Curvature Coupling Makes Langevin Bias Condition-Number Sharp

Pahan Dewasurendra
Johns Hopkins University

Abstract

The unadjusted Langevin algorithm (ULA) has stationary Wasserstein bias at most $6h\sqrt{\kappa\beta d}$ for a potential with strong convexity α, smoothness β, and condition number $\kappa = \beta/\alpha$. Gaussian examples show that the step-size and dimension dependences are necessary, but their first-order bias is only $\sqrt{\operatorname{tr} H}/4$. They do not explain whether the remaining condition-number factor is real.

We prove that it is. For every $d \geq 4$ and $\kappa \geq 8$, we construct an explicit C^∞ potential whose best Hessian bounds are exactly α and β and whose stationary bias satisfies

$$\liminf_{h\downarrow 0} \frac{W_2(\widehat{\pi}_h, \pi)}{h} \geq c\sqrt{\kappa\beta d}.$$

The mechanism is a bounded curvature coupling between one soft and one stiff coordinate. It converts an $O(h)$ perturbation of stiff-coordinate curvature into an $O(h\sqrt{\kappa})$ soft-coordinate mean shift. Independent copies give the dimension factor. Thus the worst-case first-order bias is $\Theta(\sqrt{\kappa\beta d})$ under only strong convexity and gradient smoothness.

The same family gives a high-accuracy iteration lower bound. For one fixed, bounded scale-free initialization, every constant-step ULA sequence attaining error $\epsilon \downarrow 0$ requires $\Omega(\kappa\sqrt{d}\,\epsilon^{-1}\log(1/\epsilon))$ iterations. This matches the corresponding upper guarantee, including its initialization logarithm.

1 Introduction

Let $\pi(dx) \propto e^{-V(x)}dx$ on $\mathbb{R}^d$, where

$$\alpha I_d \preceq \nabla^2 V(x) \preceq \beta I_d, \qquad \kappa = \frac{\beta}{\alpha}. \tag{1}$$

The unadjusted Langevin algorithm uses

$$X_{k+1} = X_k - h\nabla V(X_k) + \sqrt{2h}\, Z_{k+1}, \quad Z_{k+1} \sim N(0, I_d). \tag{2}$$

For $h \leq 1/\beta$, its transition is contractive and has a unique invariant law $\widehat{\pi}_h$. The difference between $\widehat{\pi}_h$ and π is the price of avoiding an accept-reject correction.

Nonasymptotic analyses first obtained total-variation and then Wasserstein guarantees with polynomial dependence on d, h, and κ under (1) (Dalalyan, 2017; Durmus and Moulines, 2017, 2019; Durmus et al., 2019). Linear dependence on h with improved dimension dependence was later proved under additional derivative or interaction assumptions (Li et al., 2022; Durmus and Eberle, 2024; Chen et al., 2026a; Dalalyan and Karagulyan, 2026; Yang and Wang, 2025). Most recently, Pedrotti and Whalley (2026) removed those extra assumptions and proved

$$W_2(\widehat{\pi}_h, \pi) \leq 6h\sqrt{\kappa\beta d}. \tag{3}$$

This implies the scale-free mixing guarantee

$$k = O\left(\frac{\kappa\sqrt{d}}{\epsilon}\log\left(1 + \frac{\sqrt{\alpha}\,W_2(\mu, \widehat{\pi}_h)}{\epsilon}\right)\right) \tag{4}$$

for $\sqrt{\alpha}W_2(\mathrm{Law}(X_k), \pi) \leq \epsilon$.

The known optimality evidence leaves a gap. If $V(x) = x^\top Hx/2$, direct calculation gives

$$\lim_{h\downarrow 0} \frac{W_2(\widehat{\pi}_h, \pi)}{h} = \frac{1}{4}\sqrt{\operatorname{tr} H}. \tag{5}$$

This proves the $h\sqrt{d}$ dependence when $H = \beta I_d$, while anisotropic Gaussians separately prove that contraction can require κ timescale. Equation (5) does not produce the $\sqrt{\kappa}$ factor in the stationary bias. It is therefore possible that the bias and contraction obstructions cannot occur simultaneously.

We show that they do. Our main theorem identifies the worst-case first-order bias over the exact curvature class. The lower target is a product of two-dimensional nonlinear blocks, so dense high-dimensional interactions are not needed. Each block couples a coordinate of scale $1/\sqrt{\alpha}$ to a coordinate of scale $1/\sqrt{\beta}$. Euler discretization changes the stiff-coordinate curvature statistic by order $h\beta$. The coupling reads this change as a soft mean shift of order $h\sqrt{\kappa\beta}$. This mechanism is absent for every Gaussian target.

Our second theorem makes the bias and contraction barriers simultaneous from a fixed initialization. We start at the target and translate an independent soft Gaussian coordinate. Its mean contracts by exactly $(1-\alpha h)^k$. By the time that translation is small, the nonlinear blocks are close to their biased ULA invariant laws. Any high-accuracy run must therefore use $\beta h = O(\epsilon/\sqrt{d})$ and wait $\Omega((\alpha h)^{-1}\log(1/\epsilon))$ iterations. This matches (4).

Table 1 separates the relevant guarantees. Bounds before 2026 either have slower step-size dependence under (1), or obtain linear dependence using stronger derivative or interaction conditions. The new upper bound is the first linear-in-h result under the Hessian sandwich alone. The open point addressed here is not whether each factor can be necessary in isolation. It is whether one target can force the full product.

Relation to bias expansions. Weak invariant-measure expansions for numerical SDE schemes are classical (Talay and Tubaro, 1990; Kopec, 2015). In particular, Chen et al. (2026a, Proposition 3.1) give the first-order observable formula used below and exhibit tensor products with $\sqrt{d}$ mean bias. We do not claim that calculus as new. Our contribution is the curvature-coupled target that amplifies its coefficient by $\sqrt{\kappa}$, the resulting minimax characterization, and the simultaneous iteration lower bound. Existing sharp examples for product-like models focus on dimension dependence (Durmus and Eberle, 2024). Average-coordinate smoothness can refine earlier square-root-in-h bounds (Dalalyan and Karagulyan, 2026). Our construction has average smoothness of order β, so its average and maximal condition numbers have the same order.

Recent entropy arguments sharpen delocalization for low-dimensional marginals under sparse or weak interactions (Lacker and Zhou, 2025). Related delocalization results now cover unadjusted HMC and underdamped Langevin (Chen et al., 2026b). These works impose interaction structure or study different integrators. They do not give a full-W_2 condition-number lower bound for ULA under the Hessian sandwich alone.

2 First-order bias geometry

We first isolate the analytic object that the construction enlarges. For the regularity class used below, the standard first invariant-density correction is

$$\begin{aligned} \frac{d\widehat{\pi}_h}{d\pi} &= 1 + h(q_V - \pi q_V) + o(h), \\ q_V &= \frac{1}{4}\|\nabla V\|^2 - \frac{1}{2}\Delta V. \end{aligned} \tag{6}$$

The remainder is in a weighted weak topology. It is enough for all linear observables and products considered here.

When (6) holds in a weighted L^2 topology, there is also a geometric reading. Let $L = \Delta - \nabla V \cdot \nabla$, and let ψ be the centered solution of

$$-L\psi = q_V - \pi q_V. \tag{7}$$

The tangent-space description of W_2 gives

$$\lim_{h\downarrow 0} \frac{W_2(\widehat{\pi}_h, \pi)}{h} = \|\nabla\psi\|_{L^2(\pi)}. \tag{8}$$

This is the weighted negative-Sobolev norm of the first density correction. It follows from the standard metric derivative of an absolutely continuous curve in Wasserstein space (Ambrosio et al., 2008). We use it only for interpretation. For every smooth g, integration by parts gives

$$|\operatorname{Cov}_\pi(g, q_V)| \le \|\nabla g\|_{L^2(\pi)} \lim_{h\downarrow 0} \frac{W_2(\widehat{\pi}_h, \pi)}{h}. \tag{9}$$

Choosing a linear g recovers the elementary lower bound by the difference of means. Appendix B records the argument.

The following calculation identifies a broad curvature-transfer mechanism. It also explains why the trigonometric block in the next section is not isolated. Let

$$\begin{aligned} U_c(s,y) &= \frac{C}{2}s^2 + \frac{m}{2}y^2 + cf(s)g(y), \\ V_{a,c}(x,y) &= U_c(\sqrt{a}\,x, y), \end{aligned} \tag{10}$$

where f, g are smooth with bounded derivatives. Let P_c be the target in (s,y) coordinates.

Proposition 1 (Curvature transfer)**.** *Suppose* (10) *is strongly convex with Lipschitz gradient for* $0 < a \le a_0$*, and suppose* $c \neq 0$*. Then its soft-coordinate first-order mean bias is*

$$\begin{aligned} &\lim_{h\downarrow 0} \frac{\mathbb{E}_{\widehat{\pi}_h} X - \mathbb{E}_\pi X}{h} \\ &\quad = -\frac{c}{4\sqrt{a}} \operatorname{Cov}_{P_c}\left(S, af''(S)g(Y) + f(S)g''(Y)\right). \end{aligned} \tag{11}$$

If $\operatorname{Cov}_{P_c}(S, f(S)g''(Y)) \neq 0$*, its magnitude is* $\Theta(a^{-1/2})$ *as* $a \downarrow 0$.

Table 1: Selected stationary W_2 bias results for ULA. Constants and lower order terms are suppressed. The last column records what was known about the joint h, d, κ dependence before this work.

Result	Setting	Bias conclusion	Sharpness information
Gaussian calculation (Wibisono, 2018)	Quadratic V, exact invariant law	$h\sqrt{\operatorname{tr} H}/4 + O(h^2)$	Sharp in h, d, but no adverse $\sqrt{\kappa}$ bias
Li et al. (2022); Durmus and Eberle (2024)	Extra derivative control or limited interactions	Linear-in-h bounds with refined dimension dependence	Product and local-interaction examples explain dimension scaling
Chen et al. (2026a)	Observable expansion and sparse-interaction analysis	Exact first-order weak bias, plus marginal W_{2,ℓ_∞} bounds	Products can have $\sqrt{d}$ mean bias, condition sharpness unresolved
Dalalyan and Karagulyan (2026)	Strong convexity and smoothness, refined coordinate average	Average-condition refinement of square-root-in-h coupling bounds	Shows when global smoothness is pessimistic
Pedrotti and Whalley (2026)	Only $\alpha I \preceq \nabla^2 V \preceq \beta I$	$W_2(\widehat{\pi}_h, \pi) \le 6h\sqrt{\kappa\beta d}$	Gaussian targets certify h, d, and contraction certifies κ separately
This work	Same Hessian class, explicit smooth subclass	$W_2(\widehat{\pi}_h, \pi) \ge ch\sqrt{\kappa\beta d} + o(h)$	One target forces all factors and the matching high-accuracy complexity

Proof. The original-coordinate Laplacian is

$$\Delta V_{a,c} = aC + m + c\{af''(s)g(y) + f(s)g''(y)\}.$$

Apply the linear-observable formula in Lemma 5 below and use $X = S/\sqrt{a}$. The law P_c does not depend on a, so the second claim follows by taking $a \downarrow 0$. ☐

The condition in Proposition 1 is stable under small smooth perturbations. At $c = 0$, independence reduces its limiting covariance to $\mathbb{E}[Sf(S)]\mathbb{E}[g''(Y)]$. Thus the condition is easy to arrange without taking large interactions. The explicit choice $f = \sin$ and $g = \cos$ makes the two terms in (11) have the same sign and permits transparent global Hessian bounds.

3 Minimax first-order bias

Let $\mathcal{V}_d(\alpha, \beta)$ be the C^∞ potentials satisfying (1) whose best global lower and upper Hessian constants are exactly α and β. Define

$$\mathfrak{B}_d(\alpha, \beta) = \sup_{V \in \mathcal{V}_d(\alpha,\beta)} \liminf_{h \downarrow 0} \frac{W_2(\widehat{\pi}_{V,h}, \pi_V)}{h}. \tag{12}$$

Theorem 2 (Sharp first-order bias). *Let $d \ge 4$, $\kappa = \beta/\alpha \ge 8$, and $r = \lfloor (d-2)/2 \rfloor$. Then*

$$\frac{1}{1280}\sqrt{\kappa\beta r} \le \mathfrak{B}_d(\alpha, \beta) \le 6\sqrt{\kappa\beta d}. \tag{13}$$

In particular, $\mathfrak{B}_d(\alpha, \beta) = \Theta(\sqrt{\kappa\beta d})$ with universal constants.

Corollary 3 (A nonlinear Gaussian separation). *Let $\mathcal{Q}_d(\alpha, \beta)$ be the quadratic potentials with exact best Hessian constants α and β. Under the conditions of Theorem 2,*

$$\frac{\mathfrak{B}_d(\alpha, \beta)}{\sup_{V \in \mathcal{Q}_d(\alpha,\beta)} \lim_{h \downarrow 0} W_2(\widehat{\pi}_h, \pi)/h} \ge \frac{\sqrt{\kappa}}{320\sqrt{5}}. \tag{14}$$

Indeed, $r \ge d/5$ for $d \ge 4$, while (5) is at most $\sqrt{\beta d}/4$. The separation can therefore diverge even though both classes obey the same Hessian sandwich. Nonlinearity is not a lower-order technicality for worst-case ULA bias.

The upper bound is (3). We develop an explicit target for the lower bound. It suffices first to set $\beta = 1$ and $a = \alpha = 1/\kappa$. Fix

$$C = 2, \qquad m = \frac{1}{2}, \qquad c_0 = \frac{1}{32}. \tag{15}$$

For $s = \sqrt{a}\,x$, define the two-dimensional block

$$\begin{aligned} V_a(x, y) &= U(s, y), \\ U(s, y) &= \frac{C}{2}s^2 + \frac{m}{2}y^2 + c_0 \sin(s)\cos(y). \end{aligned} \tag{16}$$

Lemma 4 (Curvature and covariance). *For $0 < a \le 1/8$, the block* (16) *satisfies*

$$aI_2 \preceq \nabla^2 V_a \preceq I_2.$$

Let P_c have density proportional to e^{-U} in (s, y) coordinates with the interaction coefficient in (16) *replaced by c. If $F(s, y) = \sin(s)\cos(y)$, then*

$$\operatorname{Cov}_{P_{c_0}}(S, F(S, Y)) > \frac{1}{10}. \tag{17}$$

Both claims have short proofs. The Hessian is

$$\begin{pmatrix} a(C - c_0 \sin s \cos y) & -c_0\sqrt{a}\cos s \sin y \\ -c_0\sqrt{a}\cos s\sin y & m - c_0 \sin s \cos y \end{pmatrix}. \quad (18)$$

Schur complements give the stated bounds. For the covariance, under P_0 the coordinates are independent Gaussians and

$$\mathrm{Cov}_{P_0}(S,F) = \frac{1}{C}\exp\left(-\frac{1}{2C} - \frac{1}{2m}\right) = \frac{1}{2}e^{-5/4}. \quad (19)$$

The density ratio dP_{c_0}/dP_0 is uniformly close enough to one to retain a lower bound above 0.1. Appendix C gives the constants.

We use the following standard specialization of the invariant weak expansion.

Lemma 5 (Linear-observable bias)**.** *Suppose V is strongly convex, has globally Lipschitz gradient, and is smooth with polynomially growing derivatives. For a coordinate map $g(z) = z_j$,*

$$\lim_{h\downarrow 0}\frac{\widehat{\pi}_h g - \pi g}{h} = -\frac{1}{4}\mathrm{Cov}_\pi(g, \Delta V). \quad (20)$$

These hypotheses hold for (16) *and its products.*

Lemma 5 follows directly from the invariant-measure expansion of Talay and Tubaro (1990) or the equivalent observable formula of Chen et al. (2026a, Proposition 3.1). Appendix A checks the interchange conditions for our target.

The original-coordinate Laplacian of one block is

$$\Delta V_a = m + aC - c_0(1+a)\sin(s)\cos(y). \quad (21)$$

The target law in (s, y) coordinates is P_{c_0} and does not depend on a. Combining Lemmas 4 and 5 gives

$$\begin{aligned}\lim_{h\downarrow 0}\frac{\mathbb{E}_{\widehat{\pi}_h}X - \mathbb{E}_\pi X}{h} &= \frac{c_0(1+a)}{4\sqrt{a}}\mathrm{Cov}_{P_{c_0}}(S, F(S,Y)) \\ &> \frac{1}{1280\sqrt{a}}. \end{aligned} \quad (22)$$

Thus a bounded change in stiff-coordinate curvature creates a diverging soft mean coefficient as $a \downarrow 0$.

To finish the construction, let $\ell = d - 2r - 2 \in \{0, 1\}$ and write $z = ((x_i, y_i)_{i=1}^r, u, v, w_1, \dots, w_\ell)$. Define

$$\Phi_{a,d}(z) = \sum_{i=1}^r V_a(x_i, y_i) + \frac{a}{2}u^2 + \frac{1}{2}v^2 + \frac{1}{2}\sum_{j=1}^\ell w_j^2. \quad (23)$$

The u and v coordinates are soft and stiff Gaussian anchors. Lemma 4 and these anchors show that the best Hessian constants of $\Phi_{a,d}$ are exactly a and 1. The r soft block means occupy orthogonal coordinates, so (22) and the elementary inequality $W_2(\mu,\nu) \geq \|\mathbb{E}_\mu Z - \mathbb{E}_\nu Z\|$ imply

$$\liminf_{h\downarrow 0}\frac{W_2(\widehat{\pi}_h, \pi)}{h} \geq \frac{\sqrt{r}}{1280\sqrt{a}}. \quad (24)$$

For general β, use $\Phi_{\alpha,\beta,d}(z) = \Phi_{a,d}(\sqrt{\beta}z)$. Under this change of variables, a ULA step h becomes a base step βh and Wasserstein distance is divided by $\sqrt{\beta}$. Therefore (24) is multiplied by $\sqrt{\beta}$, proving Theorem 2.

4 A matching iteration lower bound

Theorem 2 concerns the invariant law. A mixing lower bound must rule out stopping before that bias develops. The product construction lets a slow transient force enough runtime for the nonlinear bias to emerge.

Let P_h denote the ULA kernel for the target $\Phi_{\alpha,\beta,d}$ above. For every $0 < h \leq 1/\beta$, let $\widehat{\pi}_h$ be its invariant law. Fix $R > 0$. Define the fixed distribution ν by translating the target π by $R/\sqrt{\alpha}$ only in its curvature-α Gaussian anchor. Translation coupling and the mean lower bound give

$$\sqrt{\alpha}\, W_2(\nu, \pi) = R. \quad (25)$$

Moreover, (3) and the triangle inequality give

$$\left|\sqrt{\alpha}\, W_2(\nu, \widehat{\pi}_h) - R\right| \leq 6\beta h\sqrt{d},$$

so the initialization quantity in (4) tends to R.

Theorem 6 (High-accuracy constant-step complexity)**.** *Fix $d \geq 4$, $\kappa \geq 8$, $R > 0$, and the target above. Let $\epsilon_n \downarrow 0$, $0 < h_n \leq 1/\beta$, and $k_n \in \mathbb{N}$ satisfy*

$$\sqrt{\alpha}\, W_2(\nu P_{h_n}^{k_n}, \pi) \leq \epsilon_n. \quad (26)$$

For $r = \lfloor (d-2)/2 \rfloor$,

$$\liminf_{n\to\infty}\frac{k_n\epsilon_n}{\kappa\sqrt{r}\log(R/\epsilon_n)} \geq \frac{1}{10240}. \quad (27)$$

Hence the worst-case high-accuracy complexity is $\Omega(\kappa\sqrt{d}\,\epsilon^{-1}\log(1/\epsilon))$ from the fixed initialization ν with $\sqrt{\alpha}W_2(\nu,\pi) = R$.

Proof. Write $t_n = \beta h_n$. Projection onto the translated Gaussian anchor gives

$$R(1 - t_n/\kappa)^{k_n} \leq \epsilon_n. \quad (28)$$

Accuracy forces $t_n \to 0$. Otherwise, along a subsequence $t_n \to t_\star > 0$, and (28) forces $k_n \to \infty$. Contractivity then places every nonlinear block arbitrarily close to its ULA invariant law. Continuity of that invariant law would make the target invariant under the positive Euler step $t_\star$. This is impossible.

For a p-dimensional target, integration by parts gives $\mathbb{E}_\pi[X^\top \nabla V(X)] = p$, while ULA stationarity gives

$$\mathbb{E}_{\widehat{\pi}_h}[X^\top \nabla V(X)] = p + \frac{h}{2}\mathbb{E}_{\widehat{\pi}_h}\|\nabla V(X)\|^2. \tag{29}$$

Appendix D supplies the continuity argument.

Let $\rho_{n,k}$ be the joint law of the r nonlinear blocks after k steps from their target law, and let $\widehat{\pi}^{\mathrm{blk}}_{h_n}$ be their joint invariant law. Set $D_n = \sqrt{\alpha}\, W_2(\rho_{n,k_n}, \widehat{\pi}^{\mathrm{blk}}_{h_n})$. By (3), product coupling, and contraction,

$$\begin{aligned} D_n &\le 6\sqrt{2r}\, t_n(1 - t_n/\kappa)^{k_n} \\ &\le \frac{6\sqrt{2r}}{R}\, t_n \epsilon_n. \end{aligned} \tag{30}$$

Since $t_n \to 0$, equation (22) says that the scaled norm of the stationary block mean difference is eventually at least $t_n\sqrt{r}/2560$. The mean lower bound, (30), and $\epsilon_n \to 0$ therefore imply, eventually,

$$\sqrt{\alpha}\, W_2(\nu P^{k_n}_{h_n}, \pi) \ge \frac{t_n\sqrt{r}}{5120}. \tag{31}$$

Thus $t_n \le 5120\epsilon_n/\sqrt{r}$. Because $t_n/\kappa \le 1/8$ and $-\log(1-u) \le 2u$ for $0 \le u \le 1/2$,

$$k_n \ge \frac{\kappa}{2t_n}\log\frac{R}{\epsilon_n} \ge \frac{\kappa\sqrt{r}}{10240\epsilon_n}\log\frac{R}{\epsilon_n}.$$

This proves (27). $\square$

The initialization quantity in (25) asymptotically equals the one in the upper bound (4). Theorem 6 therefore matches its dependence on κ, d, ϵ, and initialization. Both the target and initialization are fixed for each (κ, d, R).

The nonlinear blocks and the slow Gaussian anchor belong to one product target. The stationary obstruction and the transient obstruction are present at the same iteration, rather than being drawn from different hard instances.

5 Discussion

Gaussian calculations can certify sharp discretization dependence while missing the geometry that makes conditioning costly. Here the missing geometry is curvature modulation across scales. The stiff coordinate contributes an order-one oscillation to ΔV, while the correlated soft coordinate has scale $1/\sqrt{\alpha}$. Their covariance is therefore of order $1/\sqrt{\alpha}$, even though the Hessian remains between αI and βI. This explains the $\sqrt{\kappa}$ stationary-bias penalty in a way that separate Gaussian bias and contraction examples cannot.

The Wasserstein tangent (8) gives another view. A small density error can have a large transport cost when the Poisson inverse $(-L)^{-1}$ places it along a soft mode. In our block, the density correction contains the stiff curvature statistic $\sin(S)\cos(Y)$. Its projection onto the soft coordinate is bounded away from zero, and transporting that component costs the soft length scale. The factor $\sqrt{\kappa}$ is a cross-scale alignment effect, not simply a large trace or a slow spectral gap.

The construction is local in interaction structure. Every nonlinear coordinate has degree one in the block graph, and the high-dimensional target is a product across blocks. Dense dependence, a growing third derivative, and rough potentials are unnecessary. All derivatives of order at least two are bounded after fixing (α, β). This distinguishes the result from examples where poor conditioning is encoded through a singular interaction graph.

The result is deliberately worst-case. Additional separability, preconditioning, or derivative structure can improve ULA guarantees (Li et al., 2022; Durmus and Eberle, 2024; Chen et al., 2026a; Dalalyan and Karagulyan, 2026; Yang and Wang, 2025). Metropolis adjustment removes stationary bias at the cost of accept-reject dynamics. Its minimax mixing lower bounds use conductance and spectral-gap obstructions rather than stationary bias (Wu et al., 2022). Our theorem instead locates the exact barrier for constant-step ULA under only the classical Hessian sandwich. The minimax result is first-order in the step for fixed (κ, d). The mixing theorem strengthens this by covering every constant-step sequence in the high-accuracy limit. It does not address decreasing-step schedules or oracle lower bounds for arbitrary sampling algorithms.

A Weak invariant-measure expansion

We verify Lemma 5 for the potentials used in the paper. The observable formula of Chen et al. (2026a, Proposition 3.1), written for a centered smooth observable g, is

$$\pi g - \widehat{\pi}_h g = -\frac{h}{4}\pi\{\Delta g + g\Delta \log \pi\} + o(h). \tag{32}$$

For $g(z) = z_j - \pi z_j$, $\Delta g = 0$ and $\Delta \log \pi = -\Delta V$. Therefore (32) becomes

$$\widehat{\pi}_h g - \pi g = -\frac{h}{4}\operatorname{Cov}_\pi(g, \Delta V) + o(h),$$

which is (20).

For completeness, we check why the regularity requirements behind (32) hold here. Each constructed potential is strongly convex with globally Lipschitz gradient. All derivatives of order at least two are bounded, and the potential and gradient have at most quadratic and linear growth. The target has Gaussian tails. Standard polynomial Lyapunov recursions for the contractive ULA chain give moments of every fixed order uniformly for all sufficiently small h.

Let $L = \Delta - \nabla V \cdot \nabla$ and let $Lu = g$. Standard elliptic Poisson regularity for a smooth strongly convex diffusion gives a smooth solution whose derivatives have polynomial growth. Expanding the one-step operator yields

$$\frac{P_h u - u}{h} = Lu + hAu + O\big(h^2(1 + \|z\|^q)\big), \tag{33}$$

for a finite q, where

$$Au = \frac{1}{2}\langle \nabla V, \nabla^2 u \nabla V\rangle - \nabla V \cdot \nabla \Delta u + \frac{1}{2}\Delta^2 u. \tag{34}$$

Set $L_h = (P_h - I)/h$ and $R_h = (L_h u - Lu)/h$. Equation (33) gives $R_h = Au + O(h(1 + \|z\|^q))$. The uniform moment bounds first imply

$$\int R_h \, d\pi \longrightarrow \int Au \, d\pi. \tag{35}$$

The classical $O(\sqrt{h})$ stationary Wasserstein bound, polynomial local Lipschitz bounds for Au, and the same uniform moments also give

$$\int R_h \, d(\pi - \widehat{\pi}_h) = \int Au \, d(\pi - \widehat{\pi}_h) + o(1) \longrightarrow 0. \tag{36}$$

Equations (35) and (36), together with smoothness of u and g, are precisely the three conditions in Chen et al. (2026a, Appendix D). Classical Talay–Tubaro theory gives the same conclusion. Thus the formula applies to every block, finite product, and Gaussian anchor used here.

The equivalent first density correction is

$$\frac{d\widehat{\pi}_h}{d\pi} = 1 + h(q_V - \pi q_V) + o(h), \qquad q_V = \frac{1}{4}\|\nabla V\|^2 - \frac{1}{2}\Delta V, \tag{37}$$

in the corresponding weighted weak topology. Direct substitution gives $Lq_V = -A^* e^{-V}/e^{-V}$, which also verifies the sign and the factor in Lemma 5.

B Wasserstein tangent calculation

Assume the expansion in (6) holds in weighted $L^2(\pi)$. Set $r_V = q_V - \pi q_V$. Its first-order density curve has derivative $r_V \pi$. A velocity field v for this curve must solve the continuity equation

$$-\nabla \cdot (\pi v) = r_V \pi.$$

The minimum-$L^2(\pi)$ solution is a gradient $v = \nabla\psi$. Since $\nabla \cdot (\pi \nabla \psi) = \pi L\psi$, its potential solves (7). The metric-derivative theorem for W_2 then gives (8). Strong convexity supplies a Poincare inequality, so the centered Poisson solution is unique in the weighted Sobolev space.

For any smooth g of polynomial growth,

$$\begin{aligned}\operatorname{Cov}_\pi(g, q_V) = \int g(-L\psi)d\pi &= \int \langle \nabla g, \nabla \psi \rangle d\pi \\ &\leq \|\nabla g\|_{L^2(\pi)} \|\nabla \psi\|_{L^2(\pi)}.\end{aligned}$$

This proves (9). For $g(z) = z_j$, the left side equals $-\operatorname{Cov}_\pi(z_j, \Delta V)/4$. This follows either from Lemma 5 or by integrating the $\|\nabla V\|^2$ term in q_V by parts.

C Block calculations

C.1 Hessian sandwich

Write $q = \sin s \cos y$. To prove $\nabla^2 V_a - aI_2 \succeq 0$, note that its diagonal entries are at least

$$a(C - c_0 - 1) = a(1 - c_0), \qquad m - c_0 - a \geq \frac{11}{32}.$$

The squared off-diagonal entry is at most $c_0^2 a$. Its determinant is therefore at least

$$a\left((1 - c_0)\frac{11}{32} - c_0^2\right) > 0.$$

Both diagonal entries are positive, proving the lower bound.

For the upper bound, the diagonal entries of $I_2 - \nabla^2 V_a$ are at least

$$1 - a(C + c_0) \geq \frac{191}{256}, \qquad 1 - m - c_0 = \frac{15}{32}.$$

Its determinant is at least

$$\frac{191}{256}\frac{15}{32} - c_0^2 a > 0.$$

This proves the upper bound. The soft and stiff Gaussian anchors attain curvatures a and 1, so the best global constants for $\Phi_{a,d}$ are exactly those values.

C.2 Covariance constant

Under P_0, $S \sim N(0, 1/C)$ and $Y \sim N(0, 1/m)$ independently. Gaussian characteristic functions give

$$\mathbb{E}[S \sin S] = \frac{1}{C} e^{-1/(2C)}, \qquad \mathbb{E}[\cos Y] = e^{-1/(2m)},$$

which proves (19). Also $\mathbb{E}S = \mathbb{E}F = 0$ and $\mathbb{E}|S| = 1/\sqrt{\pi}$ for $C = 2$.

Let $R_c = dP_c/dP_0$. Since $|F| \leq 1$,

$$e^{-2c} \leq R_c \leq e^{2c}.$$

For $\delta = e^{2c} - 1$ and every integrable G,

$$|\mathbb{E}_c G - \mathbb{E}_0 G| \leq \delta \mathbb{E}_0 |G|. \tag{38}$$

Using (38) for $G = SF, S, F$ gives

$$\begin{aligned}\operatorname{Cov}_{P_c}(S, F) &\geq \operatorname{Cov}_{P_0}(S, F) - \delta \mathbb{E}_0|S| - \delta^2 \mathbb{E}_0|S| \\ &= \frac{1}{2} e^{-5/4} - \frac{\delta(1 + \delta)}{\sqrt{\pi}}.\end{aligned}$$

At $c = c_0 = 1/32$, this lower bound is $0.1045185\ldots > 1/10$.

C.3 Tensor and scaling details

The target density for one block transforms as

$$e^{-V_a(x,y)}dx\,dy = a^{-1/2}e^{-U(s,y)}ds\,dy.$$

The Jacobian cancels upon normalization, so the (S, Y) target is exactly P_{c_0} for every a. Equation (21) follows by direct differentiation. Applying Lemma 5 to $X = S/\sqrt{a}$ gives

$$\begin{aligned}-\frac{1}{4}\operatorname{Cov}_\pi(X, \Delta V_a) &= \frac{c_0(1+a)}{4\sqrt{a}}\operatorname{Cov}_{P_{c_0}}(S, F) \\ &> \frac{1}{1280\sqrt{a}},\end{aligned}$$

as claimed.

For r product blocks, the ULA kernel and both invariant laws tensorize. The mean-difference vector has r equal soft-coordinate entries and hence norm $\sqrt{r}$ times the one-block difference. Finally, if $\widetilde{V}(z) = V(\sqrt{\beta}z)$ and $u = \sqrt{\beta}z$, then ULA with potential $\widetilde{V}$ and step h becomes base ULA with step βh in u. Also

$$W_2(\widetilde{\pi}, \widehat{\widetilde{\pi}}_h) = \beta^{-1/2}W_2(\pi, \widehat{\pi}_{\beta h}).$$

Dividing by h proves the $\sqrt{\beta}$ scaling used in Theorem 2.

For completeness, the average coordinatewise smoothness of Dalalyan and Karagulyan (2026) can also be read off exactly. For the scaled full target it is

$$M_{\mathrm{av}} = \frac{\beta}{d}\left[r\{a(C + c_0) + m + c_0\} + a + 1 + \ell\right].$$

Since $r \geq d/5$ and $m + c_0 = 17/32$, this lies between $17\beta/160$ and β. Thus its average condition number and the global condition number are within universal constant factors.

D Invariant continuity and the mixing construction

We give the omitted details in Theorem 6. It is enough to work with the normalized block, whose step is $t = \beta h \in (0, 1]$. If its best strong-convexity constant is $\lambda > 0$, synchronous coupling gives

$$W_2(\mu P_t, \nu P_t) \leq (1 - \lambda t)W_2(\mu, \nu). \tag{39}$$

Thus P_t has a unique invariant law $\widehat{\pi}_t$.

We first prove continuity for t bounded away from zero. Couple one P_t and one $P_{t'}$ update from the same X and Gaussian Z. Their difference is

$$(t' - t)\nabla V(X) + \sqrt{2}(\sqrt{t} - \sqrt{t'})Z.$$

Stationary second moments are uniformly bounded on every compact subinterval of $(0, 1]$. Since the gradient is Lipschitz, the one-step L^2 difference tends to zero as $t' \to t$. Applying (39) to one side of the invariance equations gives

$$W_2(\widehat{\pi}_t, \widehat{\pi}_{t'}) \leq \frac{W_2(\widehat{\pi}_t P_t, \widehat{\pi}_t P_{t'})}{\lambda t'},$$

which proves continuity near any $t > 0$.

Next, the target cannot be invariant under any positive ULA step. For a p-dimensional ULA state X at stationarity, expanding $\mathbb{E}\|X - t\nabla V(X) + \sqrt{2t}Z\|^2 = \mathbb{E}\|X\|^2$ gives

$$\mathbb{E}[X^\top \nabla V(X)] = p + \frac{t}{2}\mathbb{E}\|\nabla V(X)\|^2.$$

Under the target, integration by parts gives $\mathbb{E}[X^\top \nabla V(X)] = p$. Strong convexity makes the remaining expectation strictly positive. Hence $\widehat{\pi}_t \neq \pi$ for every $t > 0$.

Suppose now that (26) holds but t_n does not tend to zero. Along a subsequence, $t_n \to t_\star > 0$. The translated anchor has scaled mean $R(1 - t_n/\kappa)^{k_n}$, so accuracy implies $k_n \to \infty$. On a compact subinterval containing this subsequence, invariant continuity bounds $W_2(\pi, \widehat{\pi}_{t_n})$ uniformly, while (39) gives

$$W_2(\pi P_{t_n}^{k_n}, \widehat{\pi}_{t_n}) \le (1 - at_n)^{k_n} W_2(\pi, \widehat{\pi}_{t_n}) \longrightarrow 0.$$

Projection of (26) onto a nonlinear block then forces $\widehat{\pi}_{t_n} \to \pi$. Invariant continuity would give $\widehat{\pi}_{t_\star} = \pi$, contradicting the preceding paragraph. Therefore $t_n \to 0$.

For one scaled nonlinear block, (3) gives

$$\sqrt{\alpha}\, W_2(\pi^{\mathrm{blk}}, \widehat{\pi}_h^{\mathrm{blk}}) \le 6\sqrt{2}\, t.$$

Product coupling gives $6\sqrt{2r}\, t$ for all r blocks. Their contraction factor is at most $(1 - t/\kappa)^k$. The translated-anchor mean and accuracy give $(1 - t_n/\kappa)^{k_n} \le \epsilon_n/R$, proving (30).

For each nonlinear block, the scaled stationary soft-coordinate mean difference divided by t tends to a value greater than 1/1280. Hence for all sufficiently small t, the norm over the r orthogonal soft coordinates is at least $t\sqrt{r}/2560$. The distance from the time-k_n block law to its invariant law is at most the right side of (30). Since $\epsilon_n \to 0$, subtracting these two bounds gives (31).

Finally, translating the target by $R/\sqrt{\alpha}$ in one coordinate proves (25). Under one ULA step, the translated anchor mean is multiplied by $1 - \alpha h = 1 - t/\kappa$. Induction gives the exact scaled mean $R(1 - t/\kappa)^k$, which proves (28) and completes the argument.

E Why Gaussian targets miss the condition factor

Let $V(x) = x^\top Hx/2$ with eigenvalues $\lambda_1, \ldots, \lambda_d$. In the eigenbasis, the target standard deviation in coordinate i is $\lambda_i^{-1/2}$. The ULA recursion is an autoregression with invariant standard deviation

$$\widehat{\sigma}_i = \frac{1}{\sqrt{\lambda_i(1 - h\lambda_i/2)}}.$$

The optimal transport between the centered diagonal Gaussians is coordinatewise, so

$$W_2^2(\widehat{\pi}_h, \pi) = \sum_{i=1}^{d} \frac{1}{\lambda_i} \left((1 - h\lambda_i/2)^{-1/2} - 1 \right)^2.$$

Expanding each term gives

$$\lim_{h \downarrow 0} \frac{W_2(\widehat{\pi}_h, \pi)}{h} = \frac{1}{4} \left(\sum_{i=1}^{d} \lambda_i \right)^{1/2},$$

which is (5). For one eigenvalue α and the rest at β, this coefficient remains $O(\sqrt{\beta d})$. The slow coordinate does create the κ contraction time, but it does not amplify stationary bias. The nonlinear curvature statistic in (21) is what joins the two obstructions.

References

Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, Basel, 2 edition.

Chen, Y., Cheng, X., Niles-Weed, J., and Weare, J. (2026a). Convergence of unadjusted langevin in high dimensions: Delocalization of bias. *Communications on Pure and Applied Mathematics*, 79(6):1467–1491.

Chen, Y., Cheng, X., Niles-Weed, J., and Weare, J. (2026b). Delocalization of bias in unadjusted hamiltonian monte carlo and underdamped langevin. *arXiv preprint arXiv:2607.15208*.

Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B*, 79(3):651–676.

Dalalyan, A. S. and Karagulyan, A. (2026). Improved guarantees for langevin monte carlo with average smoothness. *arXiv preprint arXiv:2605.31413*.

Durmus, A. and Eberle, A. (2024). Asymptotic bias of inexact markov chain monte carlo methods in high dimension. *The Annals of Applied Probability*, 34(4):3435–3468.

Durmus, A., Majewski, S., and Miasojedow, B. (2019). Analysis of langevin monte carlo via convex optimization. *Journal of Machine Learning Research*, 20(73):1–46.

Durmus, A. and Moulines, É. (2017). Nonasymptotic convergence analysis for the unadjusted langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587.

Durmus, A. and Moulines, É. (2019). High-dimensional bayesian inference via the unadjusted langevin algorithm. *Bernoulli*, 25(4A):2854–2882.

Kopec, M. (2015). Weak backward error analysis for overdamped langevin processes. *IMA Journal of Numerical Analysis*, 35(2):583–614.

Lacker, D. and Zhou, F. (2025). A hierarchical entropy method for the delocalization of bias in high-dimensional langevin monte carlo. *arXiv preprint arXiv:2509.08619*.

Li, R., Zha, H., and Tao, M. (2022). $\sqrt{d}$ dimension dependence of langevin monte carlo. In *International Conference on Learning Representations*.

Pedrotti, F. and Whalley, P. A. (2026). Wasserstein mixing time of the unadjusted langevin algorithm. *arXiv preprint arXiv:2608.02430*.

Talay, D. and Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509.

Wibisono, A. (2018). Sampling as optimization in the space of measures: The langevin dynamics as a composite optimization problem. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 2093–3027.

Wu, K., Schmidler, S., and Chen, Y. (2022). Minimax mixing time of the metropolis-adjusted langevin algorithm for log-concave sampling. *Journal of Machine Learning Research*, 23(270):1–63.

Yang, B. and Wang, X. (2025). Non-asymptotic error bounds in $\beth_2$-distance with $\sqrt{d}$ dimension dependence and first order convergence for langevin monte carlo beyond log-concavity. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 71358–71382.