A POISSON CORRECTOR FOR COMPLEXITY BOUNDS OF MOREAU–YOSIDA UNADJUSTED LANGEVIN SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study the classical Moreau–Yosida unadjusted Langevin algorithm (MYULA) of sampling from $\pi(\mathrm{d}x) \propto e^{-f(x)-g(x)}\,\mathrm{d}x$, where $f \in C^2(\mathbb{R}^d)$ is m-strongly convex and has a L_f-Lipschitz gradient, and $g\colon \mathbb{R}^d \to \mathbb{R}$ is convex and globally G-Lipschitz. For the Moreau-smoothed target π_λ and the MYULA invariant law $\widehat{\pi}_{\lambda,h}$, we prove

$$\sqrt{m}\, W_2(\pi_\lambda, \widehat{\pi}_{\lambda,h}) = O(h) + \widetilde{O}(h^{3/4})$$

under $0 < h(L_f + \lambda^{-1}) \le c$, with only logarithmic dependence on λ^{-1} in the error coefficients. Combining this estimate with the Moreau approximation bias yields $\widetilde{O}(\varepsilon^{-4/3})$ iterations to achieve $\sqrt{m}\, W_2(\mu_N, \pi) \le \varepsilon$, for fixed model parameters and initialization. The proof combines a discrete Poisson corrector with active-trace estimates and a shared-noise bound for the exact–Euler two-point curvature.

1 INTRODUCTION

Sampling from high-dimensional distributions is a fundamental problem in statistics and machine learning. We consider distributions of the form

$$\pi(\mathrm{d}x) \propto \exp\{-f(x) - g(x)\}\,\mathrm{d}x, \qquad x \in \mathbb{R}^d,$$

where $f \in C^2(\mathbb{R}^d)$ is m-strongly convex and its gradient is L_f-Lipschitz, and $g : \mathbb{R}^d \to \mathbb{R}$ is convex and globally G-Lipschitz, with $G > 0$. The function g needs not be differentiable; examples include ℓ_1 and total-variation penalties (Durmus et al., 2018; Mou et al., 2022).

The Moreau–Yosida unadjusted Langevin algorithm (MYULA) (Durmus et al., 2018) replaces g by its Moreau envelope g_λ and applies an Euler step to the corresponding Langevin diffusion:

$$\widehat{X}_{k+1} = \widehat{X}_k - h\nabla(f + g_\lambda)(\widehat{X}_k) + \sqrt{2h}\,\xi_{k+1}, \qquad \xi_{k+1} \overset{\text{i.i.d.}}{\sim} N(0, I).$$

Because $\nabla g_\lambda(x) = \lambda^{-1}(x - \mathrm{prox}_{\lambda g}(x))$, each iteration requires one gradient evaluation of f and one proximal evaluation of g. This makes MYULA a natural sampling method when the same proximal operations are already available for optimization.

The main difficulty is the interaction between smoothing and discretization. The smoothed target $\pi_\lambda(\mathrm{d}x) \propto e^{-f(x)-g_\lambda(x)}\,\mathrm{d}x$ approaches π as λ decreases, but the global smoothness bound

$$L_\lambda = L_f + \lambda^{-1}$$

increases. Moreover, the invariant law of the Euler scheme is generally different from π_λ. Accuracy for the original target therefore requires controlling three errors: the Moreau approximation bias, the Euler invariant-measure bias, and convergence from the initial distribution. A fixed-λ estimate is useful for this purpose only when its dependence on λ is also quantified.

Two recent works provide the main ingredients for our analysis. Xin & Zhang (2026) show that averaged weak Moreau curvature can replace a worst-case curvature bound in the analysis of MYULA. Pedrotti & Whalley (2026) use a discrete Poisson decomposition to obtain a linear-in-h invariant-measure bias for smooth, strongly log-concave targets. Direct specialization of their global-smoothness bound to $f + g_\lambda$, combined with $\lambda \asymp \varepsilon/G^2$, gives a sufficient $\widetilde{O}(\varepsilon^{-2})$ MYULA complexity; the calculation is given in Section 2. We obtain a sharper dependence on accuracy by controlling the curvature terms within the corrector argument.

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.

Main result. Let $\widehat{\pi}_{\lambda,h}$ be the invariant law of MYULA. Under $0 < hL_\lambda \le c$, Theorem 5.1 gives

$$\sqrt{m}\, W_2(\pi_\lambda, \widehat{\pi}_{\lambda,h}) \le C_1 h + C_2 h^{3/4} \left[1 + \log\left(e + \frac{\sqrt{h}}{G\sqrt{d}\,\lambda}\right)\right]^2,$$

where C_1, C_2 depend on the model parameters but not on h or λ. Their parameter dependence is explicit in the theorem. Thus, the error coefficients depend on λ^{-1} only logarithmically, although the usual step-size restriction remains.

Combining this estimate with

$$\sqrt{m}\, W_2(\pi_\lambda, \pi) \le \frac{G^2\lambda}{4}$$

from Xin & Zhang (2026, Proposition 5.9), we choose $\lambda = \varepsilon/G^2$ and the step size specified in Theorem 5.2. For fixed model parameters and a fixed initial law $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, the resulting sufficient iteration complexity is

$$N(\varepsilon) = \widetilde{O}(\varepsilon^{-4/3}) \qquad \text{for} \qquad \sqrt{m}\, W_2(\mu_N, \pi) \le \varepsilon,$$

where μ_N is the law of the Nth iterate and $\widetilde{O}$ suppresses logarithmic factors.

Proof approach. We follow the Poisson-corrector framework of Pedrotti & Whalley (2026), comparing MYULA with the exact Langevin diffusion for π_λ driven by the same noise. The corrector captures cancellations between local errors at successive steps. Our refinement combines the active-trace estimates of Xin & Zhang (2026) with a shared-noise curvature estimate for the coupled processes. This allows us to control average curvature rather than use the worst-case bound λ^{-1} for g_λ throughout the error estimate. Together with contraction from strong convexity, these estimates yield the stated invariant-measure bias bound.

The algorithm itself is unchanged: λ and h are fixed during each run, and no Metropolis correction is introduced. Our complexity counts iterations with exact proximal evaluations, not the internal cost of solving proximal subproblems. Section 2 places the result in the literature; Sections 3–5 give the definitions, assumptions, and precise statements. The proofs of the main results are given in Section 6; the technical estimates are deferred to Appendices C and D.

2 Related Work

Moreau regularization and proximal sampling. Pereyra (2016) developed proximal MCMC methods for nonsmooth log-concave models, and Durmus et al. (2018) introduced the classical MYULA construction with asymptotic and nonasymptotic guarantees. These methods connect sampling with proximal operations used in convex optimization. A complementary viewpoint is provided by Durmus et al. (2019), who interpret Langevin algorithms through optimization on Wasserstein space and analyze both smooth and nonsmooth variants. For a smoothing-based method, bounds for a fixed surrogate target must be combined with the regularization bias before their accuracy dependence can be compared with an end-to-end guarantee.

Active-trace analysis of MYULA. A closely related analysis of the same transition kernel is given by Xin & Zhang (2026). Their reference heat-path active trace averages the weak Moreau Hessian trace along the Gaussian substep of an update started from π_λ. For the structured piecewise-linear, lasso-type, group, and total-variation penalties treated there, curvature–tube estimates yield $\widetilde{O}(\varepsilon^{-2})$ end-to-end complexity. We use their weak-curvature and Moreau-bias tools, but replace the reference-path recursion by a corrected synchronous coupling. The additional shared-noise estimate controls the actual exact–Euler two-point curvature for general convex Lipschitz g, without the geometric verification required by those examples.

Poisson correctors and a direct MYULA benchmark. For an m-strongly convex potential with L-Lipschitz gradient, Pedrotti & Whalley (2026, Theorem 1) prove

$$\sqrt{m}\, W_2(\pi, \widehat{\pi}) \le 6hL\sqrt{d}, \qquad 0 < hL \le 1.$$

Specializing this bound to $U_\lambda = f + g_\lambda$ gives

$$\sqrt{m}\, W_2(\pi_\lambda, \widehat{\pi}_{\lambda,h}) \le 6hL_\lambda\sqrt{d}.$$

Here we use the extension from C^2 to $C^{1,1}$ potentials by mollification.[1] With $\lambda \asymp \varepsilon/G^2$, choosing $h \asymp \varepsilon/(L_\lambda\sqrt{d})$ and using Wasserstein contraction yield the sufficient iteration bound

$$N_{\mathrm{PW}}(\varepsilon) = \widetilde{O}\left[\frac{\sqrt{d}}{m}\left(\frac{L_f}{\varepsilon} + \frac{G^2}{\varepsilon^2}\right)\right].$$

Thus, direct use of their global-smoothness estimate gives $\widetilde{O}(\varepsilon^{-2})$ accuracy dependence for MYULA. Our contribution is a refinement of the curvature terms in the corrector analysis, leading to $\widetilde{O}(\varepsilon^{-4/3})$ under the assumptions of this paper. The gain concerns the smoothing dependence, not the order in h for a fixed smoothed target.

Other composite samplers. Different transitions can avoid a separate Moreau approximation of the target. Salim & Richtárik (2020) give a primal–dual analysis of proximal stochastic gradient Langevin sampling, including a $\widetilde{O}(\varepsilon^{-2})$ Wasserstein complexity in the strongly convex setting. Mou et al. (2022) combine proximal sampling proposals with a Metropolis–Hastings correction and obtain polylogarithmic dependence on ε^{-1} for total variation error at most ε, under their assumptions. Their method requires a proximal sampling oracle and evaluation of the associated normalizing constants, rather than only a deterministic proximal operator. These results concern different transition kernels and, in the latter case, a different error metric and oracle model. Our result concerns classical MYULA with exact proximal evaluations and accuracy in $\sqrt{m}\, W_2$ for the original composite target.

3 Preliminaries

3.1 Basic notation

All probability measures are defined on the Borel σ-algebra of $\mathbb{R}^d$. The Euclidean norm and inner product are denoted by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$. Gradients of scalar functions are column vectors. For a vector field $F : \mathbb{R}^d \to \mathbb{R}^d$, we use $(\nabla F)_{ij} := \partial_j F_i$, so that $\nabla^2 u = \nabla(\nabla u)$ for a scalar function u. For matrices, $\|\cdot\|_{\mathrm{op}}$ denotes the Euclidean operator norm, and $\operatorname{sym} A := (A + A^\top)/2$. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the quadratic Wasserstein distance is

$$W_2^2(\mu,\nu) := \inf_{\substack{\mathcal{L}(X)=\mu \\ \mathcal{L}(Y)=\nu}} \mathbb{E}\,\|X - Y\|^2. \tag{1}$$

For symmetric matrices $A, B \in \mathbb{R}^{d\times d}$, the notation $A \preceq B$ means $v^\top A v \le v^\top B v$ for every $v \in \mathbb{R}^d$. Constants $c, C > 0$ are universal and may change from line to line. For a measure μ and a vector-valued function F, we use $\|F\|_{L^p(\mu)} = (\int \|F(x)\|^p \,\mu(\mathrm{d}x))^{1/p}$.

3.2 Moreau–Yosida regularization and weak second-order structure

Let $g : \mathbb{R}^d \to \mathbb{R}$ be convex. For $\lambda > 0$, its Moreau envelope is

$$g_\lambda(x) := \inf_{y\in\mathbb{R}^d}\left\{g(y) + \frac{\|x-y\|^2}{2\lambda}\right\}. \tag{2}$$

Standard Moreau theory gives

$$\nabla g_\lambda(x) = \lambda^{-1}\big(x - \operatorname{prox}_{\lambda g}(x)\big) \tag{3}$$

and the cocoercivity inequality

$$\langle x - y, \nabla g_\lambda(x) - \nabla g_\lambda(y)\rangle \ge \lambda\,\|\nabla g_\lambda(x) - \nabla g_\lambda(y)\|^2; \tag{4}$$

[1] Apply the bound to $f + (g_\lambda * \rho_\delta)$, where ρ_δ is a standard smooth mollifier. Mollification preserves the bounds m and L_λ, and the gradients converge uniformly. For fixed λ and h, synchronous contraction gives W_2 convergence of both the diffusion and Euler invariant laws, so the estimate passes to the limit.

see, for example, Bauschke & Combettes (2017, Chapters 12 and 18). Consequently $g_\lambda \in C^{1,1}(\mathbb{R}^d)$ and $\mathrm{Lip}(\nabla g_\lambda) \le \lambda^{-1}$. If g is G-Lipschitz, then

$$\|\nabla g_\lambda(x)\| \le G, \qquad x \in \mathbb{R}^d. \tag{5}$$

Because ∇g_λ is Lipschitz, it is differentiable almost everywhere. Its almost-everywhere derivative also represents its weak derivative; we write

$$H_\lambda := \nabla^2 g_\lambda \quad \text{and} \quad a_\lambda := \operatorname{tr} H_\lambda. \tag{6}$$

The matrix field H_λ may be chosen measurable and satisfies

$$0 \preceq H_\lambda(x) \preceq \lambda^{-1} I, \qquad 0 \le a_\lambda(x) \le d/\lambda \quad \text{for a.e. } x. \tag{7}$$

Following Xin & Zhang (2026), we call a_λ the *Moreau active trace*. The weak-Hessian properties and weighted integration-by-parts identities are collected in Appendix A; the semigroup integration-by-parts identity is established in Proposition C.6.

4 Problem Setup and Standing Assumptions

We consider the composite target

$$\pi(\,\mathrm{d}x) = Z^{-1} \exp\{-f(x) - g(x)\}\,\mathrm{d}x. \tag{8}$$

Assumption 4.1 (Smooth component)**.** The function $f \in C^2(\mathbb{R}^d)$, and there are constants $0 < m \le L_f < \infty$ such that

$$mI \preceq \nabla^2 f(x) \preceq L_f I, \qquad x \in \mathbb{R}^d. \tag{9}$$

We write

$$\tau_f := \sup_{x \in \mathbb{R}^d} \operatorname{tr} \nabla^2 f(x) \le d L_f. \tag{10}$$

Assumption 4.2 (Nonsmooth component)**.** The function $g : \mathbb{R}^d \to \mathbb{R}$ is convex and globally G-Lipschitz for some $G > 0$.

For $\lambda > 0$, define

$$U_\lambda := f + g_\lambda, \qquad \pi_\lambda(\,\mathrm{d}x) := Z_\lambda^{-1} e^{-U_\lambda(x)}\,\mathrm{d}x, \qquad L_\lambda := L_f + \lambda^{-1}. \tag{11}$$

Then U_λ is m-strongly convex, belongs to $C^{1,1}(\mathbb{R}^d)$, and $\mathrm{Lip}(\nabla U_\lambda) \le L_\lambda$.

MYULA is the Euler scheme for the Langevin diffusion with potential U_λ:

$$\widehat{X}_{k+1} = \widehat{X}_k - h \nabla U_\lambda(\widehat{X}_k) + \sqrt{2h}\,\xi_{k+1}, \qquad \xi_{k+1} \overset{\text{i.i.d.}}{\sim} N(0, I). \tag{12}$$

Its transition kernel is denoted by $Q_{\lambda,h}$, and $\mu_k := \mathcal{L}(\widehat{X}_k)$. This normalization and the use of $\sqrt{m}\,W_2$ agree with the setup in Xin & Zhang (2026, Section 4).

5 Main Results

Define

$$R := \frac{G + \sqrt{G^2 + 4\tau_f}}{2}, \qquad K := 2G + \frac{L_f}{\sqrt{m}}(\sqrt{d} + 2), \tag{13}$$

and

$$\ell_{h,\lambda} := 1 + \log\left(e + \frac{\sqrt{h}}{G\sqrt{d}\,\lambda}\right). \tag{14}$$

The constants R and K are independent of λ. Lemmas C.1 and C.2 show that $\|\nabla U_\lambda\|_{L^2(\pi_\lambda)} \le R$, $\mathbb{E}_{\pi_\lambda} a_\lambda \le GR$, and $\|\nabla U_\lambda\|_{L^p(\pi_\lambda)} \le K\sqrt{p}$ for every $p \ge 2$. The factor $\ell_{h,\lambda}$ records the logarithmic dependence on the smoothing parameter.

Theorem 5.1 (Invariant-measure bias at fixed smoothing)**.** *Under Assumptions 4.1–4.2, there exist universal constants $c, C > 0$ such that the following holds. For every $\lambda > 0$ and*

$$0 < hL_\lambda \le c, \tag{15}$$

the kernel $Q_{\lambda,h}$ has a unique invariant distribution $\widehat{\pi}_{\lambda,h} \in \mathcal{P}_2(\mathbb{R}^d)$, and

$$\sqrt{m}\, W_2(\pi_\lambda, \widehat{\pi}_{\lambda,h}) \le C\Big[R(\sqrt{m} + \sqrt{L_f}) + L_f\sqrt{d}\Big]h + C\Big[G\sqrt{R} + K\sqrt{G}\, d^{1/4}\ell_{h,\lambda}^2\Big]h^{3/4}. \tag{16}$$

The stability condition still involves L_λ, but the bias coefficients in equation 16 depend on λ^{-1} only logarithmically. We next choose λ and h jointly to control the error for the original target.

Theorem 5.2 (End-to-end complexity)**.** *Under Assumptions 4.1–4.2, let $0 < \varepsilon \le 1$ and set*

$$\lambda_\varepsilon := \frac{\varepsilon}{G^2}, \qquad \ell_\varepsilon := 1 + \log(e + \varepsilon^{-1}). \tag{17}$$

For a sufficiently small universal constant $c > 0$, set

$$h_\varepsilon := c\min\left\{\frac{1}{L_f + G^2/\varepsilon}, \frac{\varepsilon}{R(\sqrt{m} + \sqrt{L_f}) + L_f\sqrt{d}}, \left(\frac{\varepsilon}{G\sqrt{R} + K\sqrt{G}\, d^{1/4}\ell_\varepsilon^2}\right)^{4/3}\right\}. \tag{18}$$

For every $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, if

$$N \ge \frac{C}{mh_\varepsilon}\log\left(2 + \frac{\sqrt{m}\, W_2(\mu_0, \pi)}{\varepsilon}\right), \tag{19}$$

then

$$\sqrt{m}\, W_2(\mu_0 Q_{\lambda_\varepsilon,h_\varepsilon}^N, \pi) \le \varepsilon. \tag{20}$$

Moreover, the smallest integer satisfying equation 19 obeys

$$N \le \frac{C}{m}\Bigg[L_f + \frac{G^2}{\varepsilon} + \frac{R(\sqrt{m} + \sqrt{L_f}) + L_f\sqrt{d}}{\varepsilon} + \frac{G^{4/3}R^{2/3} + K^{4/3}G^{2/3}d^{1/3}\ell_\varepsilon^{8/3}}{\varepsilon^{4/3}}\Bigg]\log\left(2 + \frac{\sqrt{m}\, W_2(\mu_0, \pi)}{\varepsilon}\right) + 1. \tag{21}$$

In particular, for fixed m, L_f, τ_f, G, d and fixed μ_0,

$$N(\varepsilon) = \widetilde{O}(\varepsilon^{-4/3}). \tag{22}$$

The three constraints in equation 18 ensure stability and control the h and $h^{3/4}$ terms in equation 16, respectively. For fixed model parameters, balancing the $h^{3/4}$ term with ε gives $h \asymp \varepsilon^{4/3}$ up to logarithmic factors, yielding $N = \widetilde{O}(\varepsilon^{-4/3})$ by contraction.

6 Proof of the Main Results

We follow the corrected coupling of Pedrotti & Whalley (2026, Section 2.1). Our refinements concern two inputs: the variance of the martingale increments and the quadratic form involving the corrector and the Moreau curvature.

6.1 The two-point drift matrix

We represent the drift difference between the coupled processes by a symmetric matrix, so that strong convexity can be used directly in the error recursion. For the Moreau component, the rank-one construction below uses only the gradients at the two endpoints. For $x, y \in \mathbb{R}^d$, define

$$A^f(x, y) := \int_0^1 \nabla^2 f\big(y + t(x - y)\big)\, \mathrm{d}t \tag{23}$$

and

$$A^g_\lambda(x,y) := \begin{cases} \dfrac{\big(\nabla g_\lambda(x) - \nabla g_\lambda(y)\big)\big(\nabla g_\lambda(x) - \nabla g_\lambda(y)\big)^\top}{\langle x-y, \nabla g_\lambda(x) - \nabla g_\lambda(y)\rangle}, & \nabla g_\lambda(x) \neq \nabla g_\lambda(y), \\ 0, & \nabla g_\lambda(x) = \nabla g_\lambda(y). \end{cases} \tag{24}$$

Set

$$A_\lambda(x,y) := A^f(x,y) + A^g_\lambda(x,y). \tag{25}$$

We refer to $A_\lambda(x,y)$ as the *two-point drift matrix*.

Lemma 6.1 (Exact two-point drift identity)**.** *For every $x, y \in \mathbb{R}^d$,*

$$\nabla U_\lambda(x) - \nabla U_\lambda(y) = A_\lambda(x,y)(x-y), \tag{26}$$

$$mI \preceq A_\lambda(x,y) \preceq L_\lambda I, \tag{27}$$

$$0 \preceq A^f(x,y) \preceq L_f I, \qquad 0 \preceq A^g_\lambda(x,y) \preceq \lambda^{-1} I. \tag{28}$$

Proof. The fundamental theorem of calculus gives

$$A^f(x,y)(x-y) = \nabla f(x) - \nabla f(y), \qquad mI \preceq A^f(x,y) \preceq L_f I.$$

Suppose first that $\nabla g_\lambda(x) \neq \nabla g_\lambda(y)$. The denominator in equation 24 is strictly positive by equation 4, and direct multiplication yields

$$A^g_\lambda(x,y)(x-y) = \nabla g_\lambda(x) - \nabla g_\lambda(y).$$

For any $v \in \mathbb{R}^d$, Cauchy–Schwarz and equation 4 give

$$\begin{aligned} v^\top A^g_\lambda(x,y) v &= \frac{\langle v, \nabla g_\lambda(x) - \nabla g_\lambda(y)\rangle^2}{\langle x-y, \nabla g_\lambda(x) - \nabla g_\lambda(y)\rangle} \\ &\leq \frac{\|v\|^2 \|\nabla g_\lambda(x) - \nabla g_\lambda(y)\|^2}{\lambda \|\nabla g_\lambda(x) - \nabla g_\lambda(y)\|^2} = \lambda^{-1} \|v\|^2 . \end{aligned}$$

The case of a zero gradient difference is immediate from the definition. Adding the two components proves all assertions. □

6.2 The Poisson corrector and two technical inputs

Let $(X_t)_{t\geq 0}$ be the stationary Langevin diffusion

$$\mathrm{d}X_t = -\nabla U_\lambda(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad X_0 \sim \pi_\lambda, \tag{29}$$

and let $(P_t)_{t\geq 0}$ denote its reversible Markov semigroup, acting coordinatewise on vector fields. For $k \geq 0$, define the one-step local drift error

$$D_k := \int_0^h \big\{\nabla U_\lambda(X_{kh}) - \nabla U_\lambda(X_{kh+s})\big\}\,\mathrm{d}s \tag{30}$$

and the filtration $\mathcal{F}_t := \sigma(X_0, B_s : 0 \leq s \leq t)$.

By the Markov property,

$$\mathbb{E}[D_k \mid \mathcal{F}_{kh}] = \left[\int_0^h (I - P_s)\nabla U_\lambda \,\mathrm{d}s\right](X_{kh}).$$

We choose the corrector so that its conditional one-step decrement has this same mean. Specifically, let χ_h be the mean-zero vector field satisfying

$$\int_{\mathbb{R}^d} \chi_h \,\mathrm{d}\pi_\lambda = 0, \qquad (I - P_h)\chi_h = \int_0^h (I - P_s)\nabla U_\lambda \,\mathrm{d}s. \tag{31}$$

The equation is understood in $L^2(\pi_\lambda; \mathbb{R}^d)$. Proposition C.3 in Appendix C establishes existence, uniqueness, and the L^2 bound for χ_h. The additional moment and variance estimates are included in the following proposition.

Proposition 6.2 (Corrector decomposition and moment bounds)**.** *Under Assumptions 4.1–4.2, there are random vectors M_{k+1} such that*

$$D_k = \chi_h(X_{kh}) - \chi_h(X_{(k+1)h}) + M_{k+1}, \tag{32}$$

$$\mathbb{E}[M_{k+1} \mid \mathcal{F}_{kh}] = 0, \tag{33}$$

$$\|\chi_h\|_{L^2(\pi_\lambda)} \le hR, \tag{34}$$

$$\|\chi_h\|_{L^p(\pi_\lambda;\mathbb{R}^d)} \le ChKp^2, \qquad p \ge 2, \tag{35}$$

$$\mathbb{E}\,\|M_{k+1}\|^2 \le C\left\{L_f^2(dh^3 + R^2h^4) + G^2Rh^{5/2}\right\}. \tag{36}$$

The corrector increment telescopes over successive steps, while the centered increments contribute through their variances. The L^2 bound controls the change from the original error to the corrected error; the higher-moment bound is used to establish the curvature estimate below.

Proposition 6.3 (Two-point curvature feedback)**.** *Couple equation 29 and equation 12 from $X_0 = \widehat{X}_0 \sim \pi_\lambda$ using the same Brownian motion, and set*

$$A_k := A_\lambda(X_{kh}, \widehat{X}_k), \qquad A_k^g := A_\lambda^g(X_{kh}, \widehat{X}_k). \tag{37}$$

If equation 15 holds, then, for every $k \ge 0$,

$$\mathbb{E}\left[\chi_h(X_{kh})^\top A_k^g \chi_h(X_{kh})\right] \le CK^2G\sqrt{d}\,h^{3/2}\ell_{h,\lambda}^4. \tag{38}$$

Using $A_k^g \preceq \lambda^{-1}I$ and equation 34 would bound the left-hand side of equation 38 by h^2R^2/λ. Instead, Proposition 6.3 combines a shared-noise estimate $\mathbb{E}\,\mathrm{tr}\,A_k^g \le CG\sqrt{d/h}$ with higher moments of the corrector via Hölder's inequality, leaving only logarithmic dependence on λ^{-1}. The proofs of Propositions 6.2 and 6.3 are given in Appendices C and D, respectively.

6.3 A ONE-STEP MATRIX INEQUALITY

Lemma 6.4 (One-step energy inequality)**.** *Let A be symmetric and satisfy $mI \preceq A \preceq LI$. If $0 < hL \le 1$, then, for all $u, v \in \mathbb{R}^d$,*

$$\|(I - hA)u + hAv\|^2 \le (1 - mh)\,\|u\|^2 + h\,v^\top Av. \tag{39}$$

Proof. This inequality is used in the global-error recursion of Pedrotti & Whalley (2026, Section 2.1). For completeness, diagonalize A. For every eigenvalue $a \in [m, L]$, convexity of the square gives

$$|(1 - ha)u + hav|^2 \le (1 - ha)u^2 + hav^2, \qquad 0 \le ha \le 1.$$

Summing over the eigenbasis and using $a \ge m$ proves equation 39. □

6.4 PROOF OF THE FIXED-SMOOTHING THEOREM

Proof of Theorem 5.1. Start both processes from $X_0 = \widehat{X}_0 \sim \pi_\lambda$, so that the exact diffusion remains stationary. This initialization is only a proof device for comparing invariant laws. Use the same Gaussian increments, namely,

$$\sqrt{h}\,\xi_{k+1} = B_{(k+1)h} - B_{kh}.$$

Define

$$Z_k := X_{kh} - \widehat{X}_k. \tag{40}$$

The exact diffusion satisfies

$$X_{(k+1)h} = X_{kh} - h\nabla U_\lambda(X_{kh}) + \sqrt{2}(B_{(k+1)h} - B_{kh}) + D_k. \tag{41}$$

Subtracting equation 12 from equation 41 and applying Lemma 6.1 yields

$$\begin{aligned} Z_{k+1} &= Z_k - h\big\{\nabla U_\lambda(X_{kh}) - \nabla U_\lambda(\widehat{X}_k)\big\} + D_k \\ &= (I - hA_k)Z_k + D_k. \end{aligned} \tag{42}$$

The decomposition equation 32 suggests absorbing the telescoping increment into the error. Define the corrected error

$$Y_k := Z_k + \chi_h(X_{kh}). \tag{43}$$

Then

$$\begin{aligned} Y_{k+1} &= (I - hA_k)Z_k + \chi_h(X_{kh}) + M_{k+1} \\ &= (I - hA_k)Y_k + hA_k\chi_h(X_{kh}) + M_{k+1}. \end{aligned} \tag{44}$$

By equation 27 and equation 15, after reducing c if necessary, Lemma 6.4 applies. Moreover, Y_k, A_k, and $\chi_h(X_{kh})$ are $\mathcal{F}_{kh}$-measurable. Hence equation 33 removes all cross terms involving M_{k+1} after conditional expectation, and

$$\begin{aligned} \mathbb{E}\left\|Y_{k+1}\right\|^2 &= \mathbb{E}\left\|(I - hA_k)Y_k + hA_k\chi_h(X_{kh})\right\|^2 + \mathbb{E}\left\|M_{k+1}\right\|^2 \\ &\le (1 - mh)\mathbb{E}\left\|Y_k\right\|^2 + h\Phi_k + \mathbb{E}\left\|M_{k+1}\right\|^2, \end{aligned} \tag{45}$$

where

$$\Phi_k := \mathbb{E}\left[\chi_h(X_{kh})^\top A_k \chi_h(X_{kh})\right]. \tag{46}$$

Keeping the curvature term in this quadratic form allows us to apply Proposition 6.3 to its Moreau component. Since $A_k = A^f(X_{kh}, \widehat{X}_k) + A_k^g$, the bounds equation 28, stationarity $X_{kh} \sim \pi_\lambda$, equation 34, and equation 38 imply

$$\begin{aligned} \Phi_k &\le L_f \mathbb{E}\left\|\chi_h(X_{kh})\right\|^2 + \mathbb{E}\left[\chi_h(X_{kh})^\top A_k^g \chi_h(X_{kh})\right] \\ &\le L_f h^2 R^2 + CK^2 G\sqrt{d}\, h^{3/2}\ell_{h,\lambda}^4 =: \overline{\Phi}_h. \end{aligned} \tag{47}$$

Set

$$\overline{M}_h := C\left\{L_f^2(dh^3 + R^2h^4) + G^2Rh^{5/2}\right\}. \tag{48}$$

Then equation 36 gives $\mathbb{E}\left\|M_{k+1}\right\|^2 \le \overline{M}_h$. Iterating equation 45, for $n \ge 1$,

$$\begin{aligned} \mathbb{E}\left\|Y_n\right\|^2 &\le (1 - mh)^n \mathbb{E}\left\|Y_0\right\|^2 + \sum_{j=0}^{n-1}(1 - mh)^{n-1-j}(h\overline{\Phi}_h + \overline{M}_h) \\ &\le (1 - mh)^n \mathbb{E}\left\|Y_0\right\|^2 + \frac{h\overline{\Phi}_h + \overline{M}_h}{mh}. \end{aligned} \tag{49}$$

Letting $n \to \infty$ and substituting equation 47 and equation 48 gives

$$m \limsup_{n\to\infty} \mathbb{E}\left\|Y_n\right\|^2 \le C\big\{L_f R^2 h^2 + L_f^2 dh^2 + L_f^2 R^2 h^3 + K^2 G\sqrt{d}\, h^{3/2}\ell_{h,\lambda}^4 + G^2 R h^{3/2}\big\}. \tag{50}$$

Since $hL_f \le 1$, $L_f^2 R^2 h^3 \le L_f R^2 h^2$. Taking square roots term by term yields

$$\sqrt{m} \limsup_{n\to\infty} \left\|Y_n\right\|_{L^2} \le C\big(R\sqrt{L_f} + L_f\sqrt{d}\big)h + C\big(G\sqrt{R} + K\sqrt{G}\, d^{1/4}\ell_{h,\lambda}^2\big)h^{3/4}. \tag{51}$$

Because $Z_n = Y_n - \chi_h(X_{nh})$, stationarity and equation 34 further give

$$\begin{aligned} \sqrt{m} \limsup_{n\to\infty} \left\|Z_n\right\|_{L^2} &\le \sqrt{m} \limsup_{n\to\infty} \left\|Y_n\right\|_{L^2} + \sqrt{m}\left\|\chi_h\right\|_{L^2(\pi_\lambda)} \\ &\le C\Big[R(\sqrt{m} + \sqrt{L_f}) + L_f\sqrt{d}\Big]h + C\Big[G\sqrt{R} + K\sqrt{G}\, d^{1/4}\ell_{h,\lambda}^2\Big]h^{3/4}. \end{aligned} \tag{52}$$

It remains to identify the limiting Euler law. Define $T_h(x) := x - h\nabla U_\lambda(x)$. By equation 27 and $hL_\lambda \le 1$,

$$\left\|T_h(x) - T_h(y)\right\| = \left\|(I - hA_\lambda(x, y))(x - y)\right\| \le (1 - mh)\left\|x - y\right\|. \tag{53}$$

A synchronous Gaussian coupling therefore gives

$$W_2(\mu Q_{\lambda,h}, \nu Q_{\lambda,h}) \le (1 - mh)W_2(\mu, \nu), \qquad \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d). \tag{54}$$

The metric space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is complete, and $Q_{\lambda,h}$ maps $\mathcal{P}_2(\mathbb{R}^d)$ into itself. The Banach fixed-point theorem thus gives a unique invariant law $\widehat{\pi}_{\lambda,h}$ and $\mathcal{L}(\widehat{X}_n) \to \widehat{\pi}_{\lambda,h}$ in W_2. Since $X_{nh} \sim \pi_\lambda$, the coupling above and equation 52 imply

$$W_2(\pi_\lambda, \widehat{\pi}_{\lambda,h}) \le \limsup_{n\to\infty} \left\|Z_n\right\|_{L^2}.$$

This proves equation 16. □

6.5 PROOF OF THE END-TO-END COMPLEXITY THEOREM

Proof of Theorem 5.2. We allocate $\varepsilon/4$ to the Moreau approximation bias, $\varepsilon/4$ to the invariant-measure discretization bias, and $\varepsilon/2$ to convergence from the initial law.

The universal Moreau bias of Xin & Zhang (2026, Proposition 5.9) states that

$$\sqrt{m}\, W_2(\pi_\lambda, \pi) \le \frac{G^2 \lambda}{4}. \tag{55}$$

With $\lambda_\varepsilon = \varepsilon/G^2$,

$$\sqrt{m}\, W_2(\pi_{\lambda_\varepsilon}, \pi) \le \varepsilon/4. \tag{56}$$

The first constraint in equation 18 gives $h_\varepsilon L_{\lambda_\varepsilon} \le c$. Since $L_{\lambda_\varepsilon} \ge \lambda_\varepsilon^{-1}$,

$$h_\varepsilon \le c\lambda_\varepsilon, \qquad \frac{\sqrt{h_\varepsilon}}{G\sqrt{d}\,\lambda_\varepsilon} \le \frac{\sqrt{c}}{\sqrt{d\varepsilon}}.$$

Consequently,

$$\ell_{h_\varepsilon, \lambda_\varepsilon} \le C\ell_\varepsilon. \tag{57}$$

Substituting the last two constraints in equation 18 and equation 57 into Theorem 5.1, and reducing c if necessary, yields

$$\sqrt{m}\, W_2(\pi_{\lambda_\varepsilon}, \widehat{\pi}_{\lambda_\varepsilon, h_\varepsilon}) \le \varepsilon/4. \tag{58}$$

By equation 54,

$$\begin{aligned} &\sqrt{m}\, W_2(\mu_0 Q^N_{\lambda_\varepsilon, h_\varepsilon}, \widehat{\pi}_{\lambda_\varepsilon, h_\varepsilon}) \\ &\quad \le \sqrt{m}\, e^{-m h_\varepsilon N} W_2(\mu_0, \widehat{\pi}_{\lambda_\varepsilon, h_\varepsilon}) \\ &\quad \le \sqrt{m}\, e^{-m h_\varepsilon N} \left\{ W_2(\mu_0, \pi) + \frac{\varepsilon}{2\sqrt{m}} \right\} \le \varepsilon/2, \end{aligned} \tag{59}$$

where the second inequality uses the triangle inequality, equation 56, and equation 58, and the last follows from equation 19 with $C \ge 2$. The triangle inequality together with equation 56, equation 58, and equation 59 proves equation 20.

It remains to expand the iteration bound in terms of ε and the model parameters. By equation 18,

$$h_\varepsilon^{-1} \le C \left[L_f + \frac{G^2}{\varepsilon} + \frac{R(\sqrt{m} + \sqrt{L_f}) + L_f\sqrt{d}}{\varepsilon} + \frac{(G\sqrt{R} + K\sqrt{G}\, d^{1/4} \ell_\varepsilon^2)^{4/3}}{\varepsilon^{4/3}} \right].$$

Using $(a+b)^{4/3} \le 2^{1/3}(a^{4/3} + b^{4/3})$ proves equation 21. For fixed model parameters and $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $W_2(\mu_0, \pi)$ is finite and independent of ε. Hence equation 21 gives equation 22. □

7 CONCLUSION

We established Wasserstein error bounds for classical MYULA with a smooth strongly convex component and a convex Lipschitz nonsmooth term. A discrete Poisson corrector, combined with active-trace and shared-noise curvature estimates, yields an $O(h) + \widetilde{O}(h^{3/4})$ invariant-measure bias bound. Under the stated step-size restriction, the error coefficients depend only logarithmically on the inverse smoothing parameter. Combining this estimate with the Moreau approximation bias and Wasserstein contraction gives a sufficient $\widetilde{O}(\varepsilon^{-4/3})$ iteration complexity for $\sqrt{m}\, W_2(\mu_N, \pi) \le \varepsilon$, with model parameters and initialization fixed. The analysis assumes exact proximal evaluations and excludes their internal computational cost; it does not establish optimality of the accuracy exponent $4/3$.

AI USE STATEMENT

We used generative AI for language editing and LaTeX formatting. The authors reviewed all AI-assisted edits and take responsibility for the final content.

REPRODUCIBILITY STATEMENT

All assumptions are stated in Section 4. Proofs of the main results and supporting estimates are provided in Section 6 and Appendices A–D.

REFERENCES

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2014. doi: 10.1007/978-3-319-00227-9.

Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2 edition, 2017. doi: 10.1007/978-3-319-48311-5.

Andrea Carbonaro and Oliver Dragičević. Functional calculus for generators of symmetric contraction semigroups. *Duke Mathematical Journal*, 166(5):937–974, 2017. doi: 10.1215/00127094-3774526.

Sinho Chewi. Log-concave sampling, 2026. URL https://chewisinho.github.io/main.pdf. Book draft, version of August 20, 2026.

Alain Durmus, Éric Moulines, and Marcelo Pereyra. Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau. *SIAM Journal on Imaging Sciences*, 11(1):473–506, 2018. doi: 10.1137/16M1108340.

Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *Journal of Machine Learning Research*, 20(73):1–46, 2019. URL https://jmlr.org/papers/v20/18-173.html.

Lawrence C. Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, revised edition, 2015.

Giovanni Leoni. *A First Course in Sobolev Spaces*, volume 181 of *Graduate Studies in Mathematics*. American Mathematical Society, 2 edition, 2017. doi: 10.1090/gsm/181.

Wenlong Mou, Nicolas Flammarion, Martin J. Wainwright, and Peter L. Bartlett. An efficient sampling algorithm for non-smooth composite potentials. *Journal of Machine Learning Research*, 23 (233):1–50, 2022. URL https://jmlr.org/papers/v23/20-527.html.

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

Marcelo Pereyra. Proximal Markov chain Monte Carlo algorithms. *Statistics and Computing*, 26: 745–760, 2016. doi: 10.1007/s11222-015-9567-4.

Adil Salim and Peter Richtárik. Primal dual interpretation of the proximal stochastic gradient Langevin algorithm. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2779fda014fbadb761f67dd708c1325e-Abstract.html.

Yuchen Xin and Zhihua Zhang. Active-trace complexity bounds for Moreau–Yosida unadjusted Langevin sampling. arXiv preprint arXiv:2608.13467, 2026. URL https://arxiv.org/abs/2608.13467.

A Weak Hessians and Integration by Parts

We collect the weak-Hessian and weighted integration-by-parts facts used in the stationary estimates and the subsequent semigroup argument; see Xin & Zhang (2026, Appendices A–B), Evans & Gariepy (2015), and Leoni (2017).

A.1 Weak derivatives and weak Hessians

Definition A.1 (Weak derivative)**.** Let $u, v \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$. We say that v is the weak partial derivative $\partial_i u$ if

$$\int_{\mathbb{R}^d} u(x)\, \partial_i \phi(x)\, \mathrm{d}x = -\int_{\mathbb{R}^d} v(x)\phi(x)\, \mathrm{d}x \qquad \text{for every } \phi \in C_c^\infty(\mathbb{R}^d). \tag{60}$$

The function v, if it exists, is unique up to equality on a Lebesgue-null set.

A Lipschitz function is differentiable almost everywhere by Rademacher's theorem. Moreover, its almost-everywhere derivative is its weak derivative; see Evans & Gariepy (2015, Sections 3.1–3.2). Since $g_\lambda \in C^{1,1}$, each component of ∇g_λ is Lipschitz and therefore has a bounded weak gradient.

Definition A.2 (Weak Hessian)**.** Let $u \in C^1(\mathbb{R}^d)$ and assume that ∇u is locally Lipschitz. The weak Hessian $H_u = \nabla^2 u$ is the matrix field whose (i, j) entry is the weak derivative $\partial_j(\partial_i u)$. It agrees almost everywhere with the classical derivative of ∇u supplied by Rademacher's theorem.

Different measurable representatives of H_u can differ on a null set, but all integrals against an absolutely continuous probability law are the same. This is why the choice of representative never affects the estimates in the paper.

Proposition A.3 (Basic properties of a weak Hessian)**.** *Let $u \in C^{1,1}(\mathbb{R}^d)$ be convex and suppose $\mathrm{Lip}(\nabla u) \le L$. Then its weak Hessian H_u can be chosen measurable and satisfies*

$$H_u(x) = H_u(x)^\top, \qquad 0 \preceq H_u(x) \preceq LI \quad \textit{for a.e. } x. \tag{61}$$

If $u = g_\lambda$, then $L = \lambda^{-1}$ and $a_\lambda = \operatorname{tr} H_\lambda$ satisfies equation 7.

Proof. Rademacher's theorem gives the almost-everywhere derivative of the Lipschitz map ∇u. Equality of mixed weak derivatives implies symmetry almost everywhere. A convenient way to see the matrix inequalities is to mollify. Let $u_\delta = u * \rho_\delta$, where ρ_δ is a standard nonnegative smooth mollifier. Convexity is preserved by convolution, so $\nabla^2 u_\delta \succeq 0$. Also $\mathrm{Lip}(\nabla u_\delta) \le L$, hence $\nabla^2 u_\delta \preceq LI$. The gradients ∇u_δ converge locally uniformly to ∇u, and their classical derivatives converge to H_u in the sense of distributions and, after passing to a subsequence, weakly-* in L^∞_{loc}. Testing the inequalities against nonnegative scalar test functions and fixed vectors shows that the same inequalities hold for H_u almost everywhere. □

A.2 Weighted integration by parts under the Gibbs law

Let $x_\star$ be the minimizer of U_λ. Strong convexity gives

$$U_\lambda(x) \ge U_\lambda(x_\star) + \frac{m}{2}\, \|x - x_\star\|^2 . \tag{62}$$

Hence π_λ has Gaussian tails. In particular, every polynomial in $\|x\|$ is integrable. Since ∇U_λ is globally Lipschitz and vanishes at $x_\star$, it grows at most linearly. These two observations justify the cutoff limits below.

Choose $\eta \in C_c^\infty(\mathbb{R}^d)$ such that $0 \le \eta \le 1$, $\eta = 1$ on the unit ball, and $\eta = 0$ outside the ball of radius 2. Put $\eta_R(x) = \eta(x/R)$. Then

$$\eta_R \to 1, \qquad \|\nabla \eta_R\|_\infty \le C/R, \qquad \operatorname{supp}(\nabla \eta_R) \subset \{R \le \|x\| \le 2R\}. \tag{63}$$

Proposition A.4 (Weighted weak integration by parts)**.** *Let $V \in C^{1,1}(\mathbb{R}^d)$, and assume that ∇V grows at most linearly. Let H_V denote its weak Hessian. If the terms below are integrable, then for every $1 \le i, j \le d$,*

$$\int (H_V)_{ij}\, \mathrm{d}\pi_\lambda = \int \partial_i V\, \partial_j U_\lambda\, \mathrm{d}\pi_\lambda. \tag{64}$$

Moreover,

$$\int \partial_i U_\lambda \, \mathrm{d}\pi_\lambda = 0. \tag{65}$$

In particular,

$$\int \|\nabla U_\lambda\|^2 \, \mathrm{d}\pi_\lambda = \int \mathrm{tr}(\nabla^2 f + H_\lambda) \, \mathrm{d}\pi_\lambda, \tag{66}$$

$$\int a_\lambda \, \mathrm{d}\pi_\lambda = \int \langle \nabla g_\lambda, \nabla U_\lambda \rangle \, \mathrm{d}\pi_\lambda. \tag{67}$$

Proof. We first work with the unnormalized density e^{-U_λ}. Apply the weak derivative identity equation 60 to the function $\partial_i V$ and the compactly supported test function $\eta_R e^{-U_\lambda}$. This is legitimate because $U_\lambda \in C^1$ and $\eta_R e^{-U_\lambda}$ can be approximated in the relevant Sobolev norm by smooth compactly supported test functions. We obtain

$$\begin{aligned} \int (H_V)_{ij} \eta_R e^{-U_\lambda} \, \mathrm{d}x &= -\int \partial_i V \, \partial_j (\eta_R e^{-U_\lambda}) \, \mathrm{d}x \\ &= \int \eta_R \partial_i V \, \partial_j U_\lambda e^{-U_\lambda} \, \mathrm{d}x - \int \partial_i V \, \partial_j \eta_R e^{-U_\lambda} \, \mathrm{d}x. \end{aligned} \tag{68}$$

The last integral is supported on the annulus in equation 63. Its absolute value is bounded by

$$\frac{C}{R} \int_{R \le \|x\| \le 2R} |\partial_i V(x)| \, e^{-U_\lambda(x)} \, \mathrm{d}x,$$

which tends to zero by the at-most-linear growth of ∇V and the Gaussian tail equation 62. Dominated convergence in the first two terms of equation 68, followed by normalization, proves equation 64.

For equation 65, use $\partial_i(e^{-U_\lambda}) = -\partial_i U_\lambda e^{-U_\lambda}$:

$$\int \eta_R \partial_i U_\lambda e^{-U_\lambda} \, \mathrm{d}x = -\int \eta_R \partial_i (e^{-U_\lambda}) \, \mathrm{d}x = \int \partial_i \eta_R e^{-U_\lambda} \, \mathrm{d}x.$$

The right-hand side tends to zero by equation 63 and the Gaussian tail. Finally, take $V = U_\lambda$ in equation 64, set $i = j$, and sum over i to obtain equation 66. Taking $V = g_\lambda$ gives equation 67. □

B Operator Facts Used by the Corrector

We collect the semigroup and spectral-calculus facts used by the corrector; see Bakry et al. (2014, Appendices A.1–A.2 and A.4). The L^p multiplier theorem is stated in Proposition B.3.

B.1 From the Markov semigroup to a nonnegative generator

Fix $\lambda > 0$ under Assumptions 4.1–4.2. Then $U_\lambda \in C^{1,1}$ is m-strongly convex and ∇U_λ is globally Lipschitz. For bounded measurable u, define

$$P_t u(x) := \mathbb{E}[u(X_t^x)], \tag{69}$$

where X^x denotes the Langevin diffusion equation 29 started from x. The semigroup extends to each $L^p(\pi_\lambda)$. Invariance of π_λ and Jensen's inequality give

$$\|P_t u\|_{L^p(\pi_\lambda)} \le \|u\|_{L^p(\pi_\lambda)}, \qquad 1 \le p \le \infty, \tag{70}$$

and reversibility gives $\langle P_t u, v \rangle_{L^2} = \langle u, P_t v \rangle_{L^2}$; see Chewi (2026, Section 1.2) for these Langevin semigroup facts. Set

$$\mathsf{H} := L^2(\pi_\lambda), \qquad \langle u, v \rangle_{\mathsf{H}} := \int uv \, \mathrm{d}\pi_\lambda.$$

We use real-valued functions, identified up to π_λ-null sets. The operator $\mathcal{A}_\lambda$ below acts on functions and is distinct from the two-point matrix $A_\lambda(x, y)$ in equation 25.

For a contraction semigroup, *strong continuity* means $\|P_t u - u\|_{\mathsf{H}} \to 0$ as $t \downarrow 0$ for every $u \in \mathsf{H}$. This is the continuity condition in Bakry et al. (2014, Definition A.1.1(iii)); the semigroup property then gives continuity of $t \mapsto P_t u$ at every $t \geq 0$. A bounded operator T is *self-adjoint* if $\langle Tu, v\rangle_{\mathsf{H}} = \langle u, Tv\rangle_{\mathsf{H}}$ for all $u, v \in \mathsf{H}$. Thus reversibility is exactly self-adjointness of P_t on $L^2(\pi_\lambda)$.

The generator is defined on its full L^2 domain by

$$D(\mathcal{A}_\lambda) := \left\{ u \in \mathsf{H} : \lim_{t \downarrow 0} \frac{u - P_t u}{t} \text{ exists in } \mathsf{H} \right\}, \qquad \mathcal{A}_\lambda u := \lim_{t \downarrow 0} \frac{u - P_t u}{t}. \tag{71}$$

The minus sign relative to the usual SDE generator is intentional: $\mathcal{A}_\lambda = -\mathcal{L}_\lambda$. An operator "on H" need not have domain equal to H. For a densely defined operator A, the adjoint is given by $A^* v = w$ when $\langle Au, v\rangle_{\mathsf{H}} = \langle u, w\rangle_{\mathsf{H}}$ for every $u \in D(A)$; $D(A^*)$ consists of the v for which such a $w \in \mathsf{H}$ exists. Self-adjointness means $D(A^*) = D(A)$ and $A^* = A$, not just symmetry on $D(A)$. Nonnegativity means $\langle Au, u\rangle_{\mathsf{H}} \geq 0$ on $D(A)$. See Bakry et al. (2014, Appendix A.2).

Proposition B.1 (Generator associated with the reversible diffusion)**.** *The operators $(P_t)_{t \geq 0}$ form a strongly continuous self-adjoint contraction semigroup on H. The operator $\mathcal{A}_\lambda$ in equation 71 has dense domain, is nonnegative and self-adjoint, and satisfies*

$$P_t = e^{-t\mathcal{A}_\lambda}, \qquad t \geq 0, \tag{72}$$

where the exponential is defined by the spectral calculus described below. Moreover, $1 \in D(\mathcal{A}_\lambda)$, $\mathcal{A}_\lambda 1 = 0$, and $C_c^\infty(\mathbb{R}^d) \subset D(\mathcal{A}_\lambda)$, with

$$\mathcal{A}_\lambda u = -\Delta u + \langle \nabla U_\lambda, \nabla u \rangle, \qquad u \in C_c^\infty(\mathbb{R}^d). \tag{73}$$

Proof. The Markov property gives $P_{s+t} = P_s P_t$ and $P_0 = I$. Invariance and Jensen's inequality give equation 70, and reversibility gives self-adjointness. To check strong continuity, use $X_0 \sim \pi_\lambda$. For bounded continuous u, conditional Jensen gives

$$\|P_t u - u\|_{\mathsf{H}}^2 \leq \mathbb{E}\,|u(X_t) - u(X_0)|^2 \longrightarrow 0$$

by path continuity and dominated convergence. Density of bounded continuous functions in H and contraction extend the limit to every $u \in \mathsf{H}$.

The symmetric-semigroup theorem in Bakry et al. (2014, Appendices A.1–A.2) now gives dense domain and self-adjointness of the full generator. Nonnegativity follows from $\langle u - P_t u, u\rangle_{\mathsf{H}} \geq 0$ and equation 71. The spectral semigroup in Bakry et al. (2014, Appendix A.4, immediately after Theorem A.4.2) has generator $-\mathcal{A}_\lambda$; uniqueness of the semigroup determined by its full generator gives equation 72.

Since $P_t 1 = 1$, the generator annihilates constants. For $u \in C_c^\infty(\mathbb{R}^d)$, put $\mathcal{L}_\lambda u = \Delta u - \langle \nabla U_\lambda, \nabla u\rangle \in \mathsf{H}$. Itô's formula gives

$$\frac{P_t u - u}{t} = \frac{1}{t} \int_0^t P_s \mathcal{L}_\lambda u \, \mathrm{d}s \longrightarrow \mathcal{L}_\lambda u \quad \text{in } \mathsf{H},$$

Strong continuity gives the last limit and hence both assertions on C_c^∞. Self-adjointness concerns the full domain equation 71. □

B.2 Bounded functions of the generator

We use the spectral calculus for $\mathcal{A}_\lambda$ described in Bakry et al. (2014, Appendix A.4). For bounded continuous q, the operator $q(\mathcal{A}_\lambda)$ is defined on all of H.

Proposition B.2 (Functional-calculus rules used in the paper)**.** *For every bounded continuous function $q : [0, \infty) \to \mathbb{R}$, the spectral calculus defines a bounded self-adjoint operator $q(\mathcal{A}_\lambda)$ on all of H. For bounded continuous q, q_1, q_2 and $\alpha, \beta \in \mathbb{R}$,*

$$\|q(\mathcal{A}_\lambda)\|_{L^2 \to L^2} \leq \sup_{r \geq 0} |q(r)|, \tag{74}$$

$$(\alpha q_1 + \beta q_2)(\mathcal{A}_\lambda) = \alpha q_1(\mathcal{A}_\lambda) + \beta q_2(\mathcal{A}_\lambda), \tag{75}$$

$$q_1(\mathcal{A}_\lambda) q_2(\mathcal{A}_\lambda) = (q_1 q_2)(\mathcal{A}_\lambda). \tag{76}$$

If $q_1 \le q_2$ pointwise, then

$$\langle u, q_1(\mathcal{A}_\lambda)u\rangle_{L^2} \le \langle u, q_2(\mathcal{A}_\lambda)u\rangle_{L^2}, \qquad u \in \mathsf{H}. \tag{77}$$

Furthermore, for $t \ge 0$,

$$[r \mapsto e^{-tr}](\mathcal{A}_\lambda) = P_t, \tag{78}$$

$$q(\mathcal{A}_\lambda)1 = q(0)1. \tag{79}$$

For $t > 0$,

$$P_t\mathsf{H} \subset D(\mathcal{A}_\lambda), \qquad \mathcal{A}_\lambda P_t = [r \mapsto re^{-tr}](\mathcal{A}_\lambda), \qquad \|\mathcal{A}_\lambda P_t\|_{L^2\to L^2} \le \frac{1}{et}. \tag{80}$$

For $h > 0$, define

$$b_h(r) := \int_0^h (1 - e^{-sr})\,\mathrm{d}s. \tag{81}$$

Then, for every $u \in \mathsf{H}$,

$$b_h(\mathcal{A}_\lambda)u = \int_0^h (I - P_s)u\,\mathrm{d}s, \tag{82}$$

where the right-hand side is the limit in $L^2(\pi_\lambda)$ of Riemann sums. The mean-zero space

$$\mathsf{H}_0 := \left\{u \in \mathsf{H} : \int u\,\mathrm{d}\pi_\lambda = 0\right\}$$

is invariant under $q(\mathcal{A}_\lambda)$. Finally, the usual Poincaré consequence of m-strong log-concavity gives

$$\|P_t u\|_{L^2(\pi_\lambda)} \le e^{-mt}\,\|u\|_{L^2(\pi_\lambda)}, \qquad u \in \mathsf{H}_0, \quad t \ge 0. \tag{83}$$

Proof. Write $A = \mathcal{A}_\lambda$. The cited spectral theorem associates to each $u \in \mathsf{H}$ a finite positive scalar measure ν_u on $[0,\infty)$, of total mass $\|u\|_{\mathsf{H}}^2$. Its domain and norm formulas, namely (A.4.1) and Theorem A.4.2 of Bakry et al. (2014), read

$$D(F(A)) = \left\{u \in \mathsf{H} : \int_{[0,\infty)} |F(r)|^2\,\nu_u(\mathrm{d}r) < \infty\right\}, \qquad \|F(A)u\|_{\mathsf{H}}^2 = \int_{[0,\infty)} |F(r)|^2\,\nu_u(\mathrm{d}r).$$

These formulas hold for real Borel functions F; the norm identity is understood on $D(F(A))$. The endpoint 0 is included, with any mass it carries. For bounded q, the integral is at most $\|q\|_\infty^2\,\|u\|_{\mathsf{H}}^2$, proving $D(q(A)) = \mathsf{H}$ and equation 74. The real-valued spectral construction is self-adjoint and obeys the algebra rules equation 75–equation 76. For the order rule, let $s = \sqrt{q_2 - q_1}$. Then

$$\langle u, (q_2 - q_1)(A)u\rangle_{\mathsf{H}} = \langle u, s(A)^2 u\rangle_{\mathsf{H}} = \|s(A)u\|_{\mathsf{H}}^2 \ge 0.$$

Equation equation 78 is simply equation 72 in this notation.

Since $A1 = 0$, the norm formula gives $\int r^2\nu_1(\mathrm{d}r) = 0$. Hence ν_1 is concentrated at 0, and

$$\|q(A)1 - q(0)1\|_{\mathsf{H}}^2 = \int_{[0,\infty)} |q(r) - q(0)|^2\,\nu_1(\mathrm{d}r) = 0.$$

This proves equation 79. Self-adjointness then yields

$$\int q(A)u\,\mathrm{d}\pi_\lambda = \langle u, q(A)1\rangle_{\mathsf{H}} = q(0)\int u\,\mathrm{d}\pi_\lambda,$$

so $q(A)$ preserves H_0.

For equation 80, the spectral construction gives $\nu_{P_t u}(\mathrm{d}r) = e^{-2tr}\nu_u(\mathrm{d}r)$. Thus, for every $u \in \mathsf{H}$,

$$\int_{[0,\infty)} r^2\nu_{P_t u}(\mathrm{d}r) = \int_{[0,\infty)} r^2 e^{-2tr}\nu_u(\mathrm{d}r) \le \frac{\|u\|_{\mathsf{H}}^2}{e^2t^2} < \infty.$$

The domain formula with $F(r) = r$ first proves $P_t u \in D(A)$; multiplication in the same spectral representation then gives $AP_t u = [r \mapsto re^{-tr}](A)u$ and its norm bound. In particular, $\partial_t P_t u = -AP_t u$ in H for $t > 0$, by the semigroup-generator relation in Bakry et al. (2014, Appendix A.1).

To prove equation 82, let

$$b_{h,n}(r) := \frac{h}{n}\sum_{j=0}^{n-1}(1-e^{-jhr/n}).$$

Monotonicity of $s \mapsto 1-e^{-sr}$ gives the uniform Riemann-sum estimate

$$0 \le b_h(r) - b_{h,n}(r) \le \frac{h}{n}(1-e^{-hr}) \le \frac{h}{n}. \tag{84}$$

The algebra rules and equation 78 give

$$b_{h,n}(A)u = \frac{h}{n}\sum_{j=0}^{n-1}(I-P_{jh/n})u.$$

By equation 74, the left-hand side converges to $b_h(A)u$. Strong continuity makes the right-hand side converge in H to the time integral in equation 82. This is an integral of the continuous H-valued function $s \mapsto (I-P_s)u$, not an assertion of operator-norm continuity at $s=0$. Bounded linear maps commute with these Riemann limits. In particular, stationary evaluation $v \mapsto v(X_{kh})$ is an isometry from $L^2(\pi_\lambda)$ to $L^2(\Omega)$, and conditional expectation is an L^2 contraction. Thus this interpretation agrees with the use of the integral in equation 108.

For completeness, equation 83 is a Poincaré estimate, not an additional rule of functional calculus. Strong log-concavity gives

$$\mathrm{Var}_{\pi_\lambda}(u) \le \frac{1}{m}\int \|\nabla u\|^2 \,\mathrm{d}\pi_\lambda.$$

This holds first for smooth compactly supported u, and then for $u \in L^2(\pi_\lambda)$ with a square-integrable weak gradient by cutoff and smooth approximation. An implication valid directly for the present C^1 potential is given by Chewi (2026, Theorem 2.2.12 and Corollary 2.1.2): strong convexity bounds the Bregman cost in the former result below by $m\|x-y\|^2/2$, yielding the transport inequality with constant $1/m$; the latter result gives the displayed Poincaré inequality. Its equivalence with exponential variance decay is Chewi (2026, Theorem 1.2.21), which gives $\|P_t u\|_{\mathsf{H}}^2 \le e^{-2mt}\|u\|_{\mathsf{H}}^2$ on H_0. This proves equation 83 without requiring a classical Hessian of U_λ or a smoothing limit of π_λ. □

B.3 An L^p multiplier theorem

The bounded functional calculus above is an L^2 result. To control all moments of the corrector, we need one L^p estimate. We first state the external result in the exact notation needed to check its assumptions.

For $1 < p < \infty$, let $\mathcal{A}_{\lambda,p}$ denote the generator of the same semigroup (P_t) acting on $L^p(\pi_\lambda)$; explicitly,

$$D(\mathcal{A}_{\lambda,p}) := \left\{ u \in L^p(\pi_\lambda) : \lim_{t\downarrow 0}\frac{u-P_t u}{t} \text{ exists in } L^p(\pi_\lambda)\right\},$$

and the limit is $\mathcal{A}_{\lambda,p}u$. Its range and null space are

$$R(\mathcal{A}_{\lambda,p}) := \{\mathcal{A}_{\lambda,p}u : u \in D(\mathcal{A}_{\lambda,p})\}, \qquad N(\mathcal{A}_{\lambda,p}) := \{u \in D(\mathcal{A}_{\lambda,p}) : \mathcal{A}_{\lambda,p}u = 0\}.$$

The bar in $\overline{R(\mathcal{A}_{\lambda,p})}^{L^p}$ means closure in the L^p norm.

Proposition B.3 (Laplace-transform multiplier)**.** *Let (Ω,ν) be a σ-finite measure space, let A be a nonnegative self-adjoint operator on $L^2(\nu)$, and suppose that $T_t = e^{-tA}$ satisfies*

$$\|T_t f\|_{L^q(\nu)} \le \|f\|_{L^q(\nu)}, \qquad 1 \le q \le \infty.$$

For $M \in L^\infty(0,\infty)$ define, for $r>0$,

$$\widetilde{M}(r) := r\int_0^\infty M(s)e^{-sr}\,\mathrm{d}s. \tag{85}$$

Then, for $1 < p < \infty$ *and* $f \in \overline{R(A_p)}^{L^p}$,

$$\left\| \widetilde{M}(A) f \right\|_{L^p(\nu)} \le 120(p^* - 1) \|M\|_\infty \|f\|_{L^p(\nu)}, \qquad p^* := \max\left\{ p, \frac{p}{p-1} \right\}. \tag{86}$$

Moreover, let $\Pi_0^{(2)}$ *denote the orthogonal projection of* $L^2(\nu)$ *onto* $N(A_2)$. *Then*

$$\Pi_0^{(2)} f = \lim_{t \to \infty} T_t f \quad \text{in } L^2(\nu). \tag{87}$$

For every $1 < p < \infty$, *the restriction of* $\Pi_0^{(2)}$ *to* $L^2(\nu) \cap L^p(\nu)$ *extends to a contractive projection* $\Pi_0^{(p)}$ *on* $L^p(\nu)$. *Thus, by the meaning of extension,*

$$\Pi_0^{(p)} f = \Pi_0^{(2)} f, \qquad f \in L^2(\nu) \cap L^p(\nu).$$

Moreover,

$$(I - \Pi_0^{(p)}) f \in \overline{R(A_p)}^{L^p}, \qquad f \in L^2(\nu) \cap L^p(\nu).$$

Source and notation. These statements are quoted, not reproved. The hypotheses that A is nonnegative and self-adjoint and that e^{-tA} is contractive on every L^q are exactly the standing assumptions in Carbonaro & Dragičević (2017, Introduction); that paper calls this a *symmetric contraction semigroup*. Its notation $A_p, D(A_p), R(A_p), N(A_p)$ is introduced in Carbonaro & Dragičević (2017, Section 2).

The limit formula equation 87 follows from Lemma 2 of Carbonaro & Dragičević (2017); part (i) of that lemma also gives the contractive L^p extension. The cited paper uses the same symbol P_0 for the L^2 projection and all of its L^p extensions; the superscripts in our notation only indicate the ambient space. The fact that

$$(I - \Pi_0^{(p)}) f \in \overline{R(A_p)}^{L^p}$$

is the form used explicitly in Carbonaro & Dragičević (2017, proof of Theorem 1).

The definition equation 85 is the display immediately preceding Proposition 5 of Carbonaro & Dragičević (2017); their Proposition 10 gives equation 86. □

We now verify that the proposition applies here. The measure π_λ is a probability measure and hence is σ-finite. Nonnegativity and self-adjointness of $\mathcal{A}_\lambda$ follow from Proposition B.1; contractivity on every L^q is equation 70. If $v \in L^2(\pi_\lambda) \cap L^p(\pi_\lambda)$ has zero mean, then equation 83 gives $P_t v \to 0$ in L^2. Taking $T_t = P_t$ in equation 87 therefore gives $\Pi_0^{(2)} v = 0$. Since $\Pi_0^{(p)}$ extends $\Pi_0^{(2)}$ and $v \in L^2(\pi_\lambda) \cap L^p(\pi_\lambda)$,

$$\Pi_0^{(p)} v = \Pi_0^{(2)} v = 0.$$

Consequently,

$$v = (I - \Pi_0^{(p)}) v \in \overline{R(\mathcal{A}_{\lambda,p})}^{L^p}. \tag{88}$$

Thus every hypothesis of Proposition B.3 has now been checked in the notation of the present paper.

C Stationary Estimates and the Discrete Poisson Corrector

This appendix proves Proposition 6.2. Section C.1 provides the stationary estimates used below. Section C.2 constructs the corrector and establishes the moment bounds equation 34–equation 35. The decomposition and centering identities equation 32–equation 33 are proved in Section C.3, and the variance bound equation 36 is proved in Section C.4.

C.1 Stationary active trace and drift moments

Lemma C.1 (Stationary drift and active-trace bounds). *If* $X \sim \pi_\lambda$, *then*

$$\|\nabla U_\lambda\|_{L^2(\pi_\lambda)} \le R, \qquad \mathbb{E}_{\pi_\lambda} a_\lambda \le GR, \qquad \mathbb{E}_{\pi_\lambda} \nabla U_\lambda = 0. \tag{89}$$

Proof. The weak integration-by-parts identities proved in Proposition A.4 give

$$\mathbb{E}_{\pi_\lambda} \|\nabla U_\lambda\|^2 = \mathbb{E}_{\pi_\lambda} \operatorname{tr}(\nabla^2 f + H_\lambda) \le \tau_f + \mathbb{E}_{\pi_\lambda} a_\lambda, \tag{90}$$

$$\mathbb{E}_{\pi_\lambda} a_\lambda = \mathbb{E}_{\pi_\lambda} \langle \nabla g_\lambda, \nabla U_\lambda \rangle \le G\big(\mathbb{E}_{\pi_\lambda} \|\nabla U_\lambda\|^2\big)^{1/2}. \tag{91}$$

Let $r := \|\nabla U_\lambda\|_{L^2(\pi_\lambda)}$. Then $r^2 \le \tau_f + Gr$, and therefore

$$r \le \frac{G + \sqrt{G^2 + 4\tau_f}}{2} = R.$$

Substitution in equation 91 gives $\mathbb{E}_{\pi_\lambda} a_\lambda \le GR$. The last identity in equation 89 is also contained in Proposition A.4. □

Lemma C.2 (All stationary drift moments)**.** *For every $p \ge 2$,*

$$\|\nabla U_\lambda\|_{L^p(\pi_\lambda;\mathbb{R}^d)} \le K\sqrt{p}. \tag{92}$$

Proof. Let $x_\star$ be the unique minimizer of U_λ. Since $\nabla U_\lambda(x_\star) = 0$, the L_f-Lipschitz property of ∇f and equation 5 imply

$$\begin{aligned} \|\nabla U_\lambda(x)\| &\le \|\nabla f(x) - \nabla f(x_\star)\| + \|\nabla g_\lambda(x) - \nabla g_\lambda(x_\star)\| \\ &\le L_f \|x - x_\star\| + 2G. \end{aligned} \tag{93}$$

Applying Proposition A.4 with $V(x) = \frac{1}{2}\|x - x_\star\|^2$ and summing over the coordinates gives

$$d = \mathbb{E}_{\pi_\lambda} \langle X - x_\star, \nabla U_\lambda(X) \rangle \ge m\mathbb{E}_{\pi_\lambda} \|X - x_\star\|^2,$$

where the inequality uses $\nabla U_\lambda(x_\star) = 0$ and m-strong convexity. Hence

$$\mathbb{E}_{\pi_\lambda} \|X - x_\star\|^2 \le \frac{d}{m}.$$

Moreover, the Bregman transport inequality and its concentration consequence (Chewi, 2026, Theorems 2.2.12 and 2.4.3(ii)) give

$$\mathbb{P}\left(\|X - x_\star\| - \mathbb{E}\|X - x_\star\| \ge t\right) \le e^{-mt^2/2}, \qquad t \ge 0.$$

Tail integration therefore yields

$$\big\|\|X - x_\star\|\big\|_{L^p} \le \frac{\sqrt{d} + 2\sqrt{p}}{\sqrt{m}}, \qquad p \ge 2. \tag{94}$$

Combining equation 93–equation 94 and using $1 \le \sqrt{p}$ gives

$$\|\nabla U_\lambda\|_{L^p} \le 2G + \frac{L_f}{\sqrt{m}}(\sqrt{d} + 2\sqrt{p}) \le K\sqrt{p}.$$

□

C.2 Construction and moments of the corrector

We use the operator framework in Section B to construct and estimate the corrector. The same corrector construction, under stronger smoothness assumptions, appears in Pedrotti & Whalley (2026, Sections 2–2.1).

Proposition C.3 (The discrete Poisson corrector)**.** *For $r \ge 0$, define*

$$\varphi_h(r) := \begin{cases} \dfrac{h - (1 - e^{-hr})/r}{1 - e^{-hr}}, & r > 0, \\ h/2, & r = 0. \end{cases} \tag{95}$$

Using the construction in Proposition B.2 coordinatewise, set

$$\chi_h := \varphi_h(\mathcal{A}_\lambda)\nabla U_\lambda. \tag{96}$$

Then χ_h is the unique mean-zero solution in $L^2(\pi_\lambda;\mathbb{R}^d)$ of equation 31, and

$$0 \le \varphi_h(r) \le h, \qquad r \ge 0. \tag{97}$$

Consequently,

$$\|\chi_h\|_{L^2(\pi_\lambda)} \le h\,\|\nabla U_\lambda\|_{L^2(\pi_\lambda)} \le hR. \tag{98}$$

Proof. Let

$$a_h(r) := 1 - e^{-hr}, \qquad b_h(r) := \int_0^h (1 - e^{-sr})\,\mathrm{d}s.$$

For every $r \geq 0$,

$$a_h(r)\varphi_h(r) = b_h(r). \tag{99}$$

For $r > 0$,

$$(1 - e^{-hr})\varphi_h(r) = h - \frac{1 - e^{-hr}}{r} = \int_0^h (1 - e^{-sr})\,\mathrm{d}s;$$

At $r = 0$, both sides of equation 99 are zero, and Taylor expansion shows that φ_h is continuous there. The multiplication and time-integration rules in Proposition B.2 turn equation 99 into

$$(I - P_h)\chi_h = \int_0^h (I - P_s)\nabla U_\lambda\,\mathrm{d}s. \tag{100}$$

The same proposition preserves zero means, so Lemma C.1 shows that χ_h has zero mean.

To prove equation 97, write $u = hr$. The inequalities $1 - e^{-u} \leq u$ and $e^u \geq 1 + u$ imply

$$0 \leq \frac{1}{1 - e^{-u}} - \frac{1}{u} \leq 1.$$

Thus $0 \leq \varphi_h(r) \leq h$. The norm bound equation 74 and Lemma C.1 yield

$$\|\chi_h\|_{L^2} \leq h \|\nabla U_\lambda\|_{L^2} \leq hR.$$

If $\widetilde{\chi}_h$ is another mean-zero solution, then $v := \chi_h - \widetilde{\chi}_h$ satisfies $P_h v = v$. By equation 83,

$$\|v\|_{L^2} = \|P_h v\|_{L^2} \leq e^{-mh} \|v\|_{L^2},$$

which forces $v = 0$. Hence the mean-zero solution is unique. □

Write

$$\varphi_h(r) = h\psi(hr), \qquad \psi(u) := \begin{cases} \dfrac{1}{1 - e^{-u}} - \dfrac{1}{u}, & u > 0, \\ 1/2, & u = 0. \end{cases} \tag{101}$$

Lemma C.4 (L^p bound for the corrector multiplier)**.** *There is a universal constant C such that, for every $p \geq 2$, $h > 0$, and every scalar $v \in L^2(\pi_\lambda) \cap L^p(\pi_\lambda)$ with $\int v\,\mathrm{d}\pi_\lambda = 0$,*

$$\|\psi(h\mathcal{A}_\lambda)v\|_{L^p(\pi_\lambda)} \leq Cp \|v\|_{L^p(\pi_\lambda)}. \tag{102}$$

The operator on the left is defined by the L^2 functional calculus in Proposition B.2; the estimate gives its stated L^p bound.

Proof. Define the bounded sawtooth function

$$\omega(t) := \lceil t \rceil - t, \qquad \bar{\omega}(t) := \omega(t) - \frac{1}{2}.$$

The values at the integers are irrelevant, and $\|\bar{\omega}\|_\infty \leq 1/2$. For $u > 0$, summing over the intervals $[n, n+1)$ gives

$$\begin{aligned} u \int_0^\infty \omega(t) e^{-ut}\,\mathrm{d}t &= u \sum_{n=0}^\infty e^{-nu} \int_0^1 (1 - s) e^{-us}\,\mathrm{d}s \\ &= \frac{1}{1 - e^{-u}} - \frac{1}{u} = \psi(u). \end{aligned} \tag{103}$$

Since $u \int_0^\infty \frac{1}{2} e^{-ut}\,\mathrm{d}t = 1/2$ for $u > 0$,

$$\psi(u) = \frac{1}{2} + u \int_0^\infty \bar{\omega}(t) e^{-ut}\,\mathrm{d}t, \qquad u > 0. \tag{104}$$

The restriction $u > 0$ is essential: at $u = 0$ the integral by itself is not convergent. We instead use the separately defined value $\psi(0) = 1/2$ from equation 101.

For $h > 0$, set

$$M_h(s) := \bar{\omega}(s/h), \qquad \|M_h\|_\infty \le \frac{1}{2},$$

and define

$$m_h(r) := \begin{cases} r\displaystyle\int_0^\infty M_h(s)e^{-sr}\,\mathrm{d}s, & r > 0, \\ 0, & r = 0. \end{cases} \tag{105}$$

The change of variables $s = ht$ in equation 104 gives the pointwise identity

$$\psi(hr) = \frac{1}{2} + m_h(r), \qquad r \ge 0. \tag{106}$$

Notice that equation 106 is now valid also at $r = 0$, because $m_h(0)$ was explicitly defined to be zero.

Let $v \in L^2 \cap L^p$ have zero mean. By equation 88, Proposition B.3 applies to M_h. Since $p \ge 2$ implies $p^* = p$, it gives

$$\|m_h(\mathcal{A}_\lambda)v\|_{L^p} \le 120(p-1)\,\|M_h\|_\infty\,\|v\|_{L^p} \le 60(p-1)\,\|v\|_{L^p}\,.$$

Using equation 106, first in the L^2 functional calculus and then in the L^p extension supplied by the preceding estimate,

$$\psi(h\mathcal{A}_\lambda)v = \frac{1}{2}v + m_h(\mathcal{A}_\lambda)v.$$

Consequently,

$$\|\psi(h\mathcal{A}_\lambda)v\|_{L^p} \le \left(\frac{1}{2} + 60(p-1)\right)\|v\|_{L^p} \le Cp\,\|v\|_{L^p}\,,$$

which is equation 102. □

Lemma C.5 (Corrector moments)**.** *For every $p \ge 2$,*

$$\|\chi_h\|_{L^p(\pi_\lambda;\mathbb{R}^d)} \le ChKp^2. \tag{107}$$

Proof. In this proof, L^p means $L^p(\pi_\lambda)$, and $L^p(\ell_2)$ means $L^p(\pi_\lambda;\mathbb{R}^d)$. By equation 101,

$$\chi_h = h\psi(h\mathcal{A}_\lambda)\nabla U_\lambda.$$

Let $(\varepsilon_i)_{i=1}^d$ be independent Rademacher signs. The Khintchine inequalities, Lemma C.4, and Minkowski's inequality give

$$\begin{aligned} \|\psi(h\mathcal{A}_\lambda)\nabla U_\lambda\|_{L^p(\ell_2)} &\le C\mathbb{E}_\varepsilon \left\|\psi(h\mathcal{A}_\lambda)\sum_{i=1}^d \varepsilon_i \partial_i U_\lambda\right\|_{L^p} \\ &\le Cp\mathbb{E}_\varepsilon \left\|\sum_{i=1}^d \varepsilon_i \partial_i U_\lambda\right\|_{L^p} \\ &\le Cp^{3/2}\,\|\nabla U_\lambda\|_{L^p(\ell_2)}\,. \end{aligned}$$

Lemma C.2 gives $\|\nabla U_\lambda\|_{L^p(\ell_2)} \le K\sqrt{p}$, and therefore

$$\|\chi_h\|_{L^p(\ell_2)} \le ChKp^2.$$

□

Combining Proposition C.3 and Lemma C.5 gives equation 34–equation 35.

C.3 MARTINGALE DECOMPOSITION OF THE LOCAL ERROR

The Markov property and equation 31 imply

$$\begin{aligned}\mathbb{E}[D_k \mid \mathcal{F}_{kh}] &= \int_0^h (I - P_s)\nabla U_\lambda(X_{kh})\,\mathrm{d}s \\ &= (I - P_h)\chi_h(X_{kh}).\end{aligned} \tag{108}$$

Define

$$M_{k+1} := D_k - \big\{\chi_h(X_{kh}) - \chi_h(X_{(k+1)h})\big\}. \tag{109}$$

Since

$$\mathbb{E}[\chi_h(X_{(k+1)h}) \mid \mathcal{F}_{kh}] = P_h\chi_h(X_{kh}),$$

Equation equation 108 proves equation 32–equation 33. This is the "telescoping increment plus martingale increment" decomposition used by Pedrotti & Whalley (2026, Section 2.1, equation (10)).

C.4 SEMIGROUP ESTIMATES AND MARTINGALE VARIANCE

Our goal is the variance bound equation 36. We first control the stationary increments of ∇g_λ using semigroup estimates, then bound the local error and the corrector increment.

Semigroup gradient estimate. Let $r > 0$ and let $\phi : \mathbb{R}^d \to \mathbb{R}^d$ be bounded and measurable. First suppose that U_λ is smooth.

For every unit vector $a \in \mathbb{R}^d$, the reverse local Poincaré inequality under $\mathrm{CD}(m, \infty)$ gives

$$\|\nabla P_r \langle a, \phi\rangle\|^2 \le \frac{m}{e^{2mr} - 1}\Big\{P_r\big(\langle a, \phi\rangle^2\big) - \big(P_r \langle a, \phi\rangle\big)^2\Big\} \le \frac{\|\phi\|_\infty^2}{2r};$$

see Bakry et al. (2014, Theorem 4.7.2(iv), equation (4.7.6), and the bounded-measurable extension discussed there). Since P_r acts coordinatewise,

$$\nabla P_r \langle a, \phi\rangle = (\nabla P_r\phi)^\top a.$$

Taking the supremum over $\|a\| = 1$ yields

$$\|\nabla P_r\phi\|_{\mathrm{op},\infty} \le (2r)^{-1/2}\,\|\phi\|_\infty\,. \tag{110}$$

For the stated $C^{1,1}$ potential, set $U_\lambda^{(\delta)} = U_\lambda * \rho_\delta$. This preserves strong convexity, and the gradients converge uniformly. For smooth compactly supported test functions, SDE stability gives semigroup convergence, so the uniform gradient bound passes to the limit. Approximation in $L^2(\pi_\lambda)$ with uniform L^∞ control, together with equation 70, extends the estimate to bounded measurable test functions. Thus equation 110 holds under the standing assumptions, with gradients understood weakly.

Weak generator integration by parts. The preceding gradient estimate and the operator facts in Appendix B justify the following identity.

Proposition C.6 (Weak generator integration by parts)**.** *Let $\phi = (\phi_1, \dots, \phi_d)$ be bounded and let $r > 0$. Then*

$$\int \nabla g_\lambda \cdot \big(-\mathcal{A}_\lambda P_r\phi\big)\,\mathrm{d}\pi_\lambda = -\int H_\lambda : \nabla P_r\phi\,\mathrm{d}\pi_\lambda. \tag{111}$$

Here $H_\lambda : \nabla P_r\phi = \sum_{i,j}(H_\lambda)_{ij}\partial_j(P_r\phi_i)$.

Proof. Fix a coordinate i, and write $u_i := \partial_i g_\lambda$ and $v_i := P_r\phi_i$. The function u_i is bounded and satisfies $\partial_j u_i = (H_\lambda)_{ij}$ for every j. By equation 110, $v_i = P_r\phi_i$ has a bounded weak gradient. Proposition B.2 also gives $v_i = P_r\phi_i \in D(\mathcal{A}_\lambda)$ and

$$\|\mathcal{A}_\lambda v_i\|_{L^2} \le \frac{1}{er}\,\|\phi_i\|_{L^2}\,.$$

By self-adjointness of $\mathcal{A}_\lambda$ and equation 73, for every $\zeta \in C_c^\infty(\mathbb{R}^d)$,

$$\int \zeta\, \mathcal{A}_\lambda v_i \,\mathrm{d}\pi_\lambda = \int v_i\, \mathcal{A}_\lambda \zeta \,\mathrm{d}\pi_\lambda = \int \langle \nabla v_i, \nabla \zeta \rangle \,\mathrm{d}\pi_\lambda.$$

The last equality uses $\mathcal{A}_\lambda \zeta = -e^{U_\lambda} \operatorname{div}(e^{-U_\lambda} \nabla \zeta)$ and only the weak first derivatives of v_i.

Let η_R be the cutoff from equation 63. The compactly supported Lipschitz function $u_i \eta_R$ admits smooth compactly supported approximations whose functions and weak gradients converge in $L^2(\pi_\lambda)$: this follows by mollification on a fixed compact set, where the Gibbs density is bounded above and below by positive constants. Since $\mathcal{A}_\lambda v_i \in L^2(\pi_\lambda)$ and ∇v_i is bounded, the preceding identity therefore extends to $\zeta = u_i \eta_R$. Using $\partial_j u_i = (H_\lambda)_{ij}$ gives

$$\begin{aligned}\int \eta_R u_i (-\mathcal{A}_\lambda v_i) \,\mathrm{d}\pi_\lambda = &-\sum_{j=1}^d \int \eta_R (H_\lambda)_{ij} \partial_j v_i \,\mathrm{d}\pi_\lambda \\ &- \int u_i \langle \nabla \eta_R, \nabla v_i \rangle \,\mathrm{d}\pi_\lambda. \end{aligned} \tag{112}$$

The last term in equation 112 tends to zero because $|u_i| \le G$, $\|\nabla v_i\|_\infty < \infty$, and $\|\nabla \eta_R\|_\infty \le C/R$. The other two terms converge by dominated convergence, using $\mathcal{A}_\lambda v_i \in L^2$ and $\|H_\lambda\|_{\mathrm{op}} \le \lambda^{-1}$. Thus

$$\int u_i (-\mathcal{A}_\lambda v_i) \,\mathrm{d}\pi_\lambda = -\sum_{j=1}^d \int (H_\lambda)_{ij} \partial_j v_i \,\mathrm{d}\pi_\lambda.$$

Summing over i proves equation 111. □

Short-time drift crossing.

Lemma C.7 (A BV-type semigroup estimate)**.** *For every $t > 0$,*

$$\|P_t \nabla g_\lambda - \nabla g_\lambda\|_{L^1(\pi_\lambda; \mathbb{R}^d)} \le C\sqrt{t}\, \mathbb{E}_{\pi_\lambda} a_\lambda, \tag{113}$$

$$\mathbb{E}\, \|\nabla g_\lambda(X_t) - \nabla g_\lambda(X_0)\|^2 \le C G^2 R \sqrt{t}. \tag{114}$$

Proof. Let $\phi : \mathbb{R}^d \to \mathbb{R}^d$ be bounded with $\|\phi\|_\infty \le 1$. Reversibility and integration in r of Proposition C.6 give

$$\begin{aligned}&\int \phi \cdot (P_t \nabla g_\lambda - \nabla g_\lambda) \,\mathrm{d}\pi_\lambda \\ &\quad = \int \nabla g_\lambda \cdot (P_t \phi - \phi) \,\mathrm{d}\pi_\lambda \\ &\quad = -\int_0^t \int H_\lambda : \nabla P_r \phi \,\mathrm{d}\pi_\lambda \,\mathrm{d}r. \end{aligned} \tag{115}$$

By equation 110 and $H_\lambda \succeq 0$,

$$|H_\lambda : \nabla P_r \phi| \le a_\lambda \|\nabla P_r \phi\|_{\mathrm{op}} \le \frac{a_\lambda}{\sqrt{2r}}.$$

The integrable bound $r^{-1/2}$ justifies equation 115 by first integrating over $[\varepsilon, t]$ and then letting $\varepsilon \downarrow 0$. Substitution in equation 115, integration in r, and duality between L^1 and L^∞ prove equation 113.

By stationarity and reversibility,

$$\begin{aligned}&\mathbb{E}\, \|\nabla g_\lambda(X_t) - \nabla g_\lambda(X_0)\|^2 \\ &\quad = 2 \langle \nabla g_\lambda, (I - P_t) \nabla g_\lambda \rangle_{L^2(\pi_\lambda)} \\ &\quad \le 2G \|\nabla g_\lambda - P_t \nabla g_\lambda\|_{L^1} \le C G \sqrt{t}\, \mathbb{E}_{\pi_\lambda} a_\lambda. \end{aligned}$$

Lemma C.1 completes the proof. □

Completion of the proof of Proposition 6.2. Define the stationary drift modulus

$$\Gamma_\lambda(t) := \mathbb{E}\,\|\nabla U_\lambda(X_t) - \nabla U_\lambda(X_0)\|^2 . \tag{116}$$

From equation 29, Cauchy–Schwarz, stationarity, and Lemma C.1,

$$\begin{aligned}\mathbb{E}\,\|X_t - X_0\|^2 &\le 2t\int_0^t \mathbb{E}\,\|\nabla U_\lambda(X_s)\|^2\,\mathrm{d}s + 4dt \\ &\le 2R^2t^2 + 4dt.\end{aligned} \tag{117}$$

Combining this with the L_f-Lipschitz property of ∇f and equation 114 gives

$$\begin{aligned}\Gamma_\lambda(t) &\le 2L_f^2\mathbb{E}\,\|X_t - X_0\|^2 + 2\mathbb{E}\,\|\nabla g_\lambda(X_t) - \nabla g_\lambda(X_0)\|^2 \\ &\le C\left\{L_f^2(dt + R^2t^2) + G^2R\sqrt{t}\right\}.\end{aligned} \tag{118}$$

Cauchy–Schwarz gives

$$\mathbb{E}\,\|D_k\|^2 \le h\int_0^h \Gamma_\lambda(s)\,\mathrm{d}s. \tag{119}$$

The corrector satisfies

$$\begin{aligned}\mathbb{E}\,\|\chi_h(X_h) - \chi_h(X_0)\|^2 &\le h^2\mathbb{E}\,\|\nabla U_\lambda(X_h) - \nabla U_\lambda(X_0)\|^2 \\ &= h^2\Gamma_\lambda(h).\end{aligned} \tag{120}$$

Indeed, for this calculation write $b = \nabla U_\lambda$ and $A = \mathcal{A}_\lambda$, with all operators acting coordinatewise. Stationarity, reversibility, and the product rule give

$$\begin{aligned}\mathbb{E}\,\|\chi_h(X_h) - \chi_h(X_0)\|^2 &= 2\,\langle \chi_h, (I - P_h)\chi_h\rangle_{L^2} \\ &= 2\,\big\langle b, [r \mapsto \varphi_h(r)^2(1 - e^{-hr})](A)b\big\rangle_{L^2} \\ &\le 2h^2\,\langle b, (I - P_h)b\rangle_{L^2} = h^2\mathbb{E}\,\|b(X_h) - b(X_0)\|^2 .\end{aligned}$$

The inequality is equation 77 applied to $0 \le \varphi_h(r)^2(1 - e^{-hr}) \le h^2(1 - e^{-hr})$, using equation 97.

By stationarity, equation 109, equation 119, equation 120, and $\|a + b\|^2 \le 2\,\|a\|^2 + 2\,\|b\|^2$,

$$\mathbb{E}\,\|M_{k+1}\|^2 \le 2h\int_0^h \Gamma_\lambda(s)\,\mathrm{d}s + 2h^2\Gamma_\lambda(h). \tag{121}$$

Substitution of equation 118, together with

$$h\int_0^h s\,\mathrm{d}s = \frac{h^3}{2}, \qquad h\int_0^h s^2\,\mathrm{d}s = \frac{h^4}{3}, \qquad h\int_0^h \sqrt{s}\,\mathrm{d}s = \frac{2h^{5/2}}{3},$$

proves equation 36. Proposition 6.2 is therefore established.

D Shared-Noise Two-Point Curvature and Feedback

This appendix proves Proposition 6.3. We first establish the shared-noise trace estimate in Proposition D.1. Section D.2 then combines this estimate with the corrector moment bound via Hölder's inequality to obtain equation 38.

D.1 A one-step shared-noise trace estimate

Proposition D.1 (First moment of the shared-noise two-point curvature)**.** *Under Assumptions 4.1–4.2, fix $x, y \in \mathbb{R}^d$, and let*

$$X_t = x - \int_0^t \nabla U_\lambda(X_s)\,\mathrm{d}s + \sqrt{2}B_t, \tag{122}$$

$$\widehat{X}_1 = y - h\nabla U_\lambda(y) + \sqrt{2}B_h. \tag{123}$$

There exist universal constants $c, C > 0$ such that, whenever $hL_\lambda \le c$,

$$\mathbb{E}\,\mathrm{tr}\,A_\lambda^g(X_h, \widehat{X}_1) \le CG\sqrt{\frac{d}{h}}. \tag{124}$$

Proof. **Step 1: the smooth case.** We first prove equation 124 under the additional assumption $g_\lambda \in C^2(\mathbb{R}^d)$. The standing bounds then hold pointwise:

$$\|\nabla g_\lambda(x)\| \le G, \qquad 0 \preceq \nabla^2 g_\lambda(x) \preceq \lambda^{-1} I, \qquad mI \preceq \nabla^2 U_\lambda(x) \preceq L_\lambda I, \quad x \in \mathbb{R}^d.$$

Write

$$W := B_h \sim N(0, hI), \qquad \widetilde{B}_t := B_t - \frac{t}{h} W. \tag{125}$$

The Gaussian endpoint W and the Brownian bridge $(\widetilde{B}_t)_{0 \le t \le h}$ are independent. Conditional on the entire bridge, both X_t and $\widehat{X}_1$ are deterministic functions of the same vector W. For $0 \le \theta \le 1$, define

$$T_\theta(W) := (1-\theta)\widehat{X}_1(W) + \theta X_h(W). \tag{126}$$

With the Brownian bridge fixed, $D_W F$ denotes the ordinary Jacobian matrix $(\partial F_i / \partial W_j)_{i,j=1}^d$ with respect to the endpoint vector $W \in \mathbb{R}^d$. Set $K_t := D_W X_t \in \mathbb{R}^{d \times d}$. Differentiating equation 122 with respect to W gives

$$\dot{K}_t = -\nabla^2 U_\lambda(X_t) K_t + \frac{\sqrt{2}}{h} I, \qquad K_0 = 0. \tag{127}$$

Integrating equation 127 gives

$$K_t = \frac{\sqrt{2}\, t}{h} I - \int_0^t \nabla^2 U_\lambda(X_s) K_s \, \mathrm{d}s.$$

Since $\left\|\nabla^2 U_\lambda(X_s)\right\|_{\mathrm{op}} \le L_\lambda$,

$$\|K_t\|_{\mathrm{op}} \le \frac{\sqrt{2}\, t}{h} + L_\lambda \int_0^t \|K_s\|_{\mathrm{op}} \, \mathrm{d}s.$$

Gronwall's inequality therefore yields

$$\|K_t\|_{\mathrm{op}} \le \frac{\sqrt{2}\, t}{h} e^{L_\lambda t} \le \sqrt{2} e^{h L_\lambda}, \qquad 0 \le t \le h.$$

Moreover, evaluating the integral equation at $t = h$ gives

$$K_h - \sqrt{2} I = -\int_0^h \nabla^2 U_\lambda(X_s) K_s \, \mathrm{d}s,$$

and hence

$$\left\| K_h - \sqrt{2} I \right\|_{\mathrm{op}} \le h L_\lambda \sup_{0 \le s \le h} \|K_s\|_{\mathrm{op}} \le \sqrt{2} h L_\lambda e^{h L_\lambda}.$$

Thus

$$\sup_{0 \le t \le h} \|K_t\|_{\mathrm{op}} \le \sqrt{2} e^{h L_\lambda}, \qquad \left\| K_h - \sqrt{2} I \right\|_{\mathrm{op}} \le \sqrt{2} h L_\lambda e^{h L_\lambda}. \tag{128}$$

Choose c so that $c e^c \le 1/4$. Since $D_W \widehat{X}_1 = \sqrt{2} I$,

$$J_\theta := D_W T_\theta = (1-\theta)\sqrt{2} I + \theta K_h, \qquad \operatorname{sym} J_\theta \succeq 2^{-1/2} I. \tag{129}$$

Define the Hessian averaged along the segment joining the two endpoints by

$$\overline{H}_\lambda(X_h, \widehat{X}_1) := \int_0^1 \nabla^2 g_\lambda(T_\theta(W)) \, \mathrm{d}\theta. \tag{130}$$

Then

$$\overline{H}_\lambda(X_h, \widehat{X}_1)(X_h - \widehat{X}_1) = \nabla g_\lambda(X_h) - \nabla g_\lambda(\widehat{X}_1).$$

The positive-semidefinite Cauchy–Schwarz inequality gives

$$A_\lambda^g(X_h, \widehat{X}_1) \preceq \overline{H}_\lambda(X_h, \widehat{X}_1). \tag{131}$$

Indeed, for every $v \in \mathbb{R}^d$, the quadratic form on the left is

$$\frac{\big(v^\top \overline{H}_\lambda (X_h - \widehat{X}_1)\big)^2}{(X_h - \widehat{X}_1)^\top \overline{H}_\lambda (X_h - \widehat{X}_1)} \le v^\top \overline{H}_\lambda v.$$

Consequently,

$$\operatorname{tr} A^g_\lambda(X_h, \widehat{X}_1) \le \int_0^1 \operatorname{tr} \nabla^2 g_\lambda(T_\theta(W))\, \mathrm{d}\theta. \tag{132}$$

Fix the bridge and θ. By the chain rule, symmetry of the Hessian, and equation 129,

$$\begin{aligned} \operatorname{div}_W\{\nabla g_\lambda(T_\theta(W))\} &= \operatorname{tr}\{\nabla^2 g_\lambda(T_\theta(W)) J_\theta(W)\} \\ &= \operatorname{tr}\{\nabla^2 g_\lambda(T_\theta(W)) \operatorname{sym} J_\theta(W)\} \\ &\ge 2^{-1/2} \operatorname{tr} \nabla^2 g_\lambda(T_\theta(W)). \end{aligned} \tag{133}$$

Gaussian integration by parts for $W \sim N(0, hI)$ yields

$$\begin{aligned} \mathbb{E}_W \operatorname{div}_W\{\nabla g_\lambda(T_\theta(W))\} &= \frac{1}{h} \mathbb{E}_W \langle W, \nabla g_\lambda(T_\theta(W))\rangle \\ &\le \frac{G}{h} \mathbb{E}\, \|W\| \le G\sqrt{\frac{d}{h}}. \end{aligned} \tag{134}$$

Combining equation 132–equation 134 and integrating over θ and the Brownian bridge proves equation 124 in the smooth case.

Step 2: passage to the $C^{1,1}$ Moreau envelope. Let ρ_δ be a standard smooth mollifier and, only within this proof, set

$$g_{\lambda,\delta} := g_\lambda * \rho_\delta, \qquad U_{\lambda,\delta} := f + g_{\lambda,\delta}. \tag{135}$$

Mollification preserves the uniform bounds

$$\sup_x \|\nabla g_{\lambda,\delta}(x)\| \le G, \qquad 0 \preceq \nabla^2 g_{\lambda,\delta} \preceq \lambda^{-1} I, \qquad \|\nabla g_{\lambda,\delta} - \nabla g_\lambda\|_\infty \to 0. \tag{136}$$

Let $X^{(\delta)}$ and $\widehat{X}^{(\delta)}_1$ be the exact and Euler endpoints associated with $U_{\lambda,\delta}$, driven by the same Brownian path as the original endpoints. If

$$\eta_\delta := \|\nabla U_{\lambda,\delta} - \nabla U_\lambda\|_\infty,$$

then Gronwall's inequality gives

$$\sup_{0\le t\le h} \left\| X^{(\delta)}_t - X_t \right\| \le h\eta_\delta e^{hL_\lambda} \longrightarrow 0, \qquad \left\| \widehat{X}^{(\delta)}_1 - \widehat{X}_1 \right\| \le h\eta_\delta \longrightarrow 0. \tag{137}$$

For the limiting argument, define

$$\Sigma(u,v) := \begin{cases} \|v\|^2 / \langle u, v\rangle, & v \ne 0, \\ 0, & v = 0. \end{cases} \tag{138}$$

On the closed set $\{(u,v) : \langle u, v\rangle \ge \lambda \|v\|^2\}$, the function Σ is lower semicontinuous. The trace of the matrix in equation 24 is precisely $\Sigma(x - y, \nabla g_\lambda(x) - \nabla g_\lambda(y))$. Therefore equation 136–equation 137, cocoercivity, and Fatou's lemma imply

$$\begin{aligned} \mathbb{E} \operatorname{tr} A^g_\lambda(X_h, \widehat{X}_1) &\le \liminf_{\delta \downarrow 0} \mathbb{E} \operatorname{tr} A^g_{\lambda,\delta}(X^{(\delta)}_h, \widehat{X}^{(\delta)}_1) \\ &\le CG\sqrt{\frac{d}{h}}. \end{aligned} \tag{139}$$

Thus equation 124 holds under Assumptions 4.1–4.2, completing the proof. □

D.2 From the one-step trace estimate to curvature feedback

In the global synchronous coupling, condition on the starting points of the $(k-1)$st step and apply Proposition D.1. For $k \geq 1$,

$$S_k := \operatorname{tr} A_k^g, \qquad 0 \leq S_k \leq \lambda^{-1}, \qquad \mathbb{E}[S_k \mid \mathcal{F}_{(k-1)h}] \leq CG\sqrt{\frac{d}{h}}. \tag{140}$$

Since $X_0 = \widehat{X}_0$, one has $S_0 = 0$. Hence, for every $k \geq 0$,

$$\mathbb{E}S_k \leq CG\sqrt{d/h}. \tag{141}$$

No independence between S_k and $\chi_h(X_{kh})$ is assumed.

Lemma D.2 (A Hölder product bound)**.** *Let a nonnegative random variable S and a random vector V satisfy*

$$0 \leq S \leq L, \qquad \mathbb{E}S \leq B. \tag{142}$$

Set $B_0 := \min\{B, L\}$. If $B_0 = 0$, then $S = 0$ almost surely. If $B_0 > 0$, then, for every $p > 1$,

$$\mathbb{E}[\|V\|^2 S] \leq \|V\|_{L^{2p}}^2 B_0^{1-1/p} L^{1/p}. \tag{143}$$

If, in addition, $\|V\|_{L^r} \leq Hr^a$ for every $r \geq 2$, then

$$\mathbb{E}[\|V\|^2 S] \leq C_a H^2 B_0 \left\{1 + \log\left(e + \frac{L}{B_0}\right)\right\}^{2a}. \tag{144}$$

Proof. Let $q = p/(p-1)$. Hölder's inequality and $S^q \leq L^{q-1}S$ give

$$\begin{aligned}\mathbb{E}[\|V\|^2 S] &\leq \|V\|_{L^{2p}}^2 \|S\|_{L^q} \leq \|V\|_{L^{2p}}^2 (L^{q-1}B_0)^{1/q} \\ &= \|V\|_{L^{2p}}^2 B_0^{1-1/p} L^{1/p},\end{aligned}$$

which proves equation 143. Choose $p := 2 + \log(L/B_0)$. Then

$$B_0^{1-1/p} L^{1/p} = B_0 (L/B_0)^{1/p} \leq eB_0,$$

and $p \leq C\{1 + \log(e + L/B_0)\}$. Substitution of $\|V\|_{L^{2p}} \leq H(2p)^a$ proves equation 144. □

Proof of Proposition 6.3. We now apply Lemma D.2 with

$$V = \chi_h(X_{kh}), \qquad S = S_k, \qquad L = \lambda^{-1}, \qquad B = CG\sqrt{d/h}. \tag{145}$$

By Lemma C.5, we may take $H = ChK$ and $a = 2$. Since $A_k^g \succeq 0$,

$$\chi_h(X_{kh})^\top A_k^g \chi_h(X_{kh}) \leq \|\chi_h(X_{kh})\|^2 S_k. \tag{146}$$

If $B_0 = B$, then $L/B_0 \leq C\sqrt{h}/(G\sqrt{d}\,\lambda)$; if $B_0 = L$, then $L/B_0 = 1$. Thus

$$B_0 \leq CG\sqrt{d/h}, \qquad 1 + \log\left(e + \frac{L}{B_0}\right) \leq C\ell_{h,\lambda}. \tag{147}$$

Equations equation 144–equation 147 yield

$$\begin{aligned}\mathbb{E}\left[\chi_h(X_{kh})^\top A_k^g \chi_h(X_{kh})\right] &\leq Ch^2K^2G\sqrt{d/h}\,\ell_{h,\lambda}^4 \\ &= CK^2G\sqrt{d}\,h^{3/2}\ell_{h,\lambda}^4.\end{aligned}$$

□