A NOTE ON WASSERSTEIN-2 ERROR OF LANGEVIN MONTE CARLO UNDER STRONG LOG-CONCAVITY

Anonymous authors
Paper under double-blind review

ABSTRACT

This note refines a recent nonasymptotic bound on the Wasserstein error of Langevin Monte Carlo for strongly log-concave target distributions. The refinement preserves the linear dependence on the step size while accounting more precisely for the spectrum of the Hessian. For a representative spiked Gaussian model, the resulting bound improves the dimension dependence of the previous linear-in-step-size estimate from order $p^{3/2}$ to order p.

1 INTRODUCTION

The Langevin Monte Carlo (LMC) method, also known as the unadjusted Langevin algorithm, is widely used to sample from a prescribed target distribution π. Assuming that π has a strictly positive density on $\mathbb{R}^p$, the algorithm requires only evaluations of the gradient of the potential $f(\boldsymbol{x}) = -\log \pi(\boldsymbol{x})$, up to an additive constant. Starting from an initial value $\boldsymbol{\vartheta}_0$, it iterates

$$\boldsymbol{\vartheta}_{k+1} = \boldsymbol{\vartheta}_k - h\nabla f(\boldsymbol{\vartheta}_k) + \sqrt{2h}\,\boldsymbol{\xi}_{k+1},$$

where $h > 0$ is the step size and $(\boldsymbol{\xi}_k)_{k\in\mathbb{N}}$ is a sequence of independent standard Gaussian random vectors, independent of $\boldsymbol{\vartheta}_0$. Starting from Dalalyan (2017b), properties of the LMC have been thoroughly investigated under the mathematically convenient setting of smooth and strongly convex f. In this context, the main question is to understand the dependence of the error of the algorithm on the key parameters of the problem: dimension p, step size h, and condition number $\varkappa$. The latter is usually defined as the ratio of the largest and the smallest eigenvalues of the Hessian matrix of the potential f, but we will come back to this definition shortly. The fact that the error of sampling measured in total-variation distance depends on the dimension as a low-degree polynomial was stressed by Dalalyan et al. (2014) and further improved and extended to other distances by Durmus & Moulines (2019). There is now a substantial body of literature on this topic that we will not be able to cover in this introduction. We only mention the most important milestones in close relation to the present paper.

When the error of sampling is measured by the Wasserstein distance, the first upper bound on the discretization error of the LMC, established by Durmus & Moulines (2019), was of the order $\varkappa\sqrt{hp}$. It has been improved to $\sqrt{\varkappa hp}$ by Durmus et al. (2019). More recently, (Dalalyan & Karagulyan, 2026) proved that the usual definition of the condition number makes it dimension dependent in many statistically relevant applications and advocated for using the average condition number $\varkappa_{\mathsf{av}}$ that can be informally defined as the average of the spectrum of the Hessian of f, denoted by M_{av}, divided by the strong-convexity constant. They have shown that the discretization error is at most of order $\sqrt{\varkappa_{\mathsf{av}} hp}$, which leads to an improvement up to a factor $\sqrt{p}$ for Gaussian distributions with all partial correlations of the same order. Here the spiked Gaussian model is defined by $\boldsymbol{\Sigma}^{-1} = \mathbf{I}_p + \mathbb{1}_p\mathbb{1}_p^\top$.

An important breakthrough was obtained recently by Pedrotti & Whalley (2026), who succeeded in improving the dependence on h from square-root to linear, that is of the order of $h\sqrt{\varkappa M p}$ with M being the largest possible eigenvalue of the Hessian of f. This result is remarkable because it depends optimally on h. It strictly improves the bound of Durmus et al. (2019) when $Mh < 1$, as ensured by the step-size condition used below. The comparison with the bound in Dalalyan & Karagulyan (2026) is more nuanced. While the dependence on h is much better in Pedrotti & Whalley (2026), its dependence on $\varkappa$ rather than on $\varkappa_{\mathsf{av}}$ makes it less attractive in the statistically appealing setting of large p and a Hessian matrix with a spiked spectrum. The goal of this note is to refine the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Orders of upper bounds on the Wasserstein-2 discretization error. Under the step-size condition of Theorem 1, the last row improves the preceding bounds because $\varkappa_{\mathsf{av}} \leqslant \varkappa$ and $hM < 1$.

Reference	General bound	Spiked Gaussian model
Durmus & Moulines (2019); Dalalyan (2017a)	$\varkappa(hp)^{1/2}$	$h^{1/2}p^{3/2}$
Durmus et al. (2019)	$(\varkappa hp)^{1/2}$	$h^{1/2}p$
Dalalyan & Karagulyan (2026)	$(\varkappa_{\mathsf{av}} hp)^{1/2}$	$h^{1/2}p^{1/2}$
Pedrotti & Whalley (2026)	$h(\varkappa Mp)^{1/2}$	$hp^{3/2}$
This work	$h(\varkappa_{\mathsf{av}} Mp)^{1/2}$	hp

result of Pedrotti & Whalley (2026) so that it also improves the corresponding bound of Dalalyan & Karagulyan (2026). The bound we get, reported in the last row of Table 1, is of order $h\sqrt{\varkappa_{\mathsf{av}} Mp}$, where $\varkappa_{\mathsf{av}} = M_{\mathsf{av}}/m$ with M_{av} being a suitably defined "average smoothness" constant satisfying $M_{\mathsf{av}} \leqslant M$. For a representative spiked Gaussian model, the resulting bound improves the dimension dependence of the previous linear-in-step-size estimate from order $p^{3/2}$ to order p.

Notation We denote by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^p$. For a matrix $\mathbf{A}$, we write $\|\mathbf{A}\|_{\mathrm{op}}$ and $\|\mathbf{A}\|_{\mathsf{F}}$ for its operator and Frobenius norms, respectively. If matrices $\mathbf{A}, \mathbf{A}'$ are symmetric, notation $\mathbf{A} \preccurlyeq \mathbf{A}'$ and $\mathbf{A}' \succcurlyeq \mathbf{A}$ mean that $\mathbf{A}' - \mathbf{A}$ is positive semidefinite. The identity matrix of dimension $p \times p$ is denoted by $\mathbf{I}_p$. For a differentiable function f, ∇f stands for its gradient. For a twice differentiable function f, $\nabla^2 f$ stands for its Hessian matrix. The target distribution is denoted by π and has density proportional to $\exp(-f)$ with respect to the Lebesgue measure. If μ and ν are probability measures on $\mathbb{R}^p$, then $\mathsf{W}_2(\mu, \nu)$ denotes the Wasserstein-2 distance between μ and ν. For a random vector $\boldsymbol{X}$ and for a random matrix $\mathbf{A}$, we set $\|\boldsymbol{X}\|_{\mathbb{L}_2}^2 = \mathbf{E}[\|\boldsymbol{X}\|^2]$ and $\|\mathbf{A}\|_{\mathbb{L}_2}^2 = \mathbf{E}[\|\mathbf{A}\|_{\mathsf{F}}^2]$. For a function f and a measure μ, we define $\|f\|_{\mathbb{L}_2(\mu)}^2 = \int \|f\|^2 \mathrm{d}\mu$ and $\|f\|_{\mathbb{L}_1(\mu)} = \int \|f\| \mathrm{d}\mu$.

2 Main results and discussion

We recall that a twice differentiable function $f : \mathbb{R}^p \to \mathbb{R}$ is said to be m-strongly convex for some $m > 0$ if $\nabla^2 f(\boldsymbol{x}) \succcurlyeq m\mathbf{I}_p$ for every $\boldsymbol{x} \in \mathbb{R}^p$. We denote by M a constant that upper bounds the spectrum of $\nabla^2 f$, that is $\nabla^2 f(\boldsymbol{x}) \preccurlyeq M\mathbf{I}_p$ for every $\boldsymbol{x} \in \mathbb{R}^p$. For a twice-differentiable function f, this is equivalent to saying that the gradient ∇f is M-Lipschitz continuous. Finally, we set

$$M_{\mathsf{av}} = p^{-1} \| \operatorname{tr}(\nabla^2 f) \|_{\mathbb{L}_1(\pi)}, \qquad \varkappa = \frac{M}{m}, \qquad \varkappa_{\mathsf{av}} = \frac{M_{\mathsf{av}}}{m}.$$

The rationale behind this definition is the following. If the ordered eigenvalues of $\nabla^2 f$ are $\lambda_1, \ldots, \lambda_p$, then $m \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_p \leqslant M$ and we clearly have

$$M_{\mathsf{av}} = \int_{\mathbb{R}^p} \frac{\lambda_1(\boldsymbol{x}) + \ldots + \lambda_p(\boldsymbol{x})}{p} \, \pi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \in [m, M].$$

In settings where $\nabla^2 f$ has well distributed diagonal entries but a highly variable spectrum, the average smoothness M_{av} might be significantly smaller than the worst-case smoothness M. Note that the averaging operation in the definition of M_{av} is double: we average both over the dimensions and the state variable $\boldsymbol{x}$.

Theorem 1. *Let $f : \mathbb{R}^p \to \mathbb{R}$ be twice differentiable and m-strongly convex, and suppose that its gradient is M-Lipschitz continuous, where $0 < m \leqslant M$. The distribution ν_k of the Langevin Monte Carlo algorithm with constant step size $0 < h \leqslant 1/(m + M)$ satisfies*

$$\mathsf{W}_2(\nu_k, \pi) \leqslant (1 - mh)^k \mathsf{W}_2(\nu_0, \pi) + 3.6h\sqrt{\varkappa_{\mathsf{av}} M \, p}. \tag{1}$$

Is this the end of the story? Not quite. Although the preceding result achieves the optimal dependence on the step size h and substantially improves the dependence of the discretization error on the spectrum of the Hessian, the remaining dependence on the maximal condition number $\varkappa$ may still be suboptimal. This is suggested by the Gaussian case, in which the discretization error can be

bounded solely in terms of M_{av}; equivalently, the factor $\varkappa_{\mathsf{av}} M = \varkappa M_{\mathsf{av}}$ can be replaced by M_{av}. The following standard result makes this comparison precise. We include its short proof in Section 6 for completeness.

Proposition 1. *Let $\pi = \mathcal{N}(0, \boldsymbol{\Sigma})$ with invertible covariance matrix $\boldsymbol{\Sigma}$. Let m and M be, respectively, the smallest and the largest eigenvalues of $\boldsymbol{\Sigma}^{-1}$. Let ν_k denote the distribution of the LMC iterate with constant step size $0 < hM \leqslant 1$. Then, for every $k \geqslant 0$,*

$$\frac{h}{4}\big\{1 - (1 - hm)^{2k}\big\}\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} \leqslant \mathsf{W}_2(\nu_k, \pi) \leqslant \frac{h}{2}\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})}. \tag{2}$$

In the case where π is $\mathcal{N}(0, \boldsymbol{\Sigma})$, we have $\mathrm{tr}(\nabla^2 f(\boldsymbol{x})) = \mathrm{tr}(\boldsymbol{\Sigma}^{-1})$ for every $\boldsymbol{x} \in \mathbb{R}^p$. Therefore, $\mathrm{tr}(\boldsymbol{\Sigma}^{-1})$ plays the role of $M_{\mathsf{av}} p$. It is therefore natural to question the presence of $\varkappa$ in the right-hand side of (1). It would be appealing to obtain an upper bound that does not contain this quantity and has optimal dependence on p, h, and M_{av}, or to prove that there are non-Gaussian strongly log-concave distributions with smooth potential for which the factor $\varkappa$ cannot be avoided.

3 Key lemma

The proof of the main result closely follows the arguments of (Dalalyan & Karagulyan, 2019; Dalalyan, 2017a), while incorporating two key additional ingredients. The first, drawn from the recent work (Pedrotti & Whalley, 2026), yields an improved dependence on the step size h. The second, which yields an improved dependence on the spectrum of the Hessian of the potential, is provided by the lemma below. It is not specific to the potential or to the discretization argument considered here. It controls temporal increments of any square-integrable observable of a stationary reversible Langevin diffusion through its Dirichlet energy, without requiring strong convexity or a global smoothness constant. It may therefore also be useful for other discretization arguments and for regularity or covariance estimates involving reversible Langevin diffusions.

Lemma 1. *Let $(\boldsymbol{L}_t)_{t\geq 0}$ be a stationary and time-reversible Langevin diffusion with stationary distribution π. Let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a continuously differentiable function such that ψ and $\nabla\psi$ are square-integrable with respect to π. Then, for any $t > 0$,*

$$\|\psi(\boldsymbol{L}_t) - \psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2} \leq \sqrt{2t}\,\|\nabla\psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2}. \tag{3}$$

4 Proof of the key lemma

Proof. We start by proving the desired inequality under the additional assumption $\psi \in C^2(\mathbb{R}^p; \mathbb{R})$ and that ψ has compact support. Simple algebra yields

$$\begin{aligned}\|\psi(\boldsymbol{L}_t) - \psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2}^2 &= \|\psi(\boldsymbol{L}_t)\|_{\mathbb{L}_2}^2 + \|\psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2}^2 - 2\mathbb{E}\big[\psi(\boldsymbol{L}_t)\psi(\boldsymbol{L}_0)\big] \\ &= -2\mathbb{E}\big[\big(\psi(\boldsymbol{L}_t) - \psi(\boldsymbol{L}_0)\big)\psi(\boldsymbol{L}_0)\big],\end{aligned} \tag{4}$$

where we have used the stationarity of $\boldsymbol{L}_t$ to infer that $\|\psi(\boldsymbol{L}_t)\|_{\mathbb{L}_2} = \|\psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2}$. Let $\boldsymbol{W}$ be the Brownian motion driving the Langevin diffusion $\boldsymbol{L}$ and f be its potential function. Since ψ is C^2, we can apply Itô's formula as follows:

$$\begin{aligned}\psi(\boldsymbol{L}_t) - \psi(\boldsymbol{L}_0) &= \int_0^t \nabla\psi(\boldsymbol{L}_s)^\top \,\mathrm{d}\boldsymbol{L}_s + \int_0^t \mathrm{tr}\left(\nabla^2\psi(\boldsymbol{L}_s)\right)\mathrm{d}s \\ &= \int_0^t \Big(- \nabla\psi(\boldsymbol{L}_s)^\top \nabla f(\boldsymbol{L}_s) + \mathrm{tr}\left(\nabla^2\psi(\boldsymbol{L}_s)\right)\Big)\,\mathrm{d}s + \sqrt{2}\int_0^t \nabla\psi(\boldsymbol{L}_s)^\top \,\mathrm{d}\boldsymbol{W}_s \\ &= \int_0^t \mathfrak{L}\psi(\boldsymbol{L}_s)\,\mathrm{d}s + \xi_t,\end{aligned} \tag{5}$$

where $\mathfrak{L}$ is the infinitesimal generator of the Langevin process $\boldsymbol{L}$ and ξ_t is a random process satisfying $\mathbf{E}[\xi_t|\boldsymbol{L}_0] = 0$ for every $t > 0$. Therefore, we get

$$\mathbb{E}\big[\big(\psi(\boldsymbol{L}_t) - \psi(\boldsymbol{L}_0)\big)\psi(\boldsymbol{L}_0)\big] = \int_0^t \mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_s)\psi(\boldsymbol{L}_0)\big]\,\mathrm{d}s.$$

The reversibility of the Langevin diffusion, combined with (5), yields

$$\begin{aligned}\mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_s)\psi(\boldsymbol{L}_0)\big] &= \mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_s)\big] \\ &= \mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_0)\big] + \int_0^s \mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\mathfrak{L}\psi(\boldsymbol{L}_u)\big]\,\mathrm{d}u \\ &= \mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_0)\big] + \mathbf{E}\bigg[\mathfrak{L}\psi(\boldsymbol{L}_s)\int_0^s \mathfrak{L}\psi(\boldsymbol{L}_u)\,\mathrm{d}u\bigg].\end{aligned}$$

Note that in the equalities above we have used the reversibility both in the first and in the third lines. The last two displays imply that

$$\begin{aligned}\mathbb{E}\big[\big(\psi(\boldsymbol{L}_t)-\psi(\boldsymbol{L}_0)\big)\psi(\boldsymbol{L}_0)\big] &= t\,\mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_0)\big] + \mathbf{E}\bigg[\int_0^t\int_0^s \mathfrak{L}\psi(\boldsymbol{L}_s)\mathfrak{L}\psi(\boldsymbol{L}_u)\,\mathrm{d}u\,\mathrm{d}s\bigg] \\ &= t\,\mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_0)\big] + \frac{1}{2}\mathbf{E}\bigg\{\int_0^t \mathfrak{L}\psi(\boldsymbol{L}_s)\,\mathrm{d}s\bigg\}^2 \\ &\geqslant t\,\mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_0)\big]. \qquad (6)\end{aligned}$$

Let π denote the stationary distribution of $\boldsymbol{L}$, implying thus that its density is proportional to e^{-f}. With a slight abuse of notation, we will write π both for the distribution and for its density. Using the integration by parts formula, combined with the square-integrability of both ψ and its partial derivatives $\partial_i\psi$, we get

$$\begin{aligned}\mathbf{E}\big[\mathfrak{L}\psi(\boldsymbol{L}_0)\psi(\boldsymbol{L}_0)\big] &= \int_{\mathbb{R}^p} \mathfrak{L}\psi(\boldsymbol{x})\,\psi(\boldsymbol{x})\,\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \sum_{i=1}^p \int_{\mathbb{R}^p} (\partial_i\psi\cdot\psi)(\boldsymbol{x})\,\underbrace{(-\partial_i f\cdot\pi)(\boldsymbol{x})}_{=\partial_i\pi(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^p}\big(\operatorname{tr}(\nabla^2\psi)\cdot\psi\big)\,\mathrm{d}\pi \\ &= -\sum_{i=1}^p \int_{\mathbb{R}^p} \partial_i(\partial_i\psi\cdot\psi)(\boldsymbol{x})\,\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \sum_{i=1}^p\int_{\mathbb{R}^p}\big(\partial^2_{ii}\psi\cdot\psi\big)\,\mathrm{d}\pi \\ &= -\sum_{i=1}^p \int_{\mathbb{R}^p} (\partial_i\psi\cdot\partial_i\psi)(\boldsymbol{x})\,\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = -\int_{\mathbb{R}^p}\|\nabla\psi(\boldsymbol{x})\|^2\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \qquad (7)\end{aligned}$$

Combining (4), (6) and (7), we arrive at

$$\|\psi(\boldsymbol{L}_t)-\psi(\boldsymbol{L}_0)\|^2_{\mathbb{L}_2} \leqslant 2t\int_{\mathbb{R}^p}\|\nabla\psi(\boldsymbol{x})\|^2\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

This yields the claim of the lemma for twice continuously differentiable functions ψ. The last step is to notice that the desired inequality does not involve the second-order derivatives of ψ. Therefore, if ψ is in C^1 but not in C^2, one can use the standard smoothing argument and then pass to the limit. More precisely, let us first assume that ψ has compact support. This implies that both ψ and $\nabla\psi$ are uniformly continuous. Choose a C^2 probability density function $\phi_0:\mathbb{R}^p\to\mathbb{R}$ vanishing outside $\mathbb{B}_1(0)$, the unit ball centered at the origin. Then, for any $\alpha>0$, the function $\psi_\alpha=\psi*\phi_\alpha$, where $\phi_\alpha(\boldsymbol{y})=\alpha^{-p}\phi_0(\boldsymbol{y}/\alpha)$ and $*$ stands for the convolution, is in C^2. Furthermore, for every $i\in\{1,\dots,p\}$,

$$\begin{aligned}|\partial_i\psi_\alpha(\boldsymbol{x})-\partial_i\psi(\boldsymbol{x})| &\leqslant \int_{\mathbb{R}^p}\big|\partial_i\psi(\boldsymbol{x}-\boldsymbol{y})-\partial_i\psi(\boldsymbol{x})\big|\phi_\alpha(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \\ &= \int_{\mathbb{B}_\alpha(0)}\big|\partial_i\psi(\boldsymbol{x}-\boldsymbol{y})-\partial_i\psi(\boldsymbol{x})\big|\phi_\alpha(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \\ &\leqslant \sup_{\boldsymbol{x}\in\mathbb{R}^p}\sup_{\boldsymbol{y}\in\mathbb{B}_\alpha(0)}\big|\partial_i\psi(\boldsymbol{x}-\boldsymbol{y})-\partial_i\psi(\boldsymbol{x})\big|.\end{aligned}$$

The uniform continuity of $\partial_i\psi$ implies that the right-hand side of the last display tends to zero as α tends to zero. Therefore, $\|\partial_i\psi_\alpha-\partial_i\psi\|_\infty\to 0$ as $\alpha\to 0$. This implies that $\|\nabla\psi_\alpha-\nabla\psi\|_\infty\to 0$ as $\alpha\to 0$. The same uniform convergence holds for $\psi_\alpha-\psi$. Consequently, writing inequality (3) for

ψ_α, then passing to the limit on both sides of the inequality when $\alpha \to 0$, we get the claim of the lemma for functions ψ that are in C^1 and have compact support.

The last step is to prove the claim for functions ψ that are not compactly supported. We choose a C^1 function $\theta : \mathbb{R}^p \to \mathbb{R}$ such that $\theta(\boldsymbol{y}) = 1$ if $\|\boldsymbol{y}\| \leqslant 1$ and $\theta(\boldsymbol{y}) = 0$ if $\|\boldsymbol{y}\| \geqslant 2$. Since the function $\boldsymbol{x} \mapsto \psi(\boldsymbol{x})\theta(\boldsymbol{x}/R)$ is in C^1 and compactly supported for any $R > 0$, we can write inequality (3) for this function. Then, it suffices to remark that when $R \to \infty$, $\psi_R(\boldsymbol{x}) := \psi(\boldsymbol{x})\theta(\boldsymbol{x}/R)$ converges in $L^2(\pi)$ to ψ, and the gradient of ψ_R converges to the gradient of ψ. Indeed,

$$\begin{aligned}\|\nabla\psi_R(\boldsymbol{L}_0) - \nabla\psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2} &\leqslant \|\nabla\psi(\boldsymbol{L}_0)\theta(\boldsymbol{L}_0/R) - \nabla\psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2} + R^{-1}\|\psi(\boldsymbol{L}_0)\nabla\theta(\boldsymbol{L}_0/R)\|_{\mathbb{L}_2} \\ &\leqslant \|\nabla\psi(\boldsymbol{L}_0)\mathbb{1}(\|\boldsymbol{L}_0\| > R)\|_{\mathbb{L}_2} + R^{-1}\|\psi(\boldsymbol{L}_0)\|_{\mathbb{L}_2}\|\nabla\theta\|_\infty.\end{aligned}$$

The fact that the right-hand side tends to zero when $R \to \infty$ follows from the dominated convergence theorem. This completes the proof of the lemma. □

Corollary 1. *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a twice continuously differentiable convex function such that ∇f and $\nabla^2 f$ are square-integrable with respect to $\pi \propto e^{-f}$. Then, for any $t > 0$,*

$$\|\nabla f(\boldsymbol{L}_t) - \nabla f(\boldsymbol{L}_0)\|_{\mathbb{L}_2} \leq \sqrt{2t}\,\|\nabla^2 f(\boldsymbol{L}_0)\|_{\mathbb{L}_2}.$$

Proof. It suffices to apply Lemma 1 to the function $\psi = \partial_i f$, for every fixed $i \in \{1, \dots, p\}$, square the resulting inequalities, and then sum them over i. The gradient form of the drift, together with invariance of π, guarantees reversibility. □

5 Proof of Theorem 1

We first assume that f belongs to C^∞ and that all its partial derivatives have at most polynomial growth.

We work on a probability space on which we define the initial value $\boldsymbol{\vartheta}_0$ of the LMC chain, the initial value $\boldsymbol{L}_0 \sim \pi$ of the stationary Langevin diffusion, and a Brownian motion $\boldsymbol{W}$ independent of $(\boldsymbol{\vartheta}_0, \boldsymbol{L}_0)$. We denote by $\mathscr{F}_t$ the minimal σ-algebra for which $\boldsymbol{\vartheta}_0, \boldsymbol{L}_0$ and $(\boldsymbol{W}_s)_{0\leqslant s\leqslant t}$ are measurable.

Let $\boldsymbol{\vartheta}'_{k+1} = \boldsymbol{\vartheta}'_k - h\nabla f(\boldsymbol{\vartheta}'_k) + \sqrt{2}\,(\boldsymbol{W}_{(k+1)h} - \boldsymbol{W}_{kh})$ and $\boldsymbol{L}_t = \boldsymbol{L}_0 - \int_0^t \nabla f(\boldsymbol{L}_s)\,\mathrm{d}s + \sqrt{2}\,\boldsymbol{W}_t$ for every $t > 0$, with $\boldsymbol{\vartheta}'_0 = \boldsymbol{L}_0$. Let us introduce the following random vectors

$$\begin{aligned}\boldsymbol{\Delta}'_k &= \boldsymbol{\vartheta}'_k - \boldsymbol{L}_{kh} \\ \boldsymbol{U}_k &= \nabla f(\boldsymbol{\vartheta}'_k) - \nabla f(\boldsymbol{L}_{kh}), \\ \boldsymbol{V}_k &= \int_{kh}^{(k+1)h} \Big(\nabla f(\boldsymbol{L}_t) - \nabla f(\boldsymbol{L}_{kh})\Big)\,\mathrm{d}t.\end{aligned}$$

Simple algebra yields

$$\boldsymbol{\Delta}'_{k+1} = \boldsymbol{\Delta}'_k - h\boldsymbol{U}_k + \boldsymbol{V}_k$$

Using Corollary 1, we get

$$\|\boldsymbol{V}_k\|^2_{\mathbb{L}_2} \leqslant \Big\{\int_0^h \big\|\nabla f(\boldsymbol{L}_{kh+t}) - \nabla f(\boldsymbol{L}_{kh})\big\|_{\mathbb{L}_2}\,\mathrm{d}t\Big\}^2 \leqslant \frac{8}{9}\,h^3\|\nabla^2 f\|^2_{\mathbb{L}_2(\pi)} \leqslant \frac{8}{9}\,M M_{\mathsf{av}} h^3 p.$$

It is proved in Pedrotti & Whalley (2026) that there exists a sequence of random vectors $(\boldsymbol{g}_k)_{k\geqslant 0}$ such that $\boldsymbol{g}_k$ is $\mathscr{F}_{kh}$- measurable,

$$\boldsymbol{V}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k + \boldsymbol{M}_k$$

with

$$\|\boldsymbol{g}_k\|_{\mathbb{L}_2} \leqslant h\|\nabla f\|_{\mathbb{L}_2(\pi)}, \qquad \|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|_{\mathbb{L}_2} \leqslant h\|\nabla f(\boldsymbol{L}_{(k+1)h}) - \nabla f(\boldsymbol{L}_{kh})\|_{\mathbb{L}_2},$$

and $\mathbf{E}[\boldsymbol{M}_k \mid \mathscr{F}_{kh}] = 0$. This last property and the fact that $\boldsymbol{g}_k$ is $\mathscr{F}_{kh}$-measurable imply that

$$\|\boldsymbol{\Delta}'_{k+1} - \boldsymbol{g}_{k+1}\|^2_{\mathbb{L}_2} = \|\boldsymbol{\Delta}'_k - h\boldsymbol{U}_k - \boldsymbol{g}_k\|^2_{\mathbb{L}_2} + \|\boldsymbol{M}_k\|^2_{\mathbb{L}_2}. \tag{8}$$

Define

$$\mathbf{H}_k := \int_0^1 \nabla^2 f\big(\boldsymbol{L}_{kh} + s(\boldsymbol{\vartheta}'_k - \boldsymbol{L}_{kh})\big)\,\mathrm{d}s.$$

Then, by the fundamental theorem of calculus,

$$\boldsymbol{U}_k = \nabla f(\boldsymbol{\vartheta}'_k) - \nabla f(\boldsymbol{L}_{kh}) = \mathbf{H}_k(\boldsymbol{\vartheta}'_k - \boldsymbol{L}_{kh}) = \mathbf{H}_k\boldsymbol{\Delta}'_k.$$

Moreover, the strong convexity of f and the M-Lipschitz continuity of its gradient imply that

$$m\mathbf{I}_p \preccurlyeq \mathbf{H}_k \preccurlyeq M\mathbf{I}_p.$$

The first term on the right-hand side of (8) is bounded as follows:

$$\begin{aligned}\|\boldsymbol{\Delta}'_k - h\boldsymbol{U}_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 &= \|(\mathbf{I}_p - h\mathbf{H}_k)\boldsymbol{\Delta}'_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 \\ &= \|(\mathbf{I}_p - h\mathbf{H}_k)(\boldsymbol{\Delta}'_k - \boldsymbol{g}_k) - h\mathbf{H}_k\boldsymbol{g}_k\|_{\mathbb{L}_2}^2\end{aligned}$$

Here we apply another important tweak used in (Pedrotti & Whalley, 2026): for any pair of vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p$ and for any symmetric matrix $\mathbf{A}$ satisfying $0 \preccurlyeq \mathbf{A} \preccurlyeq \mathbf{I}_p$, it holds $\|(\mathbf{I}_p - \mathbf{A})\boldsymbol{u} + \mathbf{A}\boldsymbol{v}\|^2 \leqslant \|(\mathbf{I}-\mathbf{A})^{1/2}\boldsymbol{u}\|^2 + \|\mathbf{A}^{1/2}\boldsymbol{v}\|^2$. This inequality follows from the following form of the Cauchy-Schwarz inequality:

$$2\boldsymbol{u}^\top(\mathbf{I} - \mathbf{A})\mathbf{A}\boldsymbol{v} = 2\boldsymbol{u}^\top(\mathbf{A} - \mathbf{A}^2)\boldsymbol{v} \leqslant \|(\mathbf{A} - \mathbf{A}^2)^{1/2}\boldsymbol{u}\|^2 + \|(\mathbf{A} - \mathbf{A}^2)^{1/2}\boldsymbol{v}\|^2.$$

We may apply the more precise relation

$$\|(\mathbf{I}_p - \mathbf{A})\boldsymbol{u} + \mathbf{A}\boldsymbol{v}\|^2 = \|(\mathbf{I} - \mathbf{A})^{1/2}\boldsymbol{u}\|^2 + \|\mathbf{A}^{1/2}\boldsymbol{v}\|^2 - \|(\mathbf{A} - \mathbf{A}^2)^{1/2}(\boldsymbol{u} - \boldsymbol{v})\|^2.$$

Choosing $\mathbf{A} = h\mathbf{H}_k$, $\boldsymbol{u} = \boldsymbol{\Delta}'_k - \boldsymbol{g}_k$, and $\boldsymbol{v} = -\boldsymbol{g}_k$, the last display implies that

$$\begin{aligned}\|\boldsymbol{\Delta}'_k - h\boldsymbol{U}_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 &= \|(\mathbf{I}_p - h\mathbf{H}_k)(\boldsymbol{\Delta}'_k - \boldsymbol{g}_k) - h\mathbf{H}_k\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 \\ &\leqslant (1 - mh)\|\boldsymbol{\Delta}'_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + Mh\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 - mh(1 - mh)\|\boldsymbol{\Delta}'_k\|_{\mathbb{L}_2}^2.\end{aligned}$$

In this inequality, we used $(m + M)h \leqslant 1$ to obtain the last term. Combining the last display with (8), we arrive at

$$\begin{aligned}\|\boldsymbol{\Delta}'_{k+1} - \boldsymbol{g}_{k+1}\|_{\mathbb{L}_2}^2 \leqslant (1 - mh)\|\boldsymbol{\Delta}'_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + Mh\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \|\boldsymbol{M}_k\|_{\mathbb{L}_2}^2 \\ - mh(1 - mh)\|\boldsymbol{\Delta}'_k\|_{\mathbb{L}_2}^2.\end{aligned}$$

The integration by parts formula and the Cauchy-Schwarz inequality yield

$$\|\boldsymbol{g}_k\|_{\mathbb{L}_2}^2 \leqslant h^2\|\nabla f\|_{\mathbb{L}_2(\pi)}^2 = h^2\int_{\mathbb{R}^p} \mathrm{tr}\left\{\nabla^2 f(\boldsymbol{x})\right\}\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = M_{\mathsf{av}}h^2 p.$$

We also have

$$\|\boldsymbol{M}_k\|_{\mathbb{L}_2}^2 \leqslant (\|\boldsymbol{V}_k\|_{\mathbb{L}_2} + \|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|_{\mathbb{L}_2})^2 \leqslant \frac{50}{9}\,h^3\|\nabla^2 f\|_{\mathbb{L}_2(\pi)}^2 \leqslant \frac{50}{9}\,MM_{\mathsf{av}}h^3 p.$$

Therefore, we arrive at the recursive inequality

$$\|\boldsymbol{\Delta}'_{k+1} - \boldsymbol{g}_{k+1}\|_{\mathbb{L}_2}^2 \leqslant (1 - mh)\|\boldsymbol{\Delta}'_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \frac{59}{9}MM_{\mathsf{av}}h^3 p - mh(1 - mh)\|\boldsymbol{\Delta}'_k\|_{\mathbb{L}_2}^2.$$

Using the nonpositivity of the last term, we get

$$\|\boldsymbol{\Delta}'_{k+1} - \boldsymbol{g}_{k+1}\|_{\mathbb{L}_2}^2 \leqslant (1 - mh)\|\boldsymbol{\Delta}'_k - \boldsymbol{g}_k\|_{\mathbb{L}_2}^2 + \frac{59}{9}MM_{\mathsf{av}}h^3 p. \tag{9}$$

Unfolding inequality (9), we get

$$\begin{aligned}\|\boldsymbol{\Delta}'_{k+1} - \boldsymbol{g}_{k+1}\|_{\mathbb{L}_2}^2 &\leqslant (1 - mh)^{k+1}\|\boldsymbol{\Delta}'_0 - \boldsymbol{g}_0\|_{\mathbb{L}_2}^2 + \frac{59}{9}\varkappa_{\mathsf{av}}Mh^2 p\big(1 - (1 - mh)^{k+1}\big) \\ &= (1 - mh)^{k+1}\|\boldsymbol{g}_0\|_{\mathbb{L}_2}^2 + \frac{59}{9}\varkappa_{\mathsf{av}}Mh^2 p\big(1 - (1 - mh)^{k+1}\big) \\ &\leqslant (59/9)\varkappa_{\mathsf{av}}Mh^2 p.\end{aligned}$$

This readily yields

$$\|\boldsymbol{\Delta}'_{k+1}\|_{\mathbb{L}_2} \leqslant \|\boldsymbol{g}_{k+1}\|_{\mathbb{L}_2} + \sqrt{(59/9)\varkappa_{\mathsf{av}} M p}\, h \leqslant 3.6\sqrt{\varkappa_{\mathsf{av}} M p}\, h.$$

Using the triangle inequality, we can now assess the distance between the LMC chain starting from any initial value $\boldsymbol{\vartheta}_0$ and the Langevin diffusion starting from the stationary distribution, assuming that they share the same Brownian motion (this means that $(\boldsymbol{\vartheta}_k)$ and $(\boldsymbol{\vartheta}'_k)$ have the same innovations $(\boldsymbol{\xi}_k)$):

$$\begin{aligned}\|\boldsymbol{\vartheta}_k - \boldsymbol{L}_{kh}\|_{\mathbb{L}_2} &\leqslant \|\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}'_k\|_{\mathbb{L}_2} + \|\boldsymbol{\Delta}'_k\|_{\mathbb{L}_2} \\ &\leqslant \|\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}'_k\|_{\mathbb{L}_2} + 3.6\sqrt{\varkappa_{\mathsf{av}} M p}\, h.\end{aligned}$$

Using the same arguments as in the proof of (Dalalyan & Karagulyan, 2019, Lemma 2), we get $\|\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}'_k\|_{\mathbb{L}_2} \leqslant (1 - mh)\|\boldsymbol{\vartheta}_{k-1} - \boldsymbol{\vartheta}'_{k-1}\|_{\mathbb{L}_2}$ which eventually leads to $\|\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}'_k\|_{\mathbb{L}_2} \leqslant (1 - mh)^k\|\boldsymbol{\vartheta}_0 - \boldsymbol{L}_0\|_{\mathbb{L}_2}$. Assuming that $\boldsymbol{L}_0$ is chosen to be optimally coupled with $\boldsymbol{\vartheta}_0$, we obtain

$$\begin{aligned}\mathsf{W}_2(\nu_k, \pi) &\leqslant \|\boldsymbol{\vartheta}_k - \boldsymbol{L}_{kh}\|_{\mathbb{L}_2} \\ &\leqslant (1 - mh)^k\|\boldsymbol{\vartheta}_0 - \boldsymbol{L}_0\|_{\mathbb{L}_2} + 3.6\sqrt{\varkappa_{\mathsf{av}} M p}\, h \\ &\leqslant (1 - mh)^k\, \mathsf{W}_2(\nu_0, \pi) + 3.6\sqrt{\varkappa_{\mathsf{av}} M p}\, h. \qquad (10)\end{aligned}$$

To complete the proof, we need to remove the assumption that f is in C^∞ and has all its partial derivatives of polynomial growth. To this end, we reproduce here (up to some slight adaptations) the argument in (Pedrotti & Whalley, 2026, Section 2.2). For $\alpha > 0$ we define $\pi_\alpha = \pi * \phi_\alpha$, where ϕ_α is the density of the Gaussian distribution $\mathcal{N}(0, \alpha\mathbf{I}_p)$. This new density is known to be log-concave and to satisfy

$$\frac{m}{m\alpha + 1}\,\mathbf{I}_p \preccurlyeq \nabla^2 f_\alpha(\boldsymbol{\theta}) \preccurlyeq \frac{M}{M\alpha + 1}\mathbf{I}_p, \qquad \forall \boldsymbol{\theta} \in \mathbb{R}^p, \qquad \text{where} \quad f_\alpha = -\log \pi_\alpha,$$

see (Saumard & Wellner, 2014). Thanks to the smoothing property of the convolution with a Gaussian kernel, the function f_α is in C^∞. It follows from (Klartag & Putterman, 2023, lemma 2.2) that all the partial derivatives of f_α have at most polynomial growth. This implies that we can apply (10) to the chain

$$\boldsymbol{\vartheta}^\alpha_{k+1} = \boldsymbol{\vartheta}^\alpha_k - h\nabla f_\alpha(\boldsymbol{\vartheta}^\alpha_k) + \sqrt{2}\,(\boldsymbol{W}_{(k+1)h} - \boldsymbol{W}_{kh})$$

starting from the same initial point $\boldsymbol{\vartheta}_0$ and the LMC chain $(\boldsymbol{\vartheta}_k)$. This implies that the kth-step distribution ν^α_k of the chain $(\boldsymbol{\vartheta}^\alpha_k)$ satisfies

$$\mathsf{W}_2(\nu^\alpha_k, \pi_\alpha) \leqslant (1 - m^\alpha h)^k\, \mathsf{W}_2(\nu_0, \pi_\alpha) + 3.6\sqrt{\varkappa^\alpha_{\mathsf{av}} M^\alpha p}\, h.$$

Here,

$$m^\alpha = \frac{m}{m\alpha + 1}, \quad M^\alpha = \frac{M}{M\alpha + 1}, \quad \varkappa^\alpha_{\mathsf{av}} = \frac{M^\alpha_{\mathsf{av}}}{m^\alpha}, \quad M^\alpha_{\mathsf{av}} = p^{-1}\|\operatorname{tr}(\nabla^2 f_\alpha)\|_{\mathbb{L}_1(\pi_\alpha)}.$$

The triangle inequality yields

$$\begin{aligned}\mathsf{W}_2(\nu_k, \pi) &\leqslant \mathsf{W}_2(\nu_k, \nu^\alpha_k) + \mathsf{W}_2(\nu^\alpha_k, \pi_\alpha) + \mathsf{W}_2(\pi_\alpha, \pi) \\ &\leqslant \mathsf{W}_2(\nu_k, \nu^\alpha_k) + (1 - m^\alpha h)^k\, \mathsf{W}_2(\nu_0, \pi_\alpha) + 3.6\sqrt{\varkappa^\alpha_{\mathsf{av}} M^\alpha p}\, h + \mathsf{W}_2(\pi_\alpha, \pi) \\ &\leqslant \mathsf{W}_2(\nu_k, \nu^\alpha_k) + (1 - m^\alpha h)^k\, \mathsf{W}_2(\nu_0, \pi) + 3.6\sqrt{\varkappa^\alpha_{\mathsf{av}} M^\alpha p}\, h + 2\mathsf{W}_2(\pi_\alpha, \pi) \\ &\leqslant \mathsf{W}_2(\nu_k, \nu^\alpha_k) + (1 - m^\alpha h)^k\, \mathsf{W}_2(\nu_0, \pi) + 3.6\sqrt{\varkappa^\alpha_{\mathsf{av}} M^\alpha p}\, h + 2\sqrt{\alpha p}. \qquad (11)\end{aligned}$$

The last inequality follows from the fact that if $\boldsymbol{X} \sim \pi$ and $\boldsymbol{\xi} \sim \mathcal{N}(0, \alpha\mathbf{I}_p)$ is independent of $\boldsymbol{X}$, then $\boldsymbol{X} + \boldsymbol{\xi} \sim \pi_\alpha$ and $\|\boldsymbol{X} + \boldsymbol{\xi} - \boldsymbol{X}\|^2_{\mathbb{L}_2} = \alpha p$. To complete the proof, it suffices to establish the convergences

$$\mathsf{W}_2(\nu_k, \nu^\alpha_k) \xrightarrow[\alpha\to 0]{} 0 \qquad \text{and} \qquad \|\operatorname{tr}(\nabla^2 f_\alpha)\|_{\mathbb{L}_1(\pi_\alpha)} \xrightarrow[\alpha\to 0]{} \|\operatorname{tr}(\nabla^2 f)\|_{\mathbb{L}_1(\pi)} \qquad (12)$$

and to pass to the limit in (11). For the first limit, we use the fact that

$$\begin{aligned}\|\boldsymbol{\vartheta}^\alpha_{k+1} - \boldsymbol{\vartheta}_{k+1}\|_{\mathbb{L}_2} &\leqslant \|\boldsymbol{\vartheta}^\alpha_k - \boldsymbol{\vartheta}_k - h\big(\nabla f_\alpha(\boldsymbol{\vartheta}^\alpha_k) - \nabla f_\alpha(\boldsymbol{\vartheta}_k)\big)\|_{\mathbb{L}_2} + h\|\nabla f_\alpha(\boldsymbol{\vartheta}_k) - \nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2} \\ &\leqslant (1 - m^\alpha h)\|\boldsymbol{\vartheta}^\alpha_k - \boldsymbol{\vartheta}_k\|_{\mathbb{L}_2} + h\|\nabla f_\alpha(\boldsymbol{\vartheta}_k) - \nabla f(\boldsymbol{\vartheta}_k)\|_{\mathbb{L}_2} \\ &\leqslant h\sum_{j=0}^{k}\|\nabla f_\alpha(\boldsymbol{\vartheta}_j) - \nabla f(\boldsymbol{\vartheta}_j)\|_{\mathbb{L}_2}.\end{aligned}$$

It is easy to verify that ∇f_α converges pointwise to ∇f as $\alpha \to 0$ and that

$$\|\nabla f_\alpha(\boldsymbol{x}) - \nabla f(\boldsymbol{x})\| \leqslant \|\nabla f_\alpha(0) - \nabla f(0)\| + 2M\|\boldsymbol{x}\| \leqslant C + 2M\|\boldsymbol{x}\|.$$

Since the distribution of $\boldsymbol{\vartheta}_j$, ν_j, has a finite second moment for every j, we can apply the dominated convergence theorem to infer that $\|\nabla f_\alpha(\boldsymbol{\vartheta}_j) - \nabla f(\boldsymbol{\vartheta}_j)\|_{\mathbb{L}_2} \to 0$ as $\alpha \to 0$. The same type of argument proves the second convergence in (12): the integrand converges pointwise and is bounded by a function of the form $ce^{-c\|\boldsymbol{x}\|}$. This completes the proof of the theorem.

6 Proof of Proposition 1: order of the error for Gaussian targets

Let $a_1, \dots, a_p$ denote the eigenvalues of $\mathbf{A} := \boldsymbol{\Sigma}^{-1}$, so that $m \leqslant a_i \leqslant M$. Since the LMC recursion is linear, ν_k is a centered Gaussian distribution; write $\nu_k = \mathcal{N}(0, \mathbf{C}_k)$. Its covariance satisfies $\mathbf{C}_{k+1} = (\mathbf{I}_p - h\mathbf{A})\mathbf{C}_k(\mathbf{I}_p - h\mathbf{A}) + 2h\mathbf{I}_p$, with $\mathbf{C}_0 = \mathbf{A}^{-1}$.

The matrices $\mathbf{C}_k$ and $\mathbf{A}$ are simultaneously diagonalizable. In an eigendirection of $\mathbf{A}$ corresponding to a_i, the variance of ν_k is

$$c_{k,i} = \frac{(1-ha_i)^{2k}}{a_i} + 2h\sum_{j=0}^{k-1}(1-ha_i)^{2j} = \frac{1}{a_i}\left[1 + \frac{ha_i}{2-ha_i}\left\{1-(1-ha_i)^{2k}\right\}\right].$$

Set $s_{k,i} = 1-(1-ha_i)^{2k}$ and $u_{k,i} = ha_i s_{k,i}/(2-ha_i)$. Since the covariance matrices of ν_k and π commute, the formula for the Wasserstein distance between centered Gaussian distributions gives

$$\mathsf{W}_2^2(\nu_k, \pi) = \sum_{i=1}^p \left(\sqrt{c_{k,i}} - \frac{1}{\sqrt{a_i}}\right)^2 = h^2 \sum_{i=1}^p \frac{a_i s_{k,i}^2}{(2-ha_i)^2(\sqrt{1+u_{k,i}}+1)^2}. \tag{13}$$

The assumption $hM \leqslant 1$ implies that $0 \leqslant ha_i \leqslant 1$, and hence $s_{k,i} \in [0,1]$. The denominator before squaring in (13) is bounded below by 2, since $(2-ha_i)(\sqrt{1+u_{k,i}}+1) \geqslant 2(2-ha_i) \geqslant 2$. It is also bounded above by 4: indeed, $u_{k,i} \leqslant ha_i/(2-ha_i)$, so this denominator is at most $(2-ha_i) + \{2(2-ha_i)\}^{1/2} \leqslant 4$. It follows that

$$\frac{h^2}{16}\sum_{i=1}^p a_i s_{k,i}^2 \leqslant \mathsf{W}_2^2(\nu_k, \pi) \leqslant \frac{h^2}{4}\sum_{i=1}^p a_i s_{k,i}^2.$$

Since $a_i \geqslant m$ and $0 \leqslant ha_i \leqslant 1$, we have $s_{k,i} \geqslant 1-(1-hm)^{2k}$. Therefore,

$$\sum_{i=1}^p a_i s_{k,i}^2 \geqslant \left\{1-(1-hm)^{2k}\right\}^2 \sum_{i=1}^p a_i = \left\{1-(1-hm)^{2k}\right\}^2 \operatorname{tr}(\mathbf{A}).$$

On the other hand, $s_{k,i} \leqslant 1$ implies $\sum_{i=1}^p a_i s_{k,i}^2 \leqslant \operatorname{tr}(\mathbf{A})$. Combining these bounds and taking square roots proves (2).

Reproducibility statement

This paper is theoretical. All assumptions are stated explicitly, and complete proofs of the main theorem, the key lemma, and the Gaussian calculation are included in the main body.

AI use statement

ChatGPT was used to assist with checking and presenting mathematical derivations, editing proofs and exposition, and adapting the manuscript's LaTeX formatting. It was not used to generate data, run experiments, or interpret empirical results; those tasks are not applicable to this work. All mathematical statements and proofs were independently checked, and we take responsibility for the final content of this work, including text and claims produced with the aid of generative AI.

REFERENCES

Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In Satyen Kale and Ohad Shamir (eds.), *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pp. 678–689, 07–10 Jul 2017a.

Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from a smooth and log-concave density. *J. R. Stat. Soc. B*, 79:651 – 676, 2017b.

Arnak S Dalalyan and Avetik Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 2019.

Arnak S. Dalalyan and Avetik Karagulyan. Improved guarantees for langevin monte carlo with average smoothness, 2026. URL https://arxiv.org/abs/2605.31413.

Arnak S Dalalyan, Mohamed Hebiri, and Johannes Lederer. On the prediction performance of the LASSO. *arXiv preprint arXiv:1402.1700*, 2014.

Alain Durmus and Eric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 11 2019. doi: 10.3150/18-BEJ1073. URL https://doi.org/10.3150/18-BEJ1073.

Alain Durmus, Szymon Majewski, and Blazej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *J. Mach. Learn. Res.*, 20:73–1, 2019.

Bo'az Klartag and Eli Putterman. Spectral monotonicity under Gaussian convolution. *Annales de la Faculté des sciences de Toulouse : Mathématiques*, 32(5):939–967, 2023. doi: 10.5802/afst.1759.

Francesco Pedrotti and Peter A. Whalley. Wasserstein mixing time of the unadjusted langevin algorithm, 2026. URL https://arxiv.org/abs/2608.02430.

A. Saumard and J. A. Wellner. Log-concavity and strong log-concavity: a review. *Stat. Surv.*, 8: 45–114, 2014.