

A General Framework for Metropolis-Adjusted Dikin Walks: Dimension-Square Mixing on Polytopes and Log-Det Walks on Spectrahedra

Zhao Song* Lichen Zhang[†]

August 25, 2026

Abstract

We analyze exact-metric, Metropolis-adjusted Dikin walks by keeping the proposal determinant and reverse quadratic form together. Their leading uncentered terms cancel in the complete logarithmic acceptance ratio, leaving centered fluctuations that can be controlled with second-order tools. For a polytope given by n inequalities and a convex L -Lipschitz potential, this yields warm-start mixing in $\tilde{O}((d^2 + dL^2R^2) \log(w/\delta))$ steps for the regularized Lee–Sidford walk. For a spectrahedron with $n \times n$ blocks, the log-det walk mixes in $\tilde{O}((\psi_*^+ nd + dL^2R^2) \log(w/\delta))$ steps, where ψ_*^+ measures matrix leverage.

The two analyses share an acceptance-to-mixing reduction. A proposal-comparison argument transfers the polytope bound to an appropriately padded $O(1/d)$ -accurate metric computed from high-precision Lewis weights. For spectrahedra, given $\hat{\psi} \geq \psi_*^+$, a direct-or-two-seed TensorSRHT construction gives an exact-arithmetic implementation with ψ_*^+ replaced by $\hat{\psi}$ in the mixing bound.

Contents

1	Introduction	2
1.1	Our results	2
1.2	Prior work and scope	4
2	Technical overview	4
2.1	Proof overview for Theorem 1.1	5
2.2	Proof overview for Theorem 1.2	7
2.3	From exact metrics to implementable kernels	8
3	Preliminaries	9
3.1	Notation	9
3.2	Polytope, slacks, and Lewis weights	10
3.3	First-derivative Lewis calculus	10
3.4	Analytic, metric, and geometric inputs	11
3.5	Setting and the walk	13
3.6	The matrix leverage operator	14
3.7	Toolbox	14

*Independent researcher. Email: magic.linuxkde@gmail.com.

[†]Massachusetts Institute of Technology. Email: lichenz@mit.edu.

4	Proof of Theorem 1.1	16
4.1	The metric, the walk, and the main theorem	16
4.2	The complete proposal ratio and the exact cancellation	18
4.3	Uniform local stability	20
4.4	A row-dependent contraction of Lewis Hessians	24
4.5	A Stein bound for the centered fluctuation, and the acceptance lower tail	27
4.6	From acceptance to mixing	31
5	Proof of Theorem 1.2	34
5.1	Main theorem	34
5.2	Normalization, endpoint formulas, and the ledger	35
5.3	The spectral tube, stability, and the determinant remainder	37
5.4	The Stein mean bound	38
5.5	Fluctuation bound, parameter choice, and the acceptance theorem	40
5.6	From acceptance to mixing	43
5.7	A worst-case acceptance obstruction: the arrow spectrahedron	45
A	Approximate metrics and exact-arithmetic complexity	48
A.1	High-precision Lewis weights	49
A.2	From metric approximation to mixing	50
A.3	Arithmetic complexity for polytopes	53
A.4	A sketched log-det metric	54
A.5	Arithmetic complexity for spectrahedra	57

1 Introduction

Sampling from a log-concave density constrained to a convex body is a basic task in sampling and optimization. Dikin walks adapt a Gaussian proposal to a local barrier metric, allowing the walk to exploit the geometry of the feasible region. For nearly universal barriers, this affine-invariant approach can also avoid an artificial dependence on the number of constraints.

The same adaptation creates the main analytic difficulty. The Metropolis filter compares a forward Gaussian defined at the current point with a reverse Gaussian defined at the proposal. It is therefore not enough to show that the forward step remains in a local ellipsoid: one must also control the determinant and the reverse local quadratic form at the random endpoint.

Can these two endpoint terms be controlled directly, without estimating them separately? We answer this question by keeping the complete logarithmic Metropolis ratio intact. The linear part of the log determinant cancels the uncentered mean of the reverse quadratic term, leaving centered fluctuations that can be controlled by Gaussian integration by parts and second-order concentration. This yields direct analyses for Lee–Sidford geometry on polytopes and log-det geometry on spectrahedra.

1.1 Our results

Theorem 1.1 (Polytopes; informal version of Theorem 4.2). *Let $K \subseteq B_2(0, R)$ be a bounded full-dimensional polytope in $\mathbb{R}^d$ given by n linear inequalities, let f be convex and L -Lipschitz on K , and let $\pi \propto e^{-f} \mathbf{1}_K$. Let H_{LS} denote the Lee–Sidford metric, let $r > 0$ be a universal constant,*

and choose κ as in Eq. (47). Consider the $\frac{1}{2}$ -lazy Metropolis walk with exact Lewis weights, exact acceptance ratio, and proposal covariance

$$\frac{r^2}{d}(\kappa H_{\text{LS}}(x) + L^2 I)^{-1}.$$

Then, for $w \geq 1$ and $0 < \delta < 1$, from a w -warm start the walk reaches total-variation distance δ from π in

$$\tilde{O}((d^2 + dL^2 R^2) \log(w/\delta))$$

steps.

Theorem 1.2 (Spectrahedra; informal version of Theorem 5.1). *Let*

$$K := \{x \in \mathbb{R}^d : A_0 + \sum_{i=1}^d x_i A_i \succeq 0\}$$

be bounded, strictly feasible after minimal-face reduction, and contained in $B_2(0, R)$, with $n \times n$ matrix blocks. Let f be convex and L -Lipschitz on K , and let $\pi \propto e^{-f} \mathbf{1}_K$. For the log-det metric, let ψ_*^+ be the maximum of one and the supremum, over $x \in K^\circ$, of the spectral norm of the whitened matrix leverage operator at x . Choose κ as in Eq. (74), so $\kappa = \tilde{O}(\psi_*^+)$. Then, for $w \geq 1$ and $0 < \delta < 1$, the exact-metric $\frac{1}{2}$ -lazy Metropolis walk reaches total-variation distance δ from a w -warm start in

$$\tilde{O}((\psi_*^+ nd + dL^2 R^2) \log(w/\delta))$$

steps.

For the uniform target, an arrow spectrahedron shows that $\kappa = 1$ can have exponentially small acceptance at the center when $\psi(0) = \Theta(d)$ and the proposal radius is a fixed universal constant. This is an acceptance obstruction, not a mixing lower bound.

Approximate metrics and arithmetic complexity. The appendix turns both exact-metric analyses into exact-arithmetic implementations by comparing the implemented proposals with the exact reference proposals. For polytopes, deterministic high-precision Lewis weights give a suitably padded $O(1/d)$ -accurate metric. For spectrahedra, independent TensorSRHT sketches at the forward and reverse endpoints give an exactly reversible two-seed chain. In both settings the proposal comparison transfers the corresponding mixing bound without differentiating the approximate metric. The formal arithmetic statements are Corollaries A.9 and A.16.

Theorem 1.3 (Approximate metrics and exact-arithmetic complexity; informal). *In each setting, assume the hypotheses of the corresponding theorem above and work in the exact-arithmetic model of Appendix A. Let ω denote the exponent of square-matrix multiplication. Starting from an exact draw $X_0 \sim \mu_0$, where μ_0 is w -warm, the following implementations reach total-variation error δ .*

- (i) *For polytopes, deterministic $O(1/d)$ -relative Lewis-weight approximation with one-sided metric padding gives an exactly reversible walk using $\tilde{O}((d^2 + dL^2 R^2)(nd^{\omega-1} + d^\omega) \log(ew/\delta))$ arithmetic operations and $\tilde{O}((d^2 + dL^2 R^2) \log(ew/\delta))$ evaluations of f .*
- (ii) *For spectrahedra, given a supplied upper bound $\hat{\psi} \geq \psi_*^+$, the direct-or-two-seed-sketch construction gives an exactly reversible walk using $\tilde{O}((\hat{\psi}nd + dL^2 R^2)(n^\omega + n^2 d^{3\omega-5}) \log(ew/\delta))$ arithmetic operations and $\tilde{O}((\hat{\psi}nd + dL^2 R^2) \log(ew/\delta))$ evaluations of f . The unconditional choice $\hat{\psi} = d$ replaces $\hat{\psi}nd$ by nd^2 .*

These are arithmetic-operation bounds, not bit-complexity bounds: they exclude finite-bit conditioning, rounded Gaussian sampling, inexact Metropolis decisions, and construction of the warm start.

1.2 Prior work and scope

Kannan and Narayanan introduced the logarithmic-barrier Dikin walk [KN12]. Lewis-weight metrics later reduced the dependence on the number of inequalities and led to the conjecture that $\tilde{O}(d^2)$ iterations should suffice from a warm start [CDWY18]. Laddha, Lee, and Vempala developed strong self-concordance and symmetry as a framework for this question [LLV20], while Kook and Vempala isolated average self-concordance as the condition controlling the reverse local norm at a random proposal [KV24]. The original dimension-square claim of Laddha, Lee, and Vempala leaves open this reverse-ellipsoid step; a direct repair gives a dimension-cube bound [Koo26].

Gu et al. [GKM⁺24] claimed $\tilde{O}(d^2 + dL^2R^2)$ mixing for a robust Lee–Sidford walk. Kook subsequently identified a gap in the proof of their Lemma D.5 [Koo26]: closeness of the endpoint Lewis weights gives only $O(r^2)$ control of the squared-local-norm change, whereas the average-self-concordance argument needs $O(r^2/d)$. Moreover, the endpoint weights depend on the Gaussian proposal direction, so the fixed-coefficient polynomial argument for the log barrier does not transfer directly. This is a proof gap, not a counterexample to the dimension-square theorem. Kook proved the best previously complete $\tilde{O}(d^{9/4})$ warm-start bound for exponential targets by scaling the Lee–Sidford metric by $d^{1/4}$.

Our polytope proof addresses the same Metropolis-filter obstruction without proving average self-concordance: it controls the complete acceptance ratio after determinant–quadratic cancellation. The main theorem is first proved for the exact metric. Appendix A then transfers the acceptance and mixing bounds to a padded metric constructed from high-precision approximate Lewis weights.

The same 2024 paper separately states $\tilde{O}((nd + dL^2R^2) \log(w/\delta))$ mixing for spectrahedra [GKM⁺24]. Kook’s 2026 criticism is specific to the polytope Lee–Sidford argument and does not establish a gap in the spectrahedral proof. Theorem 1.2 neither corrects nor improves that prior statement uniformly: when $\psi_*^+ > 1$, its displayed rate is weaker. We give a complete exact-metric proof with explicit leverage dependence and make no priority claim over the earlier spectrahedral theorem.

Table 1 places the two results in context. The target class is listed explicitly because the uniform, exponential, and general Lipschitz settings are not interchangeable. A dagger marks a bound whose published proof is known to contain a gap. The daggered entries identify the buggy results discussed above.

Organization. Section 2 explains the shared cancellation and the transfer to implementable kernels. Section 3 collects the common tools. Section 4 proves the polytope result via Lewis-weight stability, Stein concentration, and conductance. Section 5 proves the spectrahedral result via the noncommutative ledger, a leverage-controlled spectral tube, and conductance, and concludes with the arrow spectrahedron obstruction. Appendix A gives the formal approximate-metric kernels and their exact-arithmetic complexity bounds.

2 Technical overview

This section separates the two proof mechanisms. Section 2.1 explains the Lewis-weight argument for Theorem 1.1, while Section 2.2 explains the noncommutative log-det argument for Theorem 1.2.

Table 1: Selected warm-start bounds for Dikin-type walks. A $\dagger$ marks a result whose published proof contains a known gap. Here n denotes the number of inequalities for a polytope or the matrix size for a spectrahedron, and $\tau = \log(w/\delta)$. The final column suppresses O and $\tilde{O}$.

(a) Polytopes			
Year	Reference	Setting/metric	Warm-start iterations
2012	[KN12]	Uniform; log-barrier Dikin walk	$nd\tau$
2018	[CDWY18]	Uniform; John walk	$d^{5/2} \log^4(n/d)\tau$
2020	[LLV20] †	Uniform; Lee–Sidford Dikin walk	$d^2\tau$
2024	[GKM ⁺ 24] †	L -Lipschitz; robust Lee–Sidford walk	$d(d + L^2 R^2)\tau$
2026	[Koo26]	Exponential targets; scaled Lee–Sidford metric	$d^{9/4}\tau$
2026	Theorem 1.1	L -Lipschitz; exact regularized Lee–Sidford metric	$d(d + L^2 R^2)\tau$
(b) Spectrahedra			
Year	Reference	Setting/metric	Warm-start iterations
2024	[GKM ⁺ 24] †	L -Lipschitz; log-det metric	$d(n + L^2 R^2)\tau$
2026	Theorem 1.2	L -Lipschitz; exact regularized log-det metric	$d(\psi_*^+ n + L^2 R^2)\tau$

Section 2.3 then explains how approximate metrics give implementable versions of both walks.

2.1 Proof overview for Theorem 1.1

Exact cancellation. Write $H(x) = A_x^\top W_x^{1-2/p} A_x$ and $G(x) = \kappa H(x) + L^2 I$. At a fixed current point, an affine change of coordinates makes $H(x) = I$ and writes $G(x) = \kappa(I + \mathcal{C})$, where $\mathcal{C} \succeq 0$. Thus a proposal has the form $y = x + \eta h$, with $h \sim N(0, Q)$, $Q = (I + \mathcal{C})^{-1} \preceq I$, and $\eta = r/\sqrt{\kappa d}$. If $U = W_x^{1/2-1/p} A_x$, then $U^\top U = I$. The endpoint metric is represented exactly as $H(y) = I + U^\top \text{Diag}(b)U$, where each coordinate of b records the combined change in the corresponding slack and Lewis weight. With $c_i = (u_i^\top h)^2$, $\bar{c}_i = \mathbb{E}[c_i]$, and $E = Q^{1/2}(H(y) - I)Q^{1/2}$, Lemma 4.4 gives

$$\log \frac{q_y(x)}{q_x(y)} = -\frac{1}{2}(c - \bar{c})^\top b + \frac{1}{2}(\log \det(I + E) - \text{tr}[E]).$$

This identity is the central cancellation: the linear determinant term $\text{tr}[E] = \bar{c}^\top b$ is exactly the Gaussian mean of the uncentered quadratic term and disappears before any tail bound is applied. What remains is a centered contraction and two genuinely second-order remainders. Figure 1 summarizes the argument.

Lewis-weight stability and acceptance. The weights are implicit, so differentiating their fixed-point equation directly would obscure the cancellation. Instead Lemma 3.1 specializes the row-rescaling Jacobian of Lee–Sidford [LS19, Lemma 24] to spatial slack changes. The resulting explicit identity yields a high-probability tube on which all slacks and weights remain comparable to their values at x ; see Lemmas 4.6 and 4.8. Rowwise resolvent estimates then give the row-dependent Hessian contraction in Lemma 4.15. This stronger, leverage-sensitive bound is what prevents a factor depending on the number of inequalities from entering the fluctuation estimate.

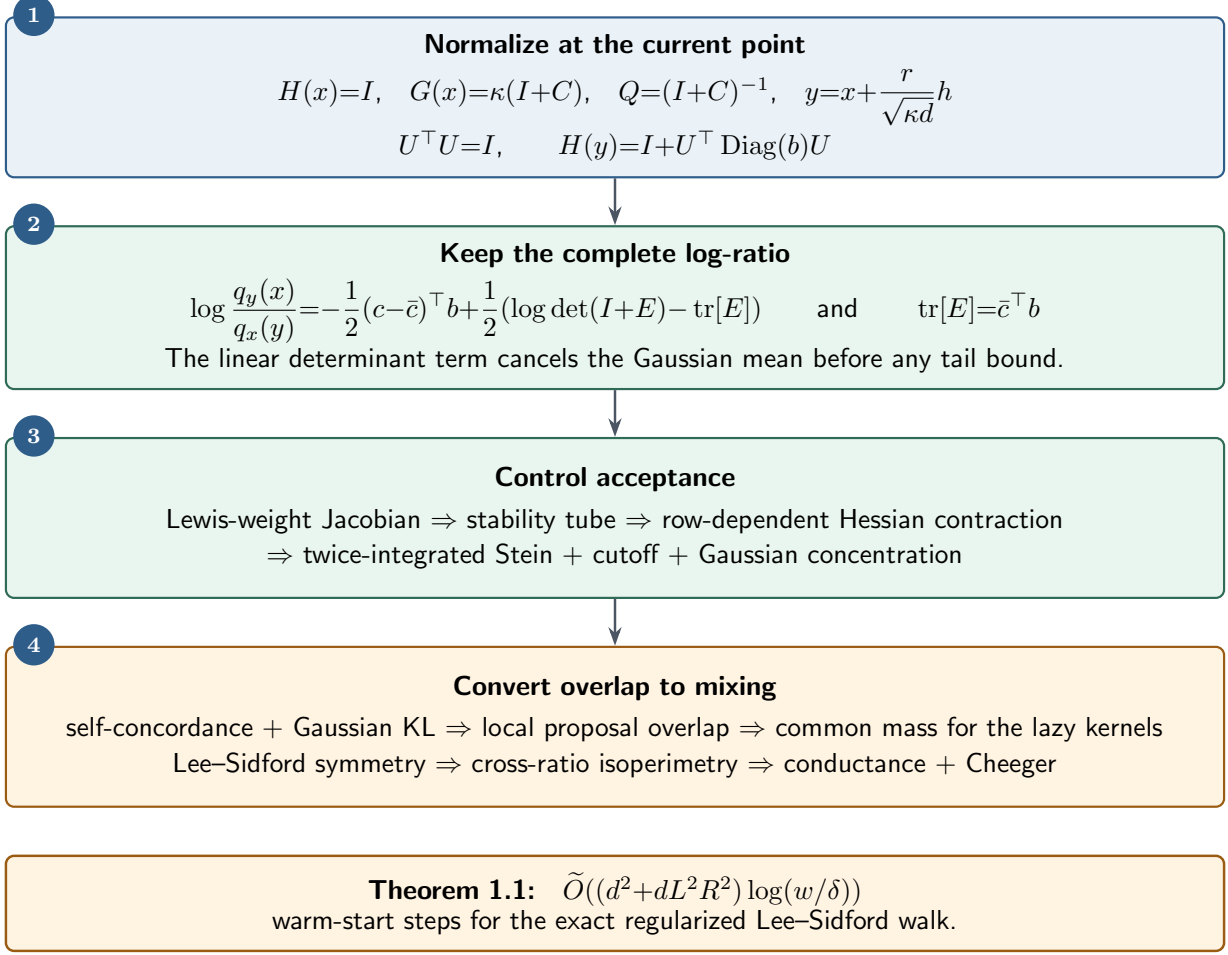


Figure 1: Main steps for Theorem 1.1: normalization, exact determinant–quadratic cancellation, Lewis-weight stability and acceptance, and conductance and mixing.

Split the centered term as its linear part $Z = (c - \bar{c})^\top a$ plus the nonlinear remainder generated by $b - 2a$. Applying Gaussian integration by parts twice controls $\mathbb{E}[Z]$ through the Hessian contraction, while Lemmas 4.9 and 4.11 control the quadratic and determinant remainders. On the stability tube, Z has a dimension-free Lipschitz constant. A cutoff, McShane extension, and Gaussian concentration [Led01] therefore give the required lower tail without conditioning the Gaussian law. Feasibility follows from the Dikin containment of the unit $H(x)$ -ellipsoid, and the potential contribution is bounded because $G(x) \succeq L^2 I$ and f is L -Lipschitz. Choosing $\rho = r/\sqrt{\kappa}$ as in Eq. (47) makes every remainder universally small, proving the uniform acceptance statement Lemma 4.23.

From overlap to mixing. Strong self-concordance of the regularized metric controls nearby covariance matrices. Substituting this distortion into the Gaussian KL formula and applying Pinsker gives proposal overlap whenever $\|x - z\|_{G(x)} = O(r/\sqrt{d})$. Combining this with the acceptance event gives a fixed common mass for the two lazy Metropolis kernels. The scaled Lee-Sidford metric is $\bar{\nu}_{\text{LS}}$ -symmetric with $\bar{\nu}_{\text{LS}} = \tilde{O}(d)$, and regularization changes the symmetry parameter

only to $\nu_G \leq \bar{\nu}_{\text{LS}} + 4L^2R^2$. Local-norm separation therefore implies chord cross-ratio separation. Lovász–Vempala isoperimetry [LV07] converts that separation into $\Phi \geq \frac{cr}{\sqrt{d(\bar{\nu}_{\text{LS}} + L^2R^2)}}$. Finally, the reversible Cheeger inequality [LS88] and the direct $L^2(\pi)$ -to-total-variation calculation for a w -warm start yield $\tilde{O}((d^2 + dL^2R^2) \log(w/\delta))$ steps, which is Theorem 1.1.

2.2 Proof overview for Theorem 1.2

Exact cancellation. For a spectrahedron, slack congruence at the current point makes $S(x) = I$. A further linear change of variables gives a Frobenius-orthonormal frame $\{\bar{A}_a\}_{a=1}^d$ for the whitened constraint directions and $H(x) = I_d$. As above, write $G(x) = \kappa(I + \mathcal{C})$, $Q = (I + \mathcal{C})^{-1}$, and $y = x + \eta h$ with $h \sim N(0, Q)$ and $\eta = r/\sqrt{\kappa d}$. Put $\Delta = \eta \sum_a h_a \bar{A}_a$ and $R = (I + \Delta)^{-1}$. Slack congruence and trace cyclicity give exact endpoint formulas for $H(y) - I =: \bar{E}$ and all of its first two derivatives. For $E = Q^{1/2} \bar{E} Q^{1/2}$, Lemma 5.3 becomes

$$\log \frac{q_y(x)}{q_x(y)} = \frac{1}{2}(\log \det(I + E) - \text{tr}[E]) - \frac{1}{2}Z(h), \quad Z(h) = h^\top \bar{E} h - \text{tr}[Q \bar{E}].$$

Thus the same determinant–quadratic cancellation is present, but the centered fluctuation is now a noncommutative trace polynomial.

The leverage-controlled spectral tube. The matrix leverage operator $\Psi(x) = \sum_a \bar{A}_a^2$ has $\psi(x) = \|\Psi(x)\|_{\text{op}}$ and $\psi_*^+ = \max(1, \sup_{x \in K^\circ} \psi(x))$. Tropp’s self-adjoint Gaussian-series inequality [Tro15, Theorem 4.1.1], applied to $\sum_a h_a \bar{A}_a$, gives a high-probability spectral tube on which $\|\Delta\|_{\text{op}}$ is small. The resolvent is then stable on a slightly larger collar, so the determinant remainder is quadratic in E . The key noncommutative estimate is the paired four-trace bound Lemma 5.8; it contracts the second derivatives arising from twice-integrated Stein to the single parameter $\psi(x)$ rather than a larger ambient-dimensional quantity. Consequently the mean of Z is $O(\rho^2 \psi)$.

The gradient formula for Z and the same trace contractions give a dimension-free Lipschitz bound on the collar. After a cutoff and McShane extension, Gaussian concentration controls $Z - \mathbb{E}[Z]$ globally; intersecting this event with the spectral tube also guarantees feasibility because $I + t\Delta \succ 0$ for $0 \leq t \leq 1$. The potential term is handled by $G(x) \succeq L^2 I$ exactly as in the polytope proof. Choosing ρ as in Eq. (74) gives $\kappa = \tilde{O}(\psi_*^+)$ and makes the determinant remainder, the Stein mean, and the fluctuation universally bounded. This proves the uniform acceptance lower tail in Lemma 5.15. Figure 2 summarizes the argument.

From overlap to mixing. Operator self-concordance of the log-det metric controls nearby Gaussian covariances. The Gaussian KL identity and Pinsker again give proposal overlap at local distance $O(r/\sqrt{d})$, and the acceptance lower tail transfers this to common mass for the lazy kernels. In normalized coordinates, define the block displacement $\Delta_{z-x} := \sum_a (z_a - x_a) \bar{A}_a$. Then $z \in K \cap (2x - K)$ is equivalent to $\|\Delta_{z-x}\|_{\text{op}} \leq 1$, whereas $\|z - x\|_{H(x)} = \|\Delta_{z-x}\|_F$. Hence $\|\Delta_{z-x}\|_F^2 \leq n \|\Delta_{z-x}\|_{\text{op}}^2$, and the regularized metric has symmetry parameter $\nu_G = \kappa n + 4L^2R^2$. Cross-ratio isoperimetry and the same deep-set conductance argument yield $\Phi \geq \frac{cr}{\sqrt{d(\kappa n + L^2R^2)}}$. Cheeger and warm-start contraction therefore give $\tilde{O}((\psi_*^+ n d + dL^2R^2) \log(w/\delta))$ steps, proving Theorem 1.2.

Why leverage scaling is necessary. For the uniform target, the arrow spectrahedron in Lemma 5.25 has $\psi(0) = \Theta(d)$ and exponentially small acceptance at the center when $\kappa = 1$ and the proposal radius is a fixed universal constant. It explains why leverage scaling is needed for uniform acceptance, while making no mixing lower-bound claim.

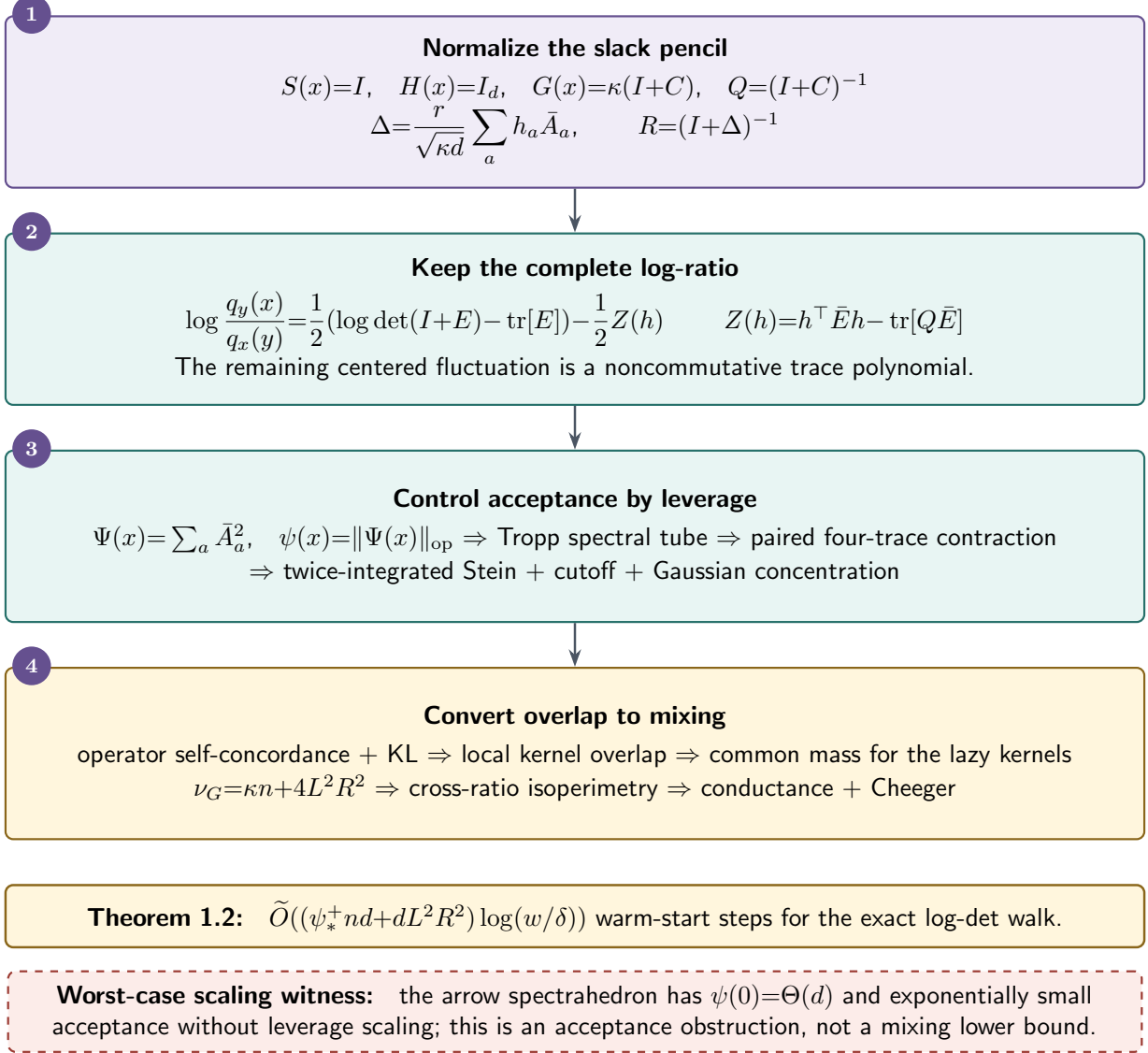


Figure 2: Main steps for Theorem 1.2: normalization, exact determinant–quadratic cancellation, leverage-controlled acceptance, and conductance and mixing.

2.3 From exact metrics to implementable kernels

A common transition. The two exact analyses use the metric G as a reference geometry. An implementation instead constructs a positive-definite metric $\bar{G}(x; \omega_s)$, where the seed is suppressed when the construction is deterministic, and proposes

$$Y \sim \bar{q}_{x, \omega_s} := N\left(x, \frac{r^2}{d} \bar{G}(x; \omega_s)^{-1}\right).$$

If Y is feasible, the algorithm constructs the reverse metric $\bar{G}(Y; \omega'_s)$ and accepts with probability

$$1 \wedge \frac{e^{-f(Y)} \bar{q}_{Y, \omega'_s}(x)}{e^{-f(x)} \bar{q}_{x, \omega_s}(Y)}.$$

Thus the implemented chain does not estimate the acceptance probability of the exact walk; it applies the exact Metropolis filter to the metrics that were actually constructed.

The two implementations. For polytopes, a deterministic fixed-schedule iteration computes the Lewis weights to coordinatewise relative accuracy $O(1/d)$. Inflating the result by its worst possible downward error gives a padded metric $\bar{G}(x)$ with $\bar{G}(x) \succeq G(x)$. Since this metric is a deterministic measurable function of the state, the usual Metropolis–Hastings identity gives exact reversibility. For spectrahedra, we use the direct log-det construction when $n < s$ and a TensorSRHT approximation when $n \geq s$. In the sketched branch, the forward and reverse endpoint sketches are fresh and independent. After averaging over the two seeds, the accepted flow is

$$\mathbb{E}_{\omega_s, \omega'_s} \left[\min \left\{ e^{-f(x)} \bar{q}_{x, \omega_s}(y), e^{-f(y)} \bar{q}_{y, \omega'_s}(x) \right\} \right],$$

which is symmetric under interchanging the endpoints and their seeds. Hence the marginal two-seed kernel is exactly reversible.

Transfer from the reference walk. In both settings, either deterministically or on a pointwise sketch-success event,

$$G(x) \preceq \bar{G}(x; \omega_s) \preceq e^\theta G(x), \quad \theta = O(1/d).$$

The centered Gaussian KL formula gives $d_{\text{TV}}(q_x, \bar{q}_{x, \omega_s}) = O(\sqrt{d}\theta)$, while on the two-endpoint event from the exact acceptance proof, each conditional Gaussian log density changes by only $O(d\theta) = O(1)$. Therefore the acceptance lower tail and nearby-proposal overlap survive with different universal constants. The conductance and warm-start arguments depend only on this common mass and the reference geometry, so the mixing bounds transfer without differentiating the approximate metric. In the randomized case, failed seeds remain part of the transition kernel and the sketch guarantee is only pointwise in the current state; no union bound over the mixing horizon is needed.

Appendix A gives the formal kernels and perturbation arguments. It combines the mixing times with the per-step costs $\tilde{O}(nd^{\omega-1} + d^\omega)$ for polytopes and $\tilde{O}(n^\omega + n^2 d^{3\omega-5})$ for spectrahedra to obtain the exact-arithmetic bounds in Corollaries A.9 and A.16.

3 Preliminaries

3.1 Notation

Throughout,

$$p := \max\{4, 2(1 + \log n)\}, \quad \beta := \frac{1}{2} - \frac{1}{p}, \quad c_p := 1 - \frac{2}{p} = 2\beta.$$

Note $\beta \in [\frac{1}{4}, \frac{1}{2})$ and $n^{2/p} \leq e$. The symbols C, c denote universal constants and C_p a quantity bounded by a fixed polynomial in p ; all may change from line to line. Polynomial factors in p and logarithmic factors in n, d are hidden in $\tilde{O}(\cdot)$. For a symmetric matrix M , $\|M\|_{\text{op}}$ and $\|M\|_F$ are the spectral and Frobenius norms; $\mathbb{S}^d$ is the space of symmetric $d \times d$ matrices with the Frobenius inner product. For a positive definite M we write $\|v\|_M^2 = v^\top M v$, and for a positive vector w we write $\|v\|_W^2 = \sum_i w_i v_i^2$ with $W = \text{Diag}(w)$. For vectors, $\circ$ denotes the entrywise product and $v^{\circ 2} = v \circ v$; for matrices, $\circ$ is the Hadamard product.

3.2 Polytope, slacks, and Lewis weights

Let $A \in \mathbb{R}^{n \times d}$ have full column rank and no zero rows, let $b \in \mathbb{R}^n$, and set

$$K := \{z \in \mathbb{R}^d : Az \geq b\}, \quad K^\circ := \{z \in \mathbb{R}^d : Az > b\}.$$

We assume K is bounded and full-dimensional and, after a translation, $K \subseteq B_2(0, R)$. For $x \in K^\circ$ define the positive slacks and the slack-rescaled constraint matrix

$$s_x := Ax - b, \quad S_x := \text{Diag}(s_x), \quad A_x := S_x^{-1}A.$$

This fixes the sign convention: feasible slacks are $Ax - b > 0$.

For a full-column-rank matrix M let $\text{lev}(M) := \text{diag}(M(M^\top M)^{-1}M^\top)$ denote its leverage scores. The ℓ_p -Lewis weights of A_x are the unique positive vector w_x , with $W_x := \text{Diag}(w_x)$, satisfying the fixed-point identity

$$w_x = \text{lev}(W_x^{1/2-1/p}A_x). \quad (1)$$

Existence and uniqueness for $p \geq 2$ go back to Lewis [Lew78]; see also Cohen–Peng [CP15]. We use one further external Lewis-weight fact: under a smooth positive diagonal rescaling of the rows, the unique weights vary smoothly and their row-rescaling Jacobian is the one quoted in Lemma 3.1 from [LS19]. This is the only imported Lewis-calculus identity. Since the associated leverage-score matrix is a rank- d orthogonal projection,

$$0 < w_{x,i} \leq 1, \quad \sum_i w_{x,i} = d. \quad (2)$$

3.3 First-derivative Lewis calculus

The following objects appear throughout. At a point $z \in K^\circ$, set

$$B_z = W_z^{1/2}A_z, \quad V_z = W_z^\beta A_z H(z)^{-1/2}, \quad \hat{B}_z = B_z H(z)^{-1/2} = W_z^{1/p}V_z.$$

Then $V_z^\top V_z = I$, and by Eq. (1) the squared norm of the i -th row $v_{z,i}$ of V_z is $w_{z,i}$. Define the row map $\Phi_z : \mathbb{S}^d \rightarrow \mathbb{R}^n$ and its associated operators by

$$(\Phi_z X)_i = \frac{v_{z,i}^\top X v_{z,i}}{\sqrt{w_{z,i}}}, \quad \bar{\Lambda}_z := I - \Phi_z \Phi_z^*, \quad \mathcal{R}_z := (I - c_p \bar{\Lambda}_z)^{-1}, \quad N_z := 2\bar{\Lambda}_z \mathcal{R}_z. \quad (3)$$

Here $\Phi_z \Phi_z^*$ is an $n \times n$ matrix depending only on the Gram matrix of the rows $v_{z,i}$, hence independent of the choice of square root $H(z)^{-1/2}$. Since $2\beta = c_p$,

$$\mathcal{R}_z = I + \beta N_z. \quad (4)$$

Two elementary norm facts will be used repeatedly. First, applying Cauchy–Schwarz in each row and using $V_z^\top V_z = I$ gives the rowwise contraction

$$\sum_i \frac{(v_{z,i}^\top X v_{z,i})^2}{w_{z,i}} \leq \sum_i v_{z,i}^\top X^2 v_{z,i} = \|X\|_F^2 \quad (X \in \mathbb{S}^d), \quad (5)$$

and hence $\|\Phi_z\|_{\text{op}} \leq 1$, and the i -th component functional has norm $\|(\Phi_z)_i\|_F = \sqrt{w_{z,i}}$. Second, $A_z^\top W_z A_z \preceq H(z)$ (because $w \leq w^{1-2/p}$ for $w \leq 1$) gives $\|\hat{B}_z\|_{\text{op}} \leq 1$, and the rows satisfy $\|(\hat{B}_z)_i\| = w_{z,i}^{1/2+1/p}$.

Lemma 3.1 (Lewis derivative formula; [LS19, Lemma 24]). *The map $z \mapsto \log w_z$ is smooth on K° with*

$$D \log w_z[v] = -W_z^{-1/2} N_z B_z v. \quad (6)$$

Moreover $0 \preceq \bar{\Lambda}_z \preceq I$, and consequently $\|\mathcal{R}_z\|_{\text{op}} \leq p/2$ and $0 \preceq N_z \preceq pI$.

Proof. The only imported step is the following row-rescaling Jacobian; we spell out its spatial specialization because that fixes both the sign and the factor two. Let $v = s_z^{-1}$ be the vector that rescales the rows of the fixed constraint matrix and let $\Upsilon = \text{Diag}(v)$. In the notation of [LS19, Lemma 24], the Jacobian of the Lewis weights under this rescaling is

$$J_w(v) = 2W(W - c_p \Lambda)^{-1} \Lambda \Upsilon^{-1}, \quad \Lambda := W^{1/2} \bar{\Lambda} W^{1/2}.$$

Along a spatial direction h , differentiation of $v_i = s_{z,i}^{-1}$ gives $D \log v[h] = -A_z h$. Therefore the chain rule, together with $Dv = \Upsilon D \log v$ and the displayed definition of Λ , gives

$$\begin{aligned} D \log w_z[h] &= W^{-1} J_w(v) Dv[h] \\ &= 2(W - c_p \Lambda)^{-1} \Lambda D \log v[h] \\ &= -2W^{-1/2} (I - c_p \bar{\Lambda})^{-1} \bar{\Lambda} W^{1/2} A_z h \\ &= -W^{-1/2} N B_z h. \end{aligned}$$

In the last step, $\bar{\Lambda}$ commutes with $(I - c_p \bar{\Lambda})^{-1}$ because the latter is a function of $\bar{\Lambda}$.

By Eq. (5), $0 \preceq \Phi \Phi^* \preceq I$, and therefore $0 \preceq \bar{\Lambda} = I - \Phi \Phi^* \preceq I$. Hence the eigenvalues of $\mathcal{R} = (I - c_p \bar{\Lambda})^{-1}$ lie in $[1, (1 - c_p)^{-1}] = [1, p/2]$, and those of $N = 2\bar{\Lambda} \mathcal{R}$ lie in $[0, p]$. $\square$

3.4 Analytic, metric, and geometric inputs

The following is the only concentration theorem used in the polytope proof. We treat it as a non-elementary black box and state the exact form needed below [Led01].

Lemma 3.2 (Gaussian Lipschitz concentration; black box). *Let $\xi \sim N(0, I_m)$ and let $F : \mathbb{R}^m \rightarrow \mathbb{R}$ be L_F -Lipschitz with $L_F > 0$. For every $t \geq 0$,*

$$\begin{aligned} \Pr[F(\xi) - \mathbb{E}[F(\xi)] \geq t] &\leq \exp\left(-\frac{t^2}{2L_F^2}\right), \\ \Pr[|F(\xi) - \mathbb{E}[F(\xi)]| \geq t] &\leq 2 \exp\left(-\frac{t^2}{2L_F^2}\right). \end{aligned}$$

In particular, if Z is centered Gaussian with $\text{Var}(Z) \leq \sigma^2$, then $\Pr[|Z| \geq t] \leq 2e^{-t^2/(2\sigma^2)}$.

Definition 3.3 (Strong self-concordance; [LLV20]). *A matrix function $M : K^\circ \rightarrow \mathbb{S}_{>0}^d$ is strongly self-concordant (SSC) if for all $x \in K^\circ$ and $h \in \mathbb{R}^d$,*

$$\|M(x)^{-1/2} D M(x)[h] M(x)^{-1/2}\|_F \leq 2\|h\|_{M(x)}.$$

Lemma 3.4 (SSC of the scaled LS metric). *The metric $2(1 + p^2)H$ is strongly self-concordant. Consequently, κH is strongly self-concordant for every $\kappa \geq 2(1 + p^2)$.*

Proof. Fix $z \in K^\circ$ and a direction h , and abbreviate $A = A_z$, $W = W_z$, and $H = H(z)$. Put

$$\zeta := Ah, \quad \dot{\ell} := D \log w_z[h], \quad \alpha := -2\zeta + c_p \dot{\ell}.$$

Since $DA_z[h] = -\text{Diag}(\zeta)A_z$ and $D(W^{c_p})[h] = c_p W^{c_p} \text{Diag}(\dot{\ell})$, differentiation of $H = A^\top W^{c_p} A$ gives

$$H^{-1/2}DH[h]H^{-1/2} = V_z^\top \text{Diag}(\alpha)V_z.$$

The columns of V_z are orthonormal and its i -th row has squared norm $w_{z,i}$, so $\|V_z^\top\|_{\text{op}} = 1$. Expanding the Frobenius norm row by row and using $\alpha = -2\zeta + c_p \dot{\ell}$, $c_p \leq 1$, and $(u+v)^2 \leq 2u^2 + 2v^2$ gives

$$\|H^{-1/2}DH[h]H^{-1/2}\|_F^2 \leq \|\text{Diag}(\alpha)V_z\|_F^2 = \sum_i w_{z,i} \alpha_i^2 \leq 8(\|\zeta\|_W^2 + \|\dot{\ell}\|_W^2).$$

For the first term, $\zeta = Ah$ and $W \preceq W^{c_p}$ imply

$$\|\zeta\|_W^2 = h^\top A^\top W A h \leq \|h\|_H^2,$$

while Eq. (6), $\|N_z\|_{\text{op}} \leq p$, and $W \preceq W^{c_p}$ give

$$\|\dot{\ell}\|_W^2 = \|N_z B_z h\|^2 \leq p^2 \|B_z h\|^2 \leq p^2 \|h\|_H^2.$$

Thus the square of the left side in the SSC inequality for H is at most $8(1+p^2)\|h\|_H^2$. Scaling H by a constant leaves that left side unchanged, whereas $4\|h\|_{2(1+p^2)H}^2 = 8(1+p^2)\|h\|_H^2$. This proves the first claim; any larger scaling only increases the right side of the SSC inequality. $\square$

Lemma 3.5 (Self-concordant distortion). *Let M be a C^1 matrix function on K° satisfying*

$$\|M(x)^{-1/2}DM(x)[h]M(x)^{-1/2}\|_{\text{op}} \leq 2\|h\|_{M(x)}.$$

Then for $x, z \in K^\circ$ with $\Delta := \|x - z\|_{M(x)} < 1$,

$$(1 - \Delta)^2 M(x) \preceq M(z), \\ M(z) \preceq (1 - \Delta)^{-2} M(x).$$

Proof. Put $u = z - x$, $\gamma(t) = x + tu$, and $\phi(t) = \|u\|_{M(\gamma(t))}$. The claim is trivial if $u = 0$. Otherwise, differentiating $\phi(t)^2 = u^\top M(\gamma(t))u$ and applying the derivative hypothesis gives

$$|\phi'(t)| = \frac{|u^\top DM(\gamma(t))[u]u|}{2\phi(t)} \leq \phi(t)^2,$$

and hence $(\phi^{-1})'(t) \geq -1$. Since $\phi(0) = \Delta$,

$$\phi(t) \leq \frac{\Delta}{1 - t\Delta} \quad (0 \leq t \leq 1). \tag{7}$$

For a fixed v , let $\psi(t) = v^\top M(\gamma(t))v$. The same derivative hypothesis and Eq. (7) give

$$|\frac{d}{dt} \log \psi(t)| \leq 2\phi(t) \leq \frac{2\Delta}{1 - t\Delta}.$$

Integration from 0 to 1 yields $|\log(\psi(1)/\psi(0))| \leq -2\log(1 - \Delta)$. Exponentiating and using that v was arbitrary proves both Loewner inequalities. $\square$

Lemma 3.6 (Cross-ratio isoperimetry; [LV07, Theorem 2.5]). *Let π be a log-concave probability measure on a convex body K and let $K = S_1 \dot{\cup} S_2 \dot{\cup} S_3$ be a measurable partition such that $\sigma(x, z) \geq \sigma_0$ for all $x \in S_1, z \in S_2$, where σ is the chord cross ratio of Definition 4.31. Then*

$$\pi(S_3) \geq c_{\text{iso}} \min\{\sigma_0, 1\} \pi(S_1) \pi(S_2)$$

for a universal $c_{\text{iso}} > 0$.

This localization/isoperimetry theorem is a genuinely non-elementary input. We apply it only after checking that K is convex and that $e^{-f} \mathbf{1}_K$ is log-concave because f is convex. Its role is exactly to turn chord cross-ratio separation between two sets into stationary mass in the intervening set.

We henceforth replace c_{iso} by $\min\{c_{\text{iso}}, 1\}$, so that $0 < c_{\text{iso}} \leq 1$.

Lemma 3.7 (Cheeger's inequality for lazy reversible chains; [LS88, Theorem 2.1]). *A $\frac{1}{2}$ -lazy reversible chain with conductance Φ has spectral gap $\text{gap}(P) \geq c_{\text{Ch}} \Phi^2$ for a universal $c_{\text{Ch}} > 0$.*

This is the other deep mixing input. Exact Metropolis detailed balance gives reversibility, and the explicit half-lazy step puts the spectrum in $[0, 1]$. The lemma converts the conductance bound proved later into an $L^2(\pi)$ spectral gap; the ensuing warm-start contraction and L^2 -to-total-variation step are derived directly in Section 4.6. To match conventions precisely, Lawler–Sokal use

$$k := \inf_A \frac{\mathcal{Q}(A, A^c)}{\pi(A)\pi(A^c)}, \quad \mathcal{Q}(A, A^c) := \int_A P_x(A^c) \pi(dx)$$

(the stationary flow), whereas Lemma 4.33 uses $\Phi := \inf_{\pi(A) \leq 1/2} \mathcal{Q}(A, A^c)/\pi(A)$. Reversibility lets us replace A by its smaller side; then $\Phi \leq k \leq 2\Phi$ because $1 \leq 1/\pi(A^c) \leq 2$. Thus their bound in terms of k gives the displayed bound in terms of Φ up to a universal constant.

3.5 Setting and the walk

Let $A_0, A_1, \dots, A_d \in \mathbb{S}^n$ and

$$K := \{x \in \mathbb{R}^d : S(x) \succeq 0\}, \quad S(x) := A_0 + \sum_{i=1}^d x_i A_i,$$

assumed bounded and full-dimensional with, after a translation, $K \subseteq B_2(0, R)$. We also assume Slater strict feasibility: $S(x_\circ) \succ 0$ for some $x_\circ \in K$; equivalently, first remove any common kernel and work with the positive-definite block on the minimal face. Write K° for the interior, on which $S(x) \succ 0$. The maps $x \mapsto S(x)^{-1}$ and everything built from it are smooth on K° . The target is $\pi \propto e^{-f} \mathbf{1}_K$ with f convex and L -Lipschitz on K .

The *log-det metric* is the Hessian of the barrier $\phi(x) = -\log \det S(x)$:

$$H(x)_{ij} = \partial_i \partial_j \phi(x) = \text{tr}[S(x)^{-1} A_i S(x)^{-1} A_j] \succ 0. \quad (8)$$

Indeed, for $v \in \mathbb{R}^d$, $v^\top H(x) v = \|S(x)^{-1/2} (\sum_i v_i A_i) S(x)^{-1/2}\|_F^2$. If this vanished for nonzero v , then $S(x + tv) = S(x)$ for every t , contradicting boundedness of K ; hence $H(x) \succ 0$. For a scale $\kappa \geq 1$ (fixed in Eq. (74) below; $\kappa = \tilde{O}(\psi_*^+)$ in the notation of Definition 3.9) set

$$G(x) := \kappa H(x) + L^2 I.$$

Definition 3.8 (Exact-metric log-det walk). *At $x \in K^\circ$, propose $Y \sim q_x := N(x, \frac{r^2}{d} G(x)^{-1})$ with $r > 0$ a universal constant, and accept with the exact Metropolis–Hastings probability $1 \wedge R_x(Y)$, $R_x(y) := e^{-f(y)} q_y(x) / e^{-f(x)} q_x(y)$. Proposals outside K are rejected, and the chain is made $\frac{1}{2}$ -lazy.*

The chain is reversible with respect to π . An initial law μ_0 is w -warm if $d\mu_0/d\pi \leq w$.

3.6 The matrix leverage operator

Definition 3.9 (Leverage operator). *At $x \in K^\circ$, let $\bar{A}_1(x), \dots, \bar{A}_d(x) \in \mathbb{S}^n$ be the slack-congruenced constraint directions $S(x)^{-1/2} A_i S(x)^{-1/2}$ after orthonormalization in the Frobenius inner product (equivalently, in the metric $H(x)$). Define*

$$\Psi(x) := \sum_{a=1}^d \bar{A}_a(x)^2, \quad \psi(x) := \|\Psi(x)\|_{\text{op}}, \quad \psi_*^+ := \max(1, \sup_{x \in K^\circ} \psi(x)).$$

Lemma 3.10 (Basic properties). *$\Psi(x)$ is well-defined (independent of the choice of orthonormal frame), $\text{tr}[\Psi(x)] = d$, and $d/n \leq \psi(x) \leq d$. If all A_i are diagonal (the polytope case), Ψ is the diagonal matrix of leverage scores of the slack-rescaled constraint matrix, so $\psi \leq 1$.*

Proof. If $\{\bar{A}_a\}$ and $\{\bar{A}'_a\}$ are two Frobenius-orthonormal bases of the same subspace of $\mathbb{S}^n$, then $\bar{A}'_a = \sum_b O_{ab} \bar{A}_b$ for an orthogonal O , and $\sum_a \bar{A}'_a \otimes \bar{A}'_a = \sum_a \bar{A}_a \otimes \bar{A}_a$; applying this bilinear identity to the product map gives $\sum_a (\bar{A}'_a)^2 = \sum_a \bar{A}_a^2$. Next $\text{tr}[\Psi] = \sum_a \|\bar{A}_a\|_F^2 = d$ by orthonormality, whence $\psi \geq \text{tr}[\Psi]/n = d/n$; and $\psi \leq \sum_a \|\bar{A}_a\|_{\text{op}}^2 = \sum_a \|\bar{A}_a\|_F^2 = d$. In the diagonal case, writing $\bar{a}_k \in \mathbb{R}^d$ for the vector of k -th diagonal entries of the $\bar{A}_a$, orthonormality makes $(\Psi)_{kk} = \|\bar{a}_k\|^2$ the k -th leverage score of the orthonormalized constraint matrix, and leverage scores are at most 1. $\square$

3.7 Toolbox

We collect the elementary and imported facts used repeatedly. Recall $\|\cdot\|_{S_1}$ is the trace norm; $|\text{tr}[XY]| \leq \|X\|_F \|Y\|_F$ and $\|XY\|_{S_1} \leq \|X\|_F \|Y\|_F$.

Lemma 3.11 (PSD trace facts). *Let $X, Y, M, T \in \mathbb{S}^n$ with $X, Y, M \succeq 0$ and $T \succeq 0$.*

(i) $\text{tr}[XY] \geq 0$; consequently $0 \preceq X \preceq Y$ implies $\text{tr}[X^2] \leq \text{tr}[XY] \leq \text{tr}[Y^2]$.

(ii) $\text{tr}[TM] \leq \|T\|_{\text{op}} \text{tr}[M]$ and, for any $B \in \mathbb{S}^n$, $BTB \preceq \|T\|_{\text{op}} B^2$.

(iii) $\|T^{1/2}MT^{1/2}\|_F \leq \|T\|_{\text{op}} \|M\|_F$.

Proof. (i) $\text{tr}[XY] = \|X^{1/2}Y^{1/2}\|_F^2 \geq 0$; then $\text{tr}[X(Y - X)] \geq 0$ and $\text{tr}[(Y - X)Y] \geq 0$ give the chain. (ii) $\text{tr}[TM] = \text{tr}[M^{1/2}TM^{1/2}] \leq \|T\|_{\text{op}} \text{tr}[M]$; and $u^\top BTBu \leq \|T\|_{\text{op}} \|Bu\|^2$ for every u . (iii) $\|T^{1/2}MT^{1/2}\|_F \leq \|T^{1/2}\|_{\text{op}}^2 \|M\|_F$. $\square$

Lemma 3.12 (Twice-integrated Gaussian Stein identity). *Let $h \sim N(0, Q)$ on $\mathbb{R}^d$ with $0 \preceq Q \preceq I$, and let $F : \mathbb{R}^d \rightarrow \mathbb{R}$ be $W^{2,\infty}$ with compact support. Then for all indices a, b ,*

$$\mathbb{E}[(h_a h_b - Q_{ab})F(h)] = \sum_{c,e} Q_{ac} Q_{be} \mathbb{E}[\partial_c \partial_e F(h)]. \quad (9)$$

Proof. Write $h = Q^{1/2}\xi$ with ξ standard. For any smooth G whose value and first two derivatives are bounded, integrating against the Gaussian density by parts twice gives $\mathbb{E}[(\xi_\alpha \xi_\beta - \delta_{\alpha\beta})G(\xi)] = \mathbb{E}[\partial_\alpha \partial_\beta G(\xi)]$. Apply this to $G = F \circ Q^{1/2}$ and contract. This G has bounded derivatives even when Q is singular, which is all the integration-by-parts argument requires:

$$\begin{aligned} h_a h_b - Q_{ab} &= \sum_{\alpha, \beta} (Q^{1/2})_{a\alpha} (Q^{1/2})_{b\beta} (\xi_\alpha \xi_\beta - \delta_{\alpha\beta}), \\ \partial_\alpha \partial_\beta (F \circ Q^{1/2}) &= \sum_{c,e} (Q^{1/2})_{c\alpha} (Q^{1/2})_{e\beta} (\partial_c \partial_e F) \circ Q^{1/2}. \end{aligned}$$

Contracting these two identities gives Eq. (9). For $F \in W^{2,\infty}$ with compact support, mollify and pass to the limit by dominated convergence. $\square$

Lemma 3.13 (Dimension-free cutoff). *Let $\mathcal{K} \subset \mathbb{R}^d$ be closed and convex. There is $\chi : \mathbb{R}^d \rightarrow [0, 1]$ with $\chi \equiv 1$ on $\mathcal{K}$, χ supported in the closed Euclidean 1-neighborhood $\mathcal{K}^+$, and $\|\nabla \chi\| \leq 12$, $\|D^2 \chi\|_{\text{op}} \leq C_0$ a.e., for a universal C_0 .*

Proof. $\phi(h) = \text{dist}(h, \mathcal{K})^2$ is convex and $C^{1,1}$ with $\nabla \phi = 2(h - \Pi_{\mathcal{K}} h)$ (2-Lipschitz), so $0 \preceq D^2 \phi \preceq 2I$ a.e. Take $\chi = \theta \circ \phi$ for a fixed smooth decreasing $\theta : [0, \infty) \rightarrow [0, 1]$ with $\theta \equiv 1$ on $[0, \frac{1}{9}]$, $\theta \equiv 0$ on $[\frac{4}{9}, \infty)$, $|\theta'| \leq 9$, and $|\theta''| \leq C_\theta$ for a fixed universal constant C_θ . On the support of $\theta' \circ \phi$ we have $\|\nabla \phi\| \leq \frac{4}{3}$, so $\|\nabla \chi\| \leq 9 \cdot \frac{4}{3} = 12$ and $\|D^2 \chi\|_{\text{op}} \leq |\theta''| \|\nabla \phi\|^2 + |\theta'| \|D^2 \phi\|_{\text{op}} \leq C_0$, all dimension-free. $\square$

Lemma 3.14 (McShane extension). *A function Z that is Λ -Lipschitz on a nonempty set $\mathcal{K} \subseteq \mathbb{R}^d$ extends to a Λ -Lipschitz function $\tilde{Z}$ on $\mathbb{R}^d$, e.g. $\tilde{Z}(h) = \inf_{g \in \mathcal{K}} \{Z(g) + \Lambda \|h - g\|\}$.*

Proof. For $h \in \mathcal{K}$, the choice $g = h$ gives $\tilde{Z}(h) \leq Z(h)$, while Lipschitzness gives $Z(g) + \Lambda \|h - g\| \geq Z(h)$ for every $g \in \mathcal{K}$; hence $\tilde{Z} = Z$ on $\mathcal{K}$. The infimum is finite: after fixing $g_0 \in \mathcal{K}$, the same Lipschitz inequality and the triangle inequality give $Z(g) + \Lambda \|h - g\| \geq Z(g_0) - \Lambda \|h - g_0\|$. Finally, for any h, h' and every $g \in \mathcal{K}$,

$$Z(g) + \Lambda \|h - g\| \leq Z(g) + \Lambda \|h' - g\| + \Lambda \|h - h'\|.$$

Taking the infimum over g gives $\tilde{Z}(h) \leq \tilde{Z}(h') + \Lambda \|h - h'\|$; exchanging h, h' proves the claimed Lipschitz bound. $\square$

Lemma 3.15 (Gaussian Lipschitz concentration: non-elementary black box). *Let $\xi \sim N(0, I_N)$ and let $F : \mathbb{R}^N \rightarrow \mathbb{R}$ be Λ -Lipschitz. Then, for every $t \geq 0$,*

$$\Pr[F(\xi) - \mathbb{E}[F(\xi)] \geq t] \leq e^{-t^2/(2\Lambda^2)}, \quad \Pr[|F(\xi) - \mathbb{E}[F(\xi)]| \geq t] \leq 2e^{-t^2/(2\Lambda^2)}.$$

The assertion for $\Lambda = 0$ has the evident interpretation.

This is the standard Gaussian isoperimetric concentration theorem. We use it as an imported, non-elementary scalar black box; all Lipschitz constants and all changes of covariance in its applications are verified explicitly below.

Lemma 3.16 (Gaussian matrix series: the noncommutative concentration black box; [Tro15, Theorem 4.1.1]). *Let $B_1, \dots, B_N \in \mathbb{S}^n$ be fixed and ξ standard Gaussian in $\mathbb{R}^N$. Then, with $\sigma^2 := \|\sum_i B_i^2\|_{\text{op}}$,*

$$\Pr[\|\sum_i \xi_i B_i\|_{\text{op}} \geq t] \leq 2n e^{-t^2/2\sigma^2} \quad (t \geq 0).$$

This is the only noncommutative concentration theorem used in the log-det analysis. Every application first writes the random matrix as a self-adjoint series $\sum_i \xi_i B_i$ and then bounds the deterministic variance operator $\sum_i B_i^2$ in Loewner order. All remaining noncommutative steps are finite-dimensional trace, congruence, and resolvent calculations proved in the text.

Lemma 3.17 (Metric distortion). *Let $M : K^\circ \rightarrow \mathbb{S}_{>0}^d$ be continuously differentiable and satisfy $\|M(x)^{-1/2} DM(x)[u] M(x)^{-1/2}\|_{\text{op}} \leq 2\|u\|_{M(x)}$ for all x, u . Then for $x, z \in K^\circ$ with $\Delta := \|x - z\|_{M(x)} < 1$, $(1 - \Delta)^2 M(x) \preceq M(z) \preceq (1 - \Delta)^{-2} M(x)$.*

Proof. The case $x = z$ is immediate. Otherwise put $v = z - x$, $x_t = x + tv$, $M_t = M(x_t)$, and $a(t) = \|v\|_{M_t} > 0$. Convexity of K° keeps the whole segment in the domain. Differentiating $a(t)^2 = v^\top M_t v$ and testing the assumed differential inequality on v gives

$$|a'(t)| = \frac{|v^\top DM_t[v]v|}{2a(t)} \leq a(t)^2.$$

Since $a(0) = \Delta$, integration of $(1/a)' = -a'/a^2 \geq -1$ yields $a(t) \leq \Delta/(1 - t\Delta)$ for $0 \leq t \leq 1$.

Now fix $w \neq 0$ and write $b(t) = w^\top M_t w$. The differential inequality and the preceding bound on $a(t)$ give

$$\left| \frac{d}{dt} \log b(t) \right| = \frac{|w^\top DM_t[w]w|}{w^\top M_t w} \leq 2a(t) \leq \frac{2\Delta}{1 - t\Delta}.$$

Thus $|\log(b(1)/b(0))| \leq -2\log(1 - \Delta)$. Exponentiating and then allowing arbitrary w gives exactly $(1 - \Delta)^2 M(x) \preceq M(z) \preceq (1 - \Delta)^{-2} M(x)$. $\square$

Lemma 3.18 (Cross-ratio isoperimetry; [LV07, Theorem 2.5]). *Let π be log-concave on a convex body K and $K = S_1 \dot{\cup} S_2 \dot{\cup} S_3$ measurable with $\sigma(x, z) \geq \sigma_0$ for all $x \in S_1, z \in S_2$, where σ is the chord cross ratio (Definition 5.21). Then $\pi(S_3) \geq c_{\text{iso}} \min\{\sigma_0, 1\} \pi(S_1)\pi(S_2)$ for a universal $c_{\text{iso}} > 0$.*

This imported localization/isoperimetry theorem is used only after two hypotheses have been checked: K is convex and the density $e^{-f} \mathbf{1}_K$ is log-concave because f is convex. Its role is to turn a lower bound on chord cross-ratio separation between two sets into stationary mass in the intervening set.

We henceforth replace c_{iso} by $\min\{c_{\text{iso}}, 1\}$, so that $0 < c_{\text{iso}} \leq 1$.

Lemma 3.19 (Cheeger for lazy reversible chains; [LS88, Theorem 2.1]). *A $\frac{1}{2}$ -lazy reversible chain with conductance Φ has spectral gap at least $c_{\text{Ch}} \Phi^2$, $c_{\text{Ch}} > 0$ universal.*

This is the second imported mixing input. It applies because the exact Metropolis–Hastings kernel is reversible with stationary law π and is made $\frac{1}{2}$ -lazy. Its sole role is to convert the conductance bound proved in Lemma 5.24 into an $L^2(\pi)$ spectral gap; the subsequent warm-start estimate is written out explicitly. The source uses $k = \inf_A \mathcal{Q}(A, A^c)/[\pi(A)\pi(A^c)]$, while this paper uses $\Phi = \inf_{\pi(A) \leq 1/2} \mathcal{Q}(A, A^c)/\pi(A)$. Reversibility permits choosing the smaller side, where $\Phi \leq k \leq 2\Phi$ because $1 \leq 1/\pi(A^c) \leq 2$. Hence the cited theorem yields the displayed $c_{\text{Ch}} \Phi^2$ form. Laziness is then what removes negative spectrum in the later L^2 contraction.

4 Proof of Theorem 1.1

4.1 The metric, the walk, and the main theorem

The unscaled *Lee–Sidford metric* is

$$H(x) = A_x^\top W_x^{1-2/p} A_x \succ 0.$$

Let κ be a sufficiently large fixed polynomial in p and $\log(nd)$ —the precise requirements are Eq. (21), Eq. (47), and $\kappa \geq 2(1 + p^2)$ from Lemma 3.4—and set

$$G(x) = \kappa H(x) + L^2 I.$$

Definition 4.1 (Exact-metric LS walk). *At $x \in K^\circ$, propose*

$$Y \sim q_x := N(x, \frac{r^2}{d} G(x)^{-1}),$$

where $r > 0$ is a universal constant, and accept with the exact Metropolis–Hastings probability

$$\alpha_x(y) = 1 \wedge R_x(y), \quad R_x(y) := \frac{e^{-f(y)} q_y(x)}{e^{-f(x)} q_x(y)}.$$

Proposals outside K are rejected, and the chain is made $\frac{1}{2}$ -lazy.

The walk is reversible with respect to $\pi \propto e^{-f} \mathbf{1}_K$: on $K^\circ \times K^\circ$ the accepted-move density satisfies

$$e^{-f(x)} q_x(y) (1 \wedge R_x(y)) = \min\{e^{-f(x)} q_x(y), e^{-f(y)} q_y(x)\},$$

which is symmetric in x, y ; rejection and laziness only add diagonal mass. An initial law μ_0 is w -warm if $d\mu_0/d\pi \leq w$.

Theorem 4.2 (Polytopes; formal version of Theorem 1.1). *Let $K \subseteq B_2(0, R)$ be a bounded full-dimensional polytope in $\mathbb{R}^d$ given by n linear inequalities, let f be convex and L -Lipschitz on K , and let $\pi \propto e^{-f} \mathbf{1}_K$. Let H_{LS} denote the Lee–Sidford metric, let $r > 0$ be a universal constant, and choose κ as in Eq. (47). Consider the $\frac{1}{2}$ -lazy Metropolis walk with exact Lewis weights, exact acceptance ratio, and proposal covariance*

$$\frac{r^2}{d} (\kappa H_{\text{LS}}(x) + L^2 I)^{-1}.$$

Then, for $w \geq 1$ and $0 < \delta < 1$, from a w -warm start the walk reaches total-variation distance δ from π in

$$\tilde{O}((d^2 + dL^2 R^2) \log(w/\delta))$$

steps.

Proof. Put $f_0 = d\mu_0/d\pi$. Since μ_0 is w -warm, $0 \leq f_0 \leq w$ and $\int f_0 d\pi = 1$. Expanding the square and using $f_0^2 \leq w f_0$ gives

$$\|f_0 - 1\|_{L^2(\pi)}^2 = \int f_0^2 d\pi - 1 \leq w - 1.$$

Write the lazy kernel as $P = (I + P_0)/2$, where P_0 is the underlying non-lazy Metropolis kernel. Since P_0 is a self-adjoint contraction on $L^2(\pi)$, its spectrum lies in $[-1, 1]$, and hence the spectrum of P lies in $[0, 1]$. Therefore Lemma 3.7 gives $\text{gap}(P) \geq c_{\text{Ch}} \Phi^2$, and the spectral theorem gives $\|P^t g\|_2 \leq (1 - \text{gap}(P))^t \|g\|_2$ for every mean-zero g . By Cauchy–Schwarz, the preceding warm-start bound, and $1 - u \leq e^{-u}$,

$$d_{\text{TV}}(\mu_0 P^t, \pi) \leq \frac{1}{2} \|P^t(f_0 - 1)\|_{L^2(\pi)} \leq \frac{1}{2} (1 - \text{gap}(P))^t \sqrt{w - 1} \leq \frac{1}{2} e^{-c_{\text{Ch}} t \Phi^2} \sqrt{w - 1}.$$

Solving for total-variation error δ and inserting Eq. (56),

$$t_{\text{mix}}(\delta) \leq C \frac{d(\bar{\nu}_{\text{LS}} + L^2 R^2)}{r^2} \log(w/\delta).$$

Since r is universal and $\bar{\nu}_{\text{LS}} = \tilde{O}(d)$ by Eq. (54), this is $\tilde{O}((d^2 + dL^2 R^2) \log(w/\delta))$. (The same conductance-to-mixing pipeline appears in [KV24, Lemma B.1 and Theorem 3.1].) $\square$

4.2 The complete proposal ratio and the exact cancellation

This section contains no estimates: it is exact algebra. We normalize coordinates at the current point, diagonalize the endpoint metric exactly, and exhibit the cancellation inside the complete Gaussian proposal ratio.

Fix a deterministic current point $x \in K^\circ$. Let $F_x = H_{\text{orig}}(x)^{-1/2}$, apply the linear change of variables $z_{\text{orig}} = F_x z$ to the entire inequality representation, and transform x , y , h , A , and A_x accordingly. Lewis weights are invariant under this substitution (leverage scores, and hence the fixed point Eq. (1), are unchanged by $A \mapsto AT$ for invertible T), while H and G transform as metrics. Throughout Sections 4.2, 4.3, 4.4 and 4.5 we suppress tildes on the transformed quantities; every slack identity below uses the transformed A in the transformed coordinates. Write

$$W := W_x, \quad w := w_x.$$

The baseline metric is then normalized:

$$H(x) = I, \quad U := W^\beta A_x, \quad U^\top U = I.$$

Write $u_i^\top$ for row i of U . The fixed point Eq. (1) gives

$$\|u_i\|^2 = w_i, \quad P := UU^\top, \quad P_{ii} = w_i, \quad \sum_i w_i = d.$$

After the same affine change, factor the (now constant) regularizer as

$$G(x) = \kappa(I + \mathcal{C}), \quad \mathcal{C} = \frac{L^2}{\kappa} F_x^\top F_x \succeq 0, \quad Q := (I + \mathcal{C})^{-1}, \quad 0 \preceq Q \preceq I.$$

Only $\mathcal{C} \succeq 0$ is used below; in particular all estimates are uniform in L and in the orientation of the regularizer. A proposal then has the representation

$$y = x + \eta h, \quad h \sim N(0, Q), \quad \eta = \frac{r}{\sqrt{\kappa d}} = \frac{\rho}{\sqrt{d}}, \quad \rho := \frac{r}{\sqrt{\kappa}}. \quad (10)$$

The role of κ is only to make ρ smaller than a sufficiently small inverse polylogarithm; it changes the LS symmetry parameter by a polylogarithmic factor only.

Lemma 4.3 (Endpoint identities). *Let $h \in \mathbb{R}^d$ be such that the segment $[x, y]$, $y = x + \eta h$, lies in K° , and put*

$$\zeta := A_x h, \quad a_i(h) := \beta \log \frac{w_{y,i}}{w_i} - \log(1 + \eta \zeta_i), \quad b_i := e^{2a_i} - 1. \quad (11)$$

Then

$$W_y^\beta A_y = \text{Diag}(e^a) U, \quad H(y) = U^\top \text{Diag}(e^{2a}) U = I + U^\top \text{Diag}(b) U. \quad (12)$$

Moreover, with

$$c_i := (u_i^\top h)^2, \quad \bar{w}_i := \mathbb{E}[c_i] = u_i^\top Q u_i,$$

one has the exact quadratic identity

$$h^\top (H(y) - I) h = c^\top b.$$

Proof. The relations $y = x + \eta h$, $A = S_x A_x$, and $\zeta = A_x h$ give

$$s_y = s_x + \eta A h = S_x(\mathbf{1} + \eta A_x h) = S_x(\mathbf{1} + \eta \zeta),$$

Hence $A_y = \text{Diag}(\mathbf{1} + \eta \zeta)^{-1} A_x$. Multiplying by $W_y^\beta = \text{Diag}((w_{y,i}/w_i)^\beta) W^\beta$ and comparing with Eq. (11) gives the first identity in Eq. (12); the second follows from $H(y) = (W_y^\beta A_y)^\top (W_y^\beta A_y)$ and $U^\top U = I$. Finally $h^\top U^\top \text{Diag}(b) U h = \sum_i b_i (u_i^\top h)^2 = c^\top b$. $\square$

No Taylor expansion of the endpoint metric has been made: the sign in the endpoint rescaling is explicit, and Eq. (12) is an identity.

Lemma 4.4 (Ledger identity). *In the setting of Lemma 4.3, let*

$$E := Q^{1/2}(H(y) - I)Q^{1/2}. \quad (13)$$

Then the complete Gaussian proposal log-ratio satisfies, exactly,

$$\begin{aligned} \Lambda_q(x, y) &:= \frac{1}{2} \log \frac{\det G(y)}{\det G(x)} - \frac{d}{2r^2} (y - x)^\top (G(y) - G(x))(y - x) \\ &= -\frac{1}{2} (c - \bar{w})^\top b + \frac{1}{2} \{\log \det(I + E) - \text{tr}[E]\}. \end{aligned} \quad (14)$$

Equivalently, $\log \frac{q_y(x)}{q_x(y)} = \Lambda_q(x, y)$.

Proof. In the normalized coordinates $G(z) = \kappa H(z) + \kappa \mathcal{C}$, so the regularizer part is constant in z and $G(y) - G(x) = \kappa(H(y) - I)$. We verify the determinant, trace, and quadratic terms separately.

(i) *Determinant part.* Since $Q = (I + \mathcal{C})^{-1}$, cancelling the common factor κ , applying congruence with $(I + \mathcal{C})^{-1/2}$ inside the determinant, and using Eq. (13) give

$$\log \frac{\det G(y)}{\det G(x)} = \log \frac{\det(H(y) + \mathcal{C})}{\det(I + \mathcal{C})} = \log \det(I + Q^{1/2}(H(y) - I)Q^{1/2}) = \log \det(I + E). \quad (15)$$

(ii) *Trace of E .* By Eqs. (13) and (12), cyclicity of the trace, expansion over the rows of U , and the definition of $\bar{w}$ give

$$\text{tr}[E] = \text{tr}[Q U^\top \text{Diag}(b) U] = \sum_i b_i u_i^\top Q u_i = \bar{w}^\top b. \quad (16)$$

(iii) *Quadratic part.* With $y - x = \eta h$ and $d\eta^2 \kappa / r^2 = 1$ from Eq. (10), the identity $G(y) - G(x) = \kappa(H(y) - I)$ and the quadratic identity in Lemma 4.3 give

$$\frac{d}{r^2} (y - x)^\top (G(y) - G(x))(y - x) = \frac{d\eta^2 \kappa}{r^2} h^\top (H(y) - I) h = c^\top b. \quad (17)$$

Combining Eq. (15), (16), and (17), and adding and subtracting $\frac{1}{2} \text{tr}[E]$, gives

$$\Lambda_q = \frac{1}{2} \log \det(I + E) - \frac{1}{2} c^\top b = \frac{1}{2} \{\log \det(I + E) - \text{tr}[E]\} + \frac{1}{2} (\bar{w}^\top b - c^\top b),$$

which is Eq. (14).

For the last claim, the Gaussian densities satisfy

$$q_z(\cdot) \propto \det G(z)^{1/2} \exp(-\frac{d}{2r^2} \|\cdot - z\|_{G(z)}^2).$$

Consequently, taking the logarithm of the density ratio gives

$$\log \frac{q_y(x)}{q_x(y)} = \frac{1}{2} \log \frac{\det G(y)}{\det G(x)} - \frac{d}{2r^2} (\|x - y\|_{G(y)}^2 - \|y - x\|_{G(x)}^2) = \Lambda_q(x, y).$$

The normalizing constants and any Jacobians of the affine substitution cancel in the ratio. $\square$

Remark 4.5 (Why the complete ratio). Eq. (14) is the key cancellation. The potentially difficult uncentered term $\bar{w}^\top b$ is *simultaneously* the linear part of the determinant and the mean of the random quadratic term. An analysis that certifies a lower tail for the determinant ratio and concentration for the quadratic form separately—in particular, any route through (regularized) average self-concordance—must estimate $\bar{w}^\top b$ on its own and therefore requires strictly stronger Lewis-weight inputs than are available at second order. After the cancellation, the first term of Eq. (14) is centered to leading order (its further reduction to the exactly centered $Z = (c - \bar{w})^\top a$ is Lemma 4.9), and the bracket is a finite determinant remainder (Lemma 4.11). The estimates of Sections 4.3, 4.4 and 4.5 bound both at constant scale.

4.3 Uniform local stability

The identities of Section 4.2 are exact but involve the endpoint weights w_y . This section constructs a convex high-probability tube of steps h on which all endpoint quantities—weights, slacks, metric—are uniformly comparable to their base-point values, with polylogarithmic constants, and then proves the two finite remainders that reduce the ledger identity Eq. (14) to the centered fluctuation $Z = (c - \bar{w})^\top a$. Throughout we work in the normalized coordinates of Section 4.2.

Let

$$k_i := \frac{u_i}{\sqrt{w_i}}, \quad \|k_i\| = 1.$$

For a fixed small failure probability ϵ_0 , choose the auxiliary tail level

$$\delta_0 := \epsilon_0^2 d^{-3},$$

define the deterministic radii

$$R_0 := \sqrt{d} + \sqrt{2 \log(2/\delta_0)}, \quad \tau_0 := \sqrt{2 \log(4n/\delta_0)},$$

and the convex good set

$$\mathcal{K}_0 := \{g \in \mathbb{R}^d : \|g\| \leq R_0, \max_i |k_i^\top g| \leq \tau_0\}.$$

Lemma 4.6. $\Pr[h \in \mathcal{K}_0] \geq 1 - \delta_0$ for $h \sim N(0, Q)$.

Proof. Write $h = Q^{1/2} \xi$ with ξ standard. Since $Q \preceq I$, each $k_i^\top h$ is Gaussian with variance $k_i^\top Q k_i \leq 1$, and $\mathbb{E}[\|h\|] \leq (\mathbb{E}[\|h\|^2])^{1/2} = (\text{tr}[Q])^{1/2} \leq \sqrt{d}$. The map $\xi \mapsto \|Q^{1/2} \xi\|$ is 1-Lipschitz, so Gaussian Lipschitz concentration (Lemma 3.2) gives $\Pr[\|h\| > R_0] \leq e^{-\log(2/\delta_0)} = \delta_0/2$, and a union bound using the centered-Gaussian tail in the same lemma gives $\Pr[\max_i |k_i^\top h| > \tau_0] \leq 2n e^{-\tau_0^2/2} = \delta_0/2$. $\square$

All estimates will be proved on the closed Euclidean one-neighborhood $\mathcal{K}_0^+$ of $\mathcal{K}_0$ (this slack is what the cutoff of Section 4.5 needs). Set

$$R_1 := R_0 + 1, \quad \tau_1 := \tau_0 + 1, \quad M_0 := \frac{R_1}{\sqrt{d}}.$$

Every $g \in \mathcal{K}_0^+$ obeys $\|g\| \leq R_1$ and $\max_i |k_i^\top g| \leq \tau_1$ (since $\|k_i\| = 1$). For $\delta_0 = \epsilon_0^2 d^{-3}$,

$$M_0 = O_{\epsilon_0}(1), \quad \tau_1^2 = O(1 + \log(nd/\epsilon_0)), \quad M_0 \geq 1. \quad (18)$$

The bootstrap requires a first-derivative bound with the *row-dependent* scaling $\sqrt{w_{z,i}}$, which we now record. It sharpens the aggregate bound $\|\mathcal{R}_z\|_{\text{op}} \leq p/2$ of Lemma 3.1 row by row; part (ii) is stated for a general coefficient because Section 4.4 needs it for $I + \frac{1}{2}N_z$ as well as for $\mathcal{R}_z = I + \beta N_z$.

Lemma 4.7 (Rowwise resolvent bound). *Let $z \in K^\circ$, recall $\widehat{B}_z = B_z H(z)^{-1/2}$, and set $L_p := \frac{p}{2}(1 + \frac{p}{2})$. Then for every row index i :*

$$(i) \quad \|e_i^\top \mathcal{R}_z \widehat{B}_z\| \leq L_p \sqrt{w_{z,i}};$$

$$(ii) \quad \text{for every } \gamma \in [0, 1], \|e_i^\top (I + \gamma N_z) \widehat{B}_z\| \leq L_p^+ \sqrt{w_{z,i}} \text{ with } L_p^+ := 1 + 2L_p + p.$$

Consequently, in arbitrary coordinates, for all $v \in \mathbb{R}^d$,

$$|e_i^\top (I + \beta N_z) B_z v| \leq L_p \sqrt{w_{z,i}} \|v\|_{H(z)}. \quad (19)$$

Proof. (i) Since $\mathcal{R}_z$ is a function of $\bar{\Lambda}_z$, it commutes with $\Phi_z \Phi_z^* = I - \bar{\Lambda}_z$, and from $\mathcal{R}_z(I - c_p \bar{\Lambda}_z) = I$ we get the resolvent identity

$$(1 - c_p) \mathcal{R}_z = I - c_p \Phi_z \Phi_z^* \mathcal{R}_z.$$

Multiply by $\widehat{B}_z$ on the right, take the i -th row, and use $(1 - c_p)^{-1} = p/2$, together with $\|(\Phi_z)_i\|_F = \sqrt{w_{z,i}}$, $\|\Phi_z\|_{\text{op}} \leq 1$, $\|\mathcal{R}_z\|_{\text{op}} \leq p/2$, and $\|\widehat{B}_z\|_{\text{op}} \leq 1$, to obtain

$$\|e_i^\top \mathcal{R}_z \widehat{B}_z\| \leq \frac{p}{2} (\|(\widehat{B}_z)_i\| + c_p \|e_i^\top \Phi_z\| \|\Phi_z^* \mathcal{R}_z \widehat{B}_z\|_{\text{op}}) \leq \frac{p}{2} (w_{z,i}^{1/2+1/p} + c_p \sqrt{w_{z,i}} \cdot \frac{p}{2}) \leq L_p \sqrt{w_{z,i}},$$

where the last inequality also uses $c_p \leq 1$, $w_{z,i}^{1/p} \leq 1$, and the definition of L_p .

(ii) Since $\bar{\Lambda}_z$ commutes with $\mathcal{R}_z$, $N_z = 2\bar{\Lambda}_z \mathcal{R}_z = 2(I - \Phi_z \Phi_z^*) \mathcal{R}_z = 2\mathcal{R}_z - 2\Phi_z \Phi_z^* \mathcal{R}_z$. Part (i) and the same operator-norm bounds give

$$\|e_i^\top N_z \widehat{B}_z\| \leq 2\|e_i^\top \mathcal{R}_z \widehat{B}_z\| + 2\|e_i^\top \Phi_z\| \|\Phi_z^* \mathcal{R}_z \widehat{B}_z\|_{\text{op}} \leq (2L_p + p) \sqrt{w_{z,i}}.$$

Consequently, $\|e_i^\top (I + \gamma N_z) \widehat{B}_z\| \leq \|(\widehat{B}_z)_i\| + \gamma \|e_i^\top N_z \widehat{B}_z\| \leq (1 + 2L_p + p) \sqrt{w_{z,i}}$. Finally, Eq. (19) follows from (i) via $(I + \beta N_z) B_z v = \mathcal{R}_z \widehat{B}_z H(z)^{1/2} v$ and Eq. (4). $\square$

Differentiating Eq. (11) in h , including the factor η from $y = x + \eta h$, gives the exact Jacobian: for $y = x + \eta g$ feasible,

$$D_h a(g)[v] = -\eta W_y^{-1/2} (I + \beta N_y) B_y v. \quad (20)$$

(The two contributions are $\beta D \log w_y[\eta v]$ from Lemma 3.1 and $-\eta(A_y v)_i = -\eta(W_y^{-1/2} B_y v)_i$ from the slack factor; they combine into Eq. (20) by Eq. (4).)

We make the bootstrap quantitative. Enlarge κ , if necessary, so that

$$\vartheta := L_p M_0 \rho \leq \frac{1}{64}, \quad \eta \tau_1 \leq \frac{1}{64}. \quad (21)$$

This still requires only a polylogarithmic κ , by Eq. (18).

Lemma 4.8 (Stability on the tube). *Let $g \in \mathcal{K}_0^+$ and $0 \leq t \leq 1$. Then the entire segment $y_t = x + t\eta g$ lies in K° , and:*

$$\|a(tg)\|_\infty \leq 2L_p M_0 \rho \leq \frac{1}{32}, \quad \|D_h a(tg)\|_{2 \rightarrow \infty} \leq e^{1/32} L_p \eta, \quad (22)$$

$$\|a(tg)\|_W \leq e^{5/32} \frac{p}{2} M_0 \rho, \quad \|W^{1/2} D_h a(tg)\|_{\text{op}} \leq e^{5/32} \frac{p}{2} \eta. \quad (23)$$

Moreover

$$e^{-1/4} W \preceq W_{y_t} \preceq e^{1/4} W, \quad e^{-2\|a\|_\infty} I \preceq H(y_t) \preceq e^{2\|a\|_\infty} I, \quad (24)$$

and the slack factors satisfy

$$|\zeta_i| = w_i^{1/p} |k_i^\top g| \leq \tau_1, \quad |\eta \zeta_i| \leq \eta \tau_1 \leq \frac{1}{64}. \quad (25)$$

Proof. Feasibility first. At the base point, $(A_x g)_i = w_i^{-\beta} u_i^\top g = w_i^{1/p} k_i^\top g$, so $|(A_x g)_i| \leq \tau_1$, which is Eq. (25). Hence $1 + t\eta(A_x g)_i \geq 1 - \eta\tau_1 > 0$ for all $t \in [0, 1]$, so the entire segment is feasible *before* any endpoint weight comparison is used, and $a(tg)$ is well-defined and smooth in t .

Dini bootstrap. Let $a_\infty(t) := \|a(tg)\|_\infty$. By Eq. (12), $H(y_t) \preceq e^{2a_\infty(t)} I$. Combine Eq. (20) with Eq. (19) at y_t to get

$$\left| \frac{d}{dt} a_i(tg) \right| = \eta w_{y_t, i}^{-1/2} |e_i^\top (I + \beta N_{y_t}) B_{y_t} g| \leq L_p \eta \|g\|_{H(y_t)} \leq L_p \eta e^{a_\infty(t)} \|g\|.$$

Hence the upper Dini derivative satisfies $D^+ a_\infty(t) \leq L_p \eta e^{a_\infty(t)} \|g\|$. Scalar comparison here is elementary: a_∞ , the maximum of finitely many absolute values of smooth functions, is absolutely continuous, and at almost every t the displayed bound implies $(e^{-a_\infty(t)})' \geq -\vartheta$. Since $a_\infty(0) = 0$ and $L_p \eta \|g\| \leq L_p \eta R_1 = L_p M_0 \rho = \vartheta$, integration gives $e^{-a_\infty(t)} \geq 1 - \vartheta t$. Taking negative logarithms and using $t \leq 1$ and $\vartheta \leq 1/64$ yields

$$a_\infty(t) \leq -\log(1 - \vartheta t) \leq -\log(1 - \vartheta) \leq 2\vartheta \leq \frac{1}{32}, \quad (26)$$

which proves the first bound in Eq. (22); the second follows from the same rowwise display and $e^{a_\infty} \leq e^{1/32}$.

Weight comparison. Since $\beta \geq \frac{1}{4}$, Eq. (11) and $|\log(1 + u)| \leq 2|u|$ for $|u| \leq 1/64$, together with Eq. (26) and $\eta\tau_1 \leq 1/64$, give

$$\left| \log \frac{w_{y_t, i}}{w_i} \right| \leq \frac{1}{\beta} (a_\infty(t) + |\log(1 + t\eta(A_x g)_i)|) \leq 4\left(\frac{1}{32} + 2\eta\tau_1\right) \leq \frac{1}{4},$$

which is the first half of Eq. (24); the second half is immediate from Eq. (12).

Base-norm bounds. We now integrate the Jacobian in the base W -norm. Eq. (20) gives

$$\|I + \beta N_{y_t}\|_{\text{op}} = \|\mathcal{R}_{y_t}\|_{\text{op}} \leq p/2, \quad B_{y_t}^\top B_{y_t} \preceq H(y_t).$$

The weight comparison gives $\|W^{1/2} W_{y_t}^{-1/2}\|_{\text{op}} \leq e^{1/8}$. Together with Eq. (26), these estimates imply

$$\|W^{1/2} D_h a(tg)[v]\| \leq e^{1/8} \frac{p}{2} \eta \|v\|_{H(y_t)} \leq e^{1/8+1/32} \frac{p}{2} \eta \|v\| = e^{5/32} \frac{p}{2} \eta \|v\|.$$

This is the operator bound in Eq. (23). Integrating along $t \mapsto tg$ from $a(0) = 0$ gives $\|a(tg)\|_W \leq e^{5/32} \frac{p}{2} \eta \|g\| t \leq e^{5/32} \frac{p}{2} M_0 \rho$. $\square$

It is the scaled quantity $\eta\tau_1$, rather than the unscaled threshold τ_1 , that controls the slack ratio in Eq. (25); this is what makes all constants in Lemma 4.8 dimension-free.

On the tube, the ledger identity Eq. (14) reduces to the centered fluctuation

$$Z(h) := (c - \bar{w})^\top a \quad (27)$$

up to two finite error terms, which we now bound. Recall from Lemma 4.8 that on $\mathcal{K}_0^+$ we have $c_i \leq \tau_1^2 w_i$ (since $c_i = w_i(k_i^\top h)^2$), $\bar{w}_i \leq \|u_i\|^2 = w_i$, and $\|a\|_\infty \leq \frac{1}{32}$.

Lemma 4.9 (Quadratic remainder). *On $\mathcal{K}_0^+$,*

$$|(c - \bar{w})^\top (b - 2a)| \leq C(\tau_1^2 + 1) \|a\|_W^2 \leq C_p M_0^2 \rho^2 (\tau_1^2 + 1). \quad (28)$$

Proof. $|c_i - \bar{w}_i| \leq (\tau_1^2 + 1)w_i$ and $|b_i - 2a_i| = |e^{2a_i} - 1 - 2a_i| \leq C a_i^2$ for $|a_i| \leq \frac{1}{32}$. Summing over i and applying Eq. (23) proves the claim. $\square$

Lemma 4.10 (Hadamard–Frobenius bound). *Let $P_Q := UQU^\top$. Then, on $\mathcal{K}_0^+$,*

$$\|E\|_F^2 = b^\top (P_Q \circ P_Q) b \leq b^\top W b \leq C \|a\|_W^2 \leq C_p M_0^2 \rho^2.$$

Proof. From Eqs. (13) and (12), cyclicity of the trace and expansion over the rows of U give

$$\|E\|_F^2 = \text{tr}[QU^\top \text{Diag}(b)UQU^\top \text{Diag}(b)U] = \sum_{i,j} b_i b_j (u_i^\top Q u_j)^2 = b^\top (P_Q \circ P_Q) b.$$

Since $0 \preceq P_Q \preceq P$, we use the following one-line proof of the Schur product fact. If positive semidefinite matrices C, D have Gram representations $C_{ij} = \langle c_i, c_j \rangle$ and $D_{ij} = \langle d_i, d_j \rangle$, then $(C \circ D)_{ij} = \langle c_i \otimes d_i, c_j \otimes d_j \rangle$, so $C \circ D \succeq 0$. Applying this fact to the pairs $P, P - P_Q$ and $P - P_Q, P_Q$ gives

$$P \circ P - P_Q \circ P_Q = P \circ (P - P_Q) + (P - P_Q) \circ P_Q \succeq 0, \quad \text{so} \quad P_Q \circ P_Q \preceq P \circ P.$$

To prove $P \circ P \preceq \text{Diag}(P) = W$, fix any $x \in \mathbb{R}^n$ and set $M := \sum_i x_i u_i u_i^\top$. Weighted Cauchy–Schwarz and Eq. (5) yield

$$x^\top (P \circ P) x = \|M\|_F^2 = \sum_i x_i u_i^\top M u_i \leq \|x\|_W \left(\sum_i \frac{(u_i^\top M u_i)^2}{w_i} \right)^{1/2} \leq \|x\|_W \|M\|_F,$$

Hence $\|M\|_F \leq \|x\|_W$ and $x^\top (P \circ P) x \leq x^\top W x$.

Finally, $|b_i| \leq C|a_i|$ for $|a_i| \leq \frac{1}{32}$, so $b^\top W b \leq C \|a\|_W^2$, and Eq. (23) concludes. $\square$

Lemma 4.11 (Determinant remainder). *On $\mathcal{K}_0^+$, every eigenvalue of $I + E$ belongs to an interval satisfying*

$$[e^{-2\|a\|_\infty}, e^{2\|a\|_\infty}] \subseteq [e^{-1/16}, e^{1/16}].$$

Moreover,

$$|\log \det(I + E) - \text{tr}[E]| \leq C \|E\|_F^2 \leq C_p M_0^2 \rho^2. \quad (29)$$

Proof. Put $a_0 = e^{-2\|a\|_\infty}$ and $b_0 = e^{2\|a\|_\infty}$. By Eq. (24), $a_0 I \preceq H(y) \preceq b_0 I$. Since $a_0 \leq 1 \leq b_0$ and $\mathcal{C} \succeq 0$, adding $a_0 \mathcal{C} \preceq \mathcal{C}$ and $\mathcal{C} \preceq b_0 \mathcal{C}$ gives

$$a_0(I + \mathcal{C}) \preceq H(y) + \mathcal{C} \preceq b_0(I + \mathcal{C}).$$

Congruence by $Q^{1/2} = (I + \mathcal{C})^{-1/2}$ gives $a_0 I \preceq I + E \preceq b_0 I$ without requiring $H(y)$ and $\mathcal{C}$ to commute. On that interval the eigenvalues λ of E obey $|\lambda| < 1/8$. The fundamental theorem of calculus then gives

$$|\log(1 + \lambda) - \lambda| = \left| \int_0^\lambda \frac{-t}{1+t} dt \right| \leq \frac{\lambda^2}{2(1-|\lambda|)} \leq C \lambda^2.$$

Summing gives $|\log \det(I + E) - \text{tr}[E]| \leq C \|E\|_F^2$, and Lemma 4.10 concludes. $\square$

One finite quadratic remainder thus controls *all* orders of the determinant: no expansion of $\log \det$ beyond second order, and no convexity of $\log \det G$, is ever used.

4.4 A row-dependent contraction of Lewis Hessians

The Stein argument of Section 4.5 converts the mean of the centered fluctuation $Z = (c - \bar{w})^\top a$ into a sum of second derivatives of $\log w$ evaluated at *row-dependent* directions of magnitude $\sqrt{w_i}$. The following deterministic lemma—the only genuinely new piece of Lewis-weight calculus in this paper—is the estimate that bounds this sum by $C_p d$ rather than the naive $C_p n d$ or $C_p d^{3/2}$.

This section takes place at an arbitrary fixed point of K° , normalized as in Section 4.2 so that $U^\top U = I$; all derivatives are with respect to the spatial point in these coordinates. (In Section 4.5 the lemma is applied at the random endpoint y , in y -normalized coordinates.) We abbreviate $\ell := \log w$, and write $u_i^\top$ for the rows of U , $B := W^{1/p} U = W^{1/2} A$, and $\Phi, \bar{A}, \mathcal{R}, N$ for the objects of Eq. (3) at this point, so that $\|\Phi\|_{\text{op}} \leq 1$, $\|B\|_{\text{op}} \leq 1$, $\|\mathcal{R}\|_{\text{op}} \leq p/2$, $\|N\|_{\text{op}} \leq p$, and rows satisfy $\|u_i\|^2 = w_i$, $\|B_i\| = w_i^{1/2+1/p}$.

The contraction proved in this subsection rests on three ingredients: an exact second-derivative identity (Lemma 4.12), a frame-derivative computation (Lemma 4.13), and an aggregate bound on the derivative of the resolvent kernel (Lemma 4.14).

For a fixed direction r , write

$$\ell_r := D\ell[r] = -W^{-1/2} N B r, \quad s_r := A r = W^{-1/2} B r, \quad t_r := \frac{1}{2} \ell_r - s_r. \quad (30)$$

Lemma 4.12 (Exact Hessian identity).

$$D^2 \ell[r, r] = -\frac{1}{2} \ell_r^{\circ 2} - W^{-1/2} D N[r] B r - W^{-1/2} N \text{Diag}(t_r) B r. \quad (31)$$

Proof. Differentiate $\ell_v = -W^{-1/2} N B v$ at fixed $v = r$, using the product rule on the three point-dependent factors. The identities $DW[r] = W \text{Diag}(\ell_r)$ and $DA[r] = -\text{Diag}(s_r) A$ give

$$D(W^{-1/2})[r] = -\frac{1}{2} \text{Diag}(\ell_r) W^{-1/2}$$

and, by the product rule,

$$\begin{aligned} DB[r] &= D(W^{1/2} A)[r] = \text{Diag}(\tfrac{1}{2} \ell_r - s_r) B \\ &= \text{Diag}(t_r) B. \end{aligned}$$

Combining these identities gives

$$D^2 \ell[r, r] = \frac{1}{2} \text{Diag}(\ell_r) W^{-1/2} N B r - W^{-1/2} D N[r] B r - W^{-1/2} N \text{Diag}(t_r) B r,$$

and the first term equals $\frac{1}{2} \text{Diag}(\ell_r)(-\ell_r) = -\frac{1}{2} \ell_r^{\circ 2}$. $\square$

The operators $\Phi, \bar{A}, N$ depend on the point only through the Gram matrix of the rows of the orthonormalized frame $V_z = W_z^\beta A_z H(z)^{-1/2}$, so they are independent of the choice of square root; but to differentiate them we fix a smooth choice. The computation is cleanest in the *horizontal* gauge.

Lemma 4.13 (Horizontal frame derivative). *Fix the symmetric square root used to define the frame $U(z)$, where*

$$U(z) := W_z^\beta A_z H(z)^{-1/2}.$$

At the normalized base point, where $H = I$, we have

$$DU[r] = P^\perp \text{Diag}(z_r) U, \quad P^\perp := I - U U^\top, \quad W^{1/2} z_r := -(I + \beta N) B r.$$

Consequently $\|DU[r]\|_F \leq \|z_r\|_W = \|(I + \beta N) B r\| \leq \frac{p}{2} \|r\|$.

Proof. Set $z_r := \beta \ell_r - s_r$; then $W^{1/2} z_r = -\beta N B r - B r = -(I + \beta N) B r$ by Eq. (30). As in Lemma 4.12, $D(W^\beta A)[r] = \text{Diag}(z_r) W^\beta A$, which at the base point equals $\text{Diag}(z_r) U$. The product rule at $H = I$ therefore gives

$$DU[r] = \text{Diag}(z_r) U + U \cdot H^{1/2} D(H^{-1/2})[r]|_{H=I} = \text{Diag}(z_r) U + U D(H^{-1/2})[r].$$

At $H = I$, differentiating $H = (W^\beta A)^\top (W^\beta A)$ gives $DH[r] = 2U^\top \text{Diag}(z_r) U$, and for the symmetric square root $D(H^{-1/2})[r] = -\frac{1}{2} DH[r] = -U^\top \text{Diag}(z_r) U$ (the derivative X of $H^{-1/2}$ solves $XH^{-1/2} + H^{-1/2}X = D(H^{-1})[r] = -DH[r]$ at $H = I$). Substituting this derivative and using $P^\perp = I - UU^\top$ yields

$$DU[r] = \text{Diag}(z_r) U - UU^\top \text{Diag}(z_r) U = P^\perp \text{Diag}(z_r) U.$$

For the norm bound, $\|DU[r]\|_F \leq \|\text{Diag}(z_r) U\|_F = (\sum_i z_{r,i}^2 \|u_i\|^2)^{1/2} = \|z_r\|_W$, and $\|(I + \beta N) B r\| \leq \|\mathcal{R}\|_{\text{op}} \|B\|_{\text{op}} \|r\| \leq \frac{p}{2} \|r\|$ by Eq. (4). $\square$

Lemma 4.14 (Aggregate resolvent-derivative bound). *For every r , $\|DN[r] B\|_F \leq C_p \|r\|$, and hence, summing over an orthonormal basis $(e_a)_{a=1}^d$,*

$$\sum_{a=1}^d \|DN[e_a] B\|_F^2 \leq C_p d. \quad (32)$$

Proof. From $\mathcal{R} = (I - c_p \bar{\Lambda})^{-1}$ we have $N = 2\bar{\Lambda} \mathcal{R} = \frac{2}{c_p} (\mathcal{R} - I)$, so $DN[r] = \frac{2}{c_p} D\mathcal{R}[r] = 2\mathcal{R} D\bar{\Lambda}[r] \mathcal{R}$. Writing $\Phi_i = u_i u_i^\top / \|u_i\| \in \mathbb{S}^d$, the derivative used below is, for $u \neq 0$,

$$D\left(\frac{uu^\top}{\|u\|}\right)[v] = \frac{vu^\top + uv^\top}{\|u\|} - \frac{uu^\top}{\|u\|^3} \langle u, v \rangle.$$

The Frobenius norms of the three displayed summands are at most $\|v\|$, $\|v\|$, and $\|v\|$, respectively. Thus the derivative has norm at most $3\|v\|$. Consequently, with $\Phi_r := D\Phi[r]$, the componentwise derivative bound and Lemma 4.13 give

$$\|\Phi_r\|_F^2 \leq 9 \sum_i \|Du_i[r]\|^2 = 9 \|DU[r]\|_F^2 \leq 9 \frac{p^2}{4} \|r\|^2.$$

The identity $D\bar{\Lambda}[r] = -(\Phi_r \Phi^* + \Phi \Phi_r^*)$ gives $\|D\bar{\Lambda}[r]\|_F \leq 2\|\Phi_r\|_F \|\Phi\|_{\text{op}} \leq 3p\|r\|$. After absorbing the fixed polynomial in p into C_p , submultiplicativity and the bounds on $\mathcal{R}$, $D\bar{\Lambda}[r]$, and $\mathcal{R}B$ give

$$\|DN[r] B\|_F \leq 2 \|\mathcal{R}\|_{\text{op}} \|D\bar{\Lambda}[r]\|_F \|\mathcal{R}B\|_{\text{op}} \leq 2 \cdot \frac{p}{2} \cdot 3p\|r\| \cdot \frac{p}{2} = C_p \|r\|.$$

$\square$

Lemma 4.15 (Row-dependent Hessian contraction). *If directions $r_i \in \mathbb{R}^d$ satisfy $\|r_i\| \leq \mu \sqrt{w_i}$ for all i , then*

$$\sum_{i=1}^n |D^2 \ell_i[r_i, r_i]| \leq C_p \mu^2 d. \quad (33)$$

Proof. Write $r_i = \sqrt{w_i} q_i$ with $\|q_i\| \leq \mu$, and take the i -th component of Eq. (31) with direction r_i ; each of the three terms is bilinear in r_i , contributing a factor w_i that combines with the prefactor $w_i^{-1/2}$.

First terms. $(\ell_{r_i})_i^2 = (e_i^\top NBq_i)^2$ after the substitution. Using $\|q_i\| \leq \mu$, Frobenius-norm submultiplicativity, $\|N\|_{\text{op}} \leq p$, $w_i \leq 1$, and $\sum_i w_i = d$ gives

$$\sum_i \frac{1}{2} (\ell_{r_i})_i^2 \leq \frac{1}{2} \mu^2 \|NB\|_F^2 \leq \frac{1}{2} \mu^2 \|N\|_{\text{op}}^2 \|B\|_F^2 \leq \frac{1}{2} \mu^2 p^2 \sum_i w_i^{1+2/p} \leq C_p \mu^2 d. \quad (34)$$

Middle terms. The i -th component is $\sqrt{w_i} e_i^\top DN[q_i]Bq_i$. By Cauchy–Schwarz with weights $\sqrt{w_i}$ and $\sum_i w_i = d$,

$$\sum_i \sqrt{w_i} |e_i^\top DN[q_i]Bq_i| \leq \sqrt{d} \left(\sum_i |e_i^\top DN[q_i]Bq_i|^2 \right)^{1/2}.$$

Expand

$$DN[q_i]Bq_i = \sum_{a,b} (q_i)_a (q_i)_b DN[e_a]Be_b, \quad T_{ab}^{(i)} := e_i^\top DN[e_a]Be_b.$$

Then Cauchy–Schwarz and $\|q_i\| \leq \mu$ imply

$$|e_i^\top DN[q_i]Bq_i| \leq \|q_i\|^2 \|T^{(i)}\|_F \leq \mu^2 \|T^{(i)}\|_F,$$

while the definition of $T^{(i)}$ and Eq. (32) give

$$\sum_i \|T^{(i)}\|_F^2 = \sum_a \sum_{i,b} (e_i^\top DN[e_a]Be_b)^2 = \sum_a \|DN[e_a]B\|_F^2 \leq C_p d.$$

Combining these bounds and absorbing the square root of C_p into C_p ,

$$\sum_i \sqrt{w_i} |e_i^\top DN[q_i]Bq_i| \leq \sqrt{d} \cdot \mu^2 \sqrt{C_p d} = C_p \mu^2 d. \quad (35)$$

Last terms. Set

$$C_0 := I + \frac{1}{2}N, \quad L_0 := -W^{-1/2}C_0B, \quad T_i := L_0^\top \text{Diag}(e_i^\top N)B = \sum_j N_{ij} (L_0)_j^\top (B)_j,$$

where $(L_0)_j, (B)_j$ denote rows and $(L_0)_j^\top (B)_j$ their rank-one outer products. Since $t_q = \frac{1}{2}\ell_q - s_q = -W^{-1/2}(I + \frac{1}{2}N)Bq = L_0q$ by Eq. (30), the i -th component of the last term of Eq. (31) with direction r_i is $\sqrt{w_i} q_i^\top T_i q_i$. Viewing $((L_0)_j^\top (B)_j)_{j \leq n}$ as a vector with entries in the Hilbert space $(\mathbb{R}^{d \times d}, \|\cdot\|_F)$, let N act on the index j . The Hilbert-space operator-norm bound, the row identities for L_0 and B , and Lemma 4.7(ii) give

$$\begin{aligned} \sum_i \|T_i\|_F^2 &\leq \|N\|_{\text{op}}^2 \sum_j \|(L_0)_j\|^2 \|(B)_j\|^2 \\ &= \|N\|_{\text{op}}^2 \sum_j w_j^{2/p} \|e_j^\top C_0 B\|^2 \\ &\leq p^2 (L_p^+)^2 \sum_j w_j^{1+2/p} \leq C_p d, \end{aligned}$$

where Lemma 4.7(ii) is used with $\gamma = \frac{1}{2}$ at the normalized point $\hat{B} = B$.

A final weighted Cauchy–Schwarz, together with $\sum_i w_i = d$, gives

$$\sum_i \sqrt{w_i} |q_i^\top T_i q_i| \leq \mu^2 \sum_i \sqrt{w_i} \|T_i\|_F \leq \mu^2 \sqrt{\sum_i w_i} \sqrt{\sum_i \|T_i\|_F^2} \leq C_p \mu^2 d. \quad (36)$$

Combining Eq. (34), (35) and (36) with Lemma 4.12 proves Eq. (33). $\square$

Remark 4.16. The lemma is sharp in its dependence on the row profile: replacing the hypothesis $\|r_i\| \leq \mu\sqrt{w_i}$ by a uniform bound $\|r_i\| \leq \mu$ would force an extra factor of $\max_i w_i^{-1}$ through Eq. (34). It is precisely the directions Qu_i , whose norms are at most $\sqrt{w_i}$, that the Stein identity produces in Section 4.5. Thus the Stein directions satisfy exactly the row-dependent hypothesis of the lemma.

4.5 A Stein bound for the centered fluctuation, and the acceptance lower tail

This section controls the centered fluctuation

$$Z(h) = (c - \bar{w})^\top a, \quad h \sim N(0, Q),$$

and assembles the uniform lower tail for the complete Metropolis ratio (Lemma 4.23). The mean of Z is treated by Gaussian integration by parts, reduced through Lemma 4.15; its deviations are controlled by a dimension-free Lipschitz bound. Both steps take place on the stability tube of Section 4.3, and a cutoff and a global extension reconcile that tube with the full-space Stein identity.

Lemma 4.17 (Twice-integrated Gaussian Stein identity). *Let $h \sim N(0, Q)$ and let $F : \mathbb{R}^d \rightarrow \mathbb{R}$ be globally $C^{1,1}$ (equivalently $W^{2,\infty}$) with compact support. Then for every $v \in \mathbb{R}^d$,*

$$\mathbb{E}[(v^\top h)^2 - v^\top Qv]F(h) = \mathbb{E}[(Qv)^\top D^2 F(h) (Qv)],$$

where $D^2 F$ is the (a.e.-defined, bounded) weak Hessian.

Proof. Write $h = Q^{1/2}\xi$ with ξ standard; this representation also covers singular Q . For $F \in C_c^\infty$, apply one-dimensional Gaussian integration by parts twice to $F(Q^{1/2}\xi)$. This composition and its first two derivatives are bounded, so the Gaussian boundary terms vanish even when the composition is not compactly supported. The chain rule gives $\mathbb{E}[h_a h_b F] = Q_{ab}\mathbb{E}[F] + \sum_{c,e} Q_{ac} Q_{be} \mathbb{E}[\partial_c \partial_e F]$; contract with $v_a v_b$. For $F \in W^{2,\infty}$ with compact support, mollify and pass to the limit using dominated convergence on both sides. $\square$

The relevant second derivative is exact: differentiating Eq. (11) twice in h (the slack term is the composition of $-\log(1 + \cdot)$ with a linear map),

$$D_h^2 a_i[v, v] = \eta^2(\beta D^2 \ell_{y,i}[v, v] + (A_y v)_i^2), \quad y = x + \eta h. \quad (37)$$

To apply Lemma 4.15 at y , pass to y -normalized coordinates: let $V = W_y^\beta A_y H(y)^{-1/2}$ with rows v_i , so $V^\top V = I$ and $\|v_i\|^2 = w_{y,i}$. From Eq. (12),

$$u_i = e^{-a_i} H(y)^{1/2} v_i, \quad \text{hence} \quad H(y)^{1/2} Q u_i = e^{-a_i} H(y)^{1/2} Q H(y)^{1/2} v_i. \quad (38)$$

On $\mathcal{K}_0^+$, Lemma 4.8 and $Q \preceq I$ give $\|H(y)^{1/2} Q H(y)^{1/2}\|_{\text{op}} \leq e^{2\|a\|_\infty} \leq e^{1/16}$ and $e^{-a_i} \leq e^{1/32}$, so

$$\|H(y)^{1/2} Q u_i\| \leq C \sqrt{w_{y,i}} \quad \text{for all } i. \quad (39)$$

The vectors $H(y)^{1/2} Q u_i$ are exactly the y -normalized versions of the Stein directions Qu_i , and Eq. (39) is the row-dependent hypothesis of Lemma 4.15 with $\mu = O(1)$. The slack part of Eq. (37) has the sharper direct bound. Set $Q_y := H(y)^{1/2} Q H(y)^{1/2} \preceq CI$. Using $\|v_i\|^2 = w_{y,i}$, $2 - 2\beta = 1 + 2/p$, $w_{y,i} \leq 1$, and $\sum_i w_{y,i} = d$, we obtain

$$(A_y Q u_i)_i = e^{-a_i} w_{y,i}^{-\beta} v_i^\top Q_y v_i \quad \implies \quad \sum_i (A_y Q u_i)_i^2 \leq C \sum_i w_{y,i}^{2-2\beta} = C \sum_i w_{y,i}^{1+2/p} \leq Cd. \quad (40)$$

Lemma 4.18 (Mean bound on the tube). *Let χ be any $[0, 1]$ -valued cutoff supported in $\mathcal{K}_0^+$. Then*

$$|\sum_i \mathbb{E}[\chi(h) (Qu_i)^\top D_h^2 a_i(h) (Qu_i)]| \leq C_p \eta^2 d = C_p \rho^2.$$

Proof. Pointwise on $\mathcal{K}_0^+$, by Eq. (37), Eq. (38), and the coordinate covariance of second derivatives ($D^2 \ell_{y,i}$ evaluated at direction Qu_i in base coordinates equals the y -normalized Hessian evaluated at $H(y)^{1/2} Qu_i$), Lemma 4.15 and Eqs. (39) and (40) give

$$\begin{aligned} & \sum_i |(Qu_i)^\top D_h^2 a_i(Qu_i)| \\ & \leq \eta^2 (\beta \sum_i |D^2 \ell_{y,i}[H(y)^{1/2} Qu_i, H(y)^{1/2} Qu_i]|_{y\text{-norm.}} + \sum_i (A_y Qu_i)_i^2) \\ & \leq \eta^2 (C_p d + Cd). \end{aligned}$$

Multiply by $\chi \leq 1$ and take expectations. $\square$

Lemma 4.19 (Dimension-free Lipschitz bound). *On the convex set $\mathcal{K}_0$,*

$$\text{Lip}(Z|\mathcal{K}_0) \leq C_p M_0 \rho (\tau_1^2 + 1), \quad (41)$$

with no power of d .

Proof. Differentiating Eq. (27),

$$\nabla Z = 2U^\top \text{Diag}(a)Uh + (D_h a)^\top (c - \bar{w}).$$

For the first term, write $U^\top \text{Diag}(a)Uh = \sum_i u_i a_i \sqrt{w_i} (k_i^\top h)$ and use $\|U^\top\|_{\text{op}} \leq 1$ and $|k_i^\top h| \leq \tau_1$ on the tube to obtain

$$\|U^\top \text{Diag}(a)Uh\| \leq (\sum_i a_i^2 w_i (k_i^\top h)^2)^{1/2} \leq \tau_1 \|a\|_W.$$

For the second term, $|c_i - \bar{w}_i| \leq (\tau_1^2 + 1)w_i$ gives $\|W^{-1/2}(c - \bar{w})\| \leq (\tau_1^2 + 1)\sqrt{d}$. Inserting $W^{1/2}W^{-1/2}$, applying Eq. (23), and using $\eta\sqrt{d} = \rho$ yields

$$\|(D_h a)^\top (c - \bar{w})\| \leq \|W^{1/2} D_h a\|_{\text{op}} \|W^{-1/2}(c - \bar{w})\| \leq e^{5/32} \frac{p}{2} \eta (\tau_1^2 + 1) \sqrt{d} = C_p \rho (\tau_1^2 + 1).$$

Combining with $\|a\|_W \leq C_p M_0 \rho$ from Eq. (23) and $M_0 \geq 1$ yields Eq. (41); convexity of $\mathcal{K}_0$ converts the gradient bound into a Lipschitz bound. $\square$

The function $a(h)$ is naturally defined only while $x + \eta h$ remains in the stable feasible tube, so Lemma 4.17 cannot be applied to a_i directly. We close the domain issue as follows.

Lemma 4.20 (Cutoff). *There is a function $\chi : \mathbb{R}^d \rightarrow [0, 1]$, equal to 1 on $\mathcal{K}_0$, supported in $\mathcal{K}_0^+$, with*

$$\|\nabla \chi\| \leq C, \quad \|D^2 \chi\|_{\text{op}} \leq C \quad \text{a.e.} \quad (42)$$

Proof. Let $\phi(h) = \text{dist}(h, \mathcal{K}_0)^2$. Since $\mathcal{K}_0$ is convex, ϕ is convex and $C^{1,1}$ with $\nabla \phi(h) = 2(h - \Pi_{\mathcal{K}_0} h)$, which is 2-Lipschitz; hence $0 \preceq D^2 \phi \preceq 2I$ a.e. For completeness, projection optimality gives, for the residual map $r = I - \Pi_{\mathcal{K}_0}$, $\langle r(x) - r(y), x - y \rangle \geq \|r(x) - r(y)\|^2$; thus r is 1-Lipschitz, and differentiating the squared distance gives $\nabla \phi = 2r$. Compose with a fixed scalar C^∞ bump $\theta : [0, \infty) \rightarrow [0, 1]$, $\theta \equiv 1$ on $[0, \frac{1}{9}]$, $\theta \equiv 0$ on $[\frac{4}{9}, \infty)$: then $\chi = \theta \circ \phi$ is 1 on $\mathcal{K}_0$, vanishes when $\text{dist}(h, \mathcal{K}_0) \geq \frac{2}{3}$, and satisfies Eq. (42) because $\|\nabla \phi\| \leq \frac{4}{3}$ on the support of $\theta' \circ \phi$. The construction is dimension-free. $\square$

Lemma 4.21 (McShane extension). *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be nonempty and let $F : \mathcal{K} \rightarrow \mathbb{R}$ be Λ -Lipschitz. Then*

$$\tilde{F}(x) := \inf_{z \in \mathcal{K}} \{F(z) + \Lambda \|x - z\|\}$$

is a Λ -Lipschitz extension of F to $\mathbb{R}^d$. If $0 \in \mathcal{K}$ and $F(0) = 0$, then $|\tilde{F}(x)| \leq \Lambda \|x\|$.

Proof. For $x \in \mathcal{K}$, Lipschitzness gives $F(z) + \Lambda \|x - z\| \geq F(x)$ for every $z \in \mathcal{K}$, while choosing $z = x$ gives the reverse inequality; hence $\tilde{F} = F$ on $\mathcal{K}$. The infimum is finite: fixing $z_0 \in \mathcal{K}$, Lipschitzness and the reverse triangle inequality give

$$F(z) + \Lambda \|x - z\| \geq F(z_0) - \Lambda \|x - z_0\| \quad (z \in \mathcal{K}),$$

while the candidate $z = z_0$ supplies a finite upper bound. The triangle inequality gives $\tilde{F}(x) \leq \tilde{F}(y) + \Lambda \|x - y\|$, and exchanging x, y proves the global Lipschitz bound. The last claim follows from $\tilde{F}(0) = F(0) = 0$. $\square$

Each product $a_i \chi$ is $C^{1,1}$ on the open tube in which $\mathcal{K}_0^+$ is compactly contained (by Lemma 4.8 the segment stays strictly feasible there) and extends by zero to a global $W^{2,\infty}$ function of compact support: the cutoff already vanishes on the open collar where $\text{dist}(h, \mathcal{K}_0) > 2/3$, before the boundary of $\mathcal{K}_0^+$ is reached. Thus Lemma 4.17 applies to $F = a_i \chi$. Summing over i ,

$$\mathbb{E}[Z\chi] = \sum_i \mathbb{E}[(Qu_i)^\top D^2(a_i \chi)(h) (Qu_i)]. \quad (43)$$

Expanding $D^2(a_i \chi) = \chi D^2 a_i + (\nabla a_i)(\nabla \chi)^\top + (\nabla \chi)(\nabla a_i)^\top + a_i D^2 \chi$, the leading term is bounded by Lemma 4.18. The three cutoff terms are supported on $\mathcal{K}_0^+ \setminus \mathcal{K}_0$, an event of probability at most δ_0 . Their deterministic sizes on the collar are as follows. The two gradient terms together are at most

$$2 \sum_i |(Qu_i)^\top \nabla a_i| |(Qu_i)^\top \nabla \chi| \leq C_p L_p \eta \sum_i w_i = C_p L_p \rho \sqrt{d},$$

where we used the rowwise bounds $|(\nabla a_i)^\top Qu_i| \leq C L_p \eta \sqrt{w_i}$ (from Eq. (20), (19) and (39)), $\|Qu_i\| \leq \sqrt{w_i}$, $\sum_i w_i = d$, and $\|\nabla \chi\| \leq C$. The Hessian-of-cutoff term is at most

$$\sum_i |a_i| \|D^2 \chi\|_{\text{op}} \|Qu_i\|^2 \leq C \sum_i |a_i| w_i \leq C \sqrt{d} \|a\|_W \leq C_p M_0 \rho \sqrt{d}.$$

Integrating on the collar and using $\delta_0 \leq \sqrt{\delta_0}$ gives the common upper bound

$$C_p M_0 \rho (1 + \sqrt{d}) \sqrt{\delta_0}. \quad (44)$$

Finally, apply Lemma 4.21 to $Z|_{\mathcal{K}_0}$ and denote the same-Lipschitz-constant extension by $\tilde{Z} : \mathbb{R}^d \rightarrow \mathbb{R}$. The means of $\tilde{Z}$ and $Z\chi$ agree up to an error supported off $\mathcal{K}_0$. Here $0 \in \mathcal{K}_0$, $a(0) = 0$, and hence $Z(0) = 0$. Because the extension preserves values on $\mathcal{K}_0$, its Lipschitz constant Λ gives the global growth bound $|\tilde{Z}(h)| \leq \Lambda \|h\|$. Using this fact and the pointwise bound $|Z| \leq (\tau_1^2 + 1) \sqrt{d} \|a\|_W$ on $\mathcal{K}_0^+$, the second moments satisfy $\mathbb{E}[\tilde{Z}(h)^2] \leq \Lambda^2 \mathbb{E}[\|h\|^2] \leq \Lambda^2 d$ and $\mathbb{E}[(Z\chi)^2] \leq C_p M_0^2 \rho^2 (\tau_1^2 + 1)^2 d$. Cauchy–Schwarz on the probability- δ_0 complement now gives

$$|\mathbb{E}[\tilde{Z}] - \mathbb{E}[Z\chi]| \leq C_p M_0 \rho (\tau_1^2 + 1) \sqrt{d} \sqrt{\delta_0}. \quad (45)$$

With $\delta_0 = \epsilon_0^2 d^{-3}$, both Eq. (44) and Eq. (45) are at most $C_p M_0 \rho (\tau_1^2 + 1) \epsilon_0 / d$.

Lemma 4.22 (Concentration of the fluctuation). *With probability at least $1 - 2\epsilon_0$ over $h \sim N(0, Q)$,*

$$|Z| \leq C_p [M_0^2 \rho^2 + M_0 \rho (\tau_1^2 + 1) (\sqrt{\log(2/\epsilon_0)} + \epsilon_0)]. \quad (46)$$

Proof. Write $h = Q^{1/2} \xi$; since $Q \preceq I$, $\tilde{Z}(Q^{1/2} \cdot)$ is Lipschitz in ξ with constant $\text{Lip}(\tilde{Z}) \leq C_p M_0 \rho (\tau_1^2 + 1)$ from Lemma 4.19. By Lemma 3.2, $|\tilde{Z} - \mathbb{E}[\tilde{Z}]| \leq \text{Lip}(\tilde{Z}) \sqrt{2 \log(2/\epsilon_0)}$ with probability $\geq 1 - \epsilon_0$. By Lemma 4.18, Eq. (43), Eq. (44) and Eq. (45), $|\mathbb{E}[\tilde{Z}]| \leq C_p \rho^2 + C_p M_0 \rho (\tau_1^2 + 1) \epsilon_0 / d$. On the event $h \in \mathcal{K}_0$ (probability $\geq 1 - \delta_0 \geq 1 - \epsilon_0$), $Z = \tilde{Z}$. Combining the preceding bounds on the intersection of these events proves the claim. $\square$

For fixed ϵ_0 , it suffices to take

$$\rho \leq \frac{c}{\Gamma_p M_0 (1 + \log(nd/\epsilon_0))}, \quad \kappa = \frac{r^2}{\rho^2}, \quad (47)$$

where Γ_p is a sufficiently large polynomial in p ; thus κ is polylogarithmic in n and d , and Eq. (21) holds as well. By Eq. (18), this choice makes the right-hand sides of Eq. (28), (29) and (46) smaller than any desired universal constant.

Lemma 4.23 (Uniform lower tail for the complete ratio). *There are universal constants $C < \infty$ and $\epsilon < \frac{1}{10}$ such that, uniformly at every deterministic $x \in K^\circ$, the walk of Definition 4.1 satisfies*

$$\Pr_{Y \sim q_x} [Y \in K \text{ and } \log R_x(Y) \geq -C] \geq 1 - \epsilon. \quad (48)$$

The constants are independent of x , n , d , L , R , and the orientation of the regularizer.

Proof. Work in the normalized coordinates; the ratio R_x is invariant under the affine substitution. Intersect three events: the tube event $\{h \in \mathcal{K}_0\}$ (probability $\geq 1 - \delta_0$, Lemma 4.6); the concentration event of Lemma 4.22 (probability $\geq 1 - 2\epsilon_0$); and the Gaussian norm event $\{\|\xi\| \leq \sqrt{d} + \sqrt{2 \log(1/\epsilon_0)}\}$ (probability $\geq 1 - \epsilon_0$ by Lemma 3.2, since $\mathbb{E}[\|\xi\|] \leq \sqrt{d}$). The failure probabilities sum to at most $\epsilon := \epsilon_0^2 + 3\epsilon_0 < \frac{1}{10}$ for a small universal ϵ_0 , since $\delta_0 = \epsilon_0^2 d^{-3} \leq \epsilon_0^2$.

Feasibility. On the norm event, $\|Y - x\|_{H(x)} = \eta \|h\| \leq \eta \|\xi\| \leq \rho M_0 < 1$. By the base-point identity $(A_x)_i H(x)^{-1} (A_x)_i^\top = w_i^{2/p} \leq 1$ (see Eq. (55) below) and Cauchy-Schwarz, the open unit $H(x)$ -ellipsoid lies in K° ; hence Y and the whole segment $[x, Y]$ are interior.

Proposal ratio. By Lemma 4.4,

$$\Lambda_q(x, Y) = -Z - \frac{1}{2}(c - \bar{w})^\top (b - 2a) + \frac{1}{2} \{\log \det(I + E) - \text{tr}[E]\},$$

and on the intersected event each of the three terms is bounded by a universal constant by Lemma 4.22, Lemma 4.9, and Lemma 4.11, using the choice Eq. (47). Hence $\Lambda_q(x, Y) \geq -C/2$.

Target ratio. If $L = 0$, Lipschitzness says that f is constant on the convex body K , so $f(Y) - f(x) = 0$. If $L > 0$, then $G(x) \succeq L^2 I$ and the proposal has covariance $\frac{r^2}{d} G(x)^{-1}$. Work on the norm event. Lipschitzness, the metric lower bound, and the Gaussian proposal representation give

$$f(Y) - f(x) \leq L \|Y - x\|_2 \leq L \cdot \frac{1}{L} \|Y - x\|_{G(x)} = \frac{r}{\sqrt{d}} \|\xi\| = O(r).$$

Combining, $\log R_x(Y) = -(f(Y) - f(x)) + \Lambda_q(x, Y) \geq -O(r) - C/2 \geq -C$ for r a small universal constant. $\square$

Remark 4.24. Lemma 4.23 holds at *every* interior point, with no dependence on the distance to ∂K . The analysis is affine-invariant, and the only property of the regularizer used in Sections 4.2, 4.3, and 4.4 is $\mathcal{C} \succeq 0$. Consequently no boundary-layer or s -conductance device is needed in Section 4.6.

4.6 From acceptance to mixing

We now return to the original coordinates and convert Lemma 4.23 into conductance through five explicit steps: local Gaussian overlap, a kernel-overlap lemma, symmetry, cross-ratio isoperimetry, and Cheeger's inequality. We execute them for the *complete* kernel directly, so that no forward/reverse asymmetry ever reappears.

Lemma 4.25 (SSC of the regularized metric). *$G = \kappa H + L^2 I$ with $\kappa \geq 2(1 + p^2)$ satisfies, for all $x \in K^\circ$ and $v \in \mathbb{R}^d$,*

$$\|G(x)^{-1/2} DG(x)[v] G(x)^{-1/2}\|_F \leq 2\|v\|_{G(x)}.$$

Proof. Let $H_\kappa := \kappa H$, which is SSC by Lemma 3.4. Set $X = H_\kappa^{-1/2} DH_\kappa[v] H_\kappa^{-1/2}$, so $\|X\|_F \leq 2\|v\|_{H_\kappa}$, and $C = H_\kappa^{1/2} G^{-1/2}$, so $\|C\|_{\text{op}} \leq 1$ because $H_\kappa \preceq G$. Also, $DG[v] = DH_\kappa[v]$. Thus the definitions of C and X , SSC of H_κ , and $H_\kappa \preceq G$ imply

$$\|G^{-1/2} DG[v] G^{-1/2}\|_F = \|C^\top X C\|_F \leq \|X\|_F \leq 2\|v\|_{H_\kappa} \leq 2\|v\|_G.$$

□

Since the Frobenius norm dominates the spectral norm, G satisfies the hypothesis of Lemma 3.5: if

$$\Delta := \|x - z\|_{G(x)} \leq \frac{c_{\text{loc}} r}{\sqrt{d}} \quad (x, z \in K^\circ), \quad (49)$$

then (for $\Delta \leq \frac{1}{2}$, ensured by $r \leq 1$)

$$(1 - \Delta)^2 G(x) \preceq G(z) \preceq (1 - \Delta)^{-2} G(x). \quad (50)$$

Lemma 4.26 (Pinsker's inequality). *For probability densities p, q ,*

$$d_{\text{TV}}(p, q)^2 \leq \frac{1}{2} D_{\text{KL}}(p \| q).$$

Proof. The claim is immediate if the relative entropy is infinite. Otherwise put $A = \{p \geq q\}$, $a = \int_A p$, and $b = \int_A q$. Then $a - b = d_{\text{TV}}(p, q)$. Jensen's inequality on A and A^c (equivalently, the two-set log-sum inequality) gives

$$D_{\text{KL}}(p \| q) \geq a \log \frac{a}{b} + (1 - a) \log \frac{1 - a}{1 - b} =: g_b(a).$$

For fixed $b \in (0, 1)$, $g_b(b) = g'_b(b) = 0$ and $g''_b(s) = 1/[s(1 - s)] \geq 4$, so $g_b(a) \geq 2(a - b)^2$. Endpoint cases follow by continuity. □

Lemma 4.27 (Proposal overlap). *There is a universal $c_{\text{loc}} > 0$ such that Eq. (49) implies*

$$d_{\text{TV}}(q_x, q_z) \leq \delta_q$$

for a fixed constant $\delta_q < 1 - 2\epsilon$, where ϵ is from Lemma 4.23.

Proof. For the Gaussians $q_x = N(x, \frac{r^2}{d} G(x)^{-1})$, $q_z = N(z, \frac{r^2}{d} G(z)^{-1})$, taking the expectation of the log density ratio gives the general identity

$$D_{\text{KL}}(N(m_0, \Sigma_0) \| N(m_1, \Sigma_1)) = \frac{1}{2} [\text{tr}[\Sigma_1^{-1} \Sigma_0] + \|m_1 - m_0\|_{\Sigma_1^{-1}}^2 - d + \log \frac{\det \Sigma_1}{\det \Sigma_0}].$$

Substitution gives

$$D_{\text{KL}}(q_x \| q_z) = \frac{1}{2} \left[\sum_{j=1}^d (\lambda_j - 1 - \log \lambda_j) + \frac{d}{r^2} \|x - z\|_{G(z)}^2 \right],$$

where λ_j are the eigenvalues of $G(x)^{-1/2} G(z) G(x)^{-1/2}$. By Eq. (50), $\lambda_j \in [(1 - \Delta)^2, (1 - \Delta)^{-2}]$. For $\Delta \leq 1/2$, this interval gives $|\lambda_j - 1| \leq C\Delta$ and $\lambda_j \geq 1/4$; moreover

$$0 \leq \lambda - 1 - \log \lambda = \int_1^\lambda \frac{t-1}{t} dt \leq C(\lambda - 1)^2.$$

Consequently $\sum_j (\lambda_j - 1 - \log \lambda_j) \leq C d \Delta^2$ and $\|x - z\|_{G(z)}^2 \leq 4\Delta^2$. With $r \leq 1$ and Eq. (49),

$$D_{\text{KL}}(q_x \| q_z) \leq C(d\Delta^2 + \frac{d\Delta^2}{r^2}) \leq C' c_{\text{loc}}^2.$$

Choosing c_{loc} sufficiently small and applying Lemma 4.26 gives the claimed bound. $\square$

The next lemma is what lets us use the complete-ratio estimate Lemma 4.23 directly, with no separate treatment of determinant and quadratic terms at the coupling stage either.

Lemma 4.28 (Kernel overlap). *Suppose Eq. (48) holds at both x and z , and $d_{\text{TV}}(q_x, q_z) \leq \delta_q$. Then the accepted (moving) parts of the two transition kernels have common mass at least*

$$\gamma_0 := e^{-C} (1 - \delta_q - 2\epsilon) > 0, \quad (51)$$

and consequently, for the $\frac{1}{2}$ -lazy kernels P_x, P_z ,

$$d_{\text{TV}}(P_x, P_z) \leq 1 - \frac{\gamma_0}{2}. \quad (52)$$

Proof. Let $E_x := \{u \in K : R_x(u) \geq e^{-C}\}$ and likewise E_z . The moving part of the non-lazy kernel at x has density $u \mapsto q_x(u) (1 \wedge R_x(u))$ on K , so the common mass of the two moving parts is at least

$$\int_{E_x \cap E_z} e^{-C} \min\{q_x(u), q_z(u)\} du.$$

The measure with density $\min(q_x, q_z)$ has total mass at least $1 - \delta_q$; its mass on E_x^c is at most $q_x(E_x^c) \leq \epsilon$ by Eq. (48), and likewise on E_z^c . This proves Eq. (51); laziness retains half of the common moving mass, giving Eq. (52). $\square$

Combining Lemmas 4.27 and 4.28: the lazy kernels have overlap at least $\gamma_0/2$ whenever

$$\|x - z\|_{G(x)} \leq \frac{c_{\text{loc}} r}{\sqrt{d}}. \quad (53)$$

Definition 4.29 (Symmetry; [LLV20]). *For a metric M write $\mathcal{D}_M(x, t) := \{z : \|z - x\|_{M(x)} \leq t\}$. The metric is $\bar{\nu}$ -symmetric if*

$$\mathcal{D}_M(x, 1) \subseteq K \cap (2x - K) \subseteq \mathcal{D}_M(x, \sqrt{\bar{\nu}}) \quad (x \in K^\circ).$$

Lemma 4.30 (Symmetry parameter). *The scaled LS metric $H_\kappa = \kappa H$ is $\bar{\nu}_{\text{LS}}$ -symmetric with*

$$\bar{\nu}_{\text{LS}} := \kappa n^{2/p} d^{1-2/p} = \tilde{O}(d), \quad (54)$$

and $G = \kappa H + L^2 I$ is ν_G -symmetric with

$$\nu_G \leq \bar{\nu}_{\text{LS}} + 4L^2 R^2.$$

Proof. Containment of the unit ellipsoid. The Lewis fixed point Eq. (1) gives, for the i -th row of A_x ,

$$(A_x)_{i,\cdot} H(x)^{-1} (A_x)_{i,\cdot}^\top = w_{x,i}^{2/p} \leq 1, \quad (55)$$

(in normalized coordinates this is $\|w_i^{-\beta} u_i\|^2 = w_i^{1-2\beta} = w_i^{2/p}$). By Cauchy–Schwarz, $\|z - x\|_{H(x)} \leq 1$ implies $|(A_x(z - x))_i| \leq 1$ for all i , and then both z and $2x - z$ have nonnegative slacks $s_x \circ (\mathbf{1} \pm A_x(z - x))$. Hence $\mathcal{D}_{H_\kappa}(x, 1) \subseteq \mathcal{D}_H(x, 1) \subseteq K \cap (2x - K)$, using $\kappa \geq 1$.

Containment in the outer ellipsoid. If $z \in K \cap (2x - K)$, then every normalized slack displacement satisfies $|(A_x(z - x))_i| \leq 1$. Hölder’s inequality, in the form $\sum_i 1 \cdot w_i^{1-2/p} \leq n^{2/p} (\sum_i w_i)^{1-2/p}$, then gives

$$\|z - x\|_{H(x)}^2 = \sum_i w_{x,i}^{1-2/p} (A_x(z - x))_i^2 \leq \sum_i w_{x,i}^{1-2/p} \leq n^{2/p} d^{1-2/p}.$$

Scaling by κ gives Eq. (54); note $n^{2/p} \leq e$ and $\kappa = \tilde{O}(1)$, so $\bar{\nu}_{\text{LS}} = \tilde{O}(d)$.

The regularized metric. $\mathcal{D}_G(x, 1) \subseteq \mathcal{D}_{H_\kappa}(x, 1) \subseteq K \cap (2x - K)$ since $G \succeq H_\kappa$; and for $z \in K \cap (2x - K)$, Eq. (54) and the diameter bound $\text{diam}(K) \leq 2R$ give

$$\|z - x\|_{G(x)}^2 = \|z - x\|_{H_\kappa(x)}^2 + L^2 \|z - x\|_2^2 \leq \bar{\nu}_{\text{LS}} + 4L^2 R^2.$$

This proves the regularized-metric bound. □

This self-contained computation agrees with [LLV20, Lemma 4.3].

Definition 4.31 (Chord cross ratio). *For distinct $x, z \in K^\circ$, let their line meet ∂K in the order $p_\partial, x, z, q_\partial$ and put $\alpha = \|x - p_\partial\|_2$, $\beta_c = \|z - x\|_2$, $\gamma = \|q_\partial - z\|_2$. The cross ratio is*

$$\sigma(x, z) := \frac{\beta_c (\alpha + \beta_c + \gamma)}{\alpha \gamma}.$$

Lemma 4.32 (Cross ratio dominates the local norm). *For all distinct $x, z \in K^\circ$,*

$$\sigma(x, z) \geq \frac{\|x - z\|_{G(x)}}{\sqrt{\nu_G}}.$$

Proof. Let $e = (z - x)/\|z - x\|_2$. The radius of the symmetric body $K \cap (2x - K)$ at x in direction $\pm e$ is $t = \min(\alpha, \beta_c + \gamma)$, so $x + te$ lies in the closure of $K \cap (2x - K)$, and ν_G -symmetry gives $t \|e\|_{G(x)} \leq \sqrt{\nu_G}$. Checking the two cases $t = \alpha$ and $t = \beta_c + \gamma$ gives

$$\sigma(x, z) = \frac{\beta_c (\alpha + \beta_c + \gamma)}{\alpha \gamma} \geq \frac{\beta_c}{t}.$$

Hence $\sigma(x, z) \geq \beta_c \|e\|_{G(x)} / \sqrt{\nu_G} = \|z - x\|_{G(x)} / \sqrt{\nu_G}$. □

Lemma 4.33 (Conductance). *The walk of Definition 4.1 has conductance*

$$\Phi := \inf_{0 < \pi(A) \leq 1/2} \frac{\mathcal{Q}(A, A^c)}{\pi(A)} \geq \frac{c_{\text{iso}} \gamma_0}{128} \ell_{\text{sep}} \geq \frac{c r}{\sqrt{d} (\bar{\nu}_{\text{LS}} + L^2 R^2)}, \quad \ell_{\text{sep}} := \frac{c_{\text{loc}} r}{\sqrt{d} \nu_G} \wedge 1, \quad (56)$$

where $\mathcal{Q}(A, A^c) := \int_A P_x(A^c) \pi(dx)$ is the stationary flow.

Proof. Fix a measurable A with $0 < \pi(A) \leq \frac{1}{2}$, put $b_0 := \gamma_0/8$, and define the deep sets

$$A_1 = \{x \in A : P_x(A^c) \leq b_0\}, \quad A_2 = \{z \in A^c : P_z(A) \leq b_0\}.$$

Because K is full-dimensional, its boundary has Lebesgue and hence π -measure zero. We may intersect A_1, A_2 with K° and put ∂K into A_3 below; stationary masses and flows are unchanged. Thus every pair to which the local and cross-ratio lemmas are applied lies in K° . For $x \in A_1$, $z \in A_2$, $d_{\text{TV}}(P_x, P_z) \geq P_z(A^c) - P_x(A^c) \geq 1 - 2b_0 > 1 - \gamma_0/2$, so by Lemma 4.28 such a pair cannot satisfy Eq. (53): $\|x - z\|_{G(x)} > c_{\text{loc}} r / \sqrt{d}$. Then Lemmas 4.32 and 4.30 give the uniform separation

$$\sigma(x, z) \geq \ell_{\text{sep}} \quad \text{for all } x \in A_1, z \in A_2. \quad (57)$$

Let $A_3 := K \setminus (A_1 \cup A_2)$, $B_1 := A \setminus A_1$, $B_2 := A^c \setminus A_2$, so $A_3 = B_1 \dot{\cup} B_2$. The definitions of B_1, B_2 and reversibility give

$$\mathcal{Q}(A, A^c) \geq b_0 \pi(B_1), \quad \mathcal{Q}(A, A^c) = \mathcal{Q}(A^c, A) \geq b_0 \pi(B_2). \quad (58)$$

Case 1: $\pi(A_1) < \pi(A)/2$ or $\pi(A_2) < \pi(A^c)/2$. Then $\pi(B_1) \geq \pi(A)/2$ or $\pi(B_2) \geq \pi(A^c)/2 \geq \pi(A)/2$, and Eq. (58) gives $\mathcal{Q}(A, A^c)/\pi(A) \geq b_0/2$.

Case 2: $\pi(A_1) \geq \pi(A)/2$ and $\pi(A_2) \geq \pi(A^c)/2 \geq \frac{1}{4}$. Since π is log-concave and Eq. (57) certifies separation $\sigma_0 = \ell_{\text{sep}} \leq 1$, Lemma 3.6 applied to the partition (A_1, A_2, A_3) gives

$$\pi(A_3) \geq c_{\text{iso}} \ell_{\text{sep}} \pi(A_1) \pi(A_2) \geq \frac{c_{\text{iso}} \ell_{\text{sep}}}{8} \pi(A).$$

Adding the two bounds in Eq. (58) and using $A_3 = B_1 \dot{\cup} B_2$, $2 \mathcal{Q}(A, A^c) \geq b_0 \pi(A_3)$, whence $\mathcal{Q}(A, A^c)/\pi(A) \geq b_0 c_{\text{iso}} \ell_{\text{sep}}/16$.

In both cases $\Phi \geq \frac{c_{\text{iso}} \gamma_0}{128} \ell_{\text{sep}}$; substituting $\nu_G \leq \bar{\nu}_{\text{LS}} + 4L^2 R^2$ from Lemma 4.30 and absorbing constants gives Eq. (56). $\square$

5 Proof of Theorem 1.2

5.1 Main theorem

Theorem 5.1 (Spectrahedra; formal version of Theorem 1.2). *Let $K := \{x \in \mathbb{R}^d : A_0 + \sum_{i=1}^d x_i A_i \succeq 0\}$ be bounded, strictly feasible after minimal-face reduction, and contained in $B_2(0, R)$, with $n \times n$ matrix blocks, let f be convex and L -Lipschitz on K , and let $\pi \propto e^{-f} \mathbf{1}_K$. For the log-det metric, let ψ_*^+ be the maximum of one and the supremum, over $x \in K^\circ$, of the spectral norm of the whitened matrix leverage operator at x . Choose κ as in Eq. (74), so $\kappa = \tilde{O}(\psi_*^+)$. Then, for $w \geq 1$ and $0 < \delta < 1$, the exact-metric $\frac{1}{2}$ -lazy Metropolis walk reaches total-variation distance δ from a w -warm start in*

$$\tilde{O}((\psi_*^+ n d + d L^2 R^2) \log(w/\delta))$$

steps.

Proof. Let $f_0 = d\mu_0/d\pi \leq w$; then $\|f_0 - 1\|_{L^2(\pi)}^2 = \int f_0^2 d\pi - 1 \leq w - 1$. The chain is reversible and $\frac{1}{2}$ -lazy, so its spectrum is nonnegative. The Cheeger inequality in Lemma 3.19 gives $\text{gap}(P) \geq c_{\text{Ch}}\Phi^2$.

Cauchy–Schwarz and spectral contraction therefore give

$$d_{\text{TV}}(\mu_0 P^t, \pi) \leq \frac{1}{2} \|P^t(f_0 - 1)\|_{L^2(\pi)} \leq \frac{1}{2} (1 - \text{gap } P)^t \sqrt{w - 1} \leq \frac{1}{2} e^{-c_{\text{Ch}} t \Phi^2} \sqrt{w}.$$

Here the warm-start bound controls the initial $L^2(\pi)$ norm, and the last inequality uses $1 - u \leq e^{-u}$, $\text{gap}(P) \geq c_{\text{Ch}}\Phi^2$, and $\sqrt{w - 1} \leq \sqrt{w}$.

Solving for error δ and inserting Lemmas 5.24 and 5.22 gives

$$t_{\text{mix}}(\delta) \leq C \frac{d\nu_G}{r^2} \log(w/\delta) = C \frac{d(\kappa n + 4L^2 R^2)}{r^2} \log(w/\delta).$$

Here the equality uses $\nu_G = \kappa n + 4L^2 R^2$.

With r universal and $\kappa = \tilde{O}(\psi_*^+)$ from Eq. (74), this is $\tilde{O}((\psi_*^+ n d + dL^2 R^2) \log(w/\delta))$. $\square$

5.2 Normalization, endpoint formulas, and the ledger

This section is exact algebra; no estimates are made.

Fix a deterministic current point $x \in K^\circ$. Replace each A_i ($i \geq 1$) by its slack congruence $S(x)^{-1/2} A_i S(x)^{-1/2}$. This gives an equivalent representation of the same constraint set, since $S(x + u) = S(x)^{1/2} (I + \sum_i u_i S(x)^{-1/2} A_i S(x)^{-1/2}) S(x)^{1/2}$. We then apply the linear change of variables in $\mathbb{R}^d$ that orthonormalizes these congruenced directions in the Frobenius inner product. In the new coordinates, which we use throughout Sections 5.2, 5.3, 5.4 and 5.5:

$$S(x) = I, \quad \bar{A}(u) := \sum_{a=1}^d u_a \bar{A}_a, \quad \|\bar{A}(u)\|_F = \|u\|, \quad H(x) = I_d, \quad (59)$$

where $\{\bar{A}_a\}$ is the orthonormal frame of Definition 3.9 and the third identity is frame orthonormality. The Metropolis ratio R_x is invariant under this substitution: the congruence is an identity on K , and under the linear change of variables both Gaussian densities in the ratio pick up the same Jacobian, while H and G transform as metrics.

After the same change of variables the regularizer becomes a constant positive semidefinite matrix, and we factor

$$G(x) = \kappa(I + \mathcal{C}), \quad \mathcal{C} \succeq 0, \quad Q := (I + \mathcal{C})^{-1}, \quad 0 \preceq Q \preceq I. \quad (60)$$

Only $\mathcal{C} \succeq 0$ is used below, so all estimates are uniform in L and in the orientation of the regularizer. A proposal has the representation

$$Y = x + \eta h, \quad h \sim N(0, Q), \quad \eta = \frac{r}{\sqrt{\kappa d}} = \frac{\rho}{\sqrt{\kappa}}, \quad \rho := \frac{r}{\sqrt{\kappa}}. \quad (61)$$

For a step $g \in \mathbb{R}^d$ write

$$\Delta = \Delta(g) := \eta \bar{A}(g), \quad R = R(g) := (I + \Delta)^{-1} \quad (\text{defined when } \|\Delta\|_{\text{op}} < 1).$$

R and Δ commute. Since $S(x + \eta g) = I + \Delta(g)$ in the normalized coordinates:

Lemma 5.2 (Endpoint identities). *Let g satisfy $\|\Delta(g)\|_{\text{op}} < 1$. Then the entire segment $[x, x + \eta g]$ lies in K° , and with $y = x + \eta g$,*

$$H(y)_{ab} = \text{tr}[\bar{R}\bar{A}_a\bar{R}\bar{A}_b], \quad \bar{E} := H(y) - I, \quad \bar{E}_{ab} = -\text{tr}[\Delta\bar{R}\bar{A}_a\bar{R}\bar{A}_b] - \text{tr}[\Delta\bar{R}\bar{A}_b\bar{R}\bar{A}_a]. \quad (62)$$

Moreover, as functions of h (with $\partial_c := \partial/\partial h_c$ and all matrices evaluated at the same h),

$$\partial_c \bar{R} = -\eta \bar{R} \bar{A}_c \bar{R}, \quad (63)$$

$$\partial_c \bar{E}_{ab} = -\eta [\text{tr}[\bar{R} \bar{A}_c \bar{R} \bar{A}_a \bar{R} \bar{A}_b] + \text{tr}[\bar{R} \bar{A}_a \bar{R} \bar{A}_c \bar{R} \bar{A}_b]], \quad (64)$$

$$\partial_e \partial_c \bar{E}_{ab} = \eta^2 [F_{ecab} + F_{ceab} + F_{caeb} + F_{eacb} + F_{aecb} + F_{aceb}], \quad (65)$$

$$F_{wxyz} := \text{tr}[\bar{R} \bar{A}_w \bar{R} \bar{A}_x \bar{R} \bar{A}_y \bar{R} \bar{A}_z],$$

where each of the four letters a, c, b, e appears exactly once in every word, and F is invariant under cyclic shifts of its four letters.

Proof. Feasibility: $S(x + t\eta g) = I + t\Delta \succ 0$ for $t \in [0, 1]$ when $\|\Delta\|_{\text{op}} < 1$. The metric formula is Eq. (8) after the congruence by $S(y)^{-1/2}$, i.e. $H(y)_{ab} = \text{tr}[S(y)^{-1} \bar{A}_a S(y)^{-1} \bar{A}_b]$ with $S(y)^{-1} = \bar{R}$. For the increment, insert $\bar{R} - I = -\Delta \bar{R}$ twice: $\text{tr}[\bar{R} \bar{A}_a \bar{R} \bar{A}_b] - \text{tr}[\bar{A}_a \bar{A}_b] = \text{tr}[(\bar{R} - I) \bar{A}_a \bar{R} \bar{A}_b] + \text{tr}[\bar{A}_a (\bar{R} - I) \bar{A}_b]$, and use cyclicity on the second term. Eq. (63) is $\partial_c (I + \Delta)^{-1} = -(I + \Delta)^{-1} (\partial_c \Delta) (I + \Delta)^{-1}$ with $\partial_c \Delta = \eta \bar{A}_c$. Differentiating $\bar{E}_{ab} = \text{tr}[\bar{R} \bar{A}_a \bar{R} \bar{A}_b] - \delta_{ab}$ by the product rule across the two $\bar{R}$ -factors gives Eq. (64); differentiating each three-letter trace in Eq. (64) across its three $\bar{R}$ -factors gives Eq. (65)—inserting $-\eta \bar{A}_e$ (with its two flanking $\bar{R}$'s) at each of the three $\bar{R}$ -slots of each of the two words, six four-letter words in total, each carrying the sign $(-\eta)(-\eta) = \eta^2$. In every resulting word the differentiated directions $\bar{A}_c$ and $\bar{A}_e$ occur once, while the original directions $\bar{A}_a$ and $\bar{A}_b$ also occur once; cyclic invariance lets us start the trace at any one of these four letters. This proves the last claim. $\square$

Lemma 5.3 (Ledger). *Let g satisfy $\|\Delta(g)\|_{\text{op}} < 1$ and $y = x + \eta g$. With $E := Q^{1/2} \bar{E} Q^{1/2}$, the complete Gaussian proposal log-ratio satisfies, exactly,*

$$\begin{aligned} \log \frac{q_y(x)}{q_x(y)} &= \frac{1}{2} \{ \log \det(I + E) - \text{tr}[E] \} - \frac{1}{2} Z(g), \\ Z(g) &:= g^\top \bar{E} g - \text{tr}[Q \bar{E}] \\ &= \sum_{a,b} (g_a g_b - Q_{ab}) \bar{E}_{ab}. \end{aligned} \quad (66)$$

Proof. In the normalized coordinates $G(z) = \kappa H(z) + \kappa \mathcal{C}$, so $G(y) - G(x) = \kappa \bar{E}$ and, with $q_z(\cdot) \propto \det G(z)^{1/2} \exp(-\frac{d}{2r^2} \|\cdot - z\|_{G(z)}^2)$,

$$\log \frac{q_y(x)}{q_x(y)} = \frac{1}{2} \log \frac{\det G(y)}{\det G(x)} - \frac{d}{2r^2} (y - x)^\top (G(y) - G(x)) (y - x).$$

For the determinant, $\det G(y)/\det G(x) = \det(I + \mathcal{C} + \bar{E})/\det(I + \mathcal{C}) = \det(I + Q^{1/2} \bar{E} Q^{1/2}) = \det(I + E)$ by Eq. (60). For the quadratic part, $y - x = \eta g$ and $d\eta^2 \kappa / r^2 = 1$ by Eq. (61), so it equals $g^\top \bar{E} g$. Finally $\text{tr}[E] = \text{tr}[Q \bar{E}]$, and adding and subtracting $\frac{1}{2} \text{tr}[E]$ gives Eq. (66). $\square$

The uncentered coupling term $\text{tr}[Q \bar{E}]$ —simultaneously the linear part of the determinant and the mean of the random quadratic form—has cancelled identically. The two braces are controlled in Section 5.3 (determinant remainder) and Sections 5.4 and 5.5 (centered fluctuation Z).

5.3 The spectral tube, stability, and the determinant remainder

Throughout, x is fixed and we work in the normalized coordinates of Section 5.2.

Fix a small universal failure parameter ϵ_0 and set

$$\delta_0 := \epsilon_0^2 d^{-3}, \quad R_0 := \sqrt{d} + \sqrt{2 \log(2/\delta_0)}, \quad \tau_0 := \sqrt{2 \psi(x) \log(8n/\delta_0)}.$$

Define the convex *spectral tube* and its collar

$$\mathcal{K}_0 := \{g \in \mathbb{R}^d : \|g\| \leq R_0, \|\bar{A}(g)\|_{\text{op}} \leq \tau_0\}, \quad \mathcal{K}_0^+ := \mathcal{K}_0 + B_2(0, 1),$$

and set $R_1 := R_0 + 1$, $\tau_1 := \tau_0 + 1$, $M_0 := R_1/\sqrt{d}$. Note $M_0 = O_{\epsilon_0}(1)$ and

$$\tau_1^2 \leq 16 \psi^+(x) (1 + \log(nd/\epsilon_0)), \quad \psi^+(x) := \max(1, \psi(x)). \quad (67)$$

Indeed, $(a+1)^2 \leq 2a^2 + 2$ and, for $n, d \geq 1$ and $0 < \epsilon_0 \leq 1$,

$$\log(8n/\delta_0) = \log(8nd^3/\epsilon_0^2) \leq 3(1 + \log(nd/\epsilon_0)).$$

Substituting the definition of τ_0 gives $\tau_1^2 \leq 12\psi(x)(1 + \log(nd/\epsilon_0)) + 2$, which is bounded by the right side of Eq. (67) because $\psi^+ \geq 1$.

Lemma 5.4 (Tube probability). *For $h \sim N(0, Q)$ with $Q \preceq I$, $\Pr[h \in \mathcal{K}_0] \geq 1 - \delta_0$.*

Proof. Write $h = Q^{1/2}\xi$, ξ standard. The map $\xi \mapsto \|Q^{1/2}\xi\|$ is 1-Lipschitz with mean at most $(\mathbb{E}[\|Q^{1/2}\xi\|^2])^{1/2} = (\text{tr}[Q])^{1/2} \leq \sqrt{d}$. Thus Lemma 3.15, with $t = \sqrt{2 \log(2/\delta_0)}$, gives $\Pr[\|h\| > R_0] \leq e^{-t^2/2} = \delta_0/2$.

For the spectral part, to apply Lemma 3.16, let $\hat{B}_i := \sum_a (Q^{1/2})_{ai} \bar{A}_a$. Then $\bar{A}(h) = \sum_i \xi_i \hat{B}_i$ is a Gaussian matrix series whose variance statistic is

$$\sum_i \hat{B}_i^2 = \sum_{a,b} Q_{ab} \bar{A}_a \bar{A}_b = V^* (Q \otimes I_n) V \preceq V^* V = \Psi(x),$$

where $V : \mathbb{R}^n \rightarrow \mathbb{R}^d \otimes \mathbb{R}^n$ stacks $u \mapsto (\bar{A}_a u)_a$. The inequality follows from $Q \otimes I_n \preceq I_{dn}$ by congruence with V . Hence $\sigma^2 \leq \psi(x)$, and Lemma 3.16 with $t = \tau_0$ gives $\Pr[\|\bar{A}(h)\|_{\text{op}} > \tau_0] \leq 2n e^{-\tau_0^2/2\psi} = 2n \cdot \delta_0/(8n) \leq \delta_0/2$. A union bound with the norm estimate completes the proof. $\square$

Lemma 5.5 (Collar). *Every $g \in \mathcal{K}_0^+$ satisfies $\|g\| \leq R_1$ and $\|\bar{A}(g)\|_{\text{op}} \leq \tau_1$.*

Proof. Write $g = g_0 + e$ with $g_0 \in \mathcal{K}_0$, $\|e\| \leq 1$. Then $\|\bar{A}(e)\|_{\text{op}} \leq \|\bar{A}(e)\|_F = \|e\| \leq 1$ by frame orthonormality Eq. (59). $\square$

We will choose ρ (equivalently $\kappa = r^2/\rho^2$) so that, in particular,

$$\eta \tau_1 \leq \frac{1}{64} \quad (\text{recall } \eta = \rho/\sqrt{d}); \quad (68)$$

this is implied by the final choice Eq. (74) with room to spare. On the collar this makes every endpoint quantity uniformly comparable to its base value—with no bootstrap: the metric is closed-form, so stability is immediate.

Lemma 5.6 (Stability on the collar). *Assume Eq. (68) and let $g \in \mathcal{K}_0^+$, $t \in [0, 1]$, $\Delta = \Delta(tg)$, $R = R(tg)$. Then:*

(i) $\|\Delta\|_{\text{op}} \leq \eta\tau_1 \leq \frac{1}{64}$; the segment $[x, x + \eta g]$ lies in K° ; $R \succ 0$ with $\frac{64}{65}I \preceq R \preceq \frac{64}{63}I$, so $\|R\|_{\text{op}} \leq \frac{64}{63}$ and $\|R\|_{\text{op}}^4 \leq \frac{13}{12}$.

(ii) The eigenvalues of $H(x + t\eta g)$ lie in $[(\frac{64}{65})^2, (\frac{64}{63})^2] \subseteq [e^{-1/16}, e^{1/16}]$.

(iii) $\|\bar{E}\|_{\text{op}} \leq 3\eta\tau_1$ and $\|\bar{E}\|_F \leq 3\eta\tau_1\sqrt{d} = 3\rho\tau_1$, where $\bar{E} = H(x + \eta g) - I$.

Proof. (i) $\|\Delta\|_{\text{op}} = t\eta\|\bar{A}(g)\|_{\text{op}} \leq \eta\tau_1$ by Lemma 5.5; feasibility and the bounds on $R = (I + \Delta)^{-1}$ follow from $\|\Delta\|_{\text{op}} \leq \frac{1}{64}$.

(ii) For any u , $u^\top H(y_t)u = \|R^{1/2}\bar{A}(u)R^{1/2}\|_F^2 \in [\lambda_{\min}(R)^2, \lambda_{\max}(R)^2] \cdot \|u\|^2$ using Eq. (62) and Eq. (59).

(iii) By Eq. (62), $|u^\top \bar{E}v| \leq |\text{tr}[\Delta R \bar{A}(u)R \bar{A}(v)]| + |\text{tr}[\Delta R \bar{A}(v)R \bar{A}(u)]|$. For the first term, $|\text{tr}[\Delta \cdot R \bar{A}(u) \cdot R \bar{A}(v)]| \leq \|\Delta\|_{\text{op}} \|\bar{A}(u)R \bar{A}(v)\|_{S_1} \leq \|\Delta\|_{\text{op}} \|R\|_{\text{op}}^2 \|\bar{A}(u)\|_F \|\bar{A}(v)\|_F$; the second is bounded the same way with one fewer R . With $\|\bar{A}(u)\|_F = \|u\|$ this gives $\|\bar{E}\|_{\text{op}} \leq \|\Delta\|_{\text{op}} (\|R\|_{\text{op}}^2 + \|R\|_{\text{op}}) \leq 3\|\Delta\|_{\text{op}}$. For the Frobenius bound, fix a and apply the frame projection inequality $\sum_b \langle M, \bar{A}_b \rangle_F^2 \leq \|M\|_F^2$ (valid for any $M \in \mathbb{S}^n$ since $\{\bar{A}_b\}$ is orthonormal; replace M by its symmetric part, which the inner products only see) to each of the two terms of $\bar{E}_{ab}$: for the first, cyclicity gives $\text{tr}[\Delta R \bar{A}_a R \bar{A}_b] = \langle \bar{A}_b, \Delta R \bar{A}_a R \rangle_F$ with $M_a = \Delta R \bar{A}_a R$ satisfying $\|M_a\|_F \leq \|\Delta\|_{\text{op}} \|R\|_{\text{op}}^2$, so $\sum_{ab} (\text{tr}[\Delta R \bar{A}_a R \bar{A}_b])^2 \leq \|\Delta\|_{\text{op}}^2 \|R\|_{\text{op}}^4 d$; for the second, fixing b and summing over a with $M'_b = \Delta R \bar{A}_b$, $\|M'_b\|_F \leq \|\Delta\|_{\text{op}} \|R\|_{\text{op}}$, gives $\sum_{ab} (\text{tr}[\Delta R \bar{A}_b R \bar{A}_a])^2 \leq \|\Delta\|_{\text{op}}^2 \|R\|_{\text{op}}^2 d$. Hence $\|\bar{E}\|_F^2 \leq 2\|\Delta\|_{\text{op}}^2 (\|R\|_{\text{op}}^4 + \|R\|_{\text{op}}^2) d \leq 9\|\Delta\|_{\text{op}}^2 d$. $\square$

Lemma 5.7 (Determinant remainder). Assume Eq. (68) and $g \in \mathcal{K}_0^+$. With $E = Q^{1/2} \bar{E} Q^{1/2}$ as in Lemma 5.3, every eigenvalue of $I + E$ lies in $[e^{-1/16}, e^{1/16}]$, and

$$|\log \det(I + E) - \text{tr}[E]| \leq \|E\|_F^2 \leq \|\bar{E}\|_F^2 \leq 9\rho^2\tau_1^2.$$

Proof. Put

$$a := \min\{1, \lambda_{\min}(H(y))\}, \quad b := \max\{1, \lambda_{\max}(H(y))\}.$$

The definitions give $aI \preceq H(y) \preceq bI$ and also $a\mathcal{C} \preceq \mathcal{C} \preceq b\mathcal{C}$, because $a \leq 1 \leq b$ and $\mathcal{C} \succeq 0$. Adding the corresponding inequalities gives

$$a(I + \mathcal{C}) \preceq \mathcal{C} + H(y) \preceq b(I + \mathcal{C}).$$

Congruence by $Q^{1/2} = (I + \mathcal{C})^{-1/2}$ shows $aI \preceq I + E \preceq bI$; no commutativity between $\mathcal{C}$ and $H(y)$ is used. By Lemma 5.6(ii), $[a, b] \subseteq [e^{-1/16}, e^{1/16}]$. On that interval $|\log(1 + \mu) - \mu| \leq \mu^2$ for the eigenvalues μ of E (indeed $|\log(1 + \mu) - \mu| \leq \frac{\mu^2}{2(1-|\mu|)} \leq \mu^2$ for $|\mu| \leq \frac{1}{2}$), and summing over eigenvalues gives the first inequality. Finally, two-sided Frobenius submultiplicativity gives $\|E\|_F = \|Q^{1/2} \bar{E} Q^{1/2}\|_F \leq \|Q^{1/2}\|_{\text{op}}^2 \|\bar{E}\|_F \leq \|\bar{E}\|_F$, and Lemma 5.6(iii) concludes. $\square$

5.4 The Stein mean bound

This section bounds the mean of the centered fluctuation $Z = \sum_{ab} (h_a h_b - Q_{ab}) \bar{E}_{ab}(h)$ from Lemma 5.3. Throughout, Eq. (68) is in force, χ is the cutoff of Lemma 3.13 for the tube $\mathcal{K}_0$, and $\psi = \psi(x)$, $\Psi = \Psi(x)$.

The heart of the argument is the following deterministic estimate, which will absorb every second-derivative term produced by the Stein identity.

Lemma 5.8 (Paired four-trace bound). Let $R \in \mathbb{S}^n$ with $0 \prec R$ and $\|R\|_{\text{op}} \leq \frac{64}{63}$, and let $\{\hat{B}_i\}_{i=1}^N \subset \mathbb{S}^n$ satisfy $\sum_i \hat{B}_i^2 \preceq \Psi$. For each pair (i, j) let $w(i, j) = (w_1, w_2, w_3, w_4)$ be any word consisting of

two copies of $\widehat{B}_i$ and two copies of $\widehat{B}_j$, and set $F_{ij} := \text{tr}[Rw_1Rw_2Rw_3Rw_4]$. Then

$$\sum_{i,j=1}^N |F_{ij}| \leq \|R\|_{\text{op}}^4 \text{tr}[\Psi^2] \leq \frac{13}{12} \psi d.$$

Proof. Insert $R = R^{1/2}R^{1/2}$ around each letter and use cyclicity: $F_{ij} = \text{tr}[\widetilde{N}_{w_1}\widetilde{N}_{w_2}\widetilde{N}_{w_3}\widetilde{N}_{w_4}]$ with $\widetilde{N}_i := R^{1/2}\widehat{B}_iR^{1/2} \in \mathbb{S}^n$. Up to cyclic rotation there are only two arrangements of the multiset $\{\widetilde{N}_i, \widetilde{N}_i, \widetilde{N}_j, \widetilde{N}_j\}$:

$$\text{adjacent: } \text{tr}[\widetilde{N}_i^2\widetilde{N}_j^2] \geq 0,$$

$$\text{alternating: } |\text{tr}[\widetilde{N}_i\widetilde{N}_j\widetilde{N}_i\widetilde{N}_j]| \leq \|\widetilde{N}_i\widetilde{N}_j\|_F^2 = \text{tr}[\widetilde{N}_i^2\widetilde{N}_j^2],$$

The adjacent identity is $\text{tr}[\widetilde{N}_i^2\widetilde{N}_j^2] = \|\widetilde{N}_i\widetilde{N}_j\|_F^2$, while the alternating bound is Frobenius Cauchy–Schwarz applied to $\widetilde{N}_i\widetilde{N}_j$ and its transpose.

Hence in either case $|F_{ij}| \leq \text{tr}[\widetilde{N}_i^2\widetilde{N}_j^2]$. Set $\widehat{\Psi} := \sum_i \widetilde{N}_i^2 = R^{1/2}(\sum_i \widehat{B}_i R \widehat{B}_i)R^{1/2}$.

$$\sum_{ij} \text{tr}[\widetilde{N}_i^2\widetilde{N}_j^2] = \text{tr}[(\sum_i \widetilde{N}_i^2)^2] = \text{tr}[\widehat{\Psi}^2].$$

By Lemma 3.11(ii), $\sum_i \widehat{B}_i R \widehat{B}_i \preceq \|R\|_{\text{op}} \sum_i \widehat{B}_i^2 \preceq \|R\|_{\text{op}} \Psi$, so $0 \preceq \widehat{\Psi} \preceq \|R\|_{\text{op}} R^{1/2} \Psi R^{1/2}$, and by Lemma 3.11(i) then (iii),

$$\text{tr}[\widehat{\Psi}^2] \leq \|R\|_{\text{op}}^2 \text{tr}[(R^{1/2} \Psi R^{1/2})^2] = \|R\|_{\text{op}}^2 \|R^{1/2} \Psi R^{1/2}\|_F^2 \leq \|R\|_{\text{op}}^4 \|\Psi\|_F^2 = \|R\|_{\text{op}}^4 \text{tr}[\Psi^2].$$

Here the first inequality is Loewner monotonicity of $\text{tr}[X^2]$ for positive-semidefinite matrices, and the second is two-sided Frobenius submultiplicativity; the equalities use $\|M\|_F^2 = \text{tr}[M^2]$ for symmetric M .

Finally $\text{tr}[\Psi^2] \leq \|\Psi\|_{\text{op}} \text{tr}[\Psi] = \psi d$ by Lemma 3.11(ii) and Lemma 3.10, and $(\frac{64}{63})^4 \leq \frac{13}{12}$. $\square$

Lemma 5.9 (Stein integrand bound). *For every h in the collar $\mathcal{K}_0^+$,*

$$|\sum_{a,b,c,e} Q_{ac}Q_{be} \partial_e \partial_c \bar{E}_{ab}(h)| \leq 7\eta^2 \psi d = 7\rho^2 \psi. \quad (69)$$

Proof. Diagonalize $Q = \sum_i \lambda_i q_i q_i^\top$ with $\lambda_i \in [0, 1]$, $\{q_i\}$ orthonormal, and set $\widehat{B}_i := \sqrt{\lambda_i} \bar{A}(q_i)$. Then

$$\sum_i \widehat{B}_i^2 = \sum_i \lambda_i \bar{A}(q_i)^2 \preceq \sum_i \bar{A}(q_i)^2 = \Psi$$

by $0 \leq \lambda_i \leq 1$ and frame independence (Lemma 3.10) applied to the orthonormal basis $\{q_i\}$.

Now contract Eq. (65): in each of the six words the letters a, c appear once each and b, e once each, so the contraction by $Q_{ac}Q_{be}$ replaces the pair $\{a, c\}$ by two copies of $\widehat{B}_i$ (summed over i) and the pair $\{b, e\}$ by two copies of $\widehat{B}_j$ (summed over j), preserving each letter's slot. Each of the six words therefore contributes $\sum_{ij} F_{ij}$ for some fixed word arrangement of $(\widehat{B}_i, \widehat{B}_i, \widehat{B}_j, \widehat{B}_j)$, and Lemma 5.8 (with $\|R\|_{\text{op}} \leq \frac{64}{63}$ from Lemma 5.6(i)) bounds each contribution by $\frac{13}{12}\psi d$. Multiplying by the prefactor η^2 and by $6 \cdot \frac{13}{12} \leq 7$ gives Eq. (69). $\square$

Apply Lemma 3.12 to $F = \bar{E}_{ab}\chi$ and sum over (a, b) . This is valid because $\bar{E}_{ab}$ is smooth on an open neighborhood of the compact $\mathcal{K}_0^+$ (the resolvent exists there by Lemma 5.6(i)), while χ is $W^{2,\infty}$ and supported in $\mathcal{K}_0^+$, so the product extends by zero:

$$\mathbb{E}[Z\chi] = \underbrace{\mathbb{E}[\chi \cdot \sum_{abce} Q_{ac}Q_{be} \partial_e \partial_c \bar{E}_{ab}]}_{\text{main}} + \underbrace{2\mathbb{E}[\sum_{be} Q_{be} \partial_e \chi \cdot G_b] + \mathbb{E}[\langle Q \bar{E} Q, D^2 \chi \rangle_F]}_{\text{collar terms}}, \quad (70)$$

where $G_b := \sum_{ac} Q_{ac} \partial_c \bar{E}_{ab}$ and we used the symmetry of $\bar{E}$ and of the two cross terms.

Lemma 5.10 (First-derivative contraction). *On $\mathcal{K}_0^+$, with $\hat{B}_i$ as above and $Y := \sum_i \hat{B}_i \hat{R} \hat{B}_i$,*

$$\begin{aligned} G_b &= -2\eta \langle \bar{A}_b, \mathbf{R} Y \mathbf{R} \rangle_F, \\ 0 &\preceq Y \preceq \|\mathbf{R}\|_{\text{op}} \Psi, \quad \text{tr}[Y] \leq \|\mathbf{R}\|_{\text{op}} d, \\ \left(\sum_b G_b^2\right)^{1/2} &\leq 3\eta \sqrt{\psi d} = 3\rho \sqrt{\psi}. \end{aligned}$$

Proof. Contract Eq. (64) with Q_{ac} : in both three-letter words the letters a and c occupy the two slots preceding $\bar{A}_b$, so after the eigen-substitution both words equal $\sum_i \text{tr}[\mathbf{R} \hat{B}_i \mathbf{R} \hat{B}_i \mathbf{R} \bar{A}_b]$. A cyclic shift gives

$$\text{tr}[\mathbf{R} \hat{B}_i \mathbf{R} \hat{B}_i \mathbf{R} \bar{A}_b] = \langle \bar{A}_b, \mathbf{R} \hat{B}_i \mathbf{R} \hat{B}_i \mathbf{R} \rangle_F,$$

and summing over i yields the stated formula. The bounds on Y are Lemma 3.11(ii) with $\sum_i \hat{B}_i^2 \preceq \Psi$ and $\text{tr}[\Psi] = d$. Finally, the frame projection inequality gives

$$\sum_b G_b^2 \leq 4\eta^2 \|\mathbf{R} Y \mathbf{R}\|_F^2 \leq 4\eta^2 \|\mathbf{R}\|_{\text{op}}^4 \|Y\|_F^2.$$

Since $Y \succeq 0$, Lemma 3.11 gives

$$\|Y\|_F^2 = \text{tr}[Y^2] \leq \|Y\|_{\text{op}} \text{tr}[Y] \leq \|\mathbf{R}\|_{\text{op}}^2 \psi d,$$

where the last inequality uses the preceding bounds on Y .

Hence $(\sum_b G_b^2)^{1/2} \leq 2\eta \|\mathbf{R}\|_{\text{op}}^3 \sqrt{\psi d} \leq 3\eta \sqrt{\psi d}$. $\square$

Lemma 5.11 (Mean bound). *Assume Eq. (68). Then*

$$|\mathbb{E}[Z\chi]| \leq 7\rho^2\psi + \delta_0(72\rho\sqrt{\psi} + 3C_0\rho\tau_1\sqrt{d}) \leq 7\rho^2\psi + \epsilon_0.$$

Proof. The main term of Eq. (70) is bounded pointwise by Lemma 5.9 (and $0 \leq \chi \leq 1$), giving $7\rho^2\psi$. The collar terms have integrands supported on $\mathcal{K}_0^+ \setminus \mathcal{K}_0$, an event of probability at most δ_0 by Lemma 5.4, and are bounded uniformly there: by Lemmas 3.13 and 5.10, $|2 \sum_{be} Q_{be} \partial_e \chi G_b| \leq 2\|\nabla \chi\| (\sum_b G_b^2)^{1/2} \leq 72\rho\sqrt{\psi}$; and by Lemma 5.6(iii), $|\langle Q \bar{E} Q, D^2 \chi \rangle_F| \leq \|D^2 \chi\|_{\text{op}} \|Q \bar{E} Q\|_{S_1} \leq C_0 \sqrt{d} \|\bar{E}\|_F \leq 3C_0 \rho \tau_1 \sqrt{d}$. Multiplying these uniform bounds by δ_0 and using $\sqrt{\psi} \leq \tau_1$, the smallness assumption gives both $\rho\sqrt{\psi} \leq \sqrt{d}/64$ and $\rho\tau_1\sqrt{d} \leq d/64$. With $\delta_0 = \epsilon_0^2 d^{-3}$, the two collar contributions are at most

$$\frac{72}{64} \epsilon_0^2 d^{-5/2} \quad \text{and} \quad \frac{3C_0}{64} \epsilon_0^2 d^{-2}.$$

Their sum is at most ϵ_0 by the universal choice of ϵ_0 in Section 5.5. $\square$

5.5 Fluctuation bound, parameter choice, and the acceptance theorem

Lemma 5.12 (Gradient of Z). *Assume Eq. (68). For h in a neighborhood of $\mathcal{K}_0$,*

$$\nabla Z(h) = 2 \bar{E}(h) h + W(h), \quad W_c(h) = -2\eta \langle \bar{A}_c, \mathbf{R}[\bar{A}(h) \mathbf{R} \bar{A}(h) - Y] \mathbf{R} \rangle_F, \quad (71)$$

with $Y = \sum_i \hat{B}_i \hat{R} \hat{B}_i$ as in Lemma 5.10, and on $\mathcal{K}_0$,

$$\|\nabla Z\| \leq 9\rho\tau_1 M_0 + 3\rho\sqrt{\psi} \leq 12\rho\tau_1 M_0 =: \Lambda_Z. \quad (72)$$

Consequently Z is Λ_Z -Lipschitz on the convex set $\mathcal{K}_0$.

Proof. Differentiating $Z = \sum_{ab}(h_a h_b - Q_{ab})\bar{E}_{ab}$ gives two contributions. By symmetry of $\bar{E}$, the product-rule term from $h_a h_b$ is $2(\bar{E}h)_c$, while the term from $\bar{E}_{ab}$ is $W_c = \sum_{ab}(h_a h_b - Q_{ab})\partial_c \bar{E}_{ab}$. For the h -part, contract Eq. (64) with $h_a h_b$: both three-letter words become $\text{tr}[\bar{R}\bar{A}(h)\bar{R}\bar{A}(h)\bar{R}\bar{A}_c]$ after a cyclic shift (exactly as in Lemma 5.10 with $\hat{B}_i$ replaced by $\bar{A}(h)$), giving $-2\eta\langle \bar{A}_c, \bar{R}\bar{A}(h)\bar{R}\bar{A}(h)\bar{R} \rangle_F$; the Q -part is $-G_c = +2\eta\langle \bar{A}_c, \bar{R}Y\bar{R} \rangle_F$ by Lemma 5.10. This is Eq. (71).

For the norm bound, the frame projection inequality, Lemma 3.11(iii), the estimates $\|\bar{A}(h)\|_{\text{op}} \leq \tau_0$ and $\|\bar{A}(h)\|_F = \|h\| \leq R_0$, and the bounds on Y from Lemma 5.10 give

$$\begin{aligned} \|W\| &\leq 2\eta\|R\|_{\text{op}}^2(\|\bar{A}(h)\bar{R}\bar{A}(h)\|_F + \|Y\|_F) \\ &\leq 2\eta\|R\|_{\text{op}}^3(\tau_0 R_0 + \sqrt{\psi d}) \\ &\leq 3\eta(\tau_1 R_1 + \sqrt{\psi d}). \end{aligned}$$

The first inequality also uses the triangle inequality, the second uses Frobenius submultiplicativity, and the last uses $\|R\|_{\text{op}} \leq 64/63$, $\tau_0 \leq \tau_1$, and $R_0 \leq R_1$.

Also, $\|2\bar{E}h\| \leq 2\|\bar{E}\|_{\text{op}}\|h\| \leq 6\eta\tau_1 R_1$ by Lemma 5.6(iii). Since $\eta R_1 = \rho M_0$ and $\eta\sqrt{\psi d} = \rho\sqrt{\psi}$, the first bound in Eq. (72) follows; the second uses $\sqrt{\psi} \leq \tau_0 \leq \tau_1$ (from $\tau_0^2 = 2\psi \log(8n/\delta_0) \geq \psi$) and $M_0 \geq 1$. A gradient bound on a convex set is a Lipschitz bound. $\square$

Lemma 5.13 (Uniform size on the collar). *On $\mathcal{K}_0^+$, $|Z| \leq 3\rho\tau_1\sqrt{d}(M_0^2 + 1) \leq 6\rho\tau_1\sqrt{d}M_0^2 =: S_Z$.*

Proof. Lemma 5.6(iii), together with the operator-norm and trace-duality bounds, gives

$$\begin{aligned} |h^\top \bar{E}h| &\leq \|\bar{E}\|_{\text{op}}\|h\|^2 \leq 3\eta\tau_1 R_1^2 = 3\rho\tau_1\sqrt{d}M_0^2, \\ |\text{tr}[Q\bar{E}]| &\leq \|Q\|_{\text{op}}\|\bar{E}\|_{S_1} \leq \sqrt{d}\|\bar{E}\|_F \leq 3\rho\tau_1\sqrt{d}. \end{aligned}$$

In the first line we used Lemma 5.6(iii) and $\eta R_1^2 = \rho\sqrt{d}M_0^2$; in the second we used $\|\bar{E}\|_{S_1} \leq \sqrt{d}\|\bar{E}\|_F$, $\|Q\|_{\text{op}} \leq 1$, and again Lemma 5.6(iii). $\square$

Lemma 5.14 (Fluctuation bound). *Assume Eq. (68). With probability at least $1 - \delta_0 - \epsilon_0$ over $h \sim N(0, Q)$,*

$$|Z| \leq 7\rho^2\psi + \Lambda_Z\sqrt{2\log(2/\epsilon_0)} + \epsilon_0 + 40\rho\tau_1 M_0^2\sqrt{d}\sqrt{\delta_0}. \quad (73)$$

Proof. Let $\tilde{Z}$ be the McShane extension (Lemma 3.14) of $Z|_{\mathcal{K}_0}$ with constant Λ_Z . Since $\tilde{Z} = Z\chi = Z$ on $\mathcal{K}_0$ and $\Pr[h \notin \mathcal{K}_0] \leq \delta_0$, the triangle inequality and Cauchy–Schwarz give

$$|\mathbb{E}[\tilde{Z}] - \mathbb{E}[Z\chi]| \leq \mathbb{E}[(|\tilde{Z}| + |Z\chi|)\mathbf{1}_{h \notin \mathcal{K}_0}] \leq (\mathbb{E}[(|\tilde{Z}| + |Z\chi|)^2])^{1/2}\sqrt{\delta_0}.$$

On all of $\mathbb{R}^d$, $|\tilde{Z}(h)| \leq S_Z + \Lambda_Z(\|h\| + R_0)$ (Lipschitz from any point of $\mathcal{K}_0$) and $|Z\chi| \leq S_Z$ (support in $\mathcal{K}_0^+$ plus Lemma 5.13). Minkowski's inequality and $\mathbb{E}[\|h\|^2] \leq d$ therefore give $|\mathbb{E}[\tilde{Z}] - \mathbb{E}[Z\chi]| \leq (2S_Z + \Lambda_Z R_0 + \Lambda_Z\sqrt{d})\sqrt{\delta_0} \leq 36\rho\tau_1 M_0^2\sqrt{d}\sqrt{\delta_0}$, where $R_0 \leq R_1 = M_0\sqrt{d}$ and $M_0 \geq 1$. Relaxing 36 to 40 and combining with Lemma 5.11, $|\mathbb{E}[\tilde{Z}]| \leq 7\rho^2\psi + \epsilon_0 + 40\rho\tau_1 M_0^2\sqrt{d}\sqrt{\delta_0}$.

The map $\xi \mapsto \tilde{Z}(Q^{1/2}\xi)$ is Λ_Z -Lipschitz because $\|Q^{1/2}\|_{\text{op}} \leq 1$. Therefore the two-sided form of Lemma 3.15 gives $\Pr[|\tilde{Z} - \mathbb{E}[\tilde{Z}]| > \Lambda_Z\sqrt{2\log(2/\epsilon_0)}] \leq \epsilon_0$. On the intersection of the Gaussian-deviation event

$$\mathcal{C}_Z := \{|\tilde{Z} - \mathbb{E}[\tilde{Z}]| \leq \Lambda_Z\sqrt{2\log(2/\epsilon_0)}\}$$

with the tube event $\mathcal{T} := \{h \in \mathcal{K}_0\}$, we have $Z = \tilde{Z}$ and Eq. (73). Here $\Pr[\mathcal{C}_Z^c] \leq \epsilon_0$ and $\Pr[\mathcal{T}^c] \leq \delta_0$. Both events are defined under the same Gaussian law, so a direct union bound gives the stated probability. $\square$

Fix once and for all a sufficiently small universal $r \in (0, 1]$ and the universal failure level

$$\epsilon_0 := \min\left\{\frac{1}{40}, \frac{32}{72 + 3C_0}\right\},$$

where C_0 is the cutoff constant in Lemma 3.13, and set

$$\rho := \frac{1}{\Gamma_0 M_0 \sqrt{\psi_*^+} (1 + \log(nd/\epsilon_0))}, \quad \kappa := \frac{r^2}{\rho^2} = \Gamma_0^2 r^2 M_0^2 \psi_*^+ (1 + \log(nd/\epsilon_0))^2, \quad (74)$$

for a sufficiently large universal constant Γ_0 . Enlarge Γ_0 if necessary so that $\kappa \geq 1$. Then $\kappa = \tilde{O}(\psi_*^+)$, and using Eq. (67) together with $\psi \leq \psi_*^+$:

- $\eta\tau_1 = \rho\tau_1/\sqrt{d} \leq 4/(\Gamma_0\sqrt{d}) \leq \frac{1}{64}$, so Eq. (68) holds;
- $7\rho^2\psi \leq 7/\Gamma_0^2$ and $9\rho^2\tau_1^2 \leq 405/\Gamma_0^2$ (Lemma 5.7);
- $\Lambda_Z \sqrt{2\log(2/\epsilon_0)} = 12\rho\tau_1 M_0 \sqrt{2\log(2/\epsilon_0)} \leq C/\Gamma_0$, and $40\rho\tau_1 M_0^2 \sqrt{d}\sqrt{\delta_0} \leq C\epsilon_0/\Gamma_0$.

Taking Γ_0 large enough (universally, since ϵ_0 is fixed) makes each displayed quantity at most $\frac{1}{10}$; we assume this henceforth.

Lemma 5.15 (Uniform lower tail for the complete ratio). *With the choice Eq. (74) there are universal constants $C_{\text{acc}} < \infty$ and $\epsilon \leq \frac{1}{10}$ such that, uniformly at every deterministic $x \in K^\circ$,*

$$\Pr_{Y \sim q_x} [Y \in K^\circ \text{ and } \log R_x(Y) \geq -C_{\text{acc}}] \geq 1 - \epsilon. \quad (75)$$

The constants are independent of x, n, d, L, R and of the orientation of the regularizer.

Proof. Work in normalized coordinates (the ratio is invariant, Section 5.2) and write the proposal as $Y = x + \eta h$, $h = Q^{1/2}\xi$. Intersect three events: the tube event $\mathcal{T} = \{h \in \mathcal{K}_0\}$ (Lemma 5.4; failure at most δ_0), the Gaussian-deviation event $\mathcal{C}_Z$ defined in the proof of Lemma 5.14 (failure at most ϵ_0), and the norm event $\{\|\xi\| \leq \sqrt{d} + \sqrt{2\log(1/\epsilon_0)}\}$ (probability $\geq 1 - \epsilon_0$ by the one-sided form of Lemma 3.15, since $\xi \mapsto \|\xi\|$ is 1-Lipschitz and $\mathbb{E}[\|\xi\|] \leq \sqrt{d}$). The total failure probability is at most $\epsilon := \epsilon_0^2 + 2\epsilon_0 \leq \frac{1}{10}$, since $\delta_0 = \epsilon_0^2 d^{-3} \leq \epsilon_0^2$.

Feasibility. On the tube event $\|\Delta(h)\|_{\text{op}} \leq \eta\tau_0 < 1$, so $Y \in K^\circ$ and the whole segment is interior (Lemma 5.2).

Proposal ratio. By Lemma 5.3, $\log \frac{q_Y(x)}{q_x(Y)} = \frac{1}{2} \{\log \det(I + E) - \text{tr}[E]\} - \frac{1}{2}Z$, and on the intersected event Lemma 5.7 and Lemma 5.14 with the parameter choice give $|\log \det(I + E) - \text{tr}[E]| \leq \frac{1}{10}$ and $|Z| \leq \frac{4}{10}$, so the proposal log-ratio is at least $-\frac{1}{2}$.

Target ratio. If $L = 0$, then f is constant on the convex body K , and the target log-ratio is zero. If $L > 0$, then in the original coordinates $G(x) \succeq L^2 I$, so $\|Y - x\|_2 \leq L^{-1} \|Y - x\|_{G(x)} = L^{-1} \frac{r}{\sqrt{d}} \|\xi\|$. On the norm event, Lipschitzness and the proposal representation give

$$f(Y) - f(x) \leq L \|Y - x\|_2 \leq \frac{r}{\sqrt{d}} (\sqrt{d} + \sqrt{2\log(1/\epsilon_0)}) \leq C_{\epsilon_0} r, \quad C_{\epsilon_0} := 1 + \sqrt{2\log(1/\epsilon_0)}.$$

Hence $\log R_x(Y) \geq -C_{\epsilon_0} r - \frac{1}{2} \geq -C_{\text{acc}}$ with $C_{\text{acc}} := C_{\epsilon_0} r + 1$, a universal constant because ϵ_0 is fixed. $\square$

Remark 5.16. Eq. (75) holds at every interior point with no dependence on the distance to ∂K : the analysis is affine-invariant and uses only $\mathcal{C} \succeq 0$. No boundary-layer or s -conductance device will be needed.

5.6 From acceptance to mixing

We now return to the original coordinates and follow the geometric half of the polytope argument, using the cross-ratio and Cheeger lemmas of Section 3.7.

Lemma 5.17 (Operator self-concordance). *For all $z \in K^\circ$ and $u \in \mathbb{R}^d$,*

$$\|H(z)^{-1/2} DH(z)[u] H(z)^{-1/2}\|_{\text{op}} \leq 2 \|u\|_{H(z)}, \quad (76)$$

and the same holds for $G = \kappa H + L^2 I$ in the G -norms. Consequently Lemma 3.17 applies to G : if $\Delta := \|x - z\|_{G(x)} < 1$ then $(1 - \Delta)^2 G(x) \preceq G(z) \preceq (1 - \Delta)^{-2} G(x)$.

Proof. Normalize at z as in Section 5.2, so $H(z) = I$ and directions v correspond to $\bar{A}(v)$ with $\|\bar{A}(v)\|_F = \|v\|$. Since $H = \nabla^2 \phi$ with $\phi = -\log \det S$, $v^\top DH(z)[u] v = D^3 \phi(z)[u, v, v] = -2 \text{tr}[\bar{A}(u) \bar{A}(v)^2]$ in these coordinates. The PSD trace bound and the normalized-coordinate isometry give

$$|\text{tr}[\bar{A}(u) \bar{A}(v)^2]| \leq \|\bar{A}(u)\|_{\text{op}} \text{tr}[\bar{A}(v)^2] \leq \|\bar{A}(u)\|_F \|\bar{A}(v)\|_F^2 = \|u\| \|v\|^2,$$

Indeed, Lemma 3.11(ii) applies with $M = \bar{A}(v)^2$, while $\|\bar{A}(u)\|_{\text{op}} \leq \|\bar{A}(u)\|_F$ and $\text{tr}[\bar{A}(v)^2] = \|\bar{A}(v)\|_F^2$.

Taking the supremum over unit v gives Eq. (76). For G , scaling by $\kappa \geq 1$ preserves the inequality (the left side is scale-invariant, the right side grows), and with $H_\kappa := \kappa H$, $C := H_\kappa^{1/2} G^{-1/2}$ ($\|C\|_{\text{op}} \leq 1$ since $H_\kappa \preceq G$), $X := H_\kappa^{-1/2} DH_\kappa[u] H_\kappa^{-1/2}$, we have

$$\|G^{-1/2} DG[u] G^{-1/2}\|_{\text{op}} = \|C^\top X C\|_{\text{op}} \leq \|X\|_{\text{op}} \leq 2 \|u\|_{H_\kappa} \leq 2 \|u\|_G,$$

Here the equality uses $DG = DH_\kappa$ and the definitions of C and X ; the inequalities use $\|C\|_{\text{op}} \leq 1$, self-concordance of H_κ , and $H_\kappa \preceq G$, respectively.

The distortion conclusion is the elementary ODE lemma Lemma 3.17; its domain hypothesis holds because the segment $[x, z]$ lies in the convex K° . $\square$

Lemma 5.18 (Pinsker's inequality). *For probability densities p, q ,*

$$d_{\text{TV}}(p, q) \leq \sqrt{D_{\text{KL}}(p\|q)/2}.$$

Proof. Let $A = \{p \geq q\}$, $a = \int_A p$, and $b = \int_A q$. Then $a - b = d_{\text{TV}}(p, q)$. Jensen's inequality on A and A^c (equivalently, the two-term log-sum inequality) gives

$$D_{\text{KL}}(p\|q) \geq a \log \frac{a}{b} + (1 - a) \log \frac{1 - a}{1 - b} =: D_{\text{Ber}}(a\|b).$$

For fixed $b \in (0, 1)$, the second derivative in a of $D_{\text{Ber}}(a\|b) - 2(a - b)^2$ is $1/[a(1 - a)] - 4 \geq 0$, and the function and its first derivative vanish at $a = b$. Hence $D_{\text{Ber}}(a\|b) \geq 2(a - b)^2$; the endpoint cases follow by limits. Thus $D_{\text{KL}}(p\|q) \geq 2d_{\text{TV}}(p, q)^2$. $\square$

Lemma 5.19 (Proposal overlap). *There is a universal $c_{\text{loc}} > 0$ such that, with $r \leq 1$,*

$$\|x - z\|_{G(x)} \leq \frac{c_{\text{loc}} r}{\sqrt{d}} \implies d_{\text{TV}}(q_x, q_z) \leq \delta_q := \frac{1}{2}. \quad (77)$$

Proof. Write $\Delta = \|x - z\|_{G(x)} \leq \frac{1}{2}$. The exact KL divergence between the Gaussians $q_x = N(x, \frac{r^2}{d}G(x)^{-1})$ and $q_z = N(z, \frac{r^2}{d}G(z)^{-1})$ is

$$D_{\text{KL}}(q_x \| q_z) = \frac{1}{2} \left[\sum_{j=1}^d (\lambda_j - 1 - \log \lambda_j) + \frac{d}{r^2} \|x - z\|_{G(z)}^2 \right],$$

with λ_j the eigenvalues of $G(x)^{-1/2}G(z)G(x)^{-1/2}$. This formula follows by integrating the log-density ratio: the centered quadratic term gives the trace of the relative covariance, the mean displacement gives $d\|x - z\|_{G(z)}^2/r^2$, and the normalizers give the log determinant. By Lemma 5.17, $\lambda_j \in [(1 - \Delta)^2, (1 - \Delta)^{-2}]$, so $\lambda_j - 1 - \log \lambda_j \leq C\Delta^2$ per eigenvalue and $\|x - z\|_{G(z)}^2 \leq (1 - \Delta)^{-2}\Delta^2 \leq 4\Delta^2$. With $\Delta \leq c_{\text{loc}}r/\sqrt{d}$ and $r \leq 1$, $D_{\text{KL}} \leq C(d\Delta^2 + d\Delta^2/r^2) \leq C'c_{\text{loc}}^2$, and the locally proved Pinsker inequality Lemma 5.18 gives $d_{\text{TV}} \leq \sqrt{D_{\text{KL}}/2}$ and hence Eq. (77) for small c_{loc} . $\square$

Lemma 5.20 (Kernel overlap). *Suppose Eq. (75) holds at x and z and $d_{\text{TV}}(q_x, q_z) \leq \delta_q = \frac{1}{2}$. Then the $\frac{1}{2}$ -lazy kernels satisfy*

$$d_{\text{TV}}(P_x, P_z) \leq 1 - \frac{\gamma_0}{2}, \quad \gamma_0 := e^{-C_{\text{acc}}}(1 - \delta_q - 2\epsilon) \geq e^{-C_{\text{acc}}} \cdot \frac{3}{10} > 0.$$

Proof. Let $E_x := \{u \in K^\circ : R_x(u) \geq e^{-C_{\text{acc}}}\}$ and likewise E_z ; by Lemma 5.15, $q_x(E_x^c) \leq \epsilon$ and $q_z(E_z^c) \leq \epsilon$. The moving part of the non-lazy kernel at x has density $u \mapsto q_x(u)(1 \wedge R_x(u))$ on K . On $E_x \cap E_z$, both acceptance ratios are at least $e^{-C_{\text{acc}}}$, while $\min\{q_x, q_z\}$ has total mass at least $1 - \delta_q$ and loses at most ϵ on each of E_x^c and E_z^c . Hence the two moving parts have common mass at least

$$e^{-C_{\text{acc}}} \int_{E_x \cap E_z} \min\{q_x, q_z\} du \geq e^{-C_{\text{acc}}}(1 - \delta_q - \epsilon - \epsilon) = \gamma_0,$$

Laziness retains half the common moving mass, and $d_{\text{TV}} \leq 1 - (\text{overlap})$. Finally $1 - \delta_q - 2\epsilon \geq 1 - \frac{1}{2} - \frac{2}{10} = \frac{3}{10}$. $\square$

Definition 5.21 (Chord cross ratio). *For distinct $x, z \in K^\circ$, let their line meet ∂K in the order $p_\partial, x, z, q_\partial$ and put $\alpha = \|x - p_\partial\|$, $\beta = \|z - x\|$, $\gamma = \|q_\partial - z\|$ (Euclidean). Then $\sigma(x, z) := \beta(\alpha + \beta + \gamma)/(\alpha\gamma)$.*

Lemma 5.22 (Symmetry). *Write $\mathcal{D}_M(x, t) = \{z : \|z - x\|_{M(x)} \leq t\}$. Then, for every $x \in K^\circ$,*

$$\mathcal{D}_G(x, 1) \subseteq K \cap (2x - K) \subseteq \mathcal{D}_G(x, \sqrt{\nu_G}), \quad \nu_G := \kappa n + 4L^2 R^2.$$

Proof. Work at x in normalized coordinates, where $\|z - x\|_{H(x)} = \|\Delta\|_F$ for $\Delta :=$ the normalized block displacement of z , and $z \in K \cap (2x - K)$ iff $I \pm \Delta \succeq 0$ iff $\|\Delta\|_{\text{op}} \leq 1$. *Inner:* if $\|z - x\|_{G(x)} \leq 1$ then, since $G \succeq \kappa H$ and $\kappa \geq 1$, $\|\Delta\|_F \leq \kappa^{-1/2} \leq 1$, hence $\|\Delta\|_{\text{op}} \leq 1$. *Outer:* if $\|\Delta\|_{\text{op}} \leq 1$ then $\|z - x\|_{\kappa H(x)}^2 = \kappa \|\Delta\|_F^2 \leq \kappa n \|\Delta\|_{\text{op}}^2 \leq \kappa n$, and $L^2 \|z - x\|_2^2 \leq 4L^2 R^2$ because $K \subseteq B_2(0, R)$ has diameter $2R$. Adding these two bounds proves the outer inclusion. $\square$

Lemma 5.23 (Cross ratio dominates the local norm). *For all distinct $x, z \in K^\circ$, $\sigma(x, z) \geq \|x - z\|_{G(x)}/\sqrt{\nu_G}$.*

Proof. Let $e = (z - x)/\beta$. The radius of the symmetric body $K \cap (2x - K)$ at x along $\pm e$ is $t = \min(\alpha, \beta + \gamma)$, so $x + te$ lies in its closure and Lemma 5.22 gives $t\|e\|_{G(x)} \leq \sqrt{\nu_G}$. Also $\sigma(x, z) \geq \beta/t$: if $t = \alpha$ this is $(\alpha + \beta + \gamma)/\gamma \geq 1$ times β/α ; if $t = \beta + \gamma$ use $(\alpha + \beta + \gamma)(\beta + \gamma) \geq \alpha\gamma$. Hence $\sigma \geq \beta\|e\|_{G(x)}/\sqrt{\nu_G} = \|z - x\|_{G(x)}/\sqrt{\nu_G}$. $\square$

Lemma 5.24 (Conductance). *The walk of Definition 3.8 has conductance*

$$\Phi := \inf_{0 < \pi(A) \leq 1/2} \frac{\mathcal{Q}(A, A^c)}{\pi(A)} \geq \frac{c_{\text{iso}} \gamma_0}{128} \ell_{\text{sep}}, \quad \ell_{\text{sep}} := \min(1, \frac{c_{\text{loc}} r}{\sqrt{d} \nu_G}). \quad (78)$$

Here $\mathcal{Q}(A, A^c) := \int_A P_x(A^c) \pi(dx)$ is stationary flow.

Proof. Fix measurable A with $0 < \pi(A) \leq \frac{1}{2}$ and put $b_0 := \gamma_0/8$,

$$A_1 = \{x \in A : P_x(A^c) \leq b_0\}, \quad A_2 = \{z \in A^c : P_z(A) \leq b_0\}, \quad A_3 = K \setminus (A_1 \cup A_2).$$

Since $\pi(\partial K) = 0$, from now on intersect A_1, A_2 with K° and absorb ∂K into A_3 ; this changes neither mass nor stationary flow and makes every pair below an interior pair. For $x \in A_1, z \in A_2$: $d_{\text{TV}}(P_x, P_z) \geq P_z(A^c) - P_x(A^c) \geq 1 - 2b_0 > 1 - \gamma_0/2$, so by Lemmas 5.20 and 5.19 the pair must violate the closeness condition: $\|x - z\|_{G(x)} > c_{\text{loc}} r / \sqrt{d}$, whence $\sigma(x, z) \geq \ell_{\text{sep}}$ by Lemma 5.23. Let $B_1 = A \setminus A_1$, $B_2 = A^c \setminus A_2$, so $A_3 = B_1 \dot{\cup} B_2$; reversibility gives $\mathcal{Q}(A, A^c) \geq b_0 \pi(B_1)$ and $\mathcal{Q}(A, A^c) = \mathcal{Q}(A^c, A) \geq b_0 \pi(B_2)$ for the stationary flow $\mathcal{Q}$.

If $\pi(A_1) < \pi(A)/2$ or $\pi(A_2) < \pi(A^c)/2$, then $\pi(B_1) \geq \pi(A)/2$ or $\pi(B_2) \geq \pi(A^c)/2 \geq \pi(A)/2$, and $\mathcal{Q}(A, A^c)/\pi(A) \geq b_0/2$. Otherwise $\pi(A_1) \geq \pi(A)/2$ and $\pi(A_2) \geq \frac{1}{4}$; π is log-concave (f convex) on the convex body K , so all hypotheses of the cross-ratio isoperimetry result in Lemma 3.18 hold. Applying it to (A_1, A_2, A_3) with certified separation $\ell_{\text{sep}} \leq 1$ gives $\pi(A_3) \geq c_{\text{iso}} \ell_{\text{sep}} \pi(A_1) \pi(A_2) \geq \frac{c_{\text{iso}} \ell_{\text{sep}}}{8} \pi(A)$, and $2\mathcal{Q}(A, A^c) \geq b_0 \pi(A_3)$ yields $\mathcal{Q}(A, A^c)/\pi(A) \geq b_0 c_{\text{iso}} \ell_{\text{sep}}/16$. Both cases are covered by Eq. (78). $\square$

5.7 A worst-case acceptance obstruction: the arrow spectrahedron

For a fixed universal proposal radius, uniform constant acceptance can require $\kappa = \Omega(\psi(x))$. The *arrow* instance below has $\psi(0) = \Theta(\psi_*^+) = \Theta(d)$, so it matches the theorem's $\kappa = \tilde{O}(\psi_*^+)$ choice up to polylogarithmic factors:

$$n = d + 1, \quad A_0 = I_n, \quad A_a = \frac{1}{\sqrt{2}}(e_1 e_{a+1}^\top + e_{a+1} e_1^\top) \quad (a = 1, \dots, d).$$

Indeed,

$$S(x) = \begin{pmatrix} 1 & x^\top/\sqrt{2} \\ x/\sqrt{2} & I_d \end{pmatrix}.$$

For $x \neq 0$, its restriction to the span of e_1 and $(0, x/\|x\|)$ is the 2×2 matrix with diagonal entries 1 and off-diagonal entries $\|x\|/\sqrt{2}$; on the orthogonal complement it is the identity. Thus its eigenvalues are $1 \pm \|x\|/\sqrt{2}$ and 1 (with multiplicity $d - 1$). Consequently $K = \{x : \|x\| \leq \sqrt{2}\}$, $\det S(x) = 1 - \|x\|^2/2$, and $S(x) \succeq 0$ iff $\|x\| \leq \sqrt{2}$. Moreover

$$A_a^2 = \frac{1}{2}(e_1 e_1^\top + e_{a+1} e_{a+1}^\top), \quad \langle A_a, A_b \rangle_F = \delta_{ab}.$$

Hence the frame is already Frobenius-orthonormal at $x = 0$ and $H(0) = I$, while summing the displayed squares gives

$$\Psi(0) = \sum_a A_a^2 = \frac{d}{2} e_1 e_1^\top + \frac{1}{2} \sum_{a=1}^d e_{a+1} e_{a+1}^\top, \quad \psi(0) = \frac{d}{2}.$$

Lemma 5.25 (Acceptance collapse for the unscaled walk). *Consider the walk of Definition 3.8 on the arrow instance with $f \equiv 0$, $L = 0$, and unit scale $\kappa = 1$, started at $x = 0$. There are universal constants $r_0 > 0$, $c > 0$, and d_0 such that for all $r \leq r_0$ and $d \geq d_0$, the probability that the proposed step is accepted is at most*

$$e^{-dr^2/10} + 2e^{-cd}.$$

Consequently, constant acceptance requires $r \lesssim d^{-1/2} \asymp \psi(0)^{-1/2}$: the ψ -scaling of Eq. (74) is necessary up to logarithmic factors.

Proof. At $x = 0$ the metric is $H(0) = I$ and the proposal is $Y \sim N(0, \frac{r^2}{d}I)$; with $f \equiv 0$ the acceptance probability of a proposal $Y \in K^\circ$ is $1 \wedge e^{\Lambda(Y)}$ with $\Lambda = \log \frac{q_Y(0)}{q_0(Y)}$. Since $\phi(x) = -\log \det S(x) = -\log(1 - \|x\|^2/2)$, writing $t = \|y\|^2$ and differentiating first in t gives $\nabla \phi(y) = y/(1 - t/2)$ and hence

$$H(y) = \frac{I}{1 - t/2} + \frac{yy^\top}{(1 - t/2)^2}, \quad t := \|y\|^2.$$

The matrix determinant lemma $\det(aI + byy^\top) = a^d(1 + (b/a)t)$ and direct multiplication by y show that both ledger ingredients are exact functions of t :

$$\begin{aligned} \log \frac{\det H(y)}{\det H(0)} &= -d \log(1 - \frac{t}{2}) + \log(1 + \frac{t}{1 - t/2}), \\ \frac{d}{r^2} y^\top (H(y) - I)y &= \frac{d}{r^2} [\frac{t}{1 - t/2} + \frac{t^2}{(1 - t/2)^2} - t] = \frac{d}{r^2} [\frac{t^2/2}{1 - t/2} + \frac{t^2}{(1 - t/2)^2}]. \end{aligned}$$

Thus $\Lambda = \frac{1}{2} \log \frac{\det H(y)}{\det H(0)} - \frac{d}{2r^2} [\cdot]$. For $0 \leq t \leq \frac{1}{2}$, the inequalities $-\log(1 - u) \leq u + u^2$ on $[0, \frac{1}{4}]$ and $\log(1 + v) \leq v$, together with $\frac{1}{1 - t/2} \leq \frac{4}{3}$, give

$$\Lambda \leq \frac{d}{2} (\frac{t}{2} + \frac{t^2}{4}) + \frac{t}{1 - t/2} - \frac{d}{2r^2} \cdot \frac{3t^2}{2} \leq \frac{dt}{4} + \frac{dt^2}{8} + 2t - \frac{3dt^2}{4r^2}.$$

Let $\mathcal{A} := \{\frac{3}{4}r^2 \leq t \leq \frac{5}{4}r^2\}$. On $\mathcal{A}$ (and with $r \leq r_0$ small enough that $t \leq \frac{5}{4}r_0^2 \leq \frac{1}{2}$), the upper endpoints in the positive terms and the lower endpoint in the negative term give

$$\Lambda \leq dr^2 (\frac{5}{16} + \frac{25r^2}{128} - \frac{3}{4} \cdot \frac{9}{16}) + \frac{5r^2}{2} = dr^2 (\frac{5}{16} - \frac{27}{64} + O(r^2)) + \frac{5r^2}{2} \leq -\frac{dr^2}{10}.$$

Here $5/16 - 27/64 = -7/64$, so the last inequality holds for universal $r \leq r_0$ and $d \geq d_0$, after absorbing the additive $\frac{5}{2}r^2$.

Since $t = \frac{r^2}{d} \|\xi\|^2$ for standard ξ , the needed chi-square estimate follows directly from $\mathbb{E}[e^{\lambda \|\xi\|^2}] = (1 - 2\lambda)^{-d/2}$: Chernoff's bound with $\lambda = 1/10$ for the upper tail and $\lambda = -1/10$ for the lower tail gives

$$\Pr[\|\xi\|^2 \geq 5d/4] \leq e^{-d(1/8 + \frac{1}{2} \log(4/5))}, \quad \Pr[\|\xi\|^2 \leq 3d/4] \leq e^{-d(\frac{1}{2} \log(6/5) - 3/40)}.$$

Both constants in parentheses are positive, so for a universal $c > 0$, $\Pr[\mathcal{A}^c] = \Pr[\|\xi\|^2/d - 1 > \frac{1}{4}] \leq 2e^{-cd}$.

Splitting the expectation over $\mathcal{A}$ and $\mathcal{A}^c$ now gives

$$\mathbb{E}[1 \wedge e^\Lambda] \leq \Pr[\mathcal{A}^c] + \sup_{\mathcal{A}} e^\Lambda \leq 2e^{-cd} + e^{-dr^2/10}.$$

□

Remark 5.26 (Reading the bound). Lemma 5.25 proves a family-specific necessity statement: at the center of the arrow instance, constant acceptance of the unscaled walk requires $r \lesssim \psi(0)^{-1/2}$. Together with Lemmas 5.9 and 5.14, this matches the sufficient leverage scale up to logarithmic factors on that family. It does not prove a pointwise “if and only if” for all spectrahedra, nor does it prove a mixing-time lower bound. For completeness, scaling the center metric from I to κI replaces the proposal radius in the calculation by $r_{\text{eff}} = r/\sqrt{\kappa}$. Thus, at a fixed universal r , the same lemma forces $dr^2/\kappa = O(1)$, i.e. $\kappa = \Omega(d) = \Omega(\psi(0))$ on this family. At the other extreme, for polytopes ($\psi \leq 1$ by Lemma 3.10) and for incoherent instances ($\psi = \tilde{O}(1)$) alike, Theorem 5.1 reads $\tilde{O}(nd)$ —dimension times symmetry; in the diagonal case this recovers the Kannan–Narayanan bound [KN12]. Since $\psi \leq d$ always, the theorem never exceeds the generic barrier-scaling bound $\tilde{O}(d^2n)$ (cf. [Nar16, KV24]).

References

- [CDWY18] Yuansi Chen, Raaz Dwivedi, Martin J. Wainwright, and Bin Yu. Fast MCMC sampling algorithms on polytopes. *Journal of Machine Learning Research*, 19(55):1–86, 2018.
- [CP15] Michael B. Cohen and Richard Peng. ℓ_p row sampling by Lewis weights. In *Symposium on Theory of Computing (STOC)*, pages 183–192, 2015.
- [GKM⁺24] Yuzhou Gu, Nikki Lijing Kuang, Yi-An Ma, Zhao Song, and Lichen Zhang. Log-concave sampling from a convex body with a barrier: a robust and unified Dikin walk. In *Advances in Neural Information Processing Systems*, volume 37, pages 69230–69298, 2024. arXiv:2410.05700.
- [GSZ26] Sander Gribling, Aaron Sidford, and Chenyi Zhang. Computing Lewis weights to high precision using local relative smoothness. arXiv:2606.29186, 2026. To appear at the Conference on Learning Theory (COLT).
- [KN12] Ravindran Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37(1):1–20, 2012.
- [Koo26] Yunbum Kook. Beyond the $d^{2.5}$ -mixing bound for Dikin walks on polytopes. arXiv:2607.13943, July 2026.
- [KV24] Yunbum Kook and Santosh S. Vempala. Gaussian cooling and Dikin walks: the interior-point method for logconcave sampling. In *Conference on Learning Theory (COLT)*, volume 247 of *Proceedings of Machine Learning Research*, pages 3137–3240. PMLR, 2024.
- [Led01] Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [Lew78] D. R. Lewis. Finite dimensional subspaces of L_p . *Studia Mathematica*, 63(2):207–212, 1978.
- [LLV20] Aditi Laddha, Yin Tat Lee, and Santosh S. Vempala. Strong self-concordance and sampling. In *Symposium on Theory of Computing (STOC)*, pages 1212–1222, 2020.

- [LS88] Gregory F. Lawler and Alan D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger’s inequality. *Transactions of the American Mathematical Society*, 309(2):557–580, 1988.
- [LS19] Yin Tat Lee and Aaron Sidford. Solving linear programs with $\tilde{O}(\sqrt{\text{rank}})$ linear system solves. arXiv:1910.08033, 2019.
- [LV07] László Lovász and Santosh S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.
- [Nar16] Hariharan Narayanan. Randomized interior point methods for sampling and optimization. *Annals of Applied Probability*, 26(1):597–641, 2016.
- [Tro15] Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015.

Acknowledgments and AI disclosure

The proof ideas in this paper were generated by two large language models, GPT-5.6 Sol (OpenAI) and Claude Fable 5 (Anthropic), and the resulting proofs were verified jointly by the same two models, including independent adversarial line-by-line reviews of every argument. The authors reviewed the manuscript and take full responsibility for its content.

A Approximate metrics and exact-arithmetic complexity

This appendix proves the implementation guarantees in Theorem 1.3. The preceding analysis uses exact metrics for a reason: the derivative of the exact Lewis map produces the cancellation in Section 4.2, while the exact log-det Hessian gives the endpoint identities in Section 5.2. An arithmetic implementation need not reproduce either metric exactly. We retain the exact walks as reference chains and compare their Gaussian proposals with approximate-metric proposals, leaving both ledger arguments unchanged.

The comparison has the same form in both settings. A one-sided padding makes the approximate metric dominate its reference, while an $O(1/d)$ relative error keeps the two Gaussian densities within a constant factor on the acceptance event already proved above. Proposal overlap and conductance then survive with different universal constants. For polytopes we obtain the approximation from high-precision Lewis weights. For spectrahedra we use the TensorSRHT approximation of the log-det Hessian from Gu et al. [GKM⁺24, Lemma G.12] and a two-seed Metropolis filter. We first treat deterministic Lewis-weight approximations for polytopes; we then isolate the two-seed perturbation argument for randomized metrics and apply it to a sketched log-det metric for spectrahedra.

We begin with the notation of the polytope proof:

$$H(x) = A_x^\top W_x^{c_p} A_x, \quad G(x) = \kappa H(x) + L^2 I, \quad \bar{\nu}_{\text{LS}} := \kappa n^{2/p} d^{1-2/p}.$$

Here q_x , R_x , and P denote the exact proposal, ratio, and lazy kernel. For the weights we use the iteration of Gribbling–Sidford–Zhang [GSZ26], whose high-precision guarantee we now state in this notation.

A.1 High-precision Lewis weights

Fix $0 < \varepsilon_w < 1/2$. At each $x \in K^\circ$ we compute a positive vector $\tilde{w}_x$ by a deterministic measurable map $\text{ApproxLewis}_{\varepsilon_w}$, with

$$(1 - \varepsilon_w)w_{x,i} \leq \tilde{w}_{x,i} \leq (1 + \varepsilon_w)w_{x,i} \quad (i = 1, \dots, n). \quad (79)$$

To make the proposal kernel a deterministic measurable function of the current state, we use a fixed initialization and iteration count. Set

$$\alpha_{\text{LW}} := \frac{2}{p-2}, \quad b_{\text{LW}} := \alpha_{\text{LW}}^{-1}, \quad \bar{b}_{\text{LW}} := \max\{1, b_{\text{LW}}\}, \quad L_{\text{LW}} := 32p\bar{b}_{\text{LW}},$$

and

$$\bar{\varepsilon}_w := \frac{\varepsilon_w}{2(1 + \alpha_{\text{LW}})}, \quad T_{\text{LW}} := \lceil 4L_{\text{LW}} \log \frac{2n}{\bar{\varepsilon}_w} \rceil.$$

Starting from $v_{\text{LW}}^{(0)} = \mathbf{1}$, write $D_{\text{LW}}^{(t)} = \text{Diag}(v_{\text{LW}}^{(t)})$ and, for $t = 0, \dots, T_{\text{LW}} - 1$, compute the row-quadratic vector (obtainable from one full leverage-score-vector computation)

$$u_{\text{LW},i}^{(t)} := (A_x)_{i,\cdot} (A_x^\top D_{\text{LW}}^{(t)} A_x)^{-1} (A_x)_{i,\cdot}^\top, \quad \varphi_{\text{LW},i}^{(t)} := \frac{(u_{\text{LW},i}^{(t)})^{b_{\text{LW}}}}{v_{\text{LW},i}^{(t)}},$$

and update

$$v_{\text{LW},i}^{(t+1)} := \left(1 + \frac{\varphi_{\text{LW},i}^{(t)} - 1}{L_{\text{LW}}}\right) v_{\text{LW},i}^{(t)} \quad (i = 1, \dots, n). \quad (80)$$

After one final computation of $u_{\text{LW}}^{(T_{\text{LW}})}$, return

$$\tilde{w}_{x,i} := (u_{\text{LW},i}^{(T_{\text{LW}})})^{1+b_{\text{LW}}}. \quad (81)$$

The high-precision guarantee of Gribling–Sidford–Zhang proves Eq. (79). The prescribed iteration count and all-ones initialization make the output a deterministic Borel function of A_x , hence of x .

The estimate in Eq. (79) is two-sided. We turn it into the one-sided comparison needed below by inflating the approximate metric by its worst possible downward error.

Definition A.1 (Approximate-metric LS walk). *For $\tilde{W}_x := \text{Diag}(\tilde{w}_x)$, define*

$$\tilde{H}(x) := A_x^\top \tilde{W}_x^{c_p} A_x, \quad \bar{H}(x) := (1 - \varepsilon_w)^{-c_p} \tilde{H}(x), \quad \bar{G}(x) := \kappa \bar{H}(x) + L^2 I. \quad (82)$$

At x , propose

$$Y \sim \bar{q}_x := N\left(x, \frac{r^2}{d} \bar{G}(x)^{-1}\right).$$

When $Y \in K^\circ$, use the exact conditional Metropolis ratio

$$\bar{R}_x(Y) := \frac{e^{-f(Y)} \bar{q}_Y(x)}{e^{-f(x)} \bar{q}_x(Y)}.$$

Proposals outside K are rejected, and the chain is $\frac{1}{2}$ -lazy. Write $\bar{P}$ for the resulting lazy kernel.

Algorithm 1 One step of the approximate-metric LS walk

Require: Current cache $(x, s_x, f(x), \bar{G}(x), C_x, \log \det \bar{G}(x))$ with $C_x C_x^\top = \bar{G}(x)$; accuracy ε_w .

- 1: With probability $1/2$, **return** the current cache. $\triangleright$ laziness
 - 2: Draw $\xi \sim N(0, I_d)$, set $h \leftarrow \frac{r}{\sqrt{d}} C_x^{-\top} \xi$, and set $y \leftarrow x + h$ and $s_y \leftarrow s_x + Ah$.
 - 3: **if** some coordinate of s_y is nonpositive **then**
 - 4: **return** the current cache.
 - 5: **end if**
 - 6: $\tilde{w}_y \leftarrow \text{APPROXLEWIS}(A_y, p, \varepsilon_w)$.
 - 7: Form $\bar{G}(y)$ from Eq. (82), factor it, and evaluate $f(y)$.
 - 8: Compute
$$\ell \leftarrow f(x) - f(y) + \frac{1}{2} \log \frac{\det \bar{G}(y)}{\det \bar{G}(x)} - \frac{d}{2r^2} h^\top (\bar{G}(y) - \bar{G}(x)) h.$$
 - 9: Draw $U_{\text{acc}} \sim \text{Unif}(0, 1)$; if $\log U_{\text{acc}} \leq \min\{0, \ell\}$, **return** the cache at y ; otherwise **return** the current cache.
-

The perturbation argument does not differentiate the approximate weights. It first compares $\bar{G}$ with G pointwise and then transfers the acceptance and overlap bounds of the exact walk.

Lemma A.2 (Weight error implies metric error). *Let*

$$\gamma_w := \left(\frac{1 + \varepsilon_w}{1 - \varepsilon_w} \right)^{c_p} = e^{\theta_w}, \quad \theta_w := c_p \log \frac{1 + \varepsilon_w}{1 - \varepsilon_w}.$$

Then, uniformly on K° ,

$$H(x) \preceq \bar{H}(x) \preceq \gamma_w H(x), \quad G(x) \preceq \bar{G}(x) \preceq \gamma_w G(x). \quad (83)$$

For $\varepsilon_w \leq 1/2$, $0 \leq \theta_w \leq 4\varepsilon_w$.

Proof. Because $c_p \geq 0$, the coordinatewise estimate Eq. (79) and the positive semidefiniteness of each row outer product give

$$(1 - \varepsilon_w)^{c_p} H(x) \preceq \tilde{H}(x) \preceq (1 + \varepsilon_w)^{c_p} H(x),$$

so multiplication by $(1 - \varepsilon_w)^{-c_p}$ proves the first sandwich. Adding $L^2 I$, and using $\gamma_w \geq 1$, gives

$$\kappa \bar{H} + L^2 I \preceq \gamma_w \kappa H + L^2 I \preceq \gamma_w G,$$

which proves the second sandwich. Finally, $c_p \leq 1$ and $\log((1+t)/(1-t)) \leq 4t$ for $0 \leq t \leq 1/2$. $\square$

A.2 From metric approximation to mixing

The sandwich in Eq. (83) has two consequences. It makes the centered proposals close in total variation, and it controls the change of each Gaussian density at the two endpoints of an accepted step. The second consequence is why the accuracy scale is $1/d$: a Gaussian log density contains a determinant contribution from all d directions.

Lemma A.3 (Gaussian proposal perturbation). *There is a universal C_q such that, whenever $\theta_w \leq 1$,*

$$\sup_{x \in K^\circ} d_{\text{TV}}(q_x, \bar{q}_x) \leq C_q \sqrt{d} \theta_w =: \tau_w.$$

Proof. Let $\lambda_1, \dots, \lambda_d \in [1, e^{\theta_w}]$ be the eigenvalues of $G(x)^{-1/2} \bar{G}(x) G(x)^{-1/2}$. The centered Gaussian KL formula gives

$$D_{\text{KL}}(q_x \| \bar{q}_x) = \frac{1}{2} \sum_{j=1}^d (\lambda_j - 1 - \log \lambda_j) \leq C d \theta_w^2,$$

where the inequality uses $0 \leq \lambda - 1 - \log \lambda \leq C(\log \lambda)^2$ on $[1, e]$. Lemma 4.26 now proves the claim. $\square$

The density comparison is needed at both endpoints. The exact acceptance proof already provides the required event.

Lemma A.4 (Two-endpoint reference event). *There are universal constants $C_A, C_N < \infty$ and the same $\epsilon < 1/10$ as in Lemma 4.23 such that, for every $x \in K^\circ$, the set*

$$\mathcal{A}_x := \{y \in K^\circ : \log R_x(y) \geq -C_A, \quad \|y - x\|_{G(x)}^2 \leq C_N r^2, \quad \|y - x\|_{G(y)}^2 \leq C_N r^2\}$$

satisfies $q_x(\mathcal{A}_x) \geq 1 - \epsilon$.

Proof. Intersect the tube, concentration, and Gaussian norm events from the proof of Lemma 4.23. On this event, the proposal is feasible, $\log R_x(Y) \geq -C_A$, and

$$\|Y - x\|_{G(x)}^2 = \frac{r^2}{d} \|\xi\|^2 \leq C r^2,$$

by the Gaussian proposal representation and the norm event. For the universal choice of r , this local norm is below $1/2$. Lemmas 4.25 and 3.5, applied along the feasible segment $[x, Y]$, give $G(Y) \preceq C G(x)$ and hence the reverse norm bound after enlarging C_N . $\square$

Lemma A.5 (Acceptance survives high-precision weights). *There is a universal $c_{\text{app}} > 0$ such that, if $\epsilon_w \leq c_{\text{app}}/d$, then uniformly in $x \in K^\circ$,*

$$\Pr_{Y \sim \bar{q}_x} [Y \in K \text{ and } \log \bar{R}_x(Y) \geq -\bar{C}_A] \geq 1 - \bar{\epsilon},$$

where $\bar{C}_A < \infty$ is a universal constant and $\bar{\epsilon} := \epsilon + \tau_w \leq \epsilon + 4C_q c_{\text{app}}$ is bounded by a universal constant.

Proof. Reduce c_{app} , if necessary, so that $4c_{\text{app}} \leq 1$. Then $\theta_w \leq 4\epsilon_w \leq 1$, and $e^{\theta_w} - 1 \leq (e - 1)\theta_w$. For $z, u \in K^\circ$, put

$$\mathcal{E}_z(u) := \log \frac{\bar{q}_z(u)}{q_z(u)} = \frac{1}{2} \log \frac{\det \bar{G}(z)}{\det G(z)} - \frac{d}{2r^2} (\|u - z\|_{G(z)}^2 - \|u - z\|_{G(z)}^2).$$

Eq. (83) yields

$$|\mathcal{E}_z(u)| \leq \frac{d\theta_w}{2} + \frac{d(e^{\theta_w} - 1)}{2r^2} \|u - z\|_{G(z)}^2.$$

Consequently, for $y \in \mathcal{A}_x$, $|\mathcal{E}_x(y)| + |\mathcal{E}_y(x)| \leq C d \theta_w$. Since

$$\log \bar{R}_x(y) = \log R_x(y) + \mathcal{E}_y(x) - \mathcal{E}_x(y).$$

Lemma A.4 gives $\log \bar{R}_x(y) \geq -C_A - C d \theta_w$. The condition $\epsilon_w \leq c_{\text{app}}/d$ makes this a universal lower bound. Finally, Lemma A.3 gives

$$\bar{q}_x(\mathcal{A}_x) \geq q_x(\mathcal{A}_x) - \tau_w \geq 1 - \epsilon - \tau_w.$$

The boundary of the full-dimensional convex body K has Lebesgue measure zero, so writing $Y \in K$ rather than $Y \in K^\circ$ in the statement does not change this probability. $\square$

Lemma A.6 (Approximate proposal overlap). *The constant c_{app} can be chosen so that Eq. (49) implies*

$$d_{\text{TV}}(\bar{q}_x, \bar{q}_z) \leq \bar{\delta}_q := \delta_q + 2\tau_w < 1 - 2\bar{\epsilon}.$$

Proof. The triangle inequality and Lemmas 4.27 and A.3 give the displayed upper bound. The exact proof chooses $\delta_q < 1 - 2\epsilon$, leaving a fixed positive margin $m_0 := 1 - \delta_q - 2\epsilon$. Because $\bar{\delta}_q + 2\bar{\epsilon} = \delta_q + 2\epsilon + 4\tau_w$, reducing c_{app} so that $4\tau_w < m_0$ proves the strict inequality. $\square$

Theorem A.7 (Mixing with approximate Lewis weights). *Under the hypotheses of Theorem 4.2, suppose the deterministic measurable estimate Eq. (79) satisfies $\varepsilon_w \leq c_{\text{app}}/d$. Then the padded approximate-metric LS walk of Definition A.1 is exactly reversible with stationary distribution π . For every w -warm initial law μ_0 and every $0 < \delta < 1$,*

$$d_{\text{TV}}(\mu_0 \bar{P}^t, \pi) \leq \delta \quad \text{for all} \quad t \geq \tilde{O}((d^2 + dL^2 R^2) \log(w/\delta)). \quad (84)$$

Proof. Detailed balance. Because $x \mapsto \bar{G}(x)$ is deterministic and measurable, the accepted move density satisfies

$$e^{-f(x)} \bar{q}_x(y) (1 \wedge \bar{R}_x(y)) = \min\{e^{-f(x)} \bar{q}_x(y), e^{-f(y)} \bar{q}_y(x)\}.$$

Thus rejection and laziness give an exactly π -reversible kernel.

Local kernel overlap. Lemmas A.5 and A.6 meet the hypotheses of the metric-independent kernel-overlap argument in Lemma 4.28, with barred constants. Hence the lazy kernels $\bar{P}_x, \bar{P}_z$ have a fixed positive common mass whenever Eq. (53) holds, where closeness is still measured in the exact reference metric G . The proof of Lemma 4.33 uses only this local kernel overlap, reversibility, and the ν_G -symmetry of the reference metric. It therefore applies with the barred common-mass constant and gives

$$\Phi(\bar{P}) \geq \frac{cr}{\sqrt{d(\bar{\nu}_{\text{LS}} + L^2 R^2)}}.$$

Conductance and mixing. The same lazy Cheeger and warm-start L^2 argument as in the proof of Theorem 4.2 proves Eq. (84). $\square$

The spectrahedral approximation is randomized, so we also need a seed-averaged version of the preceding perturbation argument.

Lemma A.8 (Two-seed perturbation for random metrics). *Let ω_s have a state-independent base law, let $\bar{G}(x; \omega_s)$ be a measurable positive-definite output, and let $\bar{q}_{x, \omega_s}$ be its conditional Gaussian proposal density. Draw ω_s at x , propose y , draw an independent seed ω'_s at y when $y \in K^\circ$, and accept with probability*

$$1 \wedge \frac{e^{-f(y)} \bar{q}_{y, \omega'_s}(x)}{e^{-f(x)} \bar{q}_{x, \omega_s}(y)}.$$

Proposals outside K° are rejected. The resulting marginal kernel is exactly π -reversible.

Suppose that, uniformly in x , a good-seed event $\mathcal{G}_x(\omega_s)$ has probability at least $1 - \beta_0$ and

$$d_{\text{TV}}(q_x, \bar{q}_{x, \omega_s}) \leq \tau \quad \text{on } \mathcal{G}_x(\omega_s).$$

Suppose also that the reference proposals have overlap at least $1 - \delta_q$ for the close pairs under consideration. For every x , let $\mathcal{B}_x \subseteq K^\circ$ be measurable with $q_x(\mathcal{B}_x^c) \leq \epsilon$. Assume that for $y \in \mathcal{B}_x$ the conditional Metropolis ratio is at least e^{-C} whenever the seeds at x and y are both good, where $C \geq 0$. Then, before the fixed laziness factor, the moving kernels have common mass at least

$$e^{-C}(1 - \beta_0)(1 - \delta_q - 2\epsilon - 4\tau - 2\beta_0). \quad (85)$$

Proof. Reversibility. The marginal accepted flow is

$$\mathbb{E}_{\omega_s, \omega'_s}[\min\{e^{-f(x)}\bar{q}_{x, \omega_s}(y), e^{-f(y)}\bar{q}_{y, \omega'_s}(x)\}],$$

which is symmetric after interchanging the two independent seeds. This is the two-seed device of Gu et al. [GKM⁺24, Lemma D.12].

Common mass on good seeds. Define the good-seed proposal subdensity

$$q_x^{\text{good}}(y) := \mathbb{E}_{\omega_s}[\mathbf{1}_{\mathcal{G}_x(\omega_s)}\bar{q}_{x, \omega_s}(y)].$$

The total-variation hypothesis and the identity between total variation and half the L^1 distance give

$$\|q_x^{\text{good}} - q_x\|_{L^1} \leq 2\tau + \beta_0.$$

Thus, for each close pair,

$$\int \min\{q_x^{\text{good}}(y), q_z^{\text{good}}(y)\} dy \geq 1 - \delta_q - 2\tau - 2\beta_0.$$

Indeed,

$$\begin{aligned} \int \min\{q_x^{\text{good}}, q_z^{\text{good}}\} &= \frac{1}{2}(\|q_x^{\text{good}}\|_1 + \|q_z^{\text{good}}\|_1 \\ &\quad - \|q_x^{\text{good}} - q_z^{\text{good}}\|_1). \end{aligned}$$

Here $\|q_x^{\text{good}}\|_1, \|q_z^{\text{good}}\|_1 \geq 1 - \beta_0$; the claim follows by the triangle inequality through q_x, q_z . The same total-variation hypothesis gives

$$q_x^{\text{good}}(\mathcal{B}_x^c) \leq \epsilon + \tau, \quad q_z^{\text{good}}(\mathcal{B}_z^c) \leq \epsilon + \tau.$$

Restricting the common proposal mass to $\mathcal{B}_x \cap \mathcal{B}_z$ therefore removes at most $2\epsilon + 2\tau$. At a remaining proposal y , requiring the independent reverse seed to be good contributes the factor $1 - \beta_0$. Applying the conditional lower bound e^{-C} proves Eq. (85). $\square$

Because the good-seed guarantee is pointwise in the current state, fresh independent endpoint seeds, exact conditional densities, and a positive-definite metric on every seed suffice; no event need hold simultaneously over the mixing horizon.

A.3 Arithmetic complexity for polytopes

Because the current metric is cached, each feasible proposal requires at most one new Lewis-weight computation. Combining this per-step cost with Theorem A.7 gives the total arithmetic bound below. Throughout the two arithmetic corollaries, we use $\log(ew/\delta) = 1 + \log(w/\delta)$. Since $w \geq 1$, $0 < \delta < 1$, and the displayed geometric prefactors are at least one, this absorbs the one-time metric build and evaluation of f .

Corollary A.9 (Arithmetic complexity for polytopes). *In the exact-arithmetic model described below, run Algorithm 1 with $\varepsilon_w = c_{\text{app}}/d$. It can be implemented with*

$$\tilde{O}(nd^{\omega-1} + d^{\omega}) \tag{86}$$

arithmetic operations and at most one new evaluation of f per Markov step. Here ω denotes the exponent of square-matrix multiplication. Consequently, for every $0 < \delta < 1$, starting from an exact draw $X_0 \sim \mu_0$, where μ_0 is w -warm, the walk reaches total-variation error δ using

$$\tilde{O}((d^2 + dL^2R^2)(nd^{\omega-1} + d^\omega) \log \frac{ew}{\delta}) \quad (87)$$

arithmetic operations and $\tilde{O}((d^2 + dL^2R^2) \log(ew/\delta))$ evaluations of f .

Proof. Lewis weights. The slack-scaled matrix A_x has n rows and d columns. The fixed-schedule estimator in Eqs. (80)–(81) returns the coordinatewise estimate Eq. (79) in

$$O(p^2 \log \frac{np\alpha_{\text{LW}}}{\varepsilon_w})$$

iterations, each computing the full leverage-score vector exactly. Here $p = \max\{4, 2(1 + \log n)\}$ and $\varepsilon_w = c_{\text{app}}/d$, so the number of iterations is polylogarithmic in n, d . One full leverage-score vector for an $n \times d$ matrix costs $O(nd^{\omega-1})$ arithmetic operations by forming and factoring its weighted Gram matrix and evaluating all row quadratic forms. Hence one call to $\text{ApproxLewis}_{\varepsilon_w}$ costs $\tilde{O}(nd^{\omega-1})$ arithmetic operations.

One transition. Forming slacks and checking membership costs $O(\text{nnz}(A))$; assembling the weighted Gram matrix is covered by $O(nd^{\omega-1})$; and factoring $\tilde{G}$, drawing the Gaussian from that factor, computing its determinant, and evaluating the two quadratic forms cost $O(d^\omega)$. The current metric, factorization, determinant, slacks, and target value are cached. Laziness is tested before any proposal work. Thus each nonlazy iteration computes at most one new feasible-endpoint metric and one new target value; on acceptance those become the next cache. The initial cache is built once, so there are at most $t + 1$ builds in t steps. This proves Eq. (86).

Total cost. Combining the per-step bound with Theorem A.7 and including the initial build and evaluation of f gives Eq. (87) and the stated number of evaluations of f . $\square$

Remark A.10 (Exact-arithmetic model). Field operations, comparisons, scalar powers and roots, logarithms and exponentials, and exact Gaussian and uniform variates have unit cost. The count includes the Lewis iterations, slack and membership computations, metric factorization, the two Gaussian quadratic forms, and one new target evaluation at a feasible proposal.

The initial warm draw and the parameters L, R, w are supplied. We do not count finite-bit conditioning, approximate Gaussian generation, or an inexact Metropolis decision. Thus the corollary is an arithmetic-operation bound, not a bit-complexity result. A bit-complexity result would additionally require controlling the precision of ill-conditioned diagonal scalings.

A.4 A sketched log-det metric

For spectrahedra, write κ_{ld} for the scale denoted there by κ , and put

$$H_{\text{ld}}(x) := \nabla^2(-\log \det S(x)), \quad G_{\text{ld}}(x) := \kappa_{\text{ld}} H_{\text{ld}}(x) + L^2 I.$$

Thus G_{ld} is the exact reference metric, and n denotes the pencil block size. Let $A \in \mathbb{R}^{d \times n^2}$ have i -th row $\text{vec}(A_i)^\top$, where vec denotes column stacking, and put

$$B_x := A(S(x)^{-1/2} \otimes S(x)^{-1/2}). \quad (88)$$

Then $H_{\text{ld}}(x) = B_x B_x^\top$. If $S(x) = L_x L_x^\top$, the symmetric inverse square root in Eq. (88) may be replaced by $T_x = L_x^{-\top}$, since $T_x T_x^\top = S(x)^{-1}$; the corresponding rows are $\text{vec}(T_x^\top A_i T_x)^\top$.

A TensorSRHT compresses this Gram factor before the $d \times d$ metric is formed. Fix $0 < \varepsilon_s < 1/10$ and $0 < \beta_s < 1/10$. For a fresh TensorSRHT seed ω_s , let $\Pi_{\omega_s} \in \mathbb{R}^{s \times n^2}$ be the transform of [GKM⁺24, Definition G.10], where, for a sufficiently large universal C_{TS} ,

$$s := \lceil C_{\text{TS}} \varepsilon_s^{-2} d \log^3 \frac{nd}{\varepsilon_s \beta_s} \rceil, \quad Z_{x, \omega_s} := \Pi_{\omega_s} B_x^\top \in \mathbb{R}^{s \times d}. \quad (89)$$

If n is not a Hadamard size, let n' be the next power of two and replace the pencil by $S(x) \oplus I_{n'-n}$ and every direction by $A_i \oplus 0_{n'-n}$. The feasible body and log-det Hessian are unchanged. Moreover, the whitened directions become $\bar{A}_i(x) \oplus 0$, so the leverage operator becomes $\Psi(x) \oplus 0$ and ψ_*^+ is unchanged. Since $n' < 2n$, this padding changes the mixing and arithmetic bounds by at most a universal constant. We henceforth relabel n' as n . Define a positive-definite output on every seed by

$$\hat{H}(x; \omega_s) := \begin{cases} Z_{x, \omega_s}^\top Z_{x, \omega_s}, & Z_{x, \omega_s}^\top Z_{x, \omega_s} \succ 0, \\ I_d, & \text{otherwise,} \end{cases}$$

and apply one-sided padding:

$$\bar{H}_{\text{ld}}(x; \omega_s) := (1 - \varepsilon_s)^{-2} \hat{H}(x; \omega_s), \quad \bar{G}_{\text{ld}}(x; \omega_s) := \kappa_{\text{ld}} \bar{H}_{\text{ld}}(x; \omega_s) + L^2 I.$$

The fallback makes the conditional proposal well defined for every seed; the approximation statement below is restricted to the sketch-success event.

Lemma A.11 (Tensor sketch implies metric approximation). *For each fixed $x \in K^\circ$, with probability at least $1 - \beta_s$ over ω_s ,*

$$H_{\text{ld}}(x) \preceq \bar{H}_{\text{ld}}(x; \omega_s) \preceq e^{\theta_s} H_{\text{ld}}(x), \quad G_{\text{ld}}(x) \preceq \bar{G}_{\text{ld}}(x; \omega_s) \preceq e^{\theta_s} G_{\text{ld}}(x),$$

where

$$\theta_s := 2 \log \frac{1 + \varepsilon_s}{1 - \varepsilon_s} \leq 8\varepsilon_s.$$

Proof. Lemma G.12 of [GKM⁺24] gives, simultaneously for all $v \in \mathbb{R}^d$,

$$(1 - \varepsilon_s) \|B_x^\top v\| \leq \|\Pi_{\omega_s} B_x^\top v\| \leq (1 + \varepsilon_s) \|B_x^\top v\|$$

with probability at least $1 - \beta_s$. Squaring gives

$$(1 - \varepsilon_s)^2 H_{\text{ld}}(x) \preceq \hat{H}(x; \omega_s) \preceq (1 + \varepsilon_s)^2 H_{\text{ld}}(x).$$

On this event the middle matrix is positive definite, so the fallback is not used. Padding proves the first sandwich. Adding $L^2 I$ proves the second exactly as in Lemma A.2. Finally, $2 \log((1+t)/(1-t)) \leq 8t$ for $0 \leq t \leq 1/2$. $\square$

Definition A.12 (Two-seed sketched log-det walk). *At every nonlazy step from x , draw a fresh seed ω_s and propose*

$$Y \sim \bar{q}_{x, \omega_s}^{\text{ld}} := N(x, \frac{r^2}{d} \bar{G}_{\text{ld}}(x; \omega_s)^{-1}).$$

Reject if $Y \notin K^\circ$. Otherwise draw an independent fresh seed ω'_s at Y and accept with probability

$$1 \wedge \frac{e^{-f(Y)} \bar{q}_{Y, \omega'_s}^{\text{ld}}(x)}{e^{-f(x)} \bar{q}_{x, \omega_s}^{\text{ld}}(Y)}. \quad (90)$$

Both conditional densities in Eq. (90) are evaluated exactly, and the chain is $\frac{1}{2}$ -lazy. Write $\bar{P}_{\text{ld}}$ for the resulting lazy kernel.

Remark A.13 (Fresh seeds). The sketches at the two endpoints are deliberately independent. Reusing the reverse seed after an accepted step would make the next forward seed depend on the chain history and would require a different, lifted-state analysis.

For the implementation, use the direct exact metric when $n < s$ and the sketched metric when $n \geq s$. Algorithm 2 spells out both branches. In the sketched branch, neither random metric is stored in the returned cache: the next nonlazy transition draws a fresh forward seed.

Algorithm 2 One step of the direct-or-sketched log-det walk

Require: Current cache $(x, S_x, L_x, f(x))$ with $S_x = S(x) = L_x L_x^\top$; fixed ε_s, β_s , and s from Eq. (89).

```

1: With probability  $1/2$ , return the current cache. ▷ laziness
2: if  $n < s$  then
3:   Form  $G_x \leftarrow G_{\text{ld}}(x)$ .
4: else
5:   Draw a fresh TensorSRHT seed  $\omega_s$  and form  $G_x \leftarrow \bar{G}_{\text{ld}}(x; \omega_s)$ .
6: end if
7: Factor  $G_x = C_x C_x^\top$  and compute  $\log \det G_x$ .
8: Draw  $\xi \sim N(0, I_d)$ , set  $h \leftarrow \frac{r}{\sqrt{d}} C_x^{-\top} \xi$ , and set  $y \leftarrow x + h$  and  $S_y \leftarrow S_x + \sum_{i=1}^d h_i A_i$ .
9: if  $S_y \neq 0$  then
10:  return the current cache.
11: end if
12: Factor  $S_y = L_y L_y^\top$ .
13: if  $n < s$  then
14:  Form  $G_y \leftarrow G_{\text{ld}}(y)$ .
15: else
16:  Draw an independent fresh TensorSRHT seed  $\omega'_s$  and form  $G_y \leftarrow \bar{G}_{\text{ld}}(y; \omega'_s)$ .
17: end if
18: Factor  $G_y = C_y C_y^\top$ , compute  $\log \det G_y$ , and evaluate  $f(y)$ .
19: Compute the exact conditional log Metropolis ratio
    
$$\ell \leftarrow f(x) - f(y) + \frac{1}{2} \log \frac{\det G_y}{\det G_x} - \frac{d}{2r^2} h^\top (G_y - G_x) h.$$

20: Draw  $U_{\text{acc}} \sim \text{Unif}(0, 1)$ ; if  $\log U_{\text{acc}} \leq \min\{0, \ell\}$ , return  $(y, S_y, L_y, f(y))$ ; otherwise return the current cache.
```

Theorem A.14 (Mixing with a sketched log-det metric). *Under the hypotheses of Theorem 5.1, there are universal constants $c_s, \bar{\beta}_s > 0$ with the following property. Let $\hat{\psi} \geq \psi_*^+$ be a supplied upper bound and choose κ_{ld} by replacing ψ_*^+ with $\hat{\psi}$ in Eq. (74). If $\varepsilon_s \leq c_s/d$ and $\beta_s \leq \bar{\beta}_s$, then the walk of Definition A.12 is exactly reversible with stationary distribution π . For every w -warm initial law μ_0 and every $0 < \delta < 1$, it satisfies*

$$d_{\text{TV}}(\mu_0 \bar{P}_{\text{ld}}^t, \pi) \leq \delta \quad \text{for all} \quad t \geq \tilde{O}((\hat{\psi} n d + d L^2 R^2) \log(w/\delta)). \quad (91)$$

Remark A.15 (Reading the bound). The sketch guarantee is pointwise at the current state, but the failed seeds are part of the transition kernel itself. Hence no union bound over the mixing horizon is needed. Since $\psi_*^+ \leq d$, the choice $\hat{\psi} = d$ is always valid and gives the unconditional order $\tilde{O}((n d^2 + d L^2 R^2) \log(w/\delta))$.

Proof. Reference event. Because $\hat{\psi} \geq \psi_*^+$, the proof of Lemma 5.15 applies with ψ_*^+ replaced by $\hat{\psi}$. Write q_x^{ld} and R_x^{ld} for the exact spectral reference proposal and ratio. The good event used in the

proof of Lemma 5.15 already contains the Gaussian norm event and hence

$$\|Y - x\|_{G_{\text{ld}}(x)}^2 = \frac{r^2}{d} \|\xi\|^2 \leq Cr^2.$$

Lemma 5.17 and its distortion conclusion then give $\|Y - x\|_{G_{\text{ld}}(Y)}^2 \leq Cr^2$ on the same feasible event. Thus the exact reference proposal has, with failure at most the same universal $\epsilon \leq 1/10$, a two-endpoint event on which $\log R_x^{\text{ld}}(Y) \geq -C_{\text{acc}}$ and both displayed local norms are $O(r^2)$.

Conditional proposal comparison. On a good sketch seed, Lemma A.11 and the Gaussian KL calculation of Lemma A.3 give

$$d_{\text{TV}}(q_x^{\text{ld}}, \bar{q}_{x, \omega_s}^{\text{ld}}) \leq C\sqrt{d}\theta_s =: \tau_s. \quad (92)$$

Moreover, for $z, u \in K^\circ$ and a good seed at z ,

$$\left| \log \frac{\bar{q}_{z, \omega_s}^{\text{ld}}(u)}{q_z^{\text{ld}}(u)} \right| \leq \frac{d\theta_s}{2} + \frac{d(e^{\theta_s} - 1)}{2r^2} \|u - z\|_{G_{\text{ld}}(z)}^2. \quad (93)$$

For $\varepsilon_s \leq c_s/d$, this is $O(1)$ at both endpoints of the reference event, while $\tau_s = O(c_s/\sqrt{d})$. Hence, on good endpoint seeds, the conditional log Metropolis ratio is at least $-\bar{C}_{\text{acc}}$ for a universal $\bar{C}_{\text{acc}} < \infty$.

Reversibility and common mass. The good-seed subdensity argument in Lemma A.8 is barrier-independent. Apply it with Eqs. (92) and (93), the exact proposal overlap Lemma 5.19, and the independently good reverse seed. Before the fixed laziness factor, the resulting moving-kernel common mass is at least

$$e^{-\bar{C}_{\text{acc}}}(1 - \beta_s)\left(1 - \frac{1}{2} - 2\epsilon - 4\tau_s - 2\beta_s\right).$$

Choose the universal $c_s, \bar{\beta}_s$ so that this is positive. The marginal accepted flow is exactly symmetric by the seed-interchange argument of Lemma A.8, so the chain is exactly reversible.

Conductance and mixing. The proofs of Lemmas 5.20, 5.22, and 5.24 use only a positive universal common-mass constant and the exact reference geometry. They therefore apply to the sketched kernel, and the warm-start argument in Theorem 5.1 gives Eq. (91). $\square$

A.5 Arithmetic complexity for spectrahedra

At each state, we use the direct construction when $n < s$ and the sketched construction when $n \geq s$, so every transition is charged to the cheaper metric build in its dimension regime.

For positive integers a, b, c , let $\mathcal{T}_{\text{mat}}(a, b, c)$ denote the arithmetic cost of multiplying an $a \times b$ matrix by a $b \times c$ matrix. At $\varepsilon_s = \Theta(1/d)$ and constant β_s , Eq. (89) gives $s = \tilde{O}(d^3)$. A complete fresh sketched-metric build has cost

$$\mathcal{C}_{\text{sk}}(n, d, s) := O(dn^2 + n^\omega + d\mathcal{T}_{\text{mat}}(n, n, s) + \mathcal{T}_{\text{mat}}(d, s, d) + d^\omega). \quad (94)$$

Here the first two terms form and factor the slack, the third is the TensorSRHT construction of Z from [GKM⁺24, Lemma G.12], the fourth forms $Z^\top Z$, and the last tests and factors the $d \times d$ metric. The $\tilde{O}(n^2)$ Hadamard work is absorbed. The fourth term is listed separately because [GKM⁺24, Lemma G.12] constructs Z , not its Gram matrix.

There is also the direct exact build

$$\mathcal{C}_{\text{ex}}(n, d) := O(dn^2 + dn^\omega + n^2 d^{\omega-1} + d^\omega), \quad (95)$$

obtained by forming the d slack-congruenced directions, their Frobenius Gram matrix H_{ld} , and the factorization of G_{ld} . When $n < s$, the implementation runs the exact-metric walk of Theorem 5.1 and uses this direct build. When $n \geq s$, it runs the sketched walk of Definition A.12 and uses Eq. (94). In this rule, fix $\varepsilon_s = c_s/d$ and $\beta_s = \bar{\beta}_s/2$, and take s from Eq. (89).

Corollary A.16 (Arithmetic complexity for spectrahedra). *Under the hypotheses and parameter choices of Theorem A.14, consider the implementation that uses the direct build when $n < s$ and the sketched build when $n \geq s$. In the exact-arithmetic model described below, each Markov step uses*

$$\tilde{O}(n^\omega + n^2 d^{3\omega-5}) \quad (96)$$

arithmetic operations and at most one new evaluation of f . Here ω denotes the exponent of square-matrix multiplication. For every $0 < \delta < 1$, starting from an exact draw $X_0 \sim \mu_0$, where μ_0 is w -warm, it reaches total-variation error δ using

$$\tilde{O}((\hat{\psi}nd + dL^2R^2)(n^\omega + n^2 d^{3\omega-5}) \log \frac{ew}{\delta}) \quad (97)$$

arithmetic operations and $\tilde{O}((\hat{\psi}nd + dL^2R^2) \log(ew/\delta))$ evaluations of f . With the unconditional choice $\hat{\psi} = d$, replace $\hat{\psi}nd$ by nd^2 in both total bounds.

Proof. Direct branch. Full-dimensional boundedness implies $d \leq n(n+1)/2 = O(n^2)$. If $n < s = \tilde{O}(d^3)$, use the direct build Eq. (95). Since $\omega \geq 2$ and $n \leq \tilde{O}(d^3)$,

$$dn^\omega \leq \tilde{O}(n^2 d^{3\omega-5}), \quad n^2 d^{\omega-1} \leq n^2 d^{3\omega-5},$$

and the remaining terms are smaller, so $C_{\text{ex}} = \tilde{O}(n^\omega + n^2 d^{3\omega-5})$.

Sketched branch. If $n \geq s$, use the sketch. Standard square blocking gives

$$d\mathcal{T}_{\text{mat}}(n, n, s) = \tilde{O}(n^2 d^{3\omega-5}),$$

which is the rectangular-multiplication bound used in [GKM⁺24, Section 2.4]. Also $\mathcal{T}_{\text{mat}}(d, s, d) = O(sd^{\omega-1})$; because $s = \tilde{O}(d^3)$ and $n \geq s$, this and every other term in Eq. (94) are absorbed by $\tilde{O}(n^\omega + n^2 d^{3\omega-5})$.

Transitions and total cost. A nonlazy feasible transition uses one fresh forward build and one fresh reverse build. This factor of two is constant; slack factors and the current target value may be cached, but random metrics are not reused. Hence Eq. (96) holds. Multiplying it by Eq. (91) in the sketched branch, and by the proof of Theorem 5.1 with ψ_*^+ replaced by $\hat{\psi}$ in the exact branch, gives the total transition cost. Including the initial metric build and evaluation of f proves Eq. (97). $\square$

Remark A.17 (Arithmetic model and scope). For the spectrahedral bound, field operations, comparisons, scalar elementary functions, and exact Gaussian, uniform, and Rademacher variates have unit cost. The count includes forming and factoring the slack, applying the TensorSRHT, forming its Gram matrix, factoring the proposal metric, evaluating both conditional Gaussian densities, and making the two-seed Metropolis decision. The input supplies an exact warm draw and the parameters L, R, w and $\hat{\psi}$; the choice $\hat{\psi} = d$ avoids certifying the leverage supremum.

The polytope implementation uses a deterministic metric, whereas the spectrahedral implementation uses fresh independent endpoint seeds. In both cases, the Metropolis ratio uses the actual conditional proposal densities, and the comparison above transfers acceptance, overlap, and conductance from the exact reference walk to the implemented kernel.

The bounds exclude finite-bit conditioning and rounded Gaussian sampling. They also exclude an inexact Metropolis decision and construction of the warm start. Thus they count arithmetic operations, not bit complexity.