|_{\mathbb{L}_2}^2 + 2\gamma^2\|\boldsymbol{\vartheta}_{n+1}-\bar{\boldsymbol{\vartheta}}_{n+1}\|_{\mathbb{L}_2}^2.$$

We then have

$$x_n \leqslant \varrho^n x_0 + \bigg(\frac{1.001\gamma}{m\eta}\sum_{k=1}^{n}\varrho^{n-k}(3\|\bar{\boldsymbol{v}}_k-\boldsymbol{V}'_h\|_{\mathbb{L}_2}^2 + 2\gamma^2\|\bar{\boldsymbol{\vartheta}}_k-\boldsymbol{L}'_h\|_{\mathbb{L}_2}^2)\bigg)^{1/2}$$
$$+ \bigg(\sum_{k=1}^{n}\varrho^{2(n-k)}(3\|\boldsymbol{v}_k-\bar{\boldsymbol{v}}_k\|_{\mathbb{L}_2}^2 + 2\gamma^2\|\boldsymbol{\vartheta}_k-\bar{\boldsymbol{\vartheta}}_k\|_{\mathbb{L}_2}^2)\bigg)^{1/2}. \tag{59}$$

By Corollary 4, we find

$$\|\bar{\boldsymbol{v}}_k-\boldsymbol{V}'_h\|_{\mathbb{L}_2}^2 \leqslant M_\gamma^4\eta^8\Big(0.066\eta\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2} + 0.168\|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2} + 0.137\sqrt{\eta\gamma p}\Big)^2$$
$$\leqslant 0.2^3M_\gamma\eta^8\times 0.0514\big(\eta^2\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2}^2 + \eta\gamma p\big),$$
$$\gamma^2\|\bar{\boldsymbol{\vartheta}}_k-\boldsymbol{L}'_h\|_{\mathbb{L}_2}^2 \leqslant M_\gamma^4\eta^{10}\Big(0.038\eta\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2} + 0.098\|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2} + 0.084\sqrt{\eta\gamma p}\Big)^2$$
$$\leqslant 0.2^3M_\gamma\eta^8\times 0.0002\big(\eta^2\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2}^2 + \eta\gamma p\big)$$
$$\|\boldsymbol{v}_k-\bar{\boldsymbol{v}}_k\|_{\mathbb{L}_2}^2 \leqslant \eta^4\Big(0.591\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2} + 0.164\|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2} + 0.082\sqrt{\eta\gamma p}\Big)^2$$
$$\leqslant \eta^4\times 0.39\big(\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2}^2 + \eta\gamma p\big),$$
$$\gamma^2\|\boldsymbol{\vartheta}_k-\bar{\boldsymbol{\vartheta}}_k\|_{\mathbb{L}_2}^2 \leqslant \eta^4\Big(0.578\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2} + 0.02\|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2} + 0.005\sqrt{\eta\gamma p}\Big)^2$$
$$\leqslant \eta^4\times 0.32\big(\|\boldsymbol{g}_{k-1}\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_{k-1}\|_{\mathbb{L}_2}^2 + \eta\gamma p\big)$$

[6]One can check by induction, that if for some sequences x_n, y_n, z_n and some $\varrho \in (0,1)$ it holds that $x_{n+1}^2 \leqslant (\varrho x_n + y_{n+1})^2 + z_{n+1}^2$, then necessarily $x_n \leqslant \varrho^n x_0 + \sum_{k=1}^n \varrho^{n-k}y_k + (\sum_{k=1}^n \varrho^{2(n-k)}z_k^2)^{1/2}$ for every $n \in \mathbb{N}$.

Therefore, we infer from equation 59 that

$$x_n \leqslant \varrho^n x_0 + \Big(\frac{\gamma}{m\eta}\sum_{k=0}^n 0.2^2 M_\gamma \eta^8 \times 0.031\varrho^{n-k}\big(\eta^2\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 + \eta\gamma p\big)\Big)^{1/2} + \Big(\sum_{k=0}^n 1.82\eta^4\varrho^{2(n-k)}\big(\,\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 + \eta\gamma p\big)\Big)^{1/2}.$$

From Proposition 3 it then follows that

$$\eta\sum_{k=0}^n \varrho^{n-k}\big(\eta^2\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 + \eta\gamma p\big) \leqslant 18.9\varrho^n\gamma\mathbb{E}[f_0] + 3.97(x_n + 1.5\sqrt{\gamma p})^2 + \frac{10.8\gamma^2 p}{m},$$

$$\eta\sum_{k=0}^n \varrho^{2(n-k)}\big(\,\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \|\boldsymbol{v}_k\|_{\mathbb{L}_2}^2 + \eta\gamma p\big) \leqslant 40.5\varrho^n\gamma\mathbb{E}[f_0] + 8.8(x_n + 1.5\sqrt{\gamma p})^2 + \frac{21.9\gamma^2 p}{m}.$$

This yields

$$\begin{aligned} x_n &\leqslant \varrho^n x_0 + 0.036\eta^3\sqrt{\kappa}\Big(18.9\varrho^n\gamma\mathbb{E}[f_0] + 3.97(x_n + 1.5\sqrt{\gamma p})^2 + \frac{10.8\gamma^2 p}{m}\Big)^{1/2} \\ &\quad + \eta^{3/2}\Big(74\varrho^n\gamma\mathbb{E}[f_0] + 16(x_n + 1.5\sqrt{\gamma p})^2 + \frac{40\gamma^2 p}{m}\Big)^{1/2} \\ &\leqslant \varrho^n x_0 + (0.072\eta^3\sqrt{\kappa} + 4\eta^{3/2})x_n + (0.16\eta^3\sqrt{\kappa} + 8.7\eta^{3/2})\sqrt{\varrho^n\gamma\mathbb{E}[f_0]} \\ &\quad + 0.12\eta^3\gamma\sqrt{\kappa p/m} + 6.4\eta^{3/2}\gamma\sqrt{p/m}. \end{aligned}$$

We assume that $\eta\kappa^{1/6} \leqslant 0.1$, which implies that

$$x_n \leqslant \varrho^n x_0 + 0.072 x_n + 0.16\sqrt{\varrho^n\gamma\mathbb{E}[f_0]} + 0.12\eta^3\gamma\sqrt{\kappa p/m} + 6.4\eta^{3/2}\gamma\sqrt{p/m}.$$

Rearranging the display leads to

$$x_n \leqslant 1.08\varrho^n x_0 + 0.18\sqrt{\varrho^n\gamma\mathbb{E}[f_0]} + 0.12\eta^3\gamma\sqrt{\kappa p/m} + 6.9\eta^{3/2}\gamma\sqrt{p/m}.$$

Finally, we use the fact that $x_0 = \gamma \mathsf{W}_2(\nu_0, \pi)$ and $x_n \geqslant \gamma\mathsf{W}_2(\nu_n, \pi)/\sqrt{2}$ to get the claim of the theorem.

C.3 Proofs of the technical lemmas

We now provide the proofs of the technical lemmas that we used in this section.

C.3.1 Proof of Lemma 9

By the definition of $\boldsymbol{\vartheta}_{n+U}$, we have

$$\begin{aligned} \|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\| &\leqslant \Big\|\int_0^{Uh}\big(1 - e^{-\gamma(Uh-s)}\big)\big(\nabla f(\boldsymbol{\vartheta}_n) - \nabla f(\boldsymbol{L}'_s)\big)\mathrm{d}s\Big\| \\ &\leqslant \int_0^{Uh}\Big\|\big(1 - e^{-\gamma(Uh-s)}\big)\big(\nabla f(\boldsymbol{L}'_0) - \nabla f(\boldsymbol{L}'_s)\big)\Big\|\,\mathrm{d}s \\ &= Uh\int_0^1\big(1 - e^{-U\eta(1-t)}\big)\big\|\nabla f(\boldsymbol{L}'_0) - \nabla f(\boldsymbol{L}'_{Uht})\big\|\,\mathrm{d}t \\ &\leqslant Mh\eta U^2\int_0^1 (1-t)\big\|\boldsymbol{L}'_0 - \boldsymbol{L}'_{Uht}\big\|\,\mathrm{d}t, \end{aligned}$$

where in the last inequality we have used the Lipschitz property of ∇f and the inequality $1 - e^{-U\eta(1-t)} \leqslant U\eta(1-t)$. By taking the expectation wrt to U, we get

$$\begin{aligned}\mathbb{E}_U\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|^2 &\leqslant M^2h^2\eta^2\mathbb{E}_U\bigg[U^4\bigg\{\int_0^1(1-t)\big\|\boldsymbol{L}'_0 - \boldsymbol{L}'_{Uht}\big\|\,\mathrm{d}t\bigg\}^2\bigg]\\ &\leqslant \frac{M^2h^2\eta^2}{3}\mathbb{E}_U\bigg[U^4\int_0^1\big\|\boldsymbol{L}'_0 - \boldsymbol{L}'_{Uht}\big\|^2\,\mathrm{d}t\bigg]\\ &\leqslant \frac{M^2h^2\eta^2}{3}\mathbb{E}_U[U^3]\int_0^1\big\|\boldsymbol{L}'_0 - \boldsymbol{L}'_{ht}\big\|^2\,\mathrm{d}t.\end{aligned}$$

Hence, we obtain in view of eq. (42)

$$\begin{aligned}\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|^2_{\mathbb{L}_2} &\leqslant \frac{M^2h^2\eta^2}{12}\int_0^1\|\boldsymbol{L}'_0 - \boldsymbol{L}'_{ht}\|^2_{\mathbb{L}_2}\,\mathrm{d}t\\ &\leqslant \frac{M^2h^2\eta^2e^{M_\gamma\eta^2}}{12}\int_0^1\bigg(\sqrt{\frac{2\gamma^2p(ht)^3}{3}} + ht\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \frac{\gamma(ht)^2}{2}\|\nabla f(\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}\bigg)^2\mathrm{d}t\\ &\leqslant \frac{\gamma^{-2}M_\gamma^2\eta^6e^{M_\gamma\eta^2}}{12}\Big\{\sqrt{(2/3)\eta\gamma p} + \sqrt{1/3}\,\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + \sqrt{0.05}\gamma h\|\nabla f(\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}\Big\}^2.\end{aligned}$$

Taking the square root of the two sides of the inequality, we get the claim of the lemma.

C.3.2 Proof of Lemma 10

By the definition of $\boldsymbol{\vartheta}_{n+1}$, we have

$$\begin{aligned}\|\bar{\boldsymbol{\vartheta}}_{n+1} - \boldsymbol{L}'_h\| &= \Big\|\mathbb{E}_U\Big[h\big(1 - e^{-\gamma(h-Uh)}\big)\nabla f(\boldsymbol{\vartheta}_{n+U})\Big] - \int_0^h\big(1 - e^{-\gamma(h-s)}\big)\nabla f(\boldsymbol{L}'_s)\,\mathrm{d}s\Big\|\\ &= \Big\|\mathbb{E}_U\Big[h\big(1 - e^{-\gamma(h-Uh)}\big)\nabla f_{n+U}\Big] - h\mathbb{E}_U\Big[\big(1 - e^{-\gamma(h-hU)}\big)\nabla f(\boldsymbol{L}'_{Uh})\Big]\Big\|\\ &\leqslant h\mathbb{E}_U\Big[\big(1 - e^{-\gamma(1-U)h}\big)\|\nabla f_{n+U} - \nabla f(\boldsymbol{L}'_{Uh})\|\Big]\\ &\leqslant M_\gamma\eta^2\mathbb{E}_U\Big[(1-U)\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|\Big],\end{aligned}$$

where in the last inequality follows from the smoothness of function f and the fact that $1 - e^{-\gamma(h-Uh)} \leqslant \gamma(1-U)h$. Using the Cauchy-Schwarz inequality, we get

$$\begin{aligned}\|\bar{\boldsymbol{\vartheta}}_{n+1} - \boldsymbol{L}'_h\|^2 &\leqslant M_\gamma^2\eta^4\mathbb{E}_U[(1-U)^2]\,\mathbb{E}_U\big[\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|^2\big]\\ &= \frac{M_\gamma^2\eta^4}{3}\mathbb{E}_U\big[\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|^2\big].\end{aligned}$$

By Lemma 9, we then obtain

$$\begin{aligned}\|\bar{\boldsymbol{\vartheta}}_{n+1} - \boldsymbol{L}'_h\|_{\mathbb{L}_2} &\leqslant \frac{M_\gamma\eta^2}{\sqrt{3}}\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}'_{Uh}\|_{\mathbb{L}_2}\\ &\leqslant \frac{M_\gamma^2\eta^5e^{M_\gamma\eta^2/2}}{\sqrt{3}\gamma}\Big(0.065\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + (1/6)\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + (1/7)\sqrt{\eta\gamma p}\Big).\end{aligned}$$

This completes the proof of the first claim.

Using the definition of $\boldsymbol{\vartheta}_{n+1}$, and the fact that the mean minimizes the squared integrated error, we get

$$\begin{aligned}\|\boldsymbol{\vartheta}_{n+1} - \bar{\boldsymbol{\vartheta}}_{n+1}\|_{\mathbb{L}_2} &= h\Big\|\big(1 - e^{-\gamma h(1-U)}\big)\nabla f_{n+U} - \mathbb{E}_U\big[\big(1 - e^{-\gamma h(1-U)}\big)\nabla f_{n+U}\big]\Big\|_{\mathbb{L}_2}\\ &\leqslant h\Big\|\big(1 - e^{-\gamma h(1-U)}\big)\nabla f_{n+U} - \mathbb{E}_U\big[1 - e^{-\gamma h(1-U)}\big]\nabla f_n\Big\|_{\mathbb{L}_2}.\end{aligned}$$

Recall that $\bar{U} = 1 - U$, combining this with the last display and the triangle inequality yields

$$\begin{aligned}\|\boldsymbol{\vartheta}_{n+1} - \bar{\boldsymbol{\vartheta}}_{n+1}\|_{\mathbb{L}_2} &\leqslant h\Big\|\big(1 - e^{-\eta\bar{U}}\big)\big(\nabla f_{n+U} - \nabla f_n\big)\Big\|_{\mathbb{L}_2} + h\Big\|\big(e^{-\eta\bar{U}} - \mathbb{E}[e^{-\eta\bar{U}}]\big)\nabla f_n\Big\|_{\mathbb{L}_2} \\ &\leqslant M_\gamma \eta^2 \big\|\bar{U}(\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n)\big\|_{\mathbb{L}_2} + h\eta\|\bar{U}\|_{\mathbb{L}_2}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}. \end{aligned} \tag{60}$$

In view of equation 48, we get

$$\begin{aligned}\big\|(1-U)(\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n)\big\|_{\mathbb{L}_2}^2 &= \mathbb{E}\Big[(1-U)^2\Big(\|(\eta/\gamma)(U\bar{\alpha}_1\boldsymbol{v}_n - U^2\eta\bar{\beta}_1\boldsymbol{g}_n)\|^2 + (2/3)U^3\eta^3 p/\gamma\Big)\Big] \\ &\leqslant \frac{\eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2}{15\gamma^2} + \frac{\eta^4\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2}{210\gamma^2} + \frac{\eta^3 p}{90\gamma}.\end{aligned}$$

In addition, $\|1-U\|_{\mathbb{L}_2} = \sqrt{1/3}$. Therefore, we infer from equation 60 that

$$\|\boldsymbol{\vartheta}_{n+1} - \bar{\boldsymbol{\vartheta}}_{n+1}\|_{\mathbb{L}_2} \leqslant \frac{M_\gamma\eta^3}{\gamma}\Big(\frac{\|\boldsymbol{v}_n\|_{\mathbb{L}_2}}{\sqrt{15}} + \sqrt{\frac{\eta\gamma p}{90}}\Big) + \frac{\eta^2}{\gamma}\Big(\frac{M_\gamma\eta^2}{\sqrt{210}} + \frac{1}{\sqrt{3}}\Big)\|\boldsymbol{g}_n\|_{\mathbb{L}_2}.$$

Numerical computations complete the proof of the second claim.

By the definition equation 50 of $\boldsymbol{v}_{n+1}$, we have

$$\begin{aligned}\|\bar{\boldsymbol{v}}_{n+1} - \boldsymbol{V}_h'\|_{\mathbb{L}_2} &= \gamma\Big\|\mathbb{E}_U\big[he^{-\gamma(h-Uh)}\nabla f_{n+U}\big] - \int_0^h e^{-\gamma(t-s)}\nabla f(\boldsymbol{L}_s')\,\mathrm{d}s\Big\|_{\mathbb{L}_2} \\ &\leqslant \gamma\Big\|he^{-\gamma(h-Uh)}\nabla f_{n+U} - he^{-\gamma(h-Uh)}\nabla f(\boldsymbol{L}_{Uh}')\Big\|_{\mathbb{L}_2} \\ &\leqslant M\gamma h\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{L}_{Uh}'\|_{\mathbb{L}_2}.\end{aligned}$$

By Lemma 9, we obtain the third claim of the lemma.

In view of equation 47, and the fact that the expectation minimizes the mean squared error, we have

$$\begin{aligned}\|\boldsymbol{v}_{n+1} - \bar{\boldsymbol{v}}_{n+1}\|_{\mathbb{L}_2} &= \gamma h\Big\|e^{-\gamma h(1-U)}\nabla f_{n+U} - \mathbb{E}_U\big[e^{-\gamma h(1-U)}\nabla f_{n+U}\big]\Big\|_{\mathbb{L}_2} \\ &\leqslant \gamma h\Big\|e^{-\gamma h(1-U)}\nabla f_{n+U} - \mathbb{E}_U\big[e^{-\gamma h(1-U)}\big]\nabla f_n\Big\|_{\mathbb{L}_2}.\end{aligned}$$

The last display, the notation $\bar{U} = 1 - U$ and the triangle inequality imply that

$$\begin{aligned}\|\boldsymbol{v}_{n+1} - \bar{\boldsymbol{v}}_{n+1}\|_{\mathbb{L}_2} &\leqslant \gamma h\big\|e^{-\eta\bar{U}}\big(\nabla f_{n+U} - \nabla f_n\big)\big\|_{\mathbb{L}_2} + \gamma h\big\|\big(e^{-\eta\bar{U}} - \mathbb{E}_U[e^{-\eta\bar{U}}]\big)\nabla f_n\big\|_{\mathbb{L}_2} \\ &\leqslant M\gamma h\big\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2} + \gamma h\big\|\big(e^{-\eta(1-U)} - 1\big)\nabla f_n\big\|_{\mathbb{L}_2} \\ &\leqslant M\gamma h\big\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2} + \frac{\eta^2}{\sqrt{3}}\big\|\boldsymbol{g}_n\big\|_{\mathbb{L}_2}.\end{aligned} \tag{61}$$

In view of equation 48, we get

$$\begin{aligned}\big\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n\big\|_{\mathbb{L}_2}^2 &= \mathbb{E}\Big[\|(\eta/\gamma)(U\bar{\alpha}_1\boldsymbol{v}_n - U^2\eta\bar{\beta}_1\,\boldsymbol{g}_n)\|^2 + (2/3)\eta^3U^3p/\gamma\Big] \\ &\leqslant \frac{2\eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2}{3\gamma^2} + \frac{\eta^4\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2}{10\gamma^2} + \frac{\eta^3 p}{6\gamma}.\end{aligned}$$

The last claim of the lemma follows from the previous display and equation 61.

C.3.3 Proof of Lemma 11

From equation 48, equation 49 and equation 50, it follows that

$$\begin{aligned}\gamma^2\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2 &\leqslant \|U\eta\bar{\alpha}_1\boldsymbol{v}_n - (U\eta)^2\bar{\beta}_1\,\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + (1/6)\eta^3\gamma p \\ &\leqslant (2\eta^2/3)\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + (\eta^4/10)\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + (\eta^3/6)\gamma p\end{aligned} \tag{62}$$

and

$$\begin{aligned}\gamma\|\boldsymbol{\vartheta}_{n+1} - \boldsymbol{\vartheta}_n\|_{\mathbb{L}_2} &\leqslant \eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.5\eta^2\|\boldsymbol{g}_{n+U}\|_{\mathbb{L}_2} + \sqrt{(2/3)\eta^3\gamma p} \\ &\leqslant \eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.5\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + 0.5M_\gamma\eta^2\gamma\|\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n\|_{\mathbb{L}_2} + \sqrt{(2/3)\eta^3\gamma p} \\ &\leqslant 1.001\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2} + 0.501\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2} + \sqrt{0.67\eta^3\gamma p}\end{aligned} \tag{63}$$

where in the last step we have used equation 62 and the fact that $M_\gamma\eta^2/2 \leqslant \eta^2/8 \leqslant 1/200$. A bit more precise computations also yield

$$\begin{aligned}
\gamma\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2} &\leqslant \big\{(\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\eta^2\|\bar\beta_2\boldsymbol{g}_n\|_{\mathbb{L}_2})^2+\tfrac{2}{3}\gamma\eta^3p\big\}^{1/2}+\frac{\eta^2}{2}\|\boldsymbol{g}_{n+U}-\boldsymbol{g}_n\|_{\mathbb{L}_2}\\
&\leqslant \Big\{\big(\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\tfrac{\eta^2}{\sqrt{3}}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big)^2+\tfrac{2}{3}\gamma\eta^3p\Big\}^{1/2}+\frac{M_\gamma\eta^2\gamma}{2}\|\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}\\
&\leqslant \big\{\big(\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\tfrac{\eta^2}{\sqrt{3}}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big)^2+\tfrac{2}{3}\gamma\eta^3p\big\}^{1/2}+\tfrac{1}{10}\eta^2\gamma\|\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}.
\end{aligned}$$

Taking the squares of this inequality, we get

$$\begin{aligned}
\gamma^2\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2 &\leqslant \big(\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\tfrac{\eta^2}{\sqrt{3}}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big)^2+\tfrac{2}{3}\gamma\eta^3p+\tfrac{1}{100}\eta^4\gamma^2\|\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2\\
&\quad+\tfrac{1}{5}\eta^2\gamma\Big\{\big(\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\tfrac{\eta^2}{\sqrt{3}}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big)^2+\tfrac{2}{3}\gamma\eta^3p\Big\}^{1/2}\|\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}\\
&\leqslant \big(1+\tfrac{1}{10}\eta^2\big)\eta^2\Big\{\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\tfrac{\eta}{\sqrt{3}}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big)^2+\tfrac{2}{3}\eta\gamma p+\tfrac{1}{10}\gamma^2\|\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2\Big\}\\
&\leqslant 1.01\eta^2\Big\{\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\tfrac{\eta}{\sqrt{3}}\|\boldsymbol{g}_n\|_{\mathbb{L}_2}\big)^2+\tfrac{2}{3}\eta\gamma p+0.1\gamma^2\|\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2}^2\Big\}\\
&\leqslant 0.68\eta^2\big(3\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2+\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+\eta\gamma p\big). \qquad (64)
\end{aligned}$$

This implies that for $\gamma \geqslant 5M$ and $\eta \leqslant 1/5$, we have

$$\begin{aligned}
\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}^2 &\leqslant (1-\bar\alpha_2\eta)^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-4\eta(1-\bar\alpha_2\eta)\mathbb{E}[\bar\beta_2\boldsymbol{v}_n^\top\boldsymbol{g}_{n+U}]+\eta^2\|\boldsymbol{g}_{n+U}\|_{\mathbb{L}_2}^2+2\eta\gamma p\\
&\quad+2\eta\sqrt{2\eta\gamma p}\|\boldsymbol{g}_{n+U}-\boldsymbol{g}_n\|_{\mathbb{L}_2}\\
&\leqslant (1-\bar\alpha_2\eta)^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-4\eta(1-\bar\alpha_2\eta)\mathbb{E}[\bar\beta_2\boldsymbol{v}_n^\top\boldsymbol{g}_{n+U}]+\eta^2\|\boldsymbol{g}_{n+U}\|_{\mathbb{L}_2}^2+2\eta\gamma p\\
&\quad+2M_\gamma\eta\sqrt{2\eta\gamma p}\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}\\
&\leqslant (1-\eta\bar\alpha_2)^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-4\eta(1-\eta\bar\alpha_2)\mathbb{E}[\bar\beta_2\boldsymbol{v}_n^\top\boldsymbol{g}_{n+U}]+\eta^2\|\boldsymbol{g}_{n+U}\|_{\mathbb{L}_2}^2+2.1\eta\gamma p\\
&\quad+20(M_\gamma\eta)^2\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2\\
&\leqslant (1-\bar\alpha_2\eta)^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-2\bar\alpha_2\eta(1-\eta\bar\alpha_2)\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n]+1.1\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+2.1\eta\gamma p\\
&\quad+2M_\gamma\eta\|\boldsymbol{v}_n\|_{\mathbb{L}_2}\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}+31(M_\gamma\eta)^2\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2\\
&\leqslant (1-\bar\alpha_2\eta)^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-2\bar\alpha_2\eta(1-\eta\bar\alpha_2)\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n]+1.1\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+2.1\eta\gamma p\\
&\quad+0.2M_\gamma\eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2+5M_\gamma\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2+31(M_\gamma\eta)^2\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2\\
&\leqslant (1-\bar\alpha_2\eta)^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-2\bar\alpha_2\eta(1-\bar\alpha_2\eta)\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n]+1.1\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+2.1\eta\gamma p\\
&\quad+0.2M_\gamma\eta^2\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2+5.4M_\gamma\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2.
\end{aligned}$$

Since for $\eta \leqslant 0.2$ we have $\bar\alpha_2 \geqslant 0.9$, we get $(1-\bar\alpha_2\eta)^2+0.2M_\gamma\eta^2+5.4M_\gamma(2\eta^2/3) \leqslant (1-0.9\eta)^2+0.008\eta+0.16\eta \leqslant 1-1.47\eta$. Therefore,

$$\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}^2 \leqslant (1-1.47\eta)\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2-2\bar\alpha_2\eta(1-\eta\bar\alpha_2)\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n]+1.12\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+2.12\eta\gamma p.$$

The next step is to get an upper bound on $\mathbb{E}[\boldsymbol{v}_{n+1}^\top\boldsymbol{g}_{n+1}]-\mathbb{E}[\boldsymbol{v}_n^\top\boldsymbol{g}_n]$ in order to prove equation 53. To this end, we first note that

$$\begin{aligned}
\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2} &\leqslant \|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\eta\|\boldsymbol{g}_{n+U}\|_{\mathbb{L}_2}+\sqrt{2\eta\gamma p}\\
&\leqslant \|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}+M_\gamma\eta\|\gamma(\boldsymbol{\vartheta}_{n+U}-\boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}+\sqrt{2\eta\gamma p}\\
&\leqslant 1.004\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}+\eta\|\boldsymbol{g}_n\|_{\mathbb{L}_2}+\sqrt{2\eta\gamma p}\big). \qquad (65)
\end{aligned}$$

From equation 63 and equation 65, we also infer that

$$\begin{aligned}
\gamma\|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2}\|\boldsymbol{\vartheta}_{n+1}-\boldsymbol{\vartheta}_n\|_{\mathbb{L}_2} &\leqslant \eta\sqrt{3(1.001^2+0.501^2+0.34)}\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2+\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+2\eta\gamma p\big)\\
&\leqslant 2.2\eta\big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2+\eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2+2\eta\gamma p\big).
\end{aligned}$$

Therefore, this bound and some elementary computations yield

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{v}_{n+1}^\top \boldsymbol{g}_{n+1}] &\leqslant \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] + \mathbb{E}[\boldsymbol{v}_{n+1}^\top(\boldsymbol{g}_{n+1} - \boldsymbol{g}_n)] + \mathbb{E}[(\boldsymbol{v}_{n+1} - \boldsymbol{v}_n)^\top \boldsymbol{g}_n] \\
&\leqslant \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] + M_\gamma \|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2} \|\gamma(\boldsymbol{\vartheta}_{n+1} - \boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2} - \bar{\alpha}_2 \eta \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - \eta \mathbb{E}[\boldsymbol{g}_n^\top \boldsymbol{g}_{n+U}] \\
&\leqslant (1 - \bar{\alpha}_2 \eta)\mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] + M_\gamma \|\boldsymbol{v}_{n+1}\|_{\mathbb{L}_2} \|\gamma(\boldsymbol{\vartheta}_{n+1} - \boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2} - \eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 \\
&\qquad + M_\gamma \eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2} \|\gamma(\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2} \\
&\leqslant (1 - \bar{\alpha}_2 \eta)\mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - \eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2.2 M_\gamma \eta \big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2\eta\gamma p\big) \\
&\qquad + M_\gamma \eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2} \big(\tfrac{2}{3} \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{1}{10}\eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \tfrac{1}{6}\eta\gamma p\big)^{1/2} \\
&\leqslant (1 - \bar{\alpha}_2 \eta)\mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - \eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2.2 M_\gamma \eta \big(\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2\eta\gamma p\big) \\
&\qquad + 0.5 M_\gamma \eta \big(\tfrac{2}{3} \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \tfrac{11}{10}\eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \tfrac{1}{6}\eta\gamma p\big).
\end{aligned}
$$

Grouping the terms, and using the fact that $M_\gamma \eta \leqslant 1/25$, we arrive at

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{v}_{n+1}^\top \boldsymbol{g}_{n+1}] &\leqslant (1 - \bar{\alpha}_2 \eta)\mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - 0.97\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 2.54 M_\gamma \eta \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 4.5 M_\gamma \gamma \eta^2 p \\
&\leqslant (1 - \bar{\alpha}_2 \eta)\mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - 0.97\eta \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.51\eta \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.9\gamma\eta^2 p.
\end{aligned}
$$

Similarly, using the Lipschitz property of ∇f and equation 64, we get

$$
\begin{aligned}
\gamma \mathbb{E}[f_{n+1} - f_n] &\leqslant \gamma \mathbb{E}[\boldsymbol{g}_n^\top(\boldsymbol{\vartheta}_{n+1} - \boldsymbol{\vartheta}_n)] + (M_\gamma/2)\|\gamma(\boldsymbol{\vartheta}_{n+1} - \boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2 \\
&= \mathbb{E}[\boldsymbol{g}_n^\top(\bar{\alpha}_2 \eta \boldsymbol{v}_n - \eta^2 \bar{\beta}_2 \boldsymbol{g}_{n+U})] + 0.07\eta^2(3\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \eta\gamma p) \\
&\leqslant \bar{\alpha}_2 \eta \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - \eta^2 \mathbb{E}[\bar{\beta}_2]\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.2\eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2} \|\gamma(\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2} \\
&\qquad + 0.07\eta^2(3\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \eta\gamma p) \\
&\leqslant \bar{\alpha}_2 \eta \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - \eta^2 \mathbb{E}[\bar{\beta}_2]\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.1\eta^4 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.1\|\gamma(\boldsymbol{\vartheta}_{n+U} - \boldsymbol{\vartheta}_n)\|_{\mathbb{L}_2}^2 \\
&\qquad + 0.07\eta^2(3\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \eta\gamma p) \\
&\leqslant \bar{\alpha}_2 \eta \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - 0.468\eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.1\eta^4 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 \\
&\qquad + 0.07\eta^2 \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.01\eta^4 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.02\eta^3\gamma p \\
&\qquad + 0.07\eta^2(3\|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + \eta^2\|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + \eta\gamma p).
\end{aligned}
$$

Grouping the terms, and using the fact that $M_\gamma \eta^2 \leqslant 1/50$, we arrive at

$$
\gamma \mathbb{E}[f_{n+1} - f_n] \leqslant \bar{\alpha}_2 \eta \mathbb{E}[\boldsymbol{v}_n^\top \boldsymbol{g}_n] - 0.46\eta^2 \|\boldsymbol{g}_n\|_{\mathbb{L}_2}^2 + 0.28\eta^2 \|\boldsymbol{v}_n\|_{\mathbb{L}_2}^2 + 0.09\eta^3\gamma p.
$$

This completes the proof of the lemma.