hexagon:2611.01079v1

Self-normalized moderate deviations for stochastic gradient Langevin averages with variance correction

Zhenyu Sun

November 26, 2026

Abstract

We prove a two-sided relative Gaussian tail approximation for bias-corrected time averages of stochastic gradient Langevin dynamics. The target is the invariant mean of the modified diffusion. From independent stochastic-gradient samples, values of the observable, and known model bounds, we estimate the first weak-bias term and the conditional variance per unit time of a Poisson increment. The variance estimate is nonnegative and includes the full first-order correction.

Suppose that the drift is smooth and strongly dissipative and that the gradient noise is state-independent, of the form $B + \tau G$, where B is bounded and centered and G is an independent standard Gaussian vector. As $\eta \downarrow 0$, $T = m\eta \to \infty$, and $r \to \infty$, the approximation holds uniformly for $0 \le x \le r$ if

$$r^4 \log(T + e)/T \to 0, \qquad r\eta^2\sqrt{T} \to 0.$$

It suffices to estimate the variance to relative error $o(r^{-2})$ and the corrected average to additive error $o((r\sqrt{T})^{-1})$, with combined failure probability $o(1 - \Phi(r))$. An explicit observation schedule permits $r = o(\eta^{-2/3}[\log(1/\eta)]^{-1/6})$.

The proof controls the original predictable quadratic variation under exponential tilting through a discrete Poisson equation. A residual bound for the diffusion Poisson equation gives a finite construction from gradient samples. This bound also applies to weakly dissipative drifts and state-dependent innovations with known local fourth-moment bounds. For fixed dimension and model bounds, the oracle sample count is polynomial in the trajectory radius, inverse tolerances, and the logarithms of the inverse failure probabilities, up to a factor m. For additive isotropic Gaussian innovations and evaluable coefficients, paired Euler paths also give polynomial auxiliary operation counts.

Keywords. Stochastic gradient Langevin dynamics, Cramér-type moderate deviations, self-normalization, Poisson equation, stochastic-gradient oracle.

Contents

1 Introduction **3**

2 Model and main results **4**

2.1 Sampling access and the computed statistic 5

3 Positive variance completion and weighted remainder bounds **7**
3.1 A weighted variance expansion beyond compact sets . . . 7
3.2 A pointwise nonnegative variance completion . . . 9
3.3 Global expansions under a uniform noise bound . . . 10

4 Contraction and exact clocks for the tilted chain **10**
4.1 Uniform estimates for the actual tilted kernel . . . 11
4.2 An exact discrete Poisson equation . . . 13
4.3 Cumulants of the full Poisson increment . . . 14

5 A self-normalized martingale comparison **14**
5.1 A bracket repair lemma . . . 14
5.2 Finite comparison at a random boundary . . . 15
5.3 Simultaneous perturbations of numerator and denominator . . . 19

6 Localization and polynomial occupation on the original chain **19**

7 The trajectory statistic with bias and variance corrections **23**

8 Certified Poisson gradients from a sampling interface **24**
8.1 Local coefficient certificates with fourth moments . . . 25
8.2 A global certificate from a local residual . . . 26
8.3 An effective finite family and a terminating constructor . . . 28

9 Computable bias correction and positive energy **30**
9.1 The first correction from gradient values . . . 30
9.2 Stability of the positive energy . . . 32
9.3 Absolute computation and a relative certificate . . . 32
9.4 Completion of the computed moderate-deviation theorem . . . 34

10 A Gaussian hierarchy and paired simulation **35**
10.1 Positive conditional Hermite energies . . . 35
10.2 A paired Euler estimator . . . 37
10.3 A finite confidence budget for all features . . . 38
10.4 Paired Euler estimates of the gradient and weak-bias correction . . . 39
10.5 Relative certification and the Hermite trajectory statistic . . . 41

11 Examples and scope **43**
11.1 A nonlinear model with asymmetric unbounded innovation . . . 43
11.2 A weakly dissipative oracle with no exponential moment . . . 43
11.3 What the numerical certificate establishes . . . 44
11.4 The target and the limits of the theorem . . . 44
11.5 Reproducible algebraic checks . . . 45

A Effective contraction and global Poisson estimates **45**
A.1 A common effective contraction rate . . . 46
A.2 An integrable semigroup smoothing estimate . . . 47
A.3 Canonical Poisson solutions and explicit derivative bounds . . . 50

A.4 Energy nondegeneracy and global cutoffs 51

B Limits of finite oracle sampling **52**

1 Introduction

A relative Gaussian tail approximation is sensitive to an error in the variance used to normalize an empirical average. If the variance estimate has relative error β, the normalized threshold x changes by $O(x\beta)$. By the normal tail ratio, this gives a relative error of order $x^2\beta$ for large x with $x^2\beta = o(1)$. To use the approximation up to a growing radius r, we also bound the probability of numerical failure relative to $\overline{\Phi}(r) = 1 - \Phi(r)$.

Dai, Fan and Lu [DFL26] proved self-normalized Cramér-type moderate deviations (CMD) for stochastic gradient Langevin dynamics (SGLD) using the Poisson equation of an approximating diffusion. They proposed estimating the Poisson gradient by Monte Carlo and identified the effect of that computation on the tail error as an open question [DFL26, Remark 3]. For the Euler–Maruyama scheme, Lu, Tan and Xu [LTX22] proved a central limit theorem and a self-normalized moderate-deviation estimate. Fan, Hu and Xu [FHX24] obtained refinements for normalized and self-normalized averages.

In related work [Sun26], the author used evaluations of the drift and a covariance factor to compute a leading Poisson-gradient normalizer. That result allows weak dissipativity and state-dependent noise and requires $r^2\eta \to 0$. Here we consider a strongly dissipative drift and additive gradient noise $B + \tau G$, where B is bounded and centered and G is an independent standard Gaussian vector. In this class we can estimate the first correction to the conditional variance using independent samples of $b(x) + Z$ at chosen states. The corresponding term in the sufficient tail-error conditions becomes $r^2\eta^2 \to 0$. The construction uses the Poisson regularity estimates and full-increment change of measure that also enter [Sun26]. Both papers center at θ_η, the invariant mean of the diffusion with covariance $\delta I + \eta\Sigma$.

The main result, Theorem 2.3, gives a two-sided relative Gaussian tail estimate for the centered, bias-corrected empirical mean divided by a computed standard error. The variance estimate retains the full first correction to the one-step variance of a Poisson increment, including the covariance of the gradient noise. On the event used for inference, the difference between the empirical average of this variance density and the original predictable quadratic variation per unit time is $O(\eta^2)$. A discrete Poisson equation for the exponentially tilted transition controls that original quadratic variation under the tilted law. The random-boundary comparison in Section 5 uses the martingale normal approximation of Fan, Grama and Liu [FGL13, Theorem 3.1].

The computation uses a local residual for the whole-space diffusion Poisson equation. With known regularity, dissipativity, and local fourth-moment bounds, the residual controls the gradient at every state used for the variance and bias corrections. This part of the analysis allows a state-dependent covariance and weak dissipativity. Appendix A gives the contraction and derivative estimates, using reflection coupling [Ebe16] and a semigroup derivative formula of Elworthy–Li type [EL94]. For additive isotropic Gaussian innovations with evaluable coefficients, Section 10 uses Hermite orthogonality [Hai26] and paired Euler paths to estimate the variance and bias with polynomial auxiliary work.

Sections 3–7 give the one-step expansions and the trajectory tail bound. Sections 8 and 9 construct the numerical estimators. Section 11 illustrates where the trajectory and numerical

assumptions differ.

2 Model and main results

Fix $d, \ell \in \mathbb{N}$, $\delta > 0$ and $x_0 \in \mathbb{R}^d$. For a step size $\eta > 0$, consider

$$X_{k+1} = X_k + \eta b(X_k) + \eta Z(X_k, \zeta_{k+1}) + \sqrt{\eta\delta}\,\xi_{k+1}, \qquad X_0 = x_0, \qquad T = m\eta. \tag{1}$$

The variables ξ_k are independent standard Gaussian vectors. The innovation variables ζ_k are independent and identically distributed and independent of the Gaussian sequence. No pathwise differentiability of $x \mapsto Z(x, \zeta)$ is required.

Assumption 2.1 (Smooth dissipative model)**.** The functions $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^{d\times\ell}$ belong to C^6. Known finite envelopes satisfy

$$|b(0)| \le B_0, \quad \|D^j b\|_\infty \le B_j, \qquad \|\sigma_0(0)\|_{\mathrm{HS}} \le S_0, \quad \|D^j\sigma_0\|_\infty \le S_j \quad (1 \le j \le 6).$$

For known $K_1 > 0$ and $K_2 \ge 0$,

$$\langle b(x) - b(y), x - y\rangle \le -K_1|x - y|^2 + K_2 \qquad (x, y \in \mathbb{R}^d).$$

The innovation is centered, has finite covariance, and

$$\Sigma(x) = \mathbb{E}[Z(x, \zeta)Z(x, \zeta)^\top] = \sigma_0(x)\sigma_0(x)^\top.$$

The nonconstant observable $h \in C^5(\mathbb{R}^d)$ has known bounds $\|D^j h\|_\infty \le H_j$, $1 \le j \le 5$. The values of b, σ_0, h may grow linearly.

We use $\log_+ u = \max(0, \log u)$. Vector norms are Euclidean. Derivative norms are multilinear operator norms, with Hilbert–Schmidt output for matrix-valued functions. An explicitly indicated full derivative Hilbert–Schmidt norm sums over all ordered indices. Constants denoted by C, c may change from line to line and depend on the fixed model bounds, dimension, temperature and initial point. Lower bounds on the limiting energy and constants in relative estimates may also depend on the particular observable through its positive limiting energy. All these constants are independent of m, η and the deviation radius. The numerical constructions use upper step bounds specified by the model envelopes. Statements involving nondegeneracy hold for all sufficiently small steps for the fixed model.

Write P_η for the one-step kernel of (1). The modified diffusion has generator and semigroup

$$L_\eta = b \cdot \nabla + \tfrac{1}{2}(\delta I + \eta\Sigma) : D^2, \qquad \mathcal{P}_t^\eta = e^{tL_\eta}. \tag{2}$$

Its invariant law is denoted by π_η, and $\theta_\eta = \pi_\eta h$. The canonical anchored solution is

$$f_\eta(x) = -\int_0^\infty [\mathcal{P}_t^\eta h(x) - \mathcal{P}_t^\eta h(0)]\,dt, \qquad L_\eta f_\eta = h - \theta_\eta, \quad f_\eta(0) = 0. \tag{3}$$

Lemmas A.1–A.3 prove existence and uniqueness in the anchored globally Lipschitz class, and give effective constants $C_c, \rho, G_1, \dots, G_6, M_*$ such that, uniformly for $0 \le \eta \le \eta_0$,

$$\begin{aligned} &\mathrm{Lip}(\mathcal{P}_t^\eta H) \le C_c e^{-\rho t}\,\mathrm{Lip}(H), && \|D^j f_\eta\|_\infty \le G_j \quad (1 \le j \le 6),\\ &\pi_\eta|x|^2 \le M_*, && \pi_\eta|\nabla f_\eta|^2 \longrightarrow \mu_0 > 0. \end{aligned} \tag{4}$$

The constants G_j and M_* are known upper bounds; a positive lower bound on μ_0 is not a numerical input. The notation π_η always refers to the diffusion. Discrete invariant laws will be written $\rho_{\eta,\lambda}$.

For a function v, let $\bar{v}_m = m^{-1}\sum_{k<m} v(X_k)$. The first frozen weak-bias operator is

$$\mathcal{A}_2 F = \tfrac{1}{2}D^2F[b,b] + \tfrac{\delta}{2}\sum_i D^3F[b,e_i,e_i] + \tfrac{\delta^2}{8}\sum_{i,j} D^4F[e_i,e_i,e_j,e_j]. \tag{5}$$

The term $\eta\Sigma$ is already included in L_η. Put $g = \nabla f_\eta$, $H = D^2 f_\eta$, and $\ell_\eta = Hb + (\delta/2)\nabla\Delta f_\eta$. The nonnegative variance density used for normalization is

$$e_\eta^+ = \delta|g + \eta\ell_\eta|^2 + \eta g^\top \Sigma g + \frac{\eta\delta^2}{2}\|H\|_{\mathrm{HS}}^2. \tag{6}$$

It includes the full first conditional-variance correction and may vanish at individual states. Theorem 7.2 bounds its trajectory average away from zero with the probability needed for the tail estimate. The sign convention in (3) makes the corrected empirical center $\bar{h}_m + \eta\overline{\mathcal{A}_2 f_{\eta_m}}$.

Assumption 2.2 (Trajectory class for the enlarged window)**.** Assumption 2.1 holds, and for a known $\kappa > 0$,

$$\langle b(x) - b(y), x - y\rangle \le -\kappa|x-y|^2.$$

The innovation law is state-independent and has the representation

$$Z = B + \tau G', \qquad \mathbb{E}B = 0, \quad |B| \le M, \quad G' \sim N(0, I_d), \quad B \text{ independent of } G', \tag{7}$$

where $M < \infty$ and $\tau \ge 0$ are known. The distribution of B and the values of b may be unknown to the algorithm.

The bounded component may be asymmetric or discrete. If $\tau > 0$, the innovation has unbounded support. Its covariance is constant, $\Sigma = \mathrm{Cov}(B) + \tau^2 I_d$, and a constant covariance factor satisfies the preceding structural assumptions. The directional sub-Gaussian proxy is $M^2 + \tau^2$. The innovation may be zero; the thermal noise with $\delta > 0$ still gives a positive transition density.

2.1 Sampling access and the computed statistic

At any requested state, an oracle returns an independent sample

$$O(x) = b(x) + Z(x,\zeta).$$

Queries can be made after the main trajectory has been observed. Conditional on the main path and the auxiliary queries and replies so far, the next reply has the oracle law at its requested state. Evaluations of h to a requested certified accuracy, ideal Gaussian draws, and the structural upper envelopes are available. Arithmetic and exact oracle draws are the basic operations; no bit-complexity assertion is made. Section 8 also treats non-Gaussian oracle laws with known local fourth-moment envelopes under Assumption 2.1 alone.

Theorem 2.3 (Self-normalized moderate deviations from gradient samples)**.** *Under Assumption 2.2, let $\eta \downarrow 0$, $T = m\eta \to \infty$ and $r \to \infty$. Suppose*

$$\frac{r^4\log(T+e)}{T} \to 0, \qquad \frac{\eta r^3}{\sqrt{T}} \to 0, \qquad r^2\eta^2 \to 0, \qquad r\eta^2\sqrt{T} \to 0. \tag{8}$$

Given $0 < \beta < 1/2$, $0 < e_a < 1$, and positive failure budgets p_E, p_a with $p_E + p_a < 1$, the algorithms in Sections 8–9 return a strictly positive number $\widehat{E}_m$, correction values $\widehat{a}_k$, and an acceptance flag. All their sampling and search loops are finite on every finite input trajectory. They use no value of f_η, θ_η, a covariance factor, or an invariant variance.

Let $\widehat{\widehat{h}}_m$ be computed with deterministic error at most ηe_a and put

$$\widehat{H}_m = \widehat{\widehat{h}}_m + \frac{\eta}{m} \sum_{k<m} \widehat{a}_k, \qquad \widehat{W}_{m,\eta} = \frac{\sqrt{T}(\widehat{H}_m - \theta_\eta)}{\sqrt{\widehat{E}_m}}. \tag{9}$$

If in addition

$$r^2\beta \to 0, \qquad r\eta\sqrt{T}\, e_a \to 0, \qquad \frac{p_E + p_a}{\overline{\Phi}(r)} \to 0, \tag{10}$$

then, for both choices of sign,

$$\sup_{0\le x\le r} \left| \frac{\mathbb{P}(\pm\widehat{W}_{m,\eta} > x)}{\overline{\Phi}(x)} - 1 \right| \longrightarrow 0. \tag{11}$$

The same limit holds for non-strict inequalities. The probability that the algorithm reports failure is $o(\overline{\Phi}(r))$. For fixed dimension and structural bounds, the number of auxiliary oracle samples is bounded by m times a fixed polynomial in $1 + R, \beta^{-1}, e_a^{-1}, 1 + \log(1/(p_E p_a))$, where $R = \max(1, \max_{k<m} |X_k|)$. The deterministic search over candidate functions has a finite bound.

Section 9.4 proves the theorem. The conditional numerical guarantee holds for every finite state array, so no independence between the computed denominator and the main statistic is used. Returning a positive fallback value when a certificate is unsuccessful makes the statistic defined on every transcript; the flag probability is included in the theorem.

Corollary 2.4 (Admissible radii and observation times)**.** *Under $\eta \to 0$, $T \to \infty$, $r \to \infty$, the first and fourth conditions of (8), namely*

$$\frac{r^4 \log(T + e)}{T} \to 0, \qquad r\eta^2\sqrt{T} \to 0, \tag{12}$$

imply its other two conditions. A sequence of radii $r = r_\eta \to \infty$ admits an observation schedule satisfying (12) if and only if

$$r = o(\eta^{-2/3}[\log(1/\eta)]^{-1/6}). \tag{13}$$

In that case, with $\mathsf{L} = \log(1/\eta)$, the schedule

$$T \asymp \eta^{-8/3}\mathsf{L}^{1/3}, \qquad m \asymp \eta^{-11/3}\mathsf{L}^{1/3} \tag{14}$$

works after integer rounding of m. The choices $\beta = \eta^2$, $e_a = \eta$, and $p_E + p_a = e^{-2r^2}$ meet the numerical conditions of Theorem 2.3.

Proof. The first binding condition gives $r/\sqrt{T} \to 0$ and $r^2/\sqrt{T} \to 0$. Consequently

$$\eta^2 r^2 = (r\eta^2\sqrt{T})(r/\sqrt{T}) \to 0, \qquad \eta r^3/\sqrt{T} = (\eta r)(r^2/\sqrt{T}) \to 0.$$

For necessity of (13), multiplying the first quantity in (12) by the square of the second yields $\eta^4 r^6 \log(T + e) \to 0$. Eventually $T \ge r^4$. If $r \le \eta^{-1/3}$, then $\eta^4 r^6 \mathsf{L} \le \eta^2 \mathsf{L} \to 0$. If $r > \eta^{-1/3}$, then

$$\eta^4 r^6 \mathsf{L} \le 3\eta^4 r^6 \log r \le \tfrac{3}{4}\eta^4 r^6 \log(T + e) \to 0.$$

The two cases prove (13). Conversely, substitution of (14) makes the two binding errors of orders $O((r\eta^{2/3}\mathsf{L}^{1/6})^4)$ and $O(r\eta^{2/3}\mathsf{L}^{1/6})$, proving sufficiency. For the stated numerical tolerances the first two errors in (10) reduce to $r^2\eta^2$ and $r\eta^2\sqrt{T}$, respectively. Mills' bounds give $e^{-2r^2}/\overline{\Phi}(r) \le C(1+r)e^{-3r^2/2} \to 0$. □

The equivalence in Corollary 2.4 characterizes the radii permitted by the two sufficient error bounds in (12). A particular model may admit a larger range.

The reported pair $(\widehat{H}_m, \widehat{E}_m)$ also gives confidence intervals for the unknown target:

$$I_m(z) = \left[\widehat{H}_m - z\sqrt{\widehat{E}_m/T},\ \widehat{H}_m + z\sqrt{\widehat{E}_m/T}\right].$$

For $0 \le z \le r$, the noncoverage probability is $2\overline{\Phi}(z)(1+o(1))$, uniformly in z. This follows by adding the two strict tails in (11); the interval algorithm never evaluates θ_η.

3 Positive variance completion and weighted remainder bounds

We retain the recursion

$$X' = x + \eta b(x) + \eta Z(x,\zeta) + \sqrt{\eta\delta}\,\xi, \qquad q_x = 1 + |x|^2,$$

where ξ is standard Gaussian, independent of the centered innovation, $|b(x)| \le C q_x^{1/2}$, and

$$\mathbb{E}e^{t\cdot Z(x,\zeta)} \le e^{K^2 q_x |t|^2/2}, \qquad \Sigma(x) = \mathbb{E}Z(x,\zeta)Z(x,\zeta)^\top.$$

Constants may depend on fixed dimension, positive δ, the displayed envelopes and the derivative bounds of the test function, but are independent of x and η. The results apply uniformly to any family of test functions with the stated common derivative bounds, including the canonical Poisson solutions in (3).

3.1 A weighted variance expansion beyond compact sets

Proposition 3.1 (Uniform first variance correction)**.** *Suppose $F \in C^5(\mathbb{R}^d)$ and its derivatives of orders one through five are globally bounded. Put $g = \nabla F$, $H = D^2F$, and*

$$v_1^F = g^\top \Sigma g + 2\delta H[b,g] + \frac{\delta^2}{2}\|H\|_{\mathrm{HS}}^2 + \delta^2 \sum_i D^3F[g, e_i, e_i].$$

For every fixed $c_0 < \infty$, there are fixed $\eta_0, C > 0$ such that

$$\left|\mathrm{Var}_x\, F(X') - \eta\delta|g(x)|^2 - \eta^2 v_1^F(x)\right| \le C\eta^3 q_x^{3/2}, \qquad 0 < \eta \le \eta_0, \quad \eta q_x \le c_0. \tag{15}$$

The estimate also applies to a cutoff family with common derivative bounds. The conditional variance includes all possible next states, and its coefficients are evaluated at the pre-step point x.

Proof. Expand first at the drifted point. Write

$$y = x + \eta b(x), \qquad W = \sqrt{\eta}\, U + \eta Z(x, \zeta), \qquad U = \sqrt{\delta}\, \xi.$$

The centered vector W has directional sub-Gaussian proxy $a_x = \eta\delta + \eta^2 K^2 q_x \le C\eta$. In particular $\mathbb{E}|W|^p \le C_p \eta^{p/2}$ for every fixed p. Taylor expansion to degree four gives

$$F(y + W) - F(y) = L + Q + C_3 + C_4 + R_5,$$

where

$$\begin{aligned} L &= g_y \cdot W, \qquad Q = \tfrac{1}{2} H_y[W, W], \\ C_3 &= \tfrac{1}{6} D^3 F(y)[W, W, W], \qquad C_4 = \tfrac{1}{24} D^4 F(y)[W, W, W, W], \\ \|R_5\|_2 &\le C\eta^{5/2}. \end{aligned}$$

The last estimate uses the fifth derivative bound and the tenth moment of W. Every covariance involving R_5 is therefore $O(\eta^3)$, and every polynomial product of degree at least six is also $O(\eta^3)$.

The relevant moment identities are

$$\begin{aligned} \mathbb{E}W^{\otimes 2} &= \eta\delta I + \eta^2 \Sigma, \qquad \mathbb{E}W^{\otimes 3} = \eta^3 \mathbb{E}Z^{\otimes 3}, \\ \mathbb{E}W^{\otimes 4} &= \eta^2 \mathbb{E}U^{\otimes 4} + \eta^3 \operatorname{Sym}(\mathbb{E}U^{\otimes 2} \otimes \Sigma) + \eta^4 \mathbb{E}Z^{\otimes 4}, \\ \mathbb{E}W^{\otimes 5} &= \eta^4 \operatorname{Sym}(\mathbb{E}U^{\otimes 2} \otimes \mathbb{E}Z^{\otimes 3}) + \eta^5 \mathbb{E}Z^{\otimes 5}. \end{aligned}$$

Here each symmetrization sums the distinct placements. The potential term with four Gaussian factors and one innovation vanishes because the innovation is centered. Hence

$$\|\mathbb{E}W^{\otimes 3}\| \le C\eta^3 q_x^{3/2}, \qquad \|\mathbb{E}W^{\otimes 5}\| \le C\eta^4 q_x^{3/2},$$

where the local condition absorbs the extra factor ηq_x in the pure fifth innovation moment.

The linear variance is exact:

$$\operatorname{Var} L = \eta\delta |g_y|^2 + \eta^2 g_y^\top \Sigma g_y.$$

Moreover $2\operatorname{Cov}(L, Q) = O(\eta^3 q_x^{3/2})$. The fourth-moment formula and Gaussian pairings give

$$\begin{aligned} \operatorname{Var} Q &= \frac{\eta^2\delta^2}{2} \|H_y\|_{\mathrm{HS}}^2 + O(\eta^3 q_x), \\ 2\operatorname{Cov}(L, C_3) &= \eta^2\delta^2 \sum_i D^3 F(y)[g_y, e_i, e_i] + O(\eta^3 q_x). \end{aligned}$$

The two degree-five covariances are bounded by $C\eta^4 q_x^{3/2}$. For $\operatorname{Cov}(Q, C_3)$, the product of the means has this same order. Thus, with

$$A(y) = \tfrac{1}{2} \|H_y\|_{\mathrm{HS}}^2 + \sum_i D^3 F(y)[g_y, e_i, e_i],$$

we obtain

$$\operatorname{Var}_x F(X') = \eta\delta |g_y|^2 + \eta^2 g_y^\top \Sigma g_y + \eta^2\delta^2 A(y) + O(\eta^3 q_x^{3/2}).$$

Return to x. The bounded derivatives imply

$$\begin{aligned} |g_y|^2 &= |g_x|^2 + 2\eta H_x[b(x), g_x] + O(\eta^2 |b(x)|^2), \\ |g_y^\top \Sigma g_y - g_x^\top \Sigma g_x| &\le C\eta |b(x)| \|\Sigma(x)\| \le C\eta q_x^{3/2}, \\ |A(y) - A(x)| &\le C\eta |b(x)|. \end{aligned}$$

Substitution proves the assertion. □

The exponent $3/2$ cannot be decreased for the general bounded-derivative test-function class in this proposition. For example, let $d = 1$, $F(x) = \sin x$, $b(x) = -x$, and $Z = x\zeta$ with ζ standard Gaussian. Then X' is Gaussian with mean $(1-\eta)x$ and variance $\eta\delta + \eta^2 x^2$. The coefficient of η^3 in the corrected variance contains

$$-2x^3 F'(x) F''(x) + O(1 + x^2).$$

Along $x = \pi/4 + 2\pi n$ this is $x^3 + O(x^2)$. For each such fixed state one may first send η to zero, so that the local condition is satisfied. A uniform $C\eta^3 q_x^a$ remainder with $a < 3/2$ would contradict these coefficients. This example establishes sharpness over the stated class of test functions. Sharpness for Poisson solutions of a fixed observable would require a separate argument.

3.2 A pointwise nonnegative variance completion

Proposition 3.2 (Positive completed energy)**.** *Set*

$$t_F = \nabla \Delta F, \qquad \ell_F = Hb + \frac{\delta}{2} t_F,$$

and define

$$e_\eta^{+,F} = \delta |g + \eta \ell_F|^2 + \eta |\sigma_0^\top g|^2 + \frac{\eta \delta^2}{2} \|H\|_{\mathrm{HS}}^2. \tag{16}$$

Then $e_\eta^{+,F} \ge 0$ *pointwise and*

$$e_\eta^{+,F} = \delta |g|^2 + \eta v_1^F + \eta^2 \delta |\ell_F|^2, \qquad |\ell_F(x)|^2 \le C q_x. \tag{17}$$

Under the assumptions of Proposition 3.1,

$$\left| \eta^{-1} \operatorname{Var}_x F(X') - e_\eta^{+,F}(x) \right| \le C\eta^2 q_x^{3/2}. \tag{18}$$

Proof. Expand the first square in (16). The identity

$$2\delta \langle g, \ell_F \rangle = 2\delta H[b, g] + \delta^2 \sum_i D^3 F[g, e_i, e_i]$$

produces exactly the two signed terms of v_1^F. The remaining terms are squares. Linear growth of b and the global derivative bounds give the estimate for ℓ_F; combine it with (15). □

For an Ornstein–Uhlenbeck model with constant innovation variance σ^2 and $F(x) = x$, the completion is exactly $\delta + \eta\sigma^2$. For nonlinear F, both the Hessian and the third derivative can contribute at first order.

3.3 Global expansions under a uniform noise bound

Proposition 3.3 (Global bounds under a uniform innovation proxy)**.** *Suppose that the innovation has the uniform bound*

$$\mathbb{E}e^{t\cdot Z(x,\zeta)} \le e^{K_0^2|t|^2/2} \quad \text{for every } x, t.$$

With the same derivative assumptions as above,

$$\left|\operatorname{Var}_x F(X') - \eta\delta|g|^2 - \eta^2 v_1^F\right| + \eta|\eta^{-1}\operatorname{Var}_x F(X') - e_\eta^{+,F}| \le C\eta^3(1+|b(x)|^2) \tag{19}$$

for every x, without a condition on ηq_x. If F also has a globally bounded sixth derivative, then

$$|P_\eta F - F - \eta L_\eta F - \eta^2 \mathcal{A}_2 F| \le C\eta^3(1+|b(x)|^3) \tag{20}$$

for every x. The global square-exponential bound for the centered increment of any globally Lipschitz F also holds, with scale η and constants independent of the state.

Proof. Now W has proxy $\eta\delta + \eta^2 K_0^2 \le C\eta$ for all x, all innovation moments are bounded uniformly, and $\|\Sigma(x)\| \le K_0^2$. The variance expansion at $y = x + \eta b(x)$ used above has an $O(\eta^3)$ remainder. Returning to x adds at most $C\eta^3(|b|^2 + |b|)$, proving (19), including the square completion term.

For the weak formula, the centered-noise expansion at y gives

$$\mathbb{E}F(y+W) - F(y) = \frac{\eta\delta}{2}\Delta F(y) + \frac{\eta^2}{2}\Sigma(x) : D^2F(y) + \frac{\eta^2\delta^2}{8}\Delta^2 F(y) + O(\eta^3).$$

All odd Gaussian moments vanish; the third innovation moment is of order η^3, and the fifth-degree terms and sixth-order remainder have at most that size. Expand $F(y) - F(x)$ to deterministic order two and $\Delta F(y)$ to order one. The remainders are bounded by $C\eta^3|b|^3$ and $C\eta^3|b|^2$, respectively. Moving each of the two order-η^2 terms back to x costs $C\eta^3|b|$. This is exactly (20). Finally, the centered increment is dominated, before centering, by the Lipschitz constant times $|W|$, proving the exponential assertion. □

The uniform proxy is stronger than bounded covariance. For example, put $a(x) = \sqrt{1+x^2}$ and let $Z(x)$ take $+a(x)$ and $-a(x)$ with probabilities $1/(2a(x)^2)$ each, and zero otherwise. Then

$$\mathbb{E}Z(x) = 0, \qquad \mathbb{E}Z(x)^2 = 1, \qquad \mathbb{E}Z(x)^4 = a(x)^2.$$

Its covariance factor can be the smooth constant one, and it satisfies the state-dependent proxy used at the start of this section. Nevertheless its fourth moments are unbounded, and its square-exponential bound at the innovation scale is not uniform in the state. Thus a bounded covariance hypothesis alone cannot justify Proposition 3.3.

4 Contraction and exact clocks for the tilted chain

Throughout this section Assumption 2.2 holds. Write $L = B_1$, $\Psi_\eta(x) = x + \eta b(x)$ and $\vartheta_\eta = \delta + \eta\tau^2$. Combining the thermal noise with the Gaussian component of the innovation gives the one-step law

$$Y = \Psi_\eta(x) + W, \qquad W = \eta B + \sqrt{\eta\vartheta_\eta}\, G, \qquad B \text{ independent of } G.$$

Take $F = f_\eta$ or $F = -f_\eta$ and define

$$c_\eta(x) = \eta^{-1} \operatorname{Var}_x F(X_1), \qquad V_m = \eta \sum_{k<m} c_\eta(X_k), \qquad Q_{\eta,\lambda}(x, dy) = \frac{e^{\lambda F(y)} P_\eta(x, dy)}{P_\eta e^{\lambda F}(x)}. \tag{21}$$

Both signs give the same variance. Since $L_\eta f_\eta = h - \theta_\eta$ and h is nonconstant, f_η is nonconstant. The continuity of f_η and the positive transition density on $\mathbb{R}^d$ then give $c_\eta(x) > 0$ at every x.

4.1 Uniform estimates for the actual tilted kernel

Lemma 4.1 (Contraction and concentration of the tilted kernel)**.** *There are constants $\lambda_0, \eta_0, c, C > 0$ such that for $0 < \eta \le \eta_0$ and $|\lambda| \le \lambda_0$,*

$$W_1(Q_{\eta,\lambda}(x,\cdot), Q_{\eta,\lambda}(x',\cdot)) \le q_\eta |x - x'|, \qquad q_\eta \le 1 - c\eta, \tag{22}$$

$$\mathbb{E}_x^\lambda e^{z(H(X_1) - \mathbb{E}_x^\lambda H(X_1))} \le \exp\{C\eta z^2 \operatorname{Lip}(H)^2/2\} \quad (z \in \mathbb{R}). \tag{23}$$

The second assertion holds for every Lipschitz H. The kernel has a unique invariant probability $\rho_{\eta,\lambda}$, and its chain from x_0 satisfies

$$\sup_m \mathbb{E}^\lambda |X_m|^2 \le C(1 + |x_0|^2), \qquad \mathbb{P}^\lambda\{|X_m| > C(1 + |x_0|) + u\} \le 2e^{-cu^2}, \tag{24}$$

$$W_1(\rho_{\eta,\lambda}, \rho_{\eta,\lambda'}) \le C|\lambda - \lambda'|. \tag{25}$$

The invariant second moments are uniformly bounded.

Proof. The centered noise has density

$$k_\eta(w) = \mathbb{E}\varphi_{\eta\vartheta_\eta I}(w - \eta B), \qquad U_\eta = -\log k_\eta,$$

where φ_A is the centered Gaussian density with covariance A. Differentiating under the bounded mixing law gives

$$D^2 U_\eta(w) = \frac{I}{\eta\vartheta_\eta} - \frac{\operatorname{Cov}(B \mid w)}{\vartheta_\eta^2} \succeq k_\eta^* I, \qquad k_\eta^* = (\eta\vartheta_\eta)^{-1} - M^2/\vartheta_\eta^2. \tag{26}$$

Here $\operatorname{Cov}(B \mid w)$ is computed under the conditional law of B given $W = w$. Since $|B| \le M$, reducing η_0 so that $\eta M^2/\vartheta_\eta \le 1/2$ gives $k_\eta^* \ge (2\eta\vartheta_\eta)^{-1}$. The Hessian is also bounded above by $(\eta\vartheta_\eta)^{-1} I$.

For a given location a, use the noise coordinate $w = y - a$. Its tilted potential is $U_\eta(w) - \lambda F(a + w)$, with curvature at least $k_\eta^* - |\lambda| G_2$. The auxiliary Langevin diffusion with this potential and noise coefficient $\sqrt{2}I$ is globally Lipschitz and strongly dissipative. Its invariant density is proportional to $e^{-U_\eta(w) + \lambda F(a+w)}$: strong convexity gives integrability, and substitution into the adjoint generator gives stationarity. Couple synchronously the diffusions for locations a, a'. Splitting the drift difference into its state difference and parameter difference gives

$$\frac{d}{ds}|w_s - w_s'| \le -(k_\eta^* - |\lambda| G_2)|w_s - w_s'| + |\lambda| G_2 |a - a'|.$$

Pass to the invariant laws and translate back to y to obtain

$$W_1(y_a, y_{a'}) \le \frac{k_\eta^*}{k_\eta^* - |\lambda| G_2} |a - a'| \le (1 + C\eta|\lambda|)|a - a'|. \tag{27}$$

At $\lambda = 0$ the two kernels are translates. The deterministic step satisfies

$$|\Psi_\eta(x) - \Psi_\eta(x')|^2 \le (1 - 2\kappa\eta + L^2\eta^2)|x - x'|^2.$$

For $\eta \le \kappa/L^2$, its Lipschitz constant is at most $1 - \kappa\eta/2$. Choose λ_0 sufficiently small in (27). This proves (22).

To prove one-step concentration, consider a strongly convex potential of curvature k. Its diffusion semigroup S_s satisfies $\mathrm{Lip}(S_s H) \le e^{-ks}\,\mathrm{Lip}(H)$ by synchronous coupling. Starting from a fixed point, the Doob martingale for $H(Y_u)$ has quadratic variation at most

$$2\int_0^u \mathrm{Lip}(S_{u-s}H)^2 ds \le \mathrm{Lip}(H)^2/k.$$

The exponential martingale inequality therefore bounds its centered moment generating function by $\exp(z^2\,\mathrm{Lip}(H)^2/(2k))$. Pass to the invariant law using convergence of means and Fatou's lemma. Smooth approximation extends the assertion to Lipschitz H. In our case $k = k_\eta^* - |\lambda|G_2 \ge c/\eta$, proving (23).

Integration by parts in the Gaussian component of the joint tilted law of B, G gives

$$\mathbb{E}^\lambda(Y - a) = \eta\mathbb{E}^\lambda B + \lambda\eta\vartheta_\eta\mathbb{E}^\lambda\nabla F(Y).$$

Its norm is bounded by $C\eta$, uniformly in a. Applying (23) to the coordinates gives $\mathrm{tr}\,\mathrm{Cov}^\lambda(Y) \le C\eta$. The strong drift inequality at the origin and Young's inequality now imply

$$\mathbb{E}_x^\lambda|X_1|^2 \le (1 - c\eta)|x|^2 + C\eta.$$

Iteration proves the moment bounds. Contraction on the complete W_1 space gives an invariant law. Weak convergence from x_0 and lower semicontinuity give its second moment bound. From every deterministic initial state the iterates converge to this law. Integrating a bounded Lipschitz test function against any invariant initial law and using dominated convergence proves uniqueness among all invariant probabilities.

For the state concentration, the Doob differences for $H(X_m)$ have conditional sub-Gaussian proxies $C\eta q_\eta^{2(m-j)}\,\mathrm{Lip}(H)^2$. Their sum is at most $C\eta/(1 - q_\eta^2)\,\mathrm{Lip}(H)^2 \le C'\,\mathrm{Lip}(H)^2$. Thus $H(X_m)$ is uniformly sub-Gaussian about its mean. Use $H(x) = |x|$ and the second-moment bound to obtain (24). Finally, at a fixed location a, synchronous coupling of the auxiliary diffusions with tilts λ, λ' yields

$$W_1(y_{a,\lambda}, y_{a,\lambda'}) \le \frac{G_1|\lambda - \lambda'|}{k_\eta^* - \lambda_0 G_2} \le C\eta|\lambda - \lambda'|.$$

Use invariance and divide by $1 - q_\eta \ge c\eta$ to prove (25). □

Lemma 4.2 (A bounded Lipschitz variance observable)**.** *Uniformly in sufficiently small η, $0 < c_\eta \le C$ and $\mathrm{Lip}(c_\eta) \le C$. The deterministic means $\nu_{\eta,\lambda} = \rho_{\eta,\lambda}c_\eta$ are uniformly Lipschitz in λ and, after decreasing fixed η_0, λ_0 if necessary, satisfy $0 < \nu_* \le \nu_{\eta,\lambda} \le \nu^* < \infty$.*

Proof. The noise W is state-independent and has second moment at most $C\eta$. An independent copy and the Lipschitz bound on F give $\mathrm{Var}\,F(\Psi_\eta(x) + W) \le C\eta$. Differentiation under the expectation gives, for $|v| = 1$,

$$Dc_\eta(x)[v] = \frac{2}{\eta}\,\mathrm{Cov}\big(F(Y), DF(Y)[D\Psi_\eta(x)v]\big).$$

Applying Cauchy–Schwarz to this covariance, with the independent-copy bound for each factor, proves the uniform Lipschitz estimate. Lemma 4.1 now gives the claimed continuity of $\nu_{\eta,\lambda}$.

We identify its zero-tilt limit. Couple the discrete chain on a fixed physical-time interval with the limiting diffusion $d\mathsf{X}_t = b(\mathsf{X}_t)dt + \sqrt{\delta}\, dB_t$ using the same thermal increments. The additional increments $\eta(B + \tau G')$ form a centered martingale with accumulated second moment $O(\eta)$ on that interval. The Euler interpolation identity and the Lipschitz drift, followed by the discrete Gronwall inequality, give root mean square error $C_t\sqrt{\eta}$ for initial distributions with uniformly bounded second moment. Indeed, the limiting diffusion has $\mathbb{E}|\mathsf{X}_s - \mathsf{X}_{\lfloor s/\eta \rfloor \eta}|^2 \le C_t\eta$; integrating this local error in the drift and adding the extra martingale proves the stated estimate.

Start with $\rho_{\eta,0}$ and use an integer number of steps approximating a fixed $t_* > 0$. Invariance of the discrete law and $e^{-\kappa t}$ contraction of the limiting diffusion imply

$$W_1(\rho_{\eta,0}, \pi_0) \le C_{t_*}\sqrt{\eta} + e^{-\kappa t_*/2} W_1(\rho_{\eta,0}, \pi_0)$$

for small η. Hence this distance tends to zero. Lemma A.3 gives local convergence of ∇f_η to ∇f_0. Expanding around $\Psi_\eta(x)$ with the uniform Hessian bound shows, uniformly on compact sets, $c_\eta(x) \to \delta|\nabla f_0(x)|^2$. The uniform global bound and tightness therefore imply

$$\nu_{\eta,0} \longrightarrow \delta\pi_0|\nabla f_0|^2 = \delta\mu_0 > 0.$$

Uniform Lipschitz continuity in λ completes the proof. The lower bound on $\nu_{\eta,\lambda}$ depends on the fixed observable through μ_0. The upper derivative envelopes alone do not bound μ_0 away from zero. □

4.2 An exact discrete Poisson equation

Proposition 4.3 (All-confidence concentration of the original clock)**.** *For $|\lambda| \le \lambda_0$ there is an anchored globally Lipschitz function $u_{\eta,\lambda}$ satisfying*

$$(Q_{\eta,\lambda} - I)u_{\eta,\lambda} = \eta(c_\eta - \nu_{\eta,\lambda}), \qquad u_{\eta,\lambda}(0) = 0, \qquad \mathrm{Lip}(u_{\eta,\lambda}) \le C.$$

For $T = m\eta \ge 1$ and every $t \ge 1$,

$$\mathbb{P}^\lambda\left\{\left|\frac{V_m}{T\nu_{\eta,\lambda}} - 1\right| > C\sqrt{t/T}\right\} \le 4e^{-t}. \tag{28}$$

Here V_m uses the original kernel's conditional variance, evaluated along the actual tilted path.

Proof. Define the anchored resolvent

$$u_{\eta,\lambda}(x) = -\eta\sum_{n=0}^{\infty}\{Q^n_{\eta,\lambda}c_\eta(x) - Q^n_{\eta,\lambda}c_\eta(0)\}.$$

Contraction gives absolute convergence and $\mathrm{Lip}(u) \le \eta\,\mathrm{Lip}(c_\eta)/(1-q_\eta) \le C$. For the partial sum u_N through $n = N$, anchoring cancels in the difference and $(Q_{\eta,\lambda} - I)u_N = \eta(c_\eta - Q^{N+1}_{\eta,\lambda}c_\eta)$. The uniform linear bound on u_N permits passage under the one-step expectation. Taking $N \to \infty$ gives the equation. Consequently

$$V_m - T\nu_{\eta,\lambda} = u(X_m) - u(x_0) - \sum_{k<m}\{u(X_{k+1}) - Q_{\eta,\lambda}u(X_k)\}.$$

The last sum has sub-Gaussian proxy CT by (23). Equation (24) and the uniform Lipschitz bound give

$$\mathbb{P}^\lambda\{|u(X_m) - u(x_0)| > C(1 + \sqrt{t})\} \le 2e^{-t}.$$

A union bound, $T, t \ge 1$, and the lower bound on $\nu_{\eta,\lambda}$ yield (28). □

4.3 Cumulants of the full Poisson increment

Lemma 4.4 (Uniform third-cumulant cancellation)**.** *Let* $d_{k+1} = F(X_{k+1}) - P_\eta F(X_k)$ *and* $\psi_k(\lambda) = \log \mathbb{E}(e^{\lambda d_{k+1}} \mid \mathcal{F}_k)$*. There are constants* $c, C > 0$ *such that*

$$\mathbb{E}(e^{cd_{k+1}^2/\eta} \mid \mathcal{F}_k) \le C, \qquad \sup_{|\lambda| \le \lambda_0} |\psi_k'''(\lambda)| \le C\eta^2. \tag{29}$$

The product exponential martingale for these increments has transition kernel $Q_{\eta,\lambda}$ *in* (21)*.*

Proof. Fix a state and abbreviate $a = \Psi_\eta(x)$. The Lipschitz bound on F and the global sub-Gaussian proxy of W prove the first assertion, including centering. For the second write

$$F(a+W) - F(a) = A + R, \qquad A = \nabla F(a) \cdot W, \qquad |R| \le G_2 |W|^2/2.$$

Under the linear tilt proportional to $e^{\lambda A}$, the Gaussian part is translated and remains independent of the bounded part. Thus A has third centered moment $O(\eta^3)$, centered L^j norm $O(\sqrt{\eta})$, and R has L^j norm $O(\eta)$, for every fixed j. These estimates are uniform in a and $|\lambda| \le \lambda_0$. Small η_0 also gives the same moment bounds weighted by $e^{2|\lambda R|}$, since the Gaussian variance is $O(\eta)$ and B is bounded.

The density from the linear tilt to the full tilt is $D = e^{\lambda R}/\mathbb{E}_{\mathrm{lin}} e^{\lambda R}$. The inequality $|e^z - 1| \le |z| e^{|z|}$ controls the numerator. The denominator is at least $\exp(\lambda \mathbb{E}_{\mathrm{lin}} R) \ge e^{-C\eta}$ by Jensen's inequality. Hence $\|D - 1\|_{L^2(\mathrm{lin})} \le C\eta$. Cauchy–Schwarz shows that the mean of A changes by $O(\eta^{3/2})$ and its third centered moment by $O(\eta^{5/2})$. Under the full tilt, the largest term in the centered cubic expansion of $A + R$ is bounded by

$$|\mathbb{E}[(A - \mathbb{E}A)^2 (R - \mathbb{E}R)]| \le C\eta^2;$$

the other mixed terms are smaller. The third derivative of a logarithmic moment generating function is precisely this tilted third centered moment. Conditional subtraction of $P_\eta F(x)$ leaves third cumulants unchanged, proving (29). Finally, the conditional tilt factor equals $e^{\lambda F(y)}/P_\eta e^{\lambda F}(x)$, which proves the kernel assertion. □

5 A self-normalized martingale comparison

The change of measure involves two predictable quadratic variations. The original martingale supplies the random boundary. The centered martingale under the tilt supplies the normal approximation. We estimate their difference using conditional third cumulants. Write $\overline{\Phi} = 1 - \Phi$, and let d_K denote Kolmogorov distance.

5.1 A bracket repair lemma

Lemma 5.1 (Repair of a bracket known with high probability)**.** *Let* $(e_j, \mathcal{H}_j)_{j=1}^n$ *be martingale differences on an arbitrary probability space. Suppose that* $|e_j| \le \varepsilon$ *almost surely, where* $0 < \varepsilon \le 1/2$*, and that*

$$\mathbb{P}\left(\left| \sum_{j=1}^n \mathbb{E}(e_j^2 \mid \mathcal{H}_{j-1}) - 1 \right| > D \right) \le p, \qquad 0 < D \le 1/4.$$

Then

$$d_K\left(\sum_{j=1}^n e_j, N(0,1) \right) \le C\{\varepsilon |\log \varepsilon| + \sqrt{D}\} + p. \tag{30}$$

Proof. Put $v_j = \mathbb{E}(e_j^2 \mid \mathcal{H}_{j-1})$. Retain the increments until adding the next v_j would make their accumulated predictable variance exceed $1+D$; suppress that increment and all remaining original increments. This is a predictable operation. The resulting bracket V_* is at most $1+D$. Extend the probability space by $K = \lceil \varepsilon^{-2} \rceil$ independent Rademacher variables and append the increments

$$((1 - D - V_*)_+/K)^{1/2}\,\zeta_i, \qquad 1 \le i \le K.$$

Their amplitudes are at most ε. The final bracket is $\max\{V_*, 1-D\} \in [1-D, 1+D]$ on every path. On the event that the original bracket lies in this interval no original increment is suppressed and every appended increment is zero. The two terminal sums consequently agree with probability at least $1-p$.

Every increment a_j of the repaired sequence satisfies

$$|\mathbb{E}(a_j^\ell \mid \mathcal{H}_{j-1})| \le \mathbb{E}(|a_j|^\ell \mid \mathcal{H}_{j-1}) \le \varepsilon^{\ell-2}\mathbb{E}(a_j^2 \mid \mathcal{H}_{j-1}), \qquad \ell \ge 3.$$

The martingale normal-approximation theorem of Fan–Grama–Liu [FGL13, Theorem 3.1, equation (24)] therefore applies with their Bernstein parameter ε and their bracket parameter $\delta = \sqrt{D} \le 1/2$. It gives $C(\varepsilon|\log\varepsilon| + \sqrt{D})$ for the repaired sum. The coupling just constructed contributes at most p to Kolmogorov distance. The argument is unchanged when the underlying probability is an exponentially tilted measure. □

5.2 Finite comparison at a random boundary

Let $(d_{k+1}, \mathcal{F}_{k+1})_{k=0}^{m-1}$ be real martingale differences under $\mathbb{P}$, let $T = m\eta$, and define

$$S = \sum_{k=0}^{m-1} d_{k+1}, \qquad v_k = \mathbb{E}(d_{k+1}^2 \mid \mathcal{F}_k), \qquad V = \sum_{k=0}^{m-1} v_k,$$

$$\psi_k(\lambda) = \log \mathbb{E}(e^{\lambda d_{k+1}} \mid \mathcal{F}_k), \qquad \Psi(\lambda) = \sum_{k=0}^{m-1} \psi_k(\lambda), \qquad \left.\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}}\right|_{\mathcal{F}_m} = e^{\lambda S - \Psi(\lambda)}. \tag{31}$$

The last expression is the terminal value of a positive mean-one product martingale whenever the conditional moment generating functions below are finite.

Proposition 5.2 (Finite random-boundary comparison)**.** *Fix $c_0, C_0, \lambda_0, \nu_*, \nu^* > 0$. Suppose $0 < \eta \le 1$, $T = m\eta \ge 2$, and, almost surely,*

$$\mathbb{E}\left(e^{c_0 d_{k+1}^2/\eta} \mid \mathcal{F}_k\right) \le C_0, \qquad \sup_{|\lambda| \le \lambda_0} |\psi_k'''(\lambda)| \le C_0\eta^2. \tag{32}$$

Assume that a deterministic continuous function $\lambda \mapsto \nu_\lambda \in [\nu_*, \nu^*]$ on $[-\lambda_0, \lambda_0]$ satisfies, for every $t \ge 1$,

$$\mathbb{Q}_\lambda\left(\left|\frac{V}{T\nu_\lambda} - 1\right| > C_0\sqrt{t/T}\right) \le C_0 e^{-t}. \tag{33}$$

Suppose also that $V > 0$ almost surely. For each fixed $J \ge 2$, there are constants $C_J, C, c > 0$, depending only on J and the displayed input constants, such that the following conclusion holds. Set

$$D = C_J\left\{\sqrt{\frac{\log(T+e)}{T}} + \frac{\eta r}{\sqrt{T}}\right\}, \qquad \varepsilon = C_J\sqrt{\frac{\eta\log(m+e)}{T}},$$

$$B_r = \frac{\eta r^3}{\sqrt{T}}, \qquad E_r = (1+r)\sqrt{D} + (1+r)^2 D + (1+r)\varepsilon|\log\varepsilon| + B_r. \tag{34}$$

If $r \geq 1$,

$$\frac{r}{\sqrt{T\nu_*}} \leq \lambda_0, \qquad D \leq \frac{1}{4}, \qquad \varepsilon \leq e^{-1}, \qquad r^2 \leq cT, \tag{35}$$

then

$$\sup_{0\leq x\leq r} \left| \frac{\mathbb{P}(S > x\sqrt{V})}{\overline{\Phi}(x)} - 1 \right| \leq CE_r e^{CE_r} + C_J(1+r)(T+e)^{-J} e^{CE_r}. \tag{36}$$

The same bound holds with $>$ *replaced by* $\geq$*, and for the two corresponding lower tails.*

Proof. All constants below are uniform in the permitted parameters. The square-exponential bound gives $v_k \leq C\eta$, hence $V \leq CT$ on every path. At the origin, $\psi_k(0) = \psi_k'(0) = 0$ and $\psi_k''(0) = v_k$. Integrating the third-derivative estimate gives

$$\begin{aligned} &|\Psi(\lambda) - \lambda^2 V/2| \leq C\eta T|\lambda|^3, \\ &\left|\sum_k \psi_k'(\lambda) - \lambda V\right| \leq C\eta T\lambda^2, \qquad \left|\sum_k \psi_k''(\lambda) - V\right| \leq C\eta T|\lambda|. \end{aligned} \tag{37}$$

Choice of the tilt and its centered martingale. Fix $x \in [0, r]$. By continuity and the lower bound on ν_λ, there is a solution

$$0 \leq \lambda \leq \frac{x}{\sqrt{T\nu_*}}, \qquad \lambda\sqrt{T\nu_\lambda} = x.$$

Choose any such λ and put $\sigma = \sqrt{T\nu_\lambda}$, $t_\lambda = V/\sigma^2$, and

$$U_\lambda = \frac{S - \sum_k \psi_k'(\lambda)}{\sigma}.$$

Under $\mathbb{Q}_\lambda$ the centered increments have conditional means zero and variances $\psi_k''(\lambda)$. Thus the bracket of U_λ is $\sigma^{-2}\sum_k \psi_k''(\lambda)$, whereas $t_\lambda = \sigma^{-2}\sum_k \psi_k''(0)$ is the original bracket evaluated on the same path. By (37) their difference is at most $C\eta x/\sqrt{T}$.

A Kolmogorov estimate under the tilt. Conditional Jensen gives $\psi_k(\lambda) \geq 0$. Young's inequality in the conditional change of measure implies, for fixed $c_1, C_1 > 0$,

$$\mathbb{E}^{\mathbb{Q}_\lambda}\left[\exp\left\{c_1|d_{k+1} - \psi_k'(\lambda)|^2/\eta\right\} \mid \mathcal{F}_k\right] \leq C_1.$$

Indeed, $|\lambda d| \leq c_0 d^2/(4\eta) + C\eta\lambda_0^2$; the conditional tilted first moment is $O(\sqrt{\eta})$, and centering only changes the constants. Denote these centered tilted increments by a_{k+1} and let

$$L_J = A_J\sqrt{\eta\log(m+e)}, \qquad \widetilde{a}_{k+1} = a_{k+1}\mathbf{1}_{\{|a_{k+1}|\leq L_J\}} - \mathbb{E}^{\mathbb{Q}_\lambda}(a_{k+1}\mathbf{1}_{\{|a_{k+1}|\leq L_J\}} \mid \mathcal{F}_k).$$

Taking the fixed constant A_J sufficiently large gives, for $j = 0, 1, 2$ and any prescribed fixed power K,

$$\mathbb{E}^{\mathbb{Q}_\lambda}\left[|a_{k+1}|^j \mathbf{1}_{\{|a_{k+1}|>L_J\}} \mid \mathcal{F}_k\right] \leq C_J\eta^{j/2}(m+e)^{-K}.$$

The probability of deleting an increment is therefore at most $C_J m(m+e)^{-K}$. On its complement, the difference between the two normalized sums is at most $C_J\sqrt{m}(m+e)^{-K}$. The normalized bracket difference is at most $C_J(m+e)^{-K}$: use the $j = 2$ bound and the square of the conditional

deleted first moment, sum over m, and divide by $\sigma^2 \geq T\nu_*$. Since $m \geq T$, these three errors are bounded by $C_J(T+e)^{-J-2}$ after increasing A_J.

Apply (33) at a sufficiently large fixed multiple of $\log(T+e)$, and use (37). The normalized clipped increments are bounded by ε in (34), and their bracket differs from one by at most D, except with probability $C_J(T+e)^{-J}$. Lemma 5.1, under $\mathbb{Q}_\lambda$, followed by the bounded normal density to absorb the small recentering shift, proves

$$d_K(U_\lambda, N(0,1)) \leq C_J\{\sqrt{D} + \varepsilon|\log\varepsilon| + (T+e)^{-J}\}. \tag{38}$$

The deletion probability enters as an absolute error under the tilted measure, which accounts for the $\log(m+e)$ in L_J.

The random boundary and the likelihood weight. Put $Z_\lambda = (S - x\sqrt{V})/\sigma$ and $R_\psi = \Psi(\lambda) - \lambda^2 V/2$. Completion of the square gives the exact identity

$$\begin{aligned} \mathbb{P}(S > x\sqrt{V}) &= e^{-x^2/2}\mathbb{E}^{\mathbb{Q}_\lambda}[e^{-xZ_\lambda}\mathbf{1}_{\{Z_\lambda>0\}}W_\lambda], \\ W_\lambda &= \exp\left\{R_\psi + \frac{x^2}{2}(\sqrt{t_\lambda}-1)^2\right\}. \end{aligned} \tag{39}$$

In particular, the clock-square term is positive. Also

$$Z_\lambda = U_\lambda + x(t_\lambda - \sqrt{t_\lambda}) + \frac{\sum_k \psi_k'(\lambda) - \lambda V}{\sigma}. \tag{40}$$

Let $h_J = C_J\sqrt{\log(T+e)/T}$, with the constant chosen below, and let $H = \{|t_\lambda - 1| \leq h_J\}$. Constants in D may be chosen so that $h_J \leq D$. Then $\mathbb{Q}_\lambda(H^c) \leq C_J(T+e)^{-J}$ and on H,

$$|Z_\lambda - U_\lambda| \leq C\{xD + \eta x^2/\sqrt{T}\}.$$

It follows from (38) that the Kolmogorov error of Z_λ is at most

$$\kappa_x = C_J\{\sqrt{D} + \varepsilon|\log\varepsilon| + xD + \eta x^2/\sqrt{T} + (T+e)^{-J}\}.$$

The function $z \mapsto e^{-xz}\mathbf{1}_{\{z>0\}}$ has total variation at most two. Stieltjes integration by parts gives

$$\left|\mathbb{E}^{\mathbb{Q}_\lambda}[e^{-xZ_\lambda}\mathbf{1}_{\{Z_\lambda>0\}}] - I(x)\right| \leq 2\kappa_x, \qquad I(x) = e^{x^2/2}\overline{\Phi}(x) \asymp (1+x)^{-1}. \tag{41}$$

This includes $x = 0$.

Integrating the positive clock-square term. For $U = |t_\lambda - 1|$, the all-confidence estimate (33) implies

$$\mathbb{Q}_\lambda(U > v) \leq Ce^{-c_1 T v^2}, \qquad v \geq 0.$$

For v below a constant times $T^{-1/2}$ this follows by enlarging C; above that threshold substitute a constant multiple of Tv^2 for the confidence parameter. For $0 \leq \gamma \leq c_1 T/2$, tail integration yields

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}_\lambda}[e^{\gamma U^2}\mathbf{1}_{\{U>h_J\}}] &\leq Ce^{-(c_1T-\gamma)h_J^2} + 2C\gamma\int_{h_J}^\infty ve^{-(c_1T-\gamma)v^2}\,dv \\ &\leq Ce^{-c_1Th_J^2/2}. \end{aligned}$$

Choose c in (35) small enough that $\gamma = x^2/2$ is permitted. Increasing the fixed constant in h_J gives

$$\mathbb{E}^{\mathbb{Q}_\lambda}[e^{x^2U^2/2}\mathbf{1}_{H^c}] \le C_J(T+e)^{-J}, \qquad 0 \le x \le r. \tag{42}$$

This estimate controls the positive weight on the exceptional event itself; replacing it by a supremum would create an unnecessary factor e^{Cr^2}.

By (37), $|R_\psi| \le CB_r$. On H, therefore, $|\log W_\lambda| \le w = C(B_r + r^2D^2)$. On H^c, use $|\sqrt{t_\lambda} - 1| \le U$ and (42). Since $0 \le e^{-xZ_\lambda}\mathbf{1}_{\{Z_\lambda>0\}} \le 1$, the difference of the weighted and unweighted expectations in (39) and (41) is at most

$$(e^w - 1)(I(x) + 2\kappa_x) + C_J e^{CB_r}(T+e)^{-J}.$$

Divide by $I(x) \ge c/(1+x)$. The quantities w and $(1+x)\kappa_x$ are bounded by CE_r and $CE_r + C_J(1+r)(T+e)^{-J}$, respectively. Here $r^2D^2 \le r^2D$ and $(1+x)\eta x^2/\sqrt{T} \le 2B_r$ because $r \ge 1$. This proves (36).

For a non-strict tail, replace $\mathbf{1}_{\{z>0\}}$ throughout by $\mathbf{1}_{\{z\ge0\}}$. Its total variation is the same; Stieltjes integration, using one-sided distribution functions, gives the same bound because the normal law has no atoms. The assumption $V > 0$ identifies these events with the respective tails of $S/\sqrt{V}$. Applying the argument to $-d_{k+1}$ uses the hypotheses for negative tilts and proves the lower tail assertions. □

Theorem 5.3 (CMD with the exact tilted clock)**.** *Suppose that the hypotheses of Proposition 5.2 hold uniformly along a family with $\eta \downarrow 0$, $T = m\eta \to \infty$, and $r \to \infty$. If*

$$\frac{r^4\log(T+e)}{T} \longrightarrow 0, \qquad \frac{\eta r^3}{\sqrt{T}} \longrightarrow 0, \tag{43}$$

then

$$\sup_{0\le x\le r}\left|\frac{\mathbb{P}(S/\sqrt{V} > x)}{\overline{\Phi}(x)} - 1\right| \longrightarrow 0. \tag{44}$$

The same conclusion holds for all four strict and non-strict upper and lower tails.

Proof. The first condition implies $r = o(T^{1/4})$ and $r^2\sqrt{\log(T+e)/T} \to 0$. In the notation of (34),

$$r^2D \le C_J\left\{r^2\sqrt{\frac{\log(T+e)}{T}} + \frac{\eta r^3}{\sqrt{T}}\right\} \to 0, \qquad r\sqrt{D} = \sqrt{r^2D} \to 0.$$

Since $T = m\eta$ and $\eta \le 1$, we have $m \ge T$ and $\varepsilon = C_J\sqrt{\log(m+e)/m}$. The function $u \mapsto u^{-1/2}\log^{3/2}(u+e)$ decreases for large u. Thus, for all sufficiently large T,

$$(1+r)\varepsilon|\log\varepsilon| \le C_J(1+r)\frac{\log^{3/2}(T+e)}{\sqrt{T}} \longrightarrow 0,$$

where the limit follows from $r = o(T^{1/4})$. These estimates also imply the finite restrictions in (35). Consequently $E_r \to 0$. For any fixed $J \ge 2$, the second term in (36) also vanishes. □

Remark 5.4 (What determines the window)**.** The third-cumulant error is $\eta r^3/\sqrt{T}$. The condition $r^4\log T/T \to 0$ comes from the $\sqrt{D}$ loss in the martingale Kolmogorov estimate and the first-order displacement of the random boundary. An additional error of order η in (33) would require $r^2\eta \to 0$ in this comparison.

5.3 Simultaneous perturbations of numerator and denominator

Lemma 5.5 (Dependent perturbation transfer)**.** *Let $W = U/\sqrt{V}$ and $\widehat{W} = (U+\Delta)/\sqrt{\widehat{V}}$, where $V > 0$ and $\widehat{V} > 0$ almost surely. Suppose all four strict and non-strict upper and lower tails of W have relative normal approximation error at most ϵ_0 on $[0, r+1]$. On an event G with $\mathbb{P}(G^c) \le p$, suppose*

$$\left|\frac{\widehat{V}}{V} - 1\right| \le \beta < \frac{1}{2}, \qquad |\Delta| \le a\sqrt{V}.$$

There is an absolute $c > 0$ such that, if

$$A = (1+r)a + (1+r)^2\beta \le c,$$

then

$$\sup_{0\le x\le r}\left|\frac{\mathbb{P}(\widehat{W} > x)}{\overline{\Phi}(x)} - 1\right| \le C(\epsilon_0 + A)e^{CA} + \frac{p}{\overline{\Phi}(r)}. \tag{45}$$

The same conclusion holds for the other three tail conventions. No independence is required between any of these quantities.

Proof. For $x \ge 0$, pointwise inclusions on G give

$$\begin{aligned}\mathbb{P}(W > x\sqrt{1+\beta} + a) - p &\le \mathbb{P}(\widehat{W} > x)\\ &\le \mathbb{P}(W > x\sqrt{1-\beta} - a) + p.\end{aligned}$$

Both new thresholds differ from x by at most $C(x\beta + a)$ and lie in $[-1, r+1]$ when $A \le c$. The normal hazard function satisfies $\varphi(z)/\overline{\Phi}(z) \le C(1+z_+)$, so integration over this shift gives

$$\left|\log\frac{\overline{\Phi}(x+v)}{\overline{\Phi}(x)}\right| \le C\{(1+x)|v| + v^2\} \le CA.$$

If a threshold z is negative, then $x \le Ca$, and $\overline{\Phi}(x)$ is bounded away from zero. Use $\mathbb{P}(W > z) = 1 - \mathbb{P}(-W \ge -z)$ and the assumed lower-tail approximation to obtain the same $C\epsilon_0$ error after division by $\overline{\Phi}(x)$. Positive thresholds use the upper-tail approximation directly. The inclusions and the normal-tail ratio estimate now prove (45). Replacing the strict inequalities and, if necessary, applying the argument to $-W$ proves the remaining cases. □

6 Localization and polynomial occupation on the original chain

A state-independent upper bound on the noise proxy lets us choose the step-size restriction uniformly in the localization radius. We will take that radius to grow with the range of deviations.

Lemma 6.1 (Uniform-proxy localization and occupation)**.** *Consider an adapted chain in $\mathbb{R}^d$ of the form*

$$X_{k+1} = X_k + \eta b(X_k) + \xi_{k+1}, \qquad X_0 = x_0,$$

where x_0 is deterministic. Suppose, with fixed positive constants κ, A, L, N, that

$$x \cdot b(x) \le -\kappa|x|^2 + A, \qquad |b(x)| \le L(1+|x|), \tag{46}$$

and that the conditional noise is centered and satisfies

$$\mathbb{E}(e^{u\cdot\xi_{k+1}} \mid \mathcal{F}_k) \le \exp\{N\eta|u|^2/2\}, \qquad u \in \mathbb{R}^d. \tag{47}$$

There are $\eta_0, \theta, c, C > 0$, depending only on the displayed constants and d, such that the following holds for $0 < \eta \le \eta_0$, $m \ge 1$, $T = m\eta$, and $R \ge 1 + |x_0|$. Put

$$q(x) = 1 + |x|^2, \qquad \tau_R = \inf\{k \ge 0 : |X_k| > R\}, \qquad I_k = \mathbf{1}_{\{k<\tau_R\}}.$$

Then

$$\mathbb{P}(\tau_R \le m) \le \{e^{\theta|x_0|^2} + CT\}e^{-\theta R^2}. \tag{48}$$

Moreover, for every $t \ge 0$,

$$\mathbb{P}\left(\eta \sum_{k=0}^{m-1} I_k q(X_k) > C\{T + t + q(x_0)\}\right) \le e^{-t}, \tag{49}$$

$$\mathbb{P}\left(\eta \sum_{k=0}^{m-1} I_k q(X_k)^2 > C\{T + R^2 t + q(x_0)^2\}\right) \le e^{-t}. \tag{50}$$

The constants and η_0 are independent of R, m, t.

Proof. Write $y = x + \eta b(x)$ and suppress conditional subscripts when the starting state is x. The drift assumptions give, after reducing η_0,

$$|y|^2 \le (1 - \kappa\eta)|x|^2 + C\eta, \qquad q(y) \le Cq(x). \tag{51}$$

We first record consequences of the uniform noise proxy. Introducing an independent standard Gaussian vector and integrating its moment generating function gives, for $2aN\eta < 1$,

$$\mathbb{E}e^{a|y+\xi|^2} \le (1 - 2aN\eta)^{-d/2} \exp\left\{\frac{a|y|^2}{1 - 2aN\eta}\right\}. \tag{52}$$

Indeed, $e^{a|z|^2} = \mathbb{E}_G e^{\sqrt{2a}\,G\cdot z}$, and (47) may be applied conditionally for each value of G. In particular,

$$\mathbb{E}e^{|\xi|^2/(4N\eta)} \le 2^{d/2}.$$

For each fixed $a \ge 0$ and integer $j \ge 1$, it follows that

$$\mathbb{E}[|\xi|^{2j} e^{a|\xi|^2}] \le C_{a,j}\eta^j \tag{53}$$

once η_0 is sufficiently small for that fixed a.

Quadratic Lyapunov function and exit. Choose a fixed sufficiently small $\theta > 0$ in (52). Combining it with (51) yields the global bound

$$\mathbb{E}_x e^{\theta|X_1|^2} \le e^{\theta|x|^2} \exp\{\eta(C - c|x|^2)\}. \tag{54}$$

The right-hand side is at most $e^{\theta|x|^2} + C\eta$: outside a fixed ball the extra factor is at most one, and inside that ball the exponential function is bounded uniformly. Apply this additive drift inequality to $X_{k\wedge\tau_R}$, including its value at the exit step, to obtain

$$\mathbb{E}e^{\theta|X_{m\wedge\tau_R}|^2} \le e^{\theta|x_0|^2} + CT.$$

On $\{\tau_R \le m\}$ its exponent is greater than θR^2, proving (48).

Multiplying (54) over the stopped steps shows that

$$\exp\left\{\theta|X_{j\wedge\tau_R}|^2 - \theta|x_0|^2 + c\eta\sum_{k<j} I_k|X_k|^2 - C\eta\sum_{k<j} I_k\right\}$$

is a nonnegative supermartingale of initial value one. Its terminal squared norm is nonnegative. Markov's inequality, and $\eta\sum_{k<m} I_k \le T$, therefore give (49).

A convex quartic function with quadratic growth at infinity. Choose a fixed smooth increasing convex function $\phi : [0,\infty) \to [0,\infty)$ such that $\phi(z) = z^2$ for $z \le 9$, ϕ is affine for $z \ge 16$, and ϕ', ϕ'' are bounded. For example, specify a nonnegative smooth second derivative equal to two on $[0,9]$ and zero on $[16,\infty)$, and integrate twice from zero. Define

$$\psi_R(x) = R^2\phi\big(q(x)/R^2\big).$$

This function is nonnegative and convex. On $|x| \le 2R$ it equals $q(x)^2/R^2$, and globally it satisfies

$$0 \preceq D^2\psi_R(x), \qquad \|D^2\psi_R(x)\| \le C, \qquad \|D^2\psi_R(x)\| \le Cq(x)/R^2. \tag{55}$$

These bounds follow by differentiating $R^2\phi(q/R^2)$; the second derivative of ϕ vanishes above 16, and $\phi'(z) \le C\min\{z,1\}$.

For $|x| \le R$ and small fixed η_0, the segment from x to $y = x + \eta b(x)$ lies in $\{|z| \le 2R\}$. Using (46) and Taylor's formula gives

$$\psi_R(y) - \psi_R(x) \le \frac{\eta}{R^2}\{C - cq(x)^2\}. \tag{56}$$

The first-order term is $4\eta q(x)x\cdot b(x)/R^2$ and is bounded by $\eta R^{-2}\{C - c_1 q(x)^2\}$. The second-order term is at most $C\eta^2 q(x)^2/R^2$ by linear growth of b and (55); a fixed reduction of η_0 absorbs it.

Write

$$\psi_R(y+\xi) - \psi_R(y) = \ell + Q, \qquad \ell = \nabla\psi_R(y)\cdot\xi.$$

Convexity and the two Hessian bounds imply

$$0 \le Q \le C|\xi|^2, \qquad Q \le \frac{C}{R^2}\{q(y)|\xi|^2 + |\xi|^4\}. \tag{57}$$

Also $|\nabla\psi_R(y)|^2 \le Cq(x)^3/R^4$. For a small fixed $\alpha > 0$, conditional Cauchy–Schwarz and (47) give

$$\log \mathbb{E}e^{\alpha(\ell+Q)} \le C\alpha^2\eta\frac{q(x)^3}{R^4} + \frac{1}{2}\log \mathbb{E}e^{2\alpha Q}.$$

The inequality $e^v - 1 \le ve^v$ for $v \ge 0$, (57), and (53) show that

$$\log \mathbb{E}e^{2\alpha Q} \le C\alpha\eta\frac{q(y)+\eta}{R^2} \le C\alpha\eta\frac{q(x)}{R^2}.$$

On $|x| \le R$, $q(x)/R^2 \le 2$. Thus the preceding linear-noise term is at most $C\alpha^2\eta q(x)^2/R^2$. Choose α small enough to absorb it in the negative term of α times (56); then absorb $C\alpha q(x)$ in that negative quadratic polynomial. We obtain

$$\mathbb{E}_x e^{\alpha\psi_R(X_1)} \le e^{\alpha\psi_R(x)}\exp\left\{\frac{\eta}{R^2}(C - cq(x)^2)\right\}, \qquad |x| \le R. \tag{58}$$

Every parameter used here is fixed independently of R. The globally quadratic growth of ψ_R is what permits the exit overshoot in this exponential estimate.

As in the quadratic argument,

$$\exp\left\{\alpha\psi_R(X_{j\wedge\tau_R}) - \alpha\psi_R(x_0) + \frac{c\eta}{R^2}\sum_{k<j} I_k q(X_k)^2 - \frac{C\eta}{R^2}\sum_{k<j} I_k\right\}$$

is a nonnegative supermartingale. Drop its nonnegative terminal Lyapunov value and use $R^2\psi_R(x_0) = q(x_0)^2$. Markov's inequality then gives (50). The same step-size restriction therefore works at every $R \geq 1 + |x_0|$. □

Corollary 6.2 (A simultaneous occupation bound)**.** *Under the hypotheses of Lemma 6.1, fix d, x_0, and $J \geq 1$. For $T \geq 1$ and $r \geq 1$ set*

$$s = d + (r+1)^2 + \log(T+e), \qquad R^2 = C_J s.$$

The constant C_J can be chosen so that an event G_J satisfies

$$\mathbb{P}(G_J^c) \leq C_J(T+e)^{-J}e^{-J(r+1)^2}, \tag{59}$$

$$\max_{0\leq k\leq m}|X_k| \leq R, \qquad H_2 := \eta\sum_{k=0}^{m-1} q(X_k) \leq C_J(T+s),$$

$$H_4 := \eta\sum_{k=0}^{m-1} q(X_k)^2 \leq C_J(T+s^2) \quad \text{on } G_J. \tag{60}$$

If $s^2/T \to 0$, then, eventually on G_J,

$$H_2 + H_4 + H_3 \leq C_J T, \qquad H_3 := \eta\sum_{k=0}^{m-1} q(X_k)^{3/2}. \tag{61}$$

Under the same condition $s^2/T \to 0$, if F is uniformly Lipschitz and $|j_\eta(x)| \leq C\eta^2 q(x)^{3/2}$, then on G_J,

$$|F(X_m) - F(x_0)| + \left|\eta\sum_{k=0}^{m-1} j_\eta(X_k)\right| \leq C_J(R + \eta^2 T). \tag{62}$$

Proof. Use Lemma 6.1 with $t = A_J s$ in both occupation estimates and enlarge C_J in $R^2 = C_J s$. Its exit bound is then at most $C_J(T+e)^{-J}e^{-J(r+1)^2}$, as are the two occupation failure probabilities. On the event of no exit, stopped and unstopped sums coincide. This proves (59)–(60). When $s^2/T \to 0$, the first two sums are bounded by $C_J T$, and Cauchy–Schwarz gives $H_3 \leq \sqrt{H_2 H_4}$. Finally the endpoint difference is at most $\mathrm{Lip}(F)(R + |x_0|)$ and the accumulated remainder is at most $C\eta^2 H_3$, proving (62). □

Remark 6.3 (Relation to the CMD window)**.** For fixed d, the condition $r^4\log(T+e)/T \to 0$ implies $s^2/T \to 0$ and $rR/\sqrt{T} \to 0$. The original chain uses this occupation event to bound its weak and variance remainders. Under the tilted law, Proposition 4.3 instead gives (33) for the predictable quadratic variation.

7 The trajectory statistic with bias and variance corrections

We first approximate the tails of the martingale normalized by its predictable quadratic variation. The one-step expansions then replace the martingale and its variance by quantities averaged along the observed trajectory.

Corollary 7.1 (Full-increment martingale CMD). *Under Assumption 2.2, take $F = f_\eta$ or $F = -f_\eta$,*

$$S_m = \sum_{k<m} \{F(X_{k+1}) - P_\eta F(X_k)\}, \qquad V_m = \sum_{k<m} \mathrm{Var}_{X_k} F(X_{k+1}).$$

If $\eta \to 0$, $T = m\eta \to \infty$, $r \to \infty$, and

$$r^4 \log(T+e)/T \to 0, \qquad \eta r^3/\sqrt{T} \to 0,$$

then $S_m/\sqrt{V_m}$ satisfies all four strict and non-strict relative normal-tail approximations uniformly on $[0, r]$. Moreover, the finite bound of Proposition 5.2 applies with model-dependent constants.

Proof. Lemma 4.4 proves the square-exponential and third-cumulant bounds and identifies the actual tilted measure. Lemma 4.2 proves positivity and continuity of the deterministic clock centers, and Proposition 4.3 gives the all-confidence estimate for the original variance under that measure. Theorem 5.3 applies. □

Theorem 7.2 (Corrected trajectory with positive energy). *Under Assumption 2.2 and the joint conditions (8), define*

$$E_m^+ = \bar{e}_m^+, \qquad W_m^+ = \frac{\sqrt{T}\{\bar{h}_m + \eta\overline{\mathcal{A}_2 f_{\eta_m}} - \theta_\eta\}}{\sqrt{E_m^+}}. \tag{63}$$

The quotient may be assigned any fixed value when $E_m^+ = 0$. Then W_m^+ has all four strict and non-strict relative normal-tail approximations uniformly on $[0, r]$.

More precisely, for every fixed $J \geq 2$, set $s = d + (r+1)^2 + \log(T+e)$ and $R^2 = C_J s$. Eventually there is an event $\mathcal{G}_J$ such that

$$\mathbb{P}(\mathcal{G}_J^c) \leq C_J (T+e)^{-J} e^{-J(r+1)^2}, \tag{64}$$

$$V_m/T \geq c_*, \qquad E_m^+ \geq c_*/2,$$

$$|TE_m^+ - V_m| \leq C_J \eta^2 T, \tag{65}$$

$$T\{\bar{h}_m + \eta\overline{\mathcal{A}_2 f_{\eta_m}} - \theta_\eta\} = S_m + Q_m, \qquad |Q_m| \leq C_J(R + \eta^2 T) \quad \text{on } \mathcal{G}_J, \tag{66}$$

where S_m is the full-increment martingale for $F = -f_\eta$, and $c_* > 0$ depends only on the fixed model.

Proof. The total one-step centered noise has the uniform directional proxy $\eta\delta + \eta^2(M^2 + \tau^2) \leq N\eta$. Strong dissipativity implies the radial drift condition of Lemma 6.1, after Young's inequality at the origin. Corollary 6.2 therefore gives an event with the stated failure probability on which

$$\max_{k \leq m} |X_k| \leq R, \qquad H_2 + H_3 + H_4 \leq C_J T.$$

Indeed, (8) implies $s^2/T \to 0$. Use (28) at zero tilt with confidence $A_J s$. The resulting additional failure probability has the same scale, and its tolerance $C\sqrt{s/T}$ tends to zero. Since $\nu_{\eta,0} \geq \nu_*$, this gives $V_m/T \geq c_*$ for a fixed $c_* > 0$.

Proposition 3.3, with $F = f_\eta$, gives on every state

$$|c_\eta(x) - e_\eta^+(x)| \leq C\eta^2(1+|x|^2),$$
$$|P_\eta f_\eta - f_\eta - \eta(h-\theta_\eta) - \eta^2 \mathcal{A}_2 f_\eta| \leq C\eta^3(1+|x|^3).$$

Sum the first bound using $H_2 \leq C_J T$ to prove (65), and hence the lower bound on E_m^+ for sufficiently small η. To check the sign and scale in the numerator, put

$$R_\eta = P_\eta f_\eta - f_\eta - \eta(h-\theta_\eta) - \eta^2 \mathcal{A}_2 f_\eta.$$

The full martingale for $-f_\eta$ gives the exact identity

$$T\{\bar{h}_m + \eta\overline{\mathcal{A}_2 f_\eta}_m - \theta_\eta\} = S_m + f_\eta(X_m) - f_\eta(x_0) - \sum_{k<m} R_\eta(X_k).$$

The endpoint is bounded by $C_J R$, and the last sum by $C\eta^2 H_3 \leq C_J \eta^2 T$. This proves (66).

On $\mathcal{G}_J$, the normalized numerator perturbation is at most $a = C_J(R/\sqrt{T} + \eta^2\sqrt{T})$ and the relative variance perturbation at most $\beta_0 = C_J\eta^2$. For the transfer, replace E_m^+ by 1 on its zero set, so that the denominator is strictly positive everywhere. This replacement changes any fixed zero-energy convention only on $\mathcal{G}_J^c$. Apply Lemma 5.5 to Corollary 7.1 with radius $r+1$. Its asymptotic conditions hold at this radius as well. The transfer costs vanish because

$$rR/\sqrt{T} \to 0, \qquad r\eta^2\sqrt{T} \to 0, \qquad r^2\eta^2 \to 0.$$

Mills' bounds and (64) show that the exceptional probability divided by $\overline{\Phi}(r)$ tends to zero. □

The predictable variance and the remaining weak bias impose the two conditions in (12). Balancing them gives the observation time in (14). Replacing V_m/T by E_m^+ costs a further $O(r^2\eta^2)$ in the relative tail estimate.

8 Certified Poisson gradients from a sampling interface

Under Assumption 2.1, a query at x returns an independent sample $O(x) = b(x) + Z(x)$. The algorithm can evaluate h and the fixed smooth functions below to prescribed absolute accuracy. It uses the given coefficient envelopes and dissipativity constants. A certified evaluation $|\widetilde{h}_0 - h(0)| \leq 1$ supplies the upper bound $|\widetilde{h}_0| + 1$ wherever a budget involves $|h(0)|$.

We condition on the finite array of input states and on earlier data. Each certificate uses fresh auxiliary samples. Sample counts are in the arithmetic model of Section 2, with exact oracle draws and arithmetic on the observations. The magnitude of a raw draw can be unbounded. Before deterministic residual tests, we project coefficient estimates onto known bounded sets. The accuracy guarantees require effective upper enclosures for upper bounds and strictly positive effective lower enclosures for positive lower bounds.

8.1 Local coefficient certificates with fourth moments

For every $D \geq 1$, assume known finite numbers V_D, W_D such that

$$\begin{aligned} \sup_{x \in B_D} \max_i \operatorname{Var}(O_i(x)) &\leq V_D, \\ \sup_{x \in B_D} \max_{i,j} \mathbb{E}\{U_i(x)^2 U_j(x)^2\} &\leq W_D, \qquad U(x) = \frac{O^{(1)}(x) - O^{(2)}(x)}{\sqrt{2}}. \end{aligned} \tag{67}$$

The two calls defining U are independent. For assertions of polynomial oracle complexity, we additionally require known constants C_V, C_W, a_V, a_W with $V_D \leq C_V(1+D)^{a_V}$ and $W_D \leq C_W(1+D)^{a_W}$. Finite effective envelopes without this growth restriction suffice for the termination and confidence results.

Theorem 8.1 (Local drift and covariance certificate)**.** *Fix $D \geq 1$, tolerances $e_b, e_\Sigma > 0$, and $0 < p < 1$. Put*

$$b_D = B_0 + B_1 D, \quad S_D = S_0 + S_1 D, \quad L_{\Sigma,D} = 2 S_D S_1, \quad a \leq \min\left\{1, \frac{e_b}{2(1+B_1)}, \frac{e_\Sigma}{2(1+L_{\Sigma,D})}\right\}.$$

Choose an a-net $(x_i)_{i \leq N}$ of B_D, with $N \leq \{1 + \lceil 2D\sqrt{d}/a \rceil\}^d$. Let B be the smallest positive odd integer for which $2^{-B} \leq p/[N(d+d^2)]$, and set

$$n_b = \left\lceil \max\{1, 64 d V_D / e_b^2\} \right\rceil, \qquad n_\Sigma = \left\lceil \max\{1, 64 d^2 W_D / e_\Sigma^2\} \right\rceil. \tag{68}$$

At each node, take the coordinatewise medians of B independent group means, using n_b observations per drift group and n_Σ independent pairs per covariance group. The covariance observations are $UU^\top$. Clip the drift coordinates to $[-b_D, b_D]$ and the covariance entries to $[-S_D^2, S_D^2]$. Project the resulting drift estimate onto the Euclidean ball of radius b_D, and project the symmetrized covariance estimate onto

$$\mathcal{C}_D = \{M = M^\top : M \succeq 0,\ \|M\|_{\mathrm{HS}} \leq S_D^2\}.$$

Their nearest-node extensions, denoted by $\widehat{b}, \widehat{\Sigma}$, satisfy, with probability at least $1 - p$,

$$\sup_{B_D} |\widehat{b} - b| \leq e_b, \qquad \sup_{B_D} \|\widehat{\Sigma} - \Sigma\|_{\mathrm{HS}} \leq e_\Sigma. \tag{69}$$

On every transcript, $|\widehat{b}| \leq b_D$, $\widehat{\Sigma} \succeq 0$, and $\|\widehat{\Sigma}\|_{\mathrm{HS}} \leq S_D^2$. The number of oracle calls is at most $NB(n_b + 2n_\Sigma)$.

Proof. A rectangular grid of spacing at most $a/\sqrt{d}$ in $[-D, D]^d$, projected onto B_D, gives the stated net. For a coordinate of a drift group mean, Chebyshev's inequality at tolerance $e_b/(2\sqrt{d})$ gives failure probability at most $4dV_D/(n_b e_b^2) \leq 1/16$. Independence of the paired observations cancels the unknown mean exactly: $\mathbb{E}U_iU_j = \Sigma_{ij}$. Moreover, $\operatorname{Var}(U_iU_j) \leq W_D$. A covariance group mean therefore fails at tolerance $e_\Sigma/(2d)$ with probability at most $1/16$. If a median fails, at least $B/2$ of its independent groups fail; a union bound over subsets bounds this probability by $2^B(1/16)^{B/2} = 2^{-B}$. Union over all nodes and scalar coordinates gives simultaneous node errors at most $e_b/2$ and $e_\Sigma/2$ in the displayed vector and matrix norms.

The true coefficients belong to their projection sets, since $\|\Sigma(x)\|_{\mathrm{HS}} \leq \|\sigma_0(x)\|_{\mathrm{HS}}^2 \leq S_D^2$. Coordinate clipping, symmetrization and orthogonal projection each preserve the node error

bounds. Indeed, each operation projects onto a closed convex set containing the corresponding true coefficient. On B_D, the Lipschitz constants of b and Σ are at most B_1 and $2S_DS_1$, respectively. Nearest-node extension adds at most the remaining halves of the tolerances. The all-transcript caps and the sample count follow from the construction. For completeness, the projection of a symmetric matrix M onto $\mathcal{C}_D$ is $M_+ \min\{1, S_D^2/\|M_+\|_{\mathrm{HS}}\}$, with value zero when $M_+ = 0$, where M_+ denotes the positive part. Clipping bounds the inputs to this projection on every transcript. On this bounded set one can approximate the matrix positive part uniformly by polynomial approximation to $s \mapsto \max(s, 0)$ on the known spectral interval. The subsequent radial projection is nonexpansive. Whenever rounded projections are used in a residual test, their prescribed errors are included in e_{num} in (74). □

Two moment assumptions give explicit bounds in (67). A centered coordinate fourth-moment envelope $\sup_{B_D,i} \mathbb{E}|O_i - b_i|^4 \le F_D$ permits $V_D = F_D^{1/2}$, $W_D = 2F_D$: independence and centering give $\mathbb{E}U_i^4 = \frac{1}{2}\mathbb{E}Z_i^4 + \frac{3}{2}(\mathbb{E}Z_i^2)^2 \le 2F_D$, and Cauchy–Schwarz gives the mixed bound. A directional sub-Gaussian proxy k_D^2 permits $V_D = k_D^2$, $W_D = 16k_D^4$, by integrating the tail bound for the equally sub-Gaussian difference U. In Assumption 2.2, one can take $k_D^2 = M^2 + \tau^2$, independent of D, and $S_0 = \sqrt{M^2 + d\tau^2}$, $S_j = 0$ for $j \ge 1$. The sampling interface need not expose the bounded and Gaussian components separately, or a covariance square root.

8.2 A global certificate from a local residual

The local numerical certificate needs a quantitative control on the modified diffusion outside the region where coefficients are sampled. The following moment is a property of that Brownian diffusion. It does not impose a sixth-moment assumption on the sampling innovation.

Lemma 8.2 (An explicit uniform diffusion sixth moment)**.** *Set*

$$k_* = K_1/2, \qquad A_* = K_2 + B_0^2/(2K_1), \qquad s^2 = 2(S_0^2 + S_1^2).$$

Decrease the fixed upper step bound η_0, if necessary, so that $\eta_0 s^2 \le k_*/5$. Define

$$D_0 = 2k_* + 2A_* + \delta d, \qquad D_3 = 3(D_0 + 4\delta), \qquad \lambda_6 = 3k_*/2,$$
$$C_6 = 4D_3^3/(9k_*^2), \qquad M_6(R) = (1 + R^2)^3 + C_6/\lambda_6.$$

For the diffusion with generator L_η,

$$\sup_{0<\eta\le\eta_0} \sup_{t\ge 0} \sup_{|x|\le R} \mathbb{E}_x(1 + |X_t|^2)^3 \le M_6(R). \tag{70}$$

The step restriction involving s is vacuous when $s = 0$.

Proof. Weak dissipativity with $y = 0$, followed by Young's inequality, gives $x \cdot b(x) \le -k_*|x|^2 + A_*$, while $\|\sigma_0(x)\|_{\mathrm{HS}}^2 \le s^2(1 + |x|^2)$. Write $q = 1 + |x|^2$ and $A_\eta = \delta I + \eta\Sigma$. Then

$$L_\eta q^p = pq^{p-1}(2x \cdot b + \operatorname{tr} A_\eta) + 2p(p-1)q^{p-2}x^\top A_\eta x.$$

The trace is at most $\delta d + \eta s^2 q$, and $x^\top A_\eta x \le \delta q + \eta s^2 q^2$. At $p = 3$,

$$L_\eta q^3 \le (-6k_* + 15\eta s^2)q^3 + D_3 q^2 \le -3k_* q^3 + D_3 q^2 \le -\lambda_6 q^3 + C_6.$$

Indeed, for $q \ge 2D_3/(3k_*)$ the quadratic term is at most $(3k_*/2)q^3$, and below this threshold it is at most C_6. Stopped Itô's formula and localization, justified by the global Lipschitz coefficients and their linear growth, followed by Gronwall's inequality give $\mathbb{E}_x q(X_t)^3 \le q(x)^3 + C_6/\lambda_6$. □

For a candidate $F \in C^3(\mathbb{R}^d)$, suppose known global bounds

$$\|\nabla F\|_\infty \le \Gamma_1, \qquad \|D^2 F\|_{\mathrm{HS},\infty} \le \Gamma_2, \qquad \sup_{x,|v|=1} \|D^3 F(x)[v]\|_{\mathrm{HS}} \le \Gamma_3. \tag{71}$$

For a scalar c, write $R_F = L_\eta F - h - c$. Known envelopes for its growth and local Lipschitz constant are

$$\begin{aligned} M_F &= (B_0 + B_1)\Gamma_1 + \tfrac{1}{2}(\delta\sqrt{d} + \eta_0 s^2)\Gamma_2 + |h(0)| + H_1 + |c|, \\ \Lambda_{F,D} &= B_1\Gamma_1 + (B_0 + B_1 D)\Gamma_2 + \tfrac{\delta\sqrt{d}}{2}\Gamma_3 + \eta(S_0 + S_1 D)S_1\Gamma_2 + \tfrac{\eta}{2}(S_0 + S_1 D)^2\Gamma_3 + H_1. \end{aligned} \tag{72}$$

In particular, $|R_F(x)| \le M_F(1 + |x|^2)$ and $\mathrm{Lip}(R_F|_{B_D}) \le \Lambda_{F,D}$.

Theorem 8.3 (Residual-to-gradient certificate)**.** *Let $D \ge R \ge 1$, $S > 0$, and suppose $\sup_{B_D} |R_F| \le e$. Define*

$$a_2 = 2B_1 + \eta_0 S_1^2, \qquad K_S = 2\delta^{-1/2}\sqrt{S}\exp(a_2 S/2).$$

With the explicit contraction constants C_c, ρ of Lemma A.1,

$$\sup_{B_R} |\nabla F - \nabla f_\eta| \le C_c e^{-\rho S}(\Gamma_1 + H_1/\rho) + K_S \left\{ e + \frac{M_F\sqrt{M_6(R)}}{\sqrt{1 + D^2}} \right\}. \tag{73}$$

On the coefficient-success event of Theorem 8.1, a valid residual bound is

$$\begin{aligned} e = \max_i \Big| \widehat{b}(x_i) \cdot \nabla F(x_i) + \tfrac{1}{2}(\delta I + \eta\widehat{\Sigma}(x_i)) : D^2 F(x_i) - h(x_i) - c \Big| \\ + \Gamma_1 e_b + \tfrac{\eta}{2}\Gamma_2 e_\Sigma + \Lambda_{F,D} a + e_{\mathrm{num}}, \end{aligned} \tag{74}$$

where the displayed maximum is computed with total certified deterministic error at most e_{num}.

Proof. Write $P_t = \mathcal{P}_t^\eta$ within this proof. Realize the modified diffusion as $dX_t = b(X_t)dt + \sqrt{\delta}\, dB_t + \sqrt{\eta}\,\sigma_0(X_t)dW_t$. For a unit vector v, its first variation $J_t v$ satisfies $\mathbb{E}|J_t v|^2 \le e^{a_2 t}$. The additive Brownian part gives the gradient representation

$$DP_t R_F(x)[v] = \mathbb{E}_x\{R_F(X_t)\mathcal{W}_t\}, \qquad \mathcal{W}_t = \frac{1}{t\sqrt{\delta}}\int_0^t \langle J_u v, dB_u\rangle, \quad \mathbb{E}\mathcal{W}_t^2 \le \frac{e^{a_2 t}}{\delta t}. \tag{75}$$

Here is the control argument for the identity. Add to the Brownian path an adapted infinitesimal control $J_u v/(t\sqrt{\delta})\, du$. The induced first variation of the solution at time $u \le t$ is $(u/t)J_u v$, as follows directly by substitution in its linear variation equation. At time t this equals the variation caused by the initial direction v. Brownian integration by parts gives the stochastic integral in (75); Itô isometry gives its stated bound. The calculation first applies to bounded controls and compactly supported smooth test functions. Localization of the control, followed by approximation of R_F, is justified by Cauchy–Schwarz and the first-variation bound. More explicitly, for a cutoff $\chi_N = 1$ on B_N, with values in $[0, 1]$,

$$\sup_{|x|\le R} \mathbb{E}_x|R_F(X_t)(1 - \chi_N(X_t))|^2 \le \frac{M_F^2 M_6(R)}{1 + N^2}.$$

Thus the truncated semigroup values and their BEL derivatives converge locally uniformly, identifying the limit derivative in (75) without a growth assumption on DR_F.

Outside B_D, $|R_F(y)|^2 \le M_F^2(1+|y|^2)^3/(1+D^2)$. Consequently, for $|x| \le R$,

$$\{\mathbb{E}_x|R_F(X_t)|^2\}^{1/2} \le e + M_F\sqrt{M_6(R)}/\sqrt{1+D^2}.$$

Integrating (75) from 0 to S, and using $\int_0^S t^{-1/2}e^{a_2 t/2}dt \le 2\sqrt{S}e^{a_2 S/2}$, gives the second term of (73). The semigroup identities for F and the centered Poisson solution give, with constants disappearing after differentiation,

$$\nabla F - \nabla f_\eta = \nabla P_S F - \int_0^S \nabla P_t R_F\, dt + \int_S^\infty \nabla P_t h\, dt.$$

Lemma A.1 bounds the two remaining terms by $C_c e^{-\rho S}(\Gamma_1 + H_1/\rho)$. This proves the posterior bound. Finally, coefficient errors contribute $\Gamma_1 e_b + \eta\Gamma_2 e_\Sigma/2$, interpolation contributes $\Lambda_{F,D}a$, and the certified arithmetic enclosure contributes e_{num}, proving (74). □

For a target ϵ, sufficient choices for the three terms of (73) are

$$\begin{aligned} S &= \max\left\{1, \rho^{-1}\log_+ \frac{3C_c(\Gamma_1 + H_1/\rho)}{\epsilon}\right\}, \\ D &\ge \max\left\{R, 1, \frac{3K_S M_F\sqrt{M_6(R)}}{\epsilon}\right\}, \qquad e \le \frac{\epsilon}{3K_S}. \end{aligned} \tag{76}$$

These choices use the candidate, its scalar center and its derivative bounds. The scalar center will range over a finite grid containing an approximation to $-\theta_\eta$.

8.3 An effective finite family and a terminating constructor

Lemma 8.4 (Finite smooth approximation with a sound derivative test)**.** *Let a function F^*, supported in B_{3D}, have known bounds $\|F^*\|_\infty \le M_0$ and $\|D^j F^*\|_\infty \le K_j'$, $1 \le j \le 4$. For every $0 < t \le 1/4$, one can construct a finite family of known smooth compactly supported functions containing an F whose every coordinate derivative of orders zero through three differs from that of F^* by at most t, uniformly on $\mathbb{R}^d$. There is a finite sound test for the coordinate bounds $\|\partial_{i_1}\cdots\partial_{i_j}F\|_\infty \le K_j' + 1$, $1 \le j \le 3$, which always accepts this approximant.*

Proof. Fix a computable nonnegative smooth mollifier ϕ, supported in the unit ball and of integral one. Let M_j^ϕ be known bounds on the L^1 norms of its coordinate derivatives of order $j \le 4$, and put $M^\phi = \max_{0\le j\le 3}(1, M_j^\phi)$. Choose positive dyadic scales

$$\tau \le \min\{1, t/[2(1+\max_j K_j')]\}, \quad h_0 \le \frac{t\tau^3}{4M^\phi(1+K_1')\sqrt{d}}, \quad \nu \le \frac{t\tau^3}{4M^\phi}.$$

Partition a cube containing B_{3D+1} into finitely many cells of side at most h_0. Enumerate all arrays of cell values in $\nu\mathbb{Z} \cap [-M_0 - 1, M_0 + 1]$, extend each array by zero outside the cube, and include its convolution with ϕ_τ in the family. Its cardinality is bounded by the number of permitted values raised to the cell count.

Choosing the nearest lattice value to F^* at each cell center gives a step function whose global uniform error is at most $K_1'\sqrt{d}\,h_0/2 + \nu/2 \le t\tau^3/(4M^\phi)$. The zero extension causes no boundary error because F^* vanishes outside B_{3D}. Differentiating the convolution gives error at most $t/4$

through coordinate order three. The separate mollification error is at most $\tau K'_{j+1} \le t/2$ at order $j \le 3$.

For every array, each coordinate derivative of order $j \le 4$ of its convolution has the known bound $(M_0+1)M_j^{\phi}\tau^{-j}$, and has support in a known finite cube. Evaluate derivatives of orders one through three on a finite net, with certified evaluation error at most $1/8$ and between-node variation at most $1/8$. For the latter, use $\sqrt{d}$ times the bound for the next coordinate derivative as a Lipschitz constant. Accept only when the resulting upper enclosures are at most $K'_j + 1$. This test is sound and accepts the constructed approximant, whose actual bounds are at most $K'_j + t \le K'_j + 1/4$. The finitely many cell integrals and their derivatives are computable by finite quadrature using the fixed mollifier's known derivative moduli. All comparisons use fixed-precision rational enclosures. We accept only when an upper endpoint is below the threshold, using a rational upper enclosure for each K'_j from the outset. The strict margin for the approximant guarantees acceptance without a precision search. □

Theorem 8.5 (Finite construction with unknown coefficients)**.** *Assume* (67)*, the known analytic hypotheses of Assumption 2.1, and the reduced upper step bound of Lemma 8.2. For each* $0 < \eta \le \eta_0$*, finite array* $(x_k)_{k<m}$*,* $0 < \epsilon < 1$*, and* $0 < p < 1$*, a finite algorithm returns vectors* v_k *and a success flag such that*

$$\mathbb{P}_{\mathrm{aux}}\left\{\max_{k<m}|v_k - \nabla f_\eta(x_k)| > \epsilon \text{ or an unsuccessful flag} \mid (x_k)_{k<m}\right\} \le p. \tag{77}$$

Every loop has a finite prescribed bound on every transcript. With the polynomial moment envelopes following (67)*, the oracle-call count is polynomial in* $1+\max_k|x_k|$*,* ϵ^{-1}*, and* $\log(1/p)$*, for fixed dimension and priors, uniformly in* $0 < \eta \le \eta_0$*. The deterministic enumeration is finite and can be extremely large; no polynomial running-time assertion is made for it.*

Proof. Set $R = \max(1, \max_k|x_k|)$ and reserve half the target tolerance for evaluating the final gradient vectors: write $\xi = \epsilon/2$. Lemma A.2 supplies known operator bounds G_j for $D^j f_\eta$. Lemma A.4 supplies constants K'_j, independent of $D \ge 1$, for $F^* = \chi_D f_\eta$, $1 \le j \le 4$. Anchoring $f_\eta(0) = 0$ gives $\|F^*\|_\infty \le M_0 = 3DG_1$. The known invariant second-moment bound gives

$$|\theta_\eta| \le \Theta := |h(0)| + H_1\sqrt{M_*}.$$

These are prior upper bounds, not evaluations of unknown solution values or of the invariant mean.

For every candidate passing the derivative test, use the common caps

$$\Gamma_1 = \sqrt{d}(K'_1+1), \qquad \Gamma_2 = d(K'_2+1), \qquad \Gamma_3 = d^{3/2}(K'_3+1), \qquad |c| \le \Theta + 1.$$

For the third cap, each of the d^2 entries of $D^3F[v]$ is at most $\sqrt{d}(K'_3+1)$ for a unit vector v, by Cauchy–Schwarz. This gives the stated Hilbert–Schmidt bound. In (72), replace $|c|$ by $\Theta+1$ to obtain a common constant M. Choose S, D by (76) with target ξ, and put $e_* = \xi/(3K_S)$. These choices precede the candidate family; the Γ_j and M do not depend on D. Choose

$$e_b = \frac{e_*}{16(1+\Gamma_1)}, \qquad e_\Sigma = \frac{e_*}{8\eta(1+\Gamma_2)}, \qquad a \le \frac{e_*}{8(1+\Lambda_{F,D})},$$

and impose also the mesh constraints of Theorem 8.1. The common $\Lambda_{F,D}$ uses the displayed caps. Run that theorem once with budget p. Then $\Gamma_1 e_b + \eta\Gamma_2 e_\Sigma/2 \le e_*/8$ and $\Lambda_{F,D}a \le e_*/8$. Its

single success event controls the coefficients uniformly and validates every subsequently selected candidate; it requires no union bound over the candidate family.

Put

$$C_D = \sqrt{d}(B_0 + B_1 D) + \frac{d}{2}\{\delta\sqrt{d} + \eta_0(S_0 + S_1 D)^2\}.$$

Construct the family of Lemma 8.4 with $t \le \min\{1/4, e_*/[16(1 + C_D)]\}$, and a finite scalar net in $[-\Theta - 1, \Theta + 1]$ of mesh at most $e_*/8$. For each candidate passing the derivative test and each scalar center, evaluate (74), assigning total numerical error $e_*/16$. Accept a pair if its certified residual bound is at most e_*. Return its gradients evaluated within $\epsilon/2$ at the input states, together with the success flag. If the finite list is exhausted, return zero vectors and an unsuccessful flag.

On the coefficient-success event, the family contains a candidate approximating $\chi_D f_\eta$ through order three within t and passing the derivative test. Since $\chi_D = 1$ on B_D, $\sup_{B_D} |L_\eta F - h + \theta_\eta| \le C_D t \le e_*/16$. A center within $e_*/16$ of $-\theta_\eta$ gives true residual at most $e_*/8$. The estimated node residual differs by at most $e_*/8$; adding the coefficient, spatial, and numerical allowances gives a certificate bounded by $e_*/8 + e_*/8 + e_*/8 + e_*/8 + 2e_*/16 < e_*$, where the last term allows both numerical overestimation and its explicit enclosure allowance. Hence a pair is accepted. Every accepted pair has gradient error at most ξ throughout B_R by Theorem 8.3; the final evaluation allowance proves (77).

All deterministic lists and quadrature accuracies are prescribed. Before the derivative test, family-wide convolution derivative bounds control its evaluations; after that test the common caps apply. The coefficient estimates are norm-capped on every transcript. Thus even an incorrect sampling event cannot cause an unbounded precision search or a nonterminating validation loop. Certified finite-precision implementations may use the strict slack above, enlarging caps by a known rounding allowance and reducing the corresponding error budgets.

Finally, $S = O(1 + \log \epsilon^{-1})$, K_S is bounded by a fixed power of ϵ^{-1}, and one can choose $D \le \mathrm{poly}(1 + R, \epsilon^{-1})$. The common $\Lambda_{F,D}$ grows at most quadratically in D. Choose the least integer radius above the required lower bound and a dyadic mesh width within a factor two of the largest permitted width. The tolerance and mesh reciprocals then have polynomial upper bounds. In particular, $e_\Sigma^{-1} = 8\eta(1 + \Gamma_2)/e_*$ is bounded uniformly for $0 < \eta \le \eta_0$. Substitution into $NB(n_b + 2n_\Sigma)$ proves the stated call bound. Its constants are not uniform in dimension or as δ tends to zero. □

9 Computable bias correction and positive energy

To compute the first weak-bias correction and the positive energy, we estimate higher derivatives of the Poisson solution by finite differences of its certified gradient values. This uses the analytic hypotheses and fourth-moment bounds of Section 8. The residual certificate controls the gradient uniformly on each stencil.

9.1 The first correction from gradient values

Let G_j be the known operator-norm envelopes for $D^j f_\eta$ from Lemma A.2. For a finite input array in B_R, $R \ge 1$, put

$$b_R = B_0 + B_1 R, \qquad C_A = \frac{d}{2}(b_R + 1)^2 + \frac{\delta}{2} d^{3/2}(b_R + 1) + \frac{\delta^2}{8} d^2, \qquad D_A = G_2(b_R + 1) + \frac{\delta d}{2} G_3.$$

We compute the correction

$$a_\eta(x) = \mathcal{A}_2 f_\eta(x) = \frac{1}{2} D^2 f_\eta(x)[b(x), b(x)] + \frac{\delta}{2} b(x) \cdot \nabla \Delta f_\eta(x) + \frac{\delta^2}{8} \Delta^2 f_\eta(x). \tag{78}$$

Corollary 9.1 (Finite oracle construction of the first correction)**.** *For every $0 < e_a < 1$ and $0 < p_a < 1$, a finite algorithm returns numbers $\widehat{a}_k$ and a success flag with*

$$\mathbb{P}_{\text{aux}} \left\{ \max_{k<m} |\widehat{a}_k - a_\eta(x_k)| > e_a \text{ or an unsuccessful flag} \mid (x_k)_{k<m} \right\} \le p_a. \tag{79}$$

Under the polynomial moment envelopes, its oracle-call count is polynomial in the input radius, in e_a^{-1}, and in $\log(1/p_a)$, for fixed dimension and priors and uniformly over the fixed small-step interval.

Proof. Choose

$$\vartheta = \frac{e_a}{2(1 + C_A)}, \qquad u = \min\left\{1, \frac{\vartheta}{12(1 + \max(G_3, G_4, G_5))}\right\}, \qquad e_g = \frac{\vartheta u^3}{32}.$$

For $g = \nabla f_\eta$, form the forward difference

$$\mathcal{D}_{\mathbf{i},u} g_j(x) = u^{-q} \sum_{\varepsilon \in \{0,1\}^q} (-1)^{q-|\varepsilon|} g_j\left(x + u \sum_{\ell=1}^{q} \varepsilon_\ell e_{i_\ell}\right), \qquad q = 1, 2, 3.$$

Use component g_i and direction lists (j), (j, j), and (i, j, j) to estimate respectively $D^2 f_\eta[e_i, e_j]$, $D^3 f_\eta[e_i, e_j, e_j]$, and $D^4 f_\eta[e_i, e_i, e_j, e_j]$. There are at most $14md^2$ gradient-value occurrences, all in B_{R+3}. Apply Theorem 8.5 once to their joint array, with Euclidean accuracy e_g and budget $p_a/2$. Its simultaneous guarantee avoids a separate failure allocation for each stencil.

Independently obtain simultaneous drift estimates at the input states with error $e_b \le \min\{1, e_a/[4(1 + D_A)]\}$ and failure budget $p_a/2$, using the mean part of Theorem 8.1. Use its norm-capped estimates. On an unsuccessful construction flag, return zero correction values and an unsuccessful flag. Otherwise substitute the drift estimates and the derivative stencils into (78).

The integral formula for an order-q forward difference gives coordinate derivative bias at most quG_{q+2}. The gradient evaluation errors contribute at most $2^q e_g/u^q$. By construction, each of these terms is at most $\vartheta/4$; hence each required coordinate derivative has error at most $\vartheta/2$. For drift vectors of norm at most $b_R + 1$, the sum of the absolute coordinate coefficients in (78) is at most C_A. The derivative errors therefore contribute at most $C_A\vartheta/2 \le e_a/4$. Using the exact derivatives but changing the drift contributes at most

$$\frac{G_2}{2} e_b(2b_R + e_b) + \frac{\delta d G_3}{2} e_b \le D_A e_b \le e_a/4.$$

Allocate another $e_a/4$ to the finitely many deterministic stencil and scalar operations. The strict total error bound is below e_a. A union bound proves (79). The inverse accuracies are polynomial in $1 + R, e_a^{-1}$, giving the sample-count claim by the two constructors used above. □

9.2 Stability of the positive energy

Write

$$g = \nabla f_\eta, \qquad H = D^2 f_\eta, \qquad t = \nabla \Delta f_\eta, \qquad v = g + \eta(Hb + \delta t/2).$$

The completed energy used in the statistic is

$$e_\eta^+(x) = \delta |v(x)|^2 + \eta g(x)^\top \Sigma(x) g(x) + \frac{\eta \delta^2}{2} \|H(x)\|_{\mathrm{HS}}^2. \tag{80}$$

For $x \in B_R$, define the known caps

$$G = G_1, \qquad H_* = \sqrt{d}\, G_2, \qquad t_* = d G_3,$$
$$b_R = B_0 + B_1 R, \qquad S_R = S_0 + S_1 R, \qquad V_R = G + \eta(H_* b_R + \delta t_*/2).$$

They bound $|g|, \|H\|_{\mathrm{HS}}, |t|, |b|, \|\sigma_0\|_{\mathrm{HS}}$, and $|v|$, respectively.

Lemma 9.2 (Scalar positive-energy stability)**.** *Project estimates $\widehat{g}, \widehat{H}, \widehat{t}, \widehat{b}$ onto the displayed norm balls, making $\widehat{H}$ symmetric, and project $\widehat{\Sigma}$ onto the positive-semidefinite Frobenius ball of radius S_R^2. If the resulting errors in the corresponding norms are at most $e_g, e_H, e_t, e_b, e_\Sigma$, then substitution in (80) gives $\widetilde{e}_\eta^+ \geq 0$ and*

$$\begin{aligned} |\widetilde{e}_\eta^+ - e_\eta^+| \leq\ & 2\delta V_R \{e_g + \eta(H_* e_b + b_R e_H + \delta e_t/2)\} \\ & + \eta\{2 S_R^2 G e_g + G^2 e_\Sigma\} + \eta \delta^2 H_* e_H. \end{aligned} \tag{81}$$

The projections do not increase the errors. In particular, if every error is at most ω, the right side is bounded by $C_{\mathrm{en}}(R)\omega$, where

$$C_{\mathrm{en}}(R) = 2\delta V_R\{1 + \eta(H_* + b_R + \delta/2)\} + \eta(2 S_R^2 G + G^2) + \eta \delta^2 H_*. \tag{82}$$

Proof. The bounds $\|H\|_{\mathrm{HS}} \leq \sqrt{d}\|H\|_{\mathrm{op}}$ and $t = \sum_j D^3 f_\eta[\,\cdot\,, e_j, e_j]$ place the true objects inside the stated convex sets. Projection is therefore nonexpansive relative to them. The caps imply $|v|, |\widehat{v}| \leq V_R$, and

$$|\widehat{v} - v| \leq e_g + \eta(H_* e_b + b_R e_H + \delta e_t/2),$$

by writing $\widehat{H}\widehat{b} - Hb = \widehat{H}(\widehat{b} - b) + (\widehat{H} - H)b$. The first term of the energy changes by at most $2\delta V_R |\widehat{v} - v|$. For the covariance term, insert and subtract $g^\top \widehat{\Sigma} g$; its absolute error before the factor η is at most $2 S_R^2 G e_g + G^2 e_\Sigma$. The squared Hessian norm changes by at most $2 H_* e_H$. These three estimates prove (81), and the positive-semidefinite covariance estimate proves nonnegativity. □

9.3 Absolute computation and a relative certificate

Theorem 9.3 (Finite oracle computation of positive energy)**.** *For a finite array $(x_k)_{k<m}$, put $R = \max(1, \max_k |x_k|)$. Given $0 < \gamma < 1$ and $0 < p_E < 1$, a finite algorithm using only oracle samples, evaluations of h, and the known priors returns $\widehat{E}_m \geq 0$ and a success flag. Writing $E_m^+ = m^{-1} \sum_{k<m} e_\eta^+(x_k)$, it satisfies*

$$\mathbb{P}_{\mathrm{aux}} \left\{ |\widehat{E}_m - E_m^+| > \gamma \text{ or an unsuccessful flag} \mid (x_k)_{k<m} \right\} \leq p_E. \tag{83}$$

Every search and sampling loop is finite on every transcript. Under the polynomial fourth-moment envelopes, its oracle-call count is polynomial in $1+R, \gamma^{-1}, \log(1/p_E)$, for fixed dimension and priors, uniformly over $0 < \eta \leq \eta_0$.

Proof. Choose

$$\omega = \frac{\gamma}{2(1 + C_{\mathrm{en}}(R))}, \qquad u = \min\left\{1, \frac{\omega}{8d(1 + \max(G_3, G_4))}\right\}, \qquad e_0 = \frac{\omega u^2}{16d}. \tag{84}$$

Apply Theorem 8.5 once to the joint array

$$x_k, \quad x_k + ue_j, \quad x_k + 2ue_j, \qquad k < m,\ 1 \le j \le d,$$

with Euclidean gradient error e_0 and failure budget $p_E/2$. There are at most $(1+2d)m$ gradient vectors, all in B_{R+2}. If this construction flags failure, return zero and an unsuccessful flag. Otherwise form raw estimates

$$\widetilde{H}(x)e_j = \frac{\widehat{g}(x + ue_j) - \widehat{g}(x)}{u}, \qquad \widetilde{t}(x) = \sum_{j=1}^{d} \frac{\widehat{g}(x + 2ue_j) - 2\widehat{g}(x + ue_j) + \widehat{g}(x)}{u^2}.$$

On the simultaneous gradient-success event, the vector integral formulas for first and second forward differences yield

$$\|\widetilde{H} - H\|_{\mathrm{HS}} \le \sqrt{d}\,(uG_3 + 2e_0/u) \le \omega/4,$$
$$|\widetilde{t} - t| \le d(2uG_4 + 4e_0/u^2) \le \omega/2.$$

The original gradient error is at most $e_0 \le \omega$. Symmetrize $\widetilde{H}$ and apply the three derivative norm caps. The remaining slack accommodates certified stencil and projection evaluation errors, leaving each final derivative error at most ω. Independently use Theorem 8.1 on B_R with $e_b = e_\Sigma = \omega$ and budget $p_E/2$; its estimates already have the required caps.

On the intersection of the two success events, all five errors are at most ω. Lemma 9.2 bounds each scalar energy error by $C_{\mathrm{en}}(R)\omega \le \gamma/2$. Compute their mean, reserving $\gamma/2$ for the remaining deterministic scalar evaluations and averaging, and project its numerical value onto $[0, \infty)$. The projection cannot increase distance to the true nonnegative mean. A union bound proves (83).

The caps make all deterministic error budgets effective on every transcript. Moreover, $C_{\mathrm{en}}(R)$ is polynomial in R, with coefficients bounded on the fixed small-step interval. The reciprocals of ω, u, e_0 are therefore polynomial in $1 + R, \gamma^{-1}$, and the call-count assertion follows from the preceding constructors. □

Lemma 9.4 (Harmonic relative certificate)**.** *If $0 < L \le U$, the minimax relative-error center of the interval $[L, U]$ is $H = 2LU/(L + U)$, and*

$$\sup_{E \in [L,U]} |H/E - 1| = \frac{U - L}{U + L}.$$

Consequently, for $0 < \beta < 1/2$, run Theorem 9.3 at accuracy $\gamma = \beta^2$, and set $L_m = (\widehat{E}_m - \gamma)_+$, $U_m = \widehat{E}_m + \gamma$. If its flag is successful, $L_m > 0$, and $(U_m - L_m)/(U_m + L_m) \le \beta$, return $\widehat{E}_m^H = 2L_mU_m/(L_m + U_m)$ with an acceptance flag. Otherwise return $\widehat{E}_m^H = 1$ with a rejection flag. The output is always strictly positive, and

$$\mathbb{P}_{\mathrm{aux}}\left\{|\widehat{E}_m^H/E_m^+ - 1| > \beta \text{ or rejection} \mid (x_k)_{k<m}\right\} \le p_E + \mathbf{1}_{\{E_m^+ < 2\beta\}}. \tag{85}$$

At zero true energy the relative error is interpreted as infinite. No positive lower bound on the invariant energy is an algorithmic input.

Proof. For any positive proposed center, the largest relative error over the interval occurs at an endpoint. Equating the two endpoint errors gives the harmonic center and the displayed value; moving the center in either direction increases one of them. On the absolute-energy success event, $E_m^+ \in [L_m, U_m]$. If $E_m^+ \geq 2\beta$, then $\widehat{E}_m \geq 2\beta - \beta^2 > \gamma$, and

$$\frac{U_m - L_m}{U_m + L_m} = \frac{\gamma}{\widehat{E}_m} \leq \frac{\beta}{2-\beta} < \beta.$$

Hence the certificate accepts and its relative error is at most β. The complement has conditional probability at most p_E, proving (85). □

The two procedures give a joint conditional failure bound of $p_a + p_E + \mathbf{1}_{\{E_m^+ < 2\beta\}}$ by the union bound. We next control the small-energy event along the main trajectory.

9.4 Completion of the computed moderate-deviation theorem

Proof of Theorem 2.3. For a realized finite trajectory, run Corollary 9.1 with inputs e_a, p_a. Run Theorem 9.3 with absolute tolerance $\gamma = \beta^2$ and budget p_E, then apply Lemma 9.4. In the main theorem, $\widehat{E}_m$ denotes this lemma's strictly positive harmonic output $\widehat{E}_m^H$. Accept precisely when both numerical flags are successful. A union bound combines the two conditional failure probabilities. The directional sub-Gaussian proxy $M^2+\tau^2$ supplies the fourth-moment envelopes in (67).

On $\mathcal{G}_J$ of Theorem 7.2, $E_m^+ \geq c_*/2$. Since $\beta \to 0$, the small-energy exception in (85) is absent eventually. Consequently, outside a set of probability at most

$$p_E + p_a + C_J(T+e)^{-J} e^{-J(r+1)^2}, \tag{86}$$

both flags are successful and

$$\left| \frac{\widehat{E}_m}{E_m^+} - 1 \right| \leq \beta, \qquad \left| \widehat{H}_m - \{\bar{h}_m + \eta \overline{\mathcal{A}_2 f_{\eta m}}\} \right| \leq 2\eta e_a.$$

The last bound includes the prescribed evaluation error of $\widehat{\bar{h}}_m$. To apply Lemma 5.5, set the ideal denominator squared equal to E_m^+ when $E_m^+ > 0$ and to 1 otherwise. Use the same ideal numerator on every path. This modification agrees with (63) on $\mathcal{G}_J$. Any discrepancy on its complement has probability $o(\overline{\Phi}(r+1))$ by (64). The modified statistic therefore inherits the ideal tail bounds on $[0, r+1]$. On the joint success event the normalized numerator error is at most $C\eta\sqrt{T}\, e_a$, and the relative squared-denominator error is at most β. By (10) the perturbation parameter in Lemma 5.5 tends to zero. Applying that lemma proves (11). The exceptional probability in (86) is $o(\overline{\Phi}(r))$. The same bound controls the probability of an unsuccessful flag.

Every numerical call has a deterministic finite budget once the finite input array and the tolerances are given. The constructor bounds in Sections 8 and 9 give the asserted polynomial oracle count for the joint-array construction. Candidate evaluation, enumeration and arithmetic are separately finite. The target θ_η appears only in the theoretical studentization, and is not used in either numerical output. □

10 A Gaussian hierarchy and paired simulation

Suppose $Z = \sigma G'$ with $G' \sim N(0, I_d)$, and write $\vartheta_\eta = \delta + \eta\sigma^2$. We retain strong dissipativity with constant κ and set $L = B_1$. To compute the normalizer and the correction in this section we take σ as known and evaluate b and h along auxiliary Euler paths.

10.1 Positive conditional Hermite energies

At a fixed state, the conditional variance is the sum of the squared Hermite coefficients of positive degree. Truncating this sum at degree p gives a nonnegative approximation at every state. We use the Gaussian chaos and derivative identities of [Hai26, Sections 5.1–5.2] with the following normalization.

Proposition 10.1 (Conditional Hermite energy)**.** *Suppose the conditional next step has the form*

$$Y = \Psi_\eta(x) + \sqrt{\eta\vartheta_\eta}\, G, \qquad G \sim N(0, I_d), \qquad 0 < \vartheta_\eta \le \vartheta_* < \infty.$$

Let $p \ge 1$ be an integer, and suppose that $F \in C^{p+1}$ has bounded positive-order derivatives through order $p+1$. For the probabilists' multi-index Hermite polynomials H_α, normalized by $\mathbb{E}H_\alpha(G)^2 = \alpha!$, set

$$c_\alpha(x) = \frac{\mathbb{E}[F(Y)H_\alpha(G)]}{\sqrt{\alpha!}}, \qquad E^F_{\eta,p}(x) = \eta^{-1} \sum_{1\le|\alpha|\le p} c_\alpha(x)^2.$$

Then, globally in x and without a restriction on $\eta|x|^2$,

$$0 \le \eta^{-1} \operatorname{Var} F(Y) - E^F_{\eta,p}(x) \le \frac{\eta^p \vartheta_\eta^{p+1}}{(p+1)!} \mathbb{E}\|D^{p+1}F(Y)\|^2_{\mathrm{HS}} \le \frac{\eta^p \vartheta_*^{p+1}}{(p+1)!} \|D^{p+1}F\|^2_{\mathrm{HS},\infty}. \tag{87}$$

The Hilbert–Schmidt norm sums over all ordered derivative indices. In particular,

$$E^F_{\eta,2}(x) = \vartheta_\eta |\mathbb{E}\nabla F(Y)|^2 + \frac{\eta\vartheta_\eta^2}{2} \|\mathbb{E}D^2F(Y)\|^2_{\mathrm{HS}}. \tag{88}$$

This second energy needs only a third derivative bound for its global $O(\eta^2)$ variance-density remainder.

Proof. Write $f(G) = F(\Psi_\eta(x) + \sqrt{\eta\vartheta_\eta}G)$. Its Hermite expansion in Gaussian L^2 is $f = \sum_\alpha c_\alpha H_\alpha/\sqrt{\alpha!}$. Linear growth of F gives $f \in L^2$, and the derivative assumptions give the required Gaussian Sobolev norms. Parseval gives

$$\operatorname{Var} f(G) = \sum_{|\alpha|\ge 1} c_\alpha^2.$$

For any integer $k \le p+1$, the ordered tensor derivative identity is

$$\mathbb{E}\|D^k_G f(G)\|^2_{\mathrm{HS}} = \sum_{|\alpha|\ge k} (|\alpha|)_k c_\alpha^2, \qquad (n)_k = n(n-1)\cdots(n-k+1). \tag{89}$$

Indeed, for a multi-index β of length k, differentiation of a normalized Hermite polynomial multiplies its coefficient by $\sqrt{\alpha!/(\alpha-\beta)!}$ when $\beta \le \alpha$. The ordered Hilbert–Schmidt sum weights this derivative by $k!/\beta!$. The multinomial falling-factorial identity

$$\sum_{\substack{\beta\le\alpha\\|\beta|=k}} \frac{k!}{\beta!}\frac{\alpha!}{(\alpha-\beta)!} = (|\alpha|)_k$$

proves (89), first for polynomials and then by Gaussian Sobolev approximation. On $|\alpha| \ge p+1$, the multiplier is at least $(p+1)!$. The chain rule contributes $(\eta\vartheta_\eta)^{p+1}$ to the squared derivative norm. Divide by η to obtain (87).

Gaussian integration by parts gives

$$c_{e_i} = \sqrt{\eta\vartheta_\eta}\,\mathbb{E}\partial_i F(Y), \quad c_{2e_i} = \frac{\eta\vartheta_\eta}{\sqrt{2}}\mathbb{E}\partial_{ii}F(Y), \quad c_{e_i+e_j} = \eta\vartheta_\eta\mathbb{E}\partial_{ij}F(Y) \quad (i<j).$$

Summing their squares proves (88); the factor $1/2$ accounts correctly for the two ordered occurrences of every off-diagonal Hessian entry. □

For the additive Gaussian SGLD subclass, $\Psi_\eta(x) = x + \eta b(x)$ and $\vartheta_\eta = \delta + \eta\sigma^2$. If F has bounded derivatives of orders one through five, Proposition 3.1 and (87) at $p=2$ give, uniformly on every fixed compact set,

$$E^F_{\eta,2}(x) = \delta|\nabla F(x)|^2 + \eta v_1^F(x) + O(\eta^2), \qquad \Sigma = \sigma^2 I.$$

Thus it shares the first correction of the explicit completed energy, while the remainder in (87) is global and independent of the drifted location.

There is also a location-independent coefficient variance bound useful for computation. For $1 \le |\alpha| \le p$, put

$$Z^\circ_\alpha = \frac{F(\Psi_\eta(x) + \sqrt{\eta\vartheta_\eta}G) - F(\Psi_\eta(x))}{\sqrt{\eta}}\frac{H_\alpha(G)}{\sqrt{\alpha!}}.$$

Its mean is $c_\alpha/\sqrt{\eta}$. If $\|\nabla F\|_\infty \le G_1$, then

$$\mathbb{E}(Z^\circ_\alpha)^2 \le G_1^2\vartheta_*\,(d+2|\alpha|). \tag{90}$$

To prove this, use the Lipschitz bound on the value difference and the normalized Hermite recurrence

$$G_i\frac{H_\alpha}{\sqrt{\alpha!}} = \sqrt{\alpha_i+1}\frac{H_{\alpha+e_i}}{\sqrt{(\alpha+e_i)!}} + \sqrt{\alpha_i}\frac{H_{\alpha-e_i}}{\sqrt{(\alpha-e_i)!}}.$$

Orthogonality gives $\mathbb{E}[G_i^2H_\alpha^2/\alpha!] = 2\alpha_i+1$; sum over i. The subtraction of $F(\Psi_\eta(x))$ does not change the mean because $\mathbb{E}H_\alpha = 0$. For $F = f_\eta$, the difference of the two Poisson values is an integral of differences of transition expectations, by (3).

The same Hermite argument extends to a conditional Gaussian covariance $\eta A_\eta(x)$. The last bound in (87) becomes

$$\frac{\eta^p}{(p+1)!}\|A_\eta(x)\|^{p+1}\|D^{p+1}F\|^2_{\mathrm{HS},\infty},$$

using the square-root representation and the operator norm. It is uniform when A_η is uniformly bounded. The conditional Gaussian hypothesis is essential to this particular orthogonal decomposition; bounded covariance alone does not supply it.

Let $1 \le p \le 5$ and

$$\mathcal{I}_p = \{\alpha \in \mathbb{N}_0^d : 1 \le |\alpha| \le p\}, \qquad N_p = \binom{d+p}{p} - 1, \qquad d_\alpha(x) = c_\alpha(x)/\sqrt{\eta}.$$

For $F = f_\eta$ we have $E_{\eta,p}^{f_\eta}(x) = |d(x)|^2$. One pair of paths estimates all N_p entries of $d(x)$.

10.2 A paired Euler estimator

Fix x and put $a = x + \eta b(x)$. Draw a standard Gaussian G, set $y = a + \sqrt{\eta\vartheta_\eta}G$, and independently draw the Gaussian increments for two Euler paths of the modified diffusion,

$$Y_{j+1}^z = Y_j^z + \tau b(Y_j^z) + \sqrt{\tau\vartheta_\eta}\, G_{j+1}, \qquad Y_0^z = z, \qquad \tau = S/K.$$

Use exactly the same G_{j+1} in the paths from y and a. One paired simulation supplies every coordinate of

$$Z_\alpha(x) = -\frac{\tau}{\sqrt{\eta\alpha!}} \sum_{j=0}^{K-1} [h(Y_j^y) - h(Y_j^a)] H_\alpha(G), \qquad \alpha \in \mathcal{I}_p. \tag{91}$$

No invariant mean is present in this expression.

Here are fully specified constants for its bias. Put $\vartheta_* = \delta + \eta_0\sigma^2$, $s_0^2 = d\vartheta_*$, and

$$k_0 = 1 + 4L^2, \quad c_0 = 4B_0^2 + 2s_0^2, \quad M_1(S) = (1 + c_0 S)e^{k_0 S},$$
$$D_1(S) = 2\{B_0^2 + s_0^2 + L^2 M_1(S)\}, \quad E_1(S) = \sqrt{D_1(S)}\{1 + \sqrt{LS}e^{3LS/2}\}.$$

For $|x| \le R$, set

$$A_R = (1 + \eta L)R + \eta B_0, \qquad B_E(S,R) = H_1 S E_1(S)[2 + 2A_R + \sqrt{\eta\vartheta_*}\sqrt{d}]. \tag{92}$$

Lemma 10.2 (Bias and variance of the direct sample)**.** *Let $S \ge 1$, $\tau \le 1$ and $1 \le |\alpha| \le p$. The sample in (91) satisfies*

$$|\mathbb{E}Z_\alpha(x) - d_\alpha(x)| \le \frac{H_1}{\kappa} e^{-\kappa S}\sqrt{\vartheta_* d} + B_E(S,R)\sqrt{\tau/\eta}, \qquad |x| \le R. \tag{93}$$

If also $\tau \le \tau_* := \min(1, \kappa/L^2)$, then

$$\mathbb{E}Z_\alpha(x)^2 \le V_p := \frac{4H_1^2\vartheta_*}{\kappa^2}(d + 2p) \quad \textit{for every } x. \tag{94}$$

Strong dissipativity implies $L \ge \kappa > 0$, so τ_* is well-defined. These estimates are uniform in the small step size of the main SGLD trajectory.

Proof. The anchored representation gives

$$f_\eta(y) - f_\eta(a) = -\int_0^\infty [\mathcal{P}_t^\eta h(y) - \mathcal{P}_t^\eta h(a)]dt.$$

Its part after S has absolute value at most $H_1 e^{-\kappa S}|y-a|/\kappa$, by synchronous diffusion contraction. Multiply by $|H_\alpha(G)|/\sqrt{\eta\alpha!}$ and use Cauchy–Schwarz with $\mathbb{E}H_\alpha^2/\alpha! = 1$ and $\mathbb{E}|G|^2 = d$. This gives the first term of (93).

For any deterministic initial point z, the elementary coupled Euler estimate is

$$\sup_{0\le t\le S}\mathbb{E}|\mathsf{X}^z_t - Y^z_{\lfloor t/\tau\rfloor}| \le E_1(S)(1+|z|)\sqrt{\tau}. \tag{95}$$

If $|z|\le Q$, the Euler second-moment recursion is bounded by $(1+k_0\tau)\mathbb{E}|Y_j|^2 + c_0\tau$. For $j\tau\le t\le (j+1)\tau$, couple the diffusion to the interpolation

$$\overline{Y}_t = Y_j + (t-j\tau)b(Y_j) + \sqrt{\vartheta_\eta}\,(W_t - W_{j\tau}).$$

The displacement $z_t = \overline{Y}_t - Y_{\lfloor t/\tau\rfloor}$ has second moment at most $2[B_0^2+s_0^2+L^2(Q^2+c_0S)e^{k_0S}]\tau$. For $e_t = \mathsf{X}^z_t - \overline{Y}_t$, Young's inequality gives $(\mathbb{E}|e_t|^2)' \le 3L\mathbb{E}|e_t|^2 + L\mathbb{E}|z_t|^2$. Integration and the triangle inequality prove (95), since the square root of the displacement bound at $Q=\max(1,|z|)$ is at most Q times its value at $Q=1$.

Use (95) for $z=y,a$ inside the resolvent integral over $[0,S]$. The initial G is independent of all simulation increments. Since $|a|\le A_R$ and $|y|\le A_R+\sqrt{\eta\vartheta_*}|G|$, the coefficient error is bounded by

$$H_1SE_1(S)\sqrt{\tau/\eta}\,\mathbb{E}\big[(2+|y|+|a|)|H_\alpha(G)|/\sqrt{\alpha!}\big].$$

Use $\mathbb{E}|H_\alpha|/\sqrt{\alpha!}\le 1$ and $\mathbb{E}|G||H_\alpha|/\sqrt{\alpha!}\le\sqrt{d}$ to obtain the second term of (93). This estimate also includes time discretization of the resolvent integral.

For the variance, the common Euler noise cancels pathwise in the difference. Strong dissipativity gives

$$|Y^y_{j+1}-Y^a_{j+1}|^2 \le (1-2\kappa\tau+L^2\tau^2)|Y^y_j-Y^a_j|^2 \le (1-\kappa\tau/2)^2|Y^y_j-Y^a_j|^2.$$

Consequently $\tau\sum_{j<K}|Y^y_j-Y^a_j|\le 2|y-a|/\kappa$. Square (91) and use $\mathbb{E}[|G|^2H^2_\alpha/\alpha!] = d+2|\alpha|$, proved after Proposition 10.1. This proves (94). □

10.3 A finite confidence budget for all features

Theorem 10.3 (Computed Hermite feature vectors)**.** *Let $x_0,\dots,x_{m-1}$ be an arbitrary finite input array, $R=\max(1,\max_k|x_k|)$, $0<\epsilon\le 1$ and $0<p_f<1$. Write $N=N_p$ and $\log_+ u = \max(0,\log u)$. Choose*

$$\begin{aligned} S &= \max\left\{1, \frac{1}{\kappa}\log_+\frac{3\sqrt{N}H_1\sqrt{\vartheta_* d}}{\kappa\epsilon}\right\},\\ K &= \left\lceil S\max\left\{\tau_*^{-1}, \frac{9NB_E(S,R)^2}{\eta\epsilon^2}\right\}\right\rceil, \qquad \tau = S/K,\\ n &= \left\lceil \max\{1, 144NV_p/\epsilon^2\}\right\rceil. \end{aligned} \tag{96}$$

Let B be the smallest positive odd integer with $2^{-B}\le p_f/(mN)$. At every state generate Bn independent paired simulations (91), arrange them into B groups of size n, and take a coordinatewise median of group means. The resulting vectors $\widehat{d}_k$ satisfy, conditionally on the input array,

$$\mathbb{P}_{\mathrm{aux}}\{\max_{k<m}|\widehat{d}_k - d(x_k)| > \epsilon\} \le p_f. \tag{97}$$

The number of Euler updates is $2mnBK$. Each replicate yields all N coefficients. For fixed d,p, evaluating the Hermite weights and accumulating the coefficients takes $O(N)$ additional operations per replicate. The medians can be found in linear time per group array. All budgets are finite on every finite input array.

Proof. The horizon and Euler choices bound each of the two deterministic coordinate biases by $\epsilon/(3\sqrt{N})$. For each group, Chebyshev's inequality and (94) give failure at most $1/16$ at radius $\epsilon/(3\sqrt{N})$ about the sample expectation. The probability that at least half the independent groups fail is at most $2^B(1/16)^{B/2} = 2^{-B}$. A union bound over mN coordinates proves that all coordinate errors are at most $\epsilon/\sqrt{N}$; hence the vector errors are at most ϵ. The pairing cost has no extra factor N because the path integral in (91) is shared by all its Hermite coefficients. □

For fixed model constants, dimension and p, the Euler-update count is bounded by

$$Cm\eta^{-1}\{1+\log(mN/p_f)\}P(1+R,\epsilon^{-1})$$

for a fixed polynomial P. Both the degree of this polynomial and the number N_p of coefficients depend on the fixed model or dimension.

The operation count uses exact real arithmetic. If b and h are evaluated with certified absolute errors, run the construction with tolerance $\epsilon/2$. After drawing G, put

$$H_*(G) = \max_{\alpha\in\mathcal{I}_p} |H_\alpha(G)|/\sqrt{\alpha!}.$$

Let a_b, a_h bound, respectively, the Euclidean drift error and the scalar observable error in this replicate. The drift evaluation at x perturbs both initial points by at most ηa_b. Iterating the Euler recursion then bounds either path error by $e^{LS}(\eta+S)a_b$. Hence the absolute error in each sample (91) is at most

$$\frac{2SH_*(G)}{\sqrt{\eta}}\{H_1 e^{LS}(\eta+S)a_b + a_h\}.$$

Choose the evaluation tolerances after observing G to make this at most $\epsilon/(2\sqrt{N})$. The bound passes to group means and coordinatewise medians, so the Euclidean error allowance is $\epsilon/2$. This argument counts evaluations and Euler steps for ideal Gaussian draws.

10.4 Paired Euler estimates of the gradient and weak-bias correction

The numerator correction $\mathcal{A}_2 f_\eta$ involves derivatives of orders two through four. We first estimate ∇f_η by forward differences along coupled paths, then difference those estimates to obtain the required higher derivatives. We keep the same known σ and evaluable coefficients b, h as in the preceding construction.

Retain $\vartheta_\eta = \delta + \eta\sigma^2$, $L = B_1$, and $\tau_* = \min(1, \kappa/L^2)$. Let $E_1(S)$ be the explicit Euler constant in (95). For a finite array $(x_k)_{k<m}$, set $R = \max(1, \max_k |x_k|)$. Given $u > 0$, one simulation at a state x runs the $d+1$ Euler paths starting at $x, x+ue_1, \dots, x+ue_d$, using the same noise increments in all paths. For coordinate j, its output is

$$A_j(x) = -\frac{\tau}{u}\sum_{\ell=0}^{K-1}\{h(Y_\ell^{x+ue_j}) - h(Y_\ell^x)\}, \qquad \tau = S/K. \tag{98}$$

The shared base path reduces the cost from $2d$ to $d+1$ Euler paths per vector sample.

Lemma 10.4 (Gradient estimates from coupled Euler paths)**.** *For $S \ge 1$ and $\tau \le \tau_*$, the coordinate sample satisfies*

$$|\mathbb{E}A_j(x) - \partial_j f_\eta(x)| \le \frac{G_2 u}{2} + \frac{H_1}{\kappa}e^{-\kappa S} + C_g(S,R,u)\sqrt{\tau}, \qquad |x| \le R,$$
$$C_g(S,R,u) = \frac{H_1 S E_1(S)}{u}(2+2R+u), \qquad |A_j(x)| \le \frac{2H_1}{\kappa}. \tag{99}$$

In particular its variance is at most $V_g = 4H_1^2/\kappa^2$, uniformly in the initial state and the small main-trajectory step size.

For $0 < \epsilon \le 1$, $0 < p < 1$, choose

$$u = \min\left\{1, \frac{\epsilon}{2\sqrt{d}(1+G_2)}\right\}, \qquad S = \max\left\{1, \frac{1}{\kappa}\log_+ \frac{4\sqrt{d}H_1}{\kappa\epsilon}\right\},$$
$$K = \left\lceil S\max\left\{\tau_*^{-1}, \frac{16dC_g(S,R,u)^2}{\epsilon^2}\right\}\right\rceil, \qquad \tau = S/K,$$
$$n = \left\lceil \max\{1, 256dV_g/\epsilon^2\}\right\rceil. \tag{100}$$

Let B be the smallest positive odd integer for which $2^{-B} \le p/(md)$. At each state generate Bn independent vector samples (98), divide them into B groups of size n, and take coordinatewise medians of group means. The resulting vectors $\widehat{g}_k$ satisfy

$$\mathbb{P}_{\mathrm{aux}}\{\max_{k<m} |\widehat{g}_k - \nabla f_\eta(x_k)| > \epsilon \mid (x_k)_{k<m}\} \le p. \tag{101}$$

The number of Euler updates is $(d+1)mnBK$. For fixed dimension and model bounds it is at most

$$Cm\{1+\log(md/p)\}\, P_g(1+R, \epsilon^{-1}) \tag{102}$$

for a fixed polynomial P_g, uniformly in $0 < \eta \le \eta_0$. Coefficient evaluations, Gaussian draws, and the remaining arithmetic have polynomial bounds as well.

Proof. The anchored resolvent representation gives

$$\frac{f_\eta(x+ue_j) - f_\eta(x)}{u} = -\frac{1}{u}\int_0^\infty [\mathcal{P}_t^\eta h(x+ue_j) - \mathcal{P}_t^\eta h(x)]\, dt.$$

The forward difference on the left differs from $\partial_j f_\eta(x)$ by at most $G_2 u/2$. Synchronous contraction of the strongly dissipative additive diffusion bounds the part of the integral after S by $H_1 e^{-\kappa S}/\kappa$. On $[0, S]$, apply (95) to the two starting points. Integrating the Lipschitz observable error, and dividing by u, gives at most $H_1 S E_1(S)(2 + 2R + u)\sqrt{\tau}/u$. This includes the left-endpoint time discretization of the integral.

For the pathwise bound, the common Euler noise cancels. Strong dissipativity and the Lipschitz drift imply

$$|Y_{\ell+1}^{x+ue_j} - Y_{\ell+1}^x|^2 \le (1 - 2\kappa\tau + L^2\tau^2)|Y_\ell^{x+ue_j} - Y_\ell^x|^2 \le (1-\kappa\tau/2)^2 |Y_\ell^{x+ue_j} - Y_\ell^x|^2.$$

Here $L \ge \kappa > 0$ and $\kappa\tau \le 1$, so the geometric factor is nonnegative. Consequently $\tau\sum_{\ell<K} |Y_\ell^{x+ue_j} - Y_\ell^x| \le 2u/\kappa$. The Lipschitz bound for h proves the almost sure sample bound in (99), and hence the variance bound.

The choices of u, S, K allocate at most $\epsilon/(4\sqrt{d})$ to each of the three deterministic biases. At this same tolerance around the sample expectation, the probability that one group mean fails is at most $16dV_g/(n\epsilon^2) \le 1/16$, by Chebyshev's inequality. For a given coordinate its median fails with probability at most $2^B(1/16)^{B/2} = 2^{-B}$. Union over the md coordinates, followed by the Euclidean norm bound, proves (101). Independence between the coordinates of one vector sample is unnecessary.

Each replicate uses K common Gaussian increments and $d+1$ Euler paths, giving the displayed update count and the same order of coefficient evaluations. Group sums and deterministic linear-time median selection use $O(mdnB)$ further operations. For fixed priors,

$S = O(1 + \log \epsilon^{-1})$, while $E_1(S)$ grows at most exponentially times a polynomial in S. Since u^{-1} is polynomial in ϵ^{-1}, K and n are bounded by fixed polynomials in $1 + R, \epsilon^{-1}$. Also $B \le C\{1 + \log(md/p)\}$, proving (102). All bounds use the fixed upper diffusion variance ϑ_*, so they are uniform in η. □

For coefficient evaluations with Euclidean drift error at most a_b and scalar observable error at most a_h, the additional error in each coordinate sample is at most

$$\frac{2S}{u}\{H_1 a_b S e^{LS} + a_h\},$$

by the deterministic Euler error recursion. Running the lemma at $\epsilon/2$ and choosing this bound at most $\epsilon/(2\sqrt{d})$ leaves the same final guarantee. Group means and coordinatewise medians preserve the bound. The number of Euler updates and coefficient evaluations is unchanged.

Corollary 10.5 (Estimating the first weak-bias correction)**.** *In the known-coefficient pure Gaussian model, for every finite array in B_R, $0 < e_a < 1$, and $0 < p_a < 1$, one can compute $\widehat{a}_k$ such that*

$$\mathbb{P}_{\text{aux}}\left\{\max_{k<m} |\widehat{a}_k - \mathcal{A}_2 f_\eta(x_k)| > e_a \mid (x_k)_{k<m}\right\} \le p_a.$$

The number of Euler updates, coefficient evaluations, and arithmetic operations is at most $Cm\{1 + \log(m/p_a)\}P_a(1 + R, e_a^{-1})$, for fixed dimension and priors, with a fixed polynomial P_a. No candidate search is used, and the bound is uniform over the fixed small-step interval.

Proof. Use the finite-difference stencil construction in Corollary 9.1, with its known constants C_A and G_j. More explicitly, put

$$\vartheta = \frac{e_a}{2(1 + C_A)}, \qquad s = \min\left\{1, \frac{\vartheta}{12(1 + \max(G_3, G_4, G_5))}\right\}, \qquad \epsilon_g = \frac{\vartheta s^3}{32}.$$

Apply Lemma 10.4 once to all the gradient arguments needed for the order-one, order-two, and order-three forward differences, at tolerance ϵ_g and failure budget p_a. There are at most $14md^2$ arguments, all in B_{R+3}. For an order-q stencil, its bias is at most $qsG_{q+2} \le \vartheta/4$, and its gradient-value error is at most $2^q \epsilon_g / s^q \le \vartheta/4$. Thus all the required coordinate derivative errors are at most $\vartheta/2$. Insert them and the known drift values into $\mathcal{A}_2 f_\eta$. The sum of the absolute coordinate coefficients is at most C_A, so the resulting error is at most $C_A\vartheta/2 \le e_a/4$. The remaining strict slack accommodates certified scalar evaluations: if needed, use the drift tolerance $\min\{1, e_a/[4(1+D_A)]\}$ from that corollary and a final arithmetic allowance $e_a/4$. Their total with the derivative error is at most $3e_a/4$. This proves the confidence statement.

The reciprocals of s and ϵ_g are bounded by fixed polynomials in $1 + R, e_a^{-1}$. Substitution into (102), with the joint stencil array of size at most $14md^2$, gives the claimed bound. □

10.5 Relative certification and the Hermite trajectory statistic

Lemma 10.6 (Relative accuracy of the estimated energy)**.** *Let $E = m^{-1}\sum_k |d(x_k)|^2$. Assume that, with conditional probability at least $1 - p_f$, $|\widehat{d}_k - d(x_k)| \le \epsilon$ for every k. Define*

$$L_m = \frac{1}{m}\sum_k (|\widehat{d}_k| - \epsilon)_+^2, \qquad U_m = \frac{1}{m}\sum_k (|\widehat{d}_k| + \epsilon)^2.$$

For $0 < \beta < 1/2$, use $\epsilon = \beta^2$ and accept the harmonic value $\widehat{E} = 2L_mU_m/(L_m + U_m)$ exactly when $L_m > 0$ and $(U_m - L_m)/(U_m + L_m) \le \beta$. Otherwise return 1 and a rejection flag. Then

$$\mathbb{P}_{\mathrm{aux}}\{|\widehat{E}/E - 1| > \beta \text{ or rejection} \mid (x_k)_{k<m}\} \le p_f + \mathbf{1}_{\{E<16\beta^2\}}. \tag{103}$$

As before, relative error at zero true energy is infinite.

Proof. On the feature-success event, $E \in [L_m, U_m]$. The triangle and reverse triangle inequalities in the normalized Euclidean norm on $\mathbb{R}^m$ give

$$\sqrt{U_m} \le \sqrt{E} + 2\epsilon, \qquad \sqrt{L_m} \ge (\sqrt{E} - 2\epsilon)_+.$$

If $E \ge 16\beta^2$ and $\epsilon = \beta^2$, then $L_m > 0$ and

$$\frac{U_m - L_m}{U_m + L_m} \le \frac{4\epsilon\sqrt{E}}{E + 4\epsilon^2} \le \frac{4\beta^2}{\sqrt{E}} \le \beta.$$

The relative-error identity in Lemma 9.4 proves the claim. □

Theorem 10.7 (Moderate deviations with a Hermite normalizer)**.** *In the pure additive Gaussian subclass of Assumption 2.2, let $2 \le p \le 5$ and $E_{p,m} = m^{-1}\sum_{k<m} E_{\eta,p}^{f_\eta}(X_k)$. The statistic*

$$W_{p,m} = \frac{\sqrt{T}\{\bar{h}_m + \eta\overline{\mathcal{A}_2 f_{\eta m}} - \theta_\eta\}}{\sqrt{E_{p,m}}}$$

has all four strict and non-strict relative normal-tail approximations uniformly on $[0, r]$ if

$$\frac{r^4 \log(T + e)}{T} \to 0, \qquad \frac{\eta r^3}{\sqrt{T}} \to 0, \qquad r^2\eta^p \to 0, \qquad r\eta^2\sqrt{T} \to 0, \tag{104}$$

with $\eta \to 0$, $T = m\eta \to \infty$ and $r \to \infty$. A fixed convention may be used at zero energy.

If σ is known and b and h are evaluable in the arithmetic model, compute the features by Theorem 10.3, with accuracy $\epsilon = \beta^2$ and budget p_E, and apply Lemma 10.6. Compute the correction by Corollary 10.5 with inputs e_a, p_a. The statistic (9) with this Hermite normalizer has the same conclusion under (10). For fixed d, p and model bounds, the total number of auxiliary Euler updates and arithmetic operations is at most

$$Cm\{1 + \log(m/(p_E p_a))\}P(1 + R, \eta^{-1}, \beta^{-1}, e_a^{-1}) \tag{105}$$

for a fixed polynomial P. Function evaluations and ideal Gaussian draws are counted as operations. This construction uses no candidate enumeration or invariant mean. All budgets are finite on every finite main trajectory.

Proof. Proposition 10.1 gives the deterministic, pathwise estimate

$$0 \le V_m - TE_{p,m} \le C\eta^p T.$$

The clock condition implies $s^2/T \to 0$ for the value of s in Theorem 7.2. Its occupation and zero-tilt clock estimates therefore give $V_m/T \ge c_*$ and (66) outside an event of probability $o(\overline{\Phi}(r))$. In particular $E_{p,m} > 0$ on this event for small η. The choice of a fixed value on the zero-energy event has no effect at this tail scale. Applying Lemma 5.5 to Corollary 7.1 with

normalized numerator error $C(R/\sqrt{T}+\eta^2\sqrt{T})$ and relative variance error $C\eta^p$ gives the asserted relative tails under (104).

On this event $E_{p,m}$ is bounded below by a fixed positive constant, so the exception in (103) disappears eventually. The proof of Theorem 2.3 now applies verbatim to the numerical errors. The feature and gradient constructions have explicit horizons of order the logarithm of inverse accuracy; their Euler bias constants grow at most exponentially in those horizons. The stated budgets are therefore polynomial in inverse accuracy and the input radius. Substituting the finite-difference budgets of Corollary 10.5, which are polynomial as well, proves (105). The only dependence on the confidence budgets is through the number of median groups. Every replicate uses finitely many Euler steps and fixed-dimensional arithmetic. □

Taking $p = 2$ gives the schedule in Corollary 2.4. Larger p reduces the variance-density remainder from $O(\eta^2)$ to $O(\eta^p)$. For each $p \geq 2$, the clock and weak-bias conditions already imply

$$r^2\eta^p = (r\eta^2\sqrt{T})(r/\sqrt{T})\eta^{p-2} \longrightarrow 0.$$

Thus every order $2 \leq p \leq 5$ gives the same sufficient deviation range for this first-corrected numerator.

11 Examples and scope

11.1 A nonlinear model with asymmetric unbounded innovation

In dimension one take

$$b(x) = -ax + c\tanh x, \qquad 0 < c < a, \qquad h(x) = x + \tanh x + \cos x.$$

Then $b'(x) \leq -(a-c)$, and all positive-order derivatives required in Assumption 2.1 are bounded. Let B equal 2 with probability $1/3$ and -1 with probability $2/3$, independently of $G' \sim N(0,1)$, and set $Z = B + \tau G'$. This law is centered and asymmetric, with variance $2 + \tau^2$; for $\tau > 0$ it is non-Gaussian and has unbounded support. With the stated parameter bounds, Theorem 2.3 therefore gives confidence intervals from samples of $b(x) + Z$ and evaluations of h, without an explicit solution of the Poisson equation.

For comparison, let $b(x) = -ax$, $a > 0$, and $h(x) = x$. Then $f_\eta(x) = -x/a$, $\theta_\eta = 0$ and $\mathcal{A}_2 f_\eta = 0$. For an additive innovation of variance Σ, the completed energy is exactly

$$e_\eta^+ = (\delta + \eta\Sigma)/a^2 = \eta^{-1}\operatorname{Var}_x f_\eta(X_1).$$

This checks both the covariance correction and its scaling. In the pure Gaussian case, the Hermite expansion has only a first-order term and every $E_{\eta,p}^{f_\eta}$ is exact. The paired-gradient estimator is also explicit: with inner step τ and K steps, its value is $-\{1-(1-a\tau)^K\}/a$ on every auxiliary noise realization. The deterministic paired-gradient formula provides an exact test of the sign of the estimator.

11.2 A weakly dissipative oracle with no exponential moment

Keep $b(x) = -ax + c\tanh x$ but now let $c > a > 0$, and set $\sigma_0(x) = s_0\sqrt{1+x^2}$, with $s_0 > 0$. For $z = x - y$, boundedness of tanh gives

$$(b(x) - b(y))z \leq -az^2 + 2c|z| \leq -\tfrac{a}{2}z^2 + 2c^2/a.$$

Thus the drift is weakly dissipative, while $b'(0) = c - a > 0$ precludes global strong dissipativity. The covariance factor has linear growth and known derivative bounds: one may take $S_0 = S_1 = S_2 = s_0$, $S_3 = 3s_0$, $S_4 = 15s_0$, $S_5 = 105s_0$, and $S_6 = 945s_0$. These follow by differentiating $P_j(x)/(1+x^2)^{j-1/2}$, starting with $P_1(x) = x$, and summing the absolute polynomial coefficients.

Let R have density $5t^{-6}\mathbf{1}_{\{t\geq 1\}}$, and let ε be an independent symmetric sign. Put $\zeta = \sqrt{3/5}\,\varepsilon R$ and $Z(x) = s_0\sqrt{1+x^2}\,\zeta$. Direct integration gives $\mathbb{E}\zeta = 0$, $\mathbb{E}\zeta^2 = 1$ and $\mathbb{E}\zeta^4 = 9/5$, whereas $\mathbb{E}|\zeta|^6 = \infty$ and there is no nontrivial exponential moment. On B_D, the local moment inputs can be chosen as

$$V_D = s_0^2(1+D^2), \qquad W_D = \tfrac{18}{5}s_0^4(1+D^2)^2.$$

With any nonconstant admissible h, the oracle gradient, correction, and positive-energy certificates apply to this model. Its heavy-tailed, state-dependent main chain is outside the trajectory conclusion of Theorem 2.3. In particular, this example separates the sixth moment of the modified Brownian diffusion in Lemma 8.2 from the fourth-moment requirement on the actual oracle samples.

11.3 What the numerical certificate establishes

In (73), truncating the semigroup integral at time S accounts for the first error. The other two terms come from the residual on B_D and the diffusion outside B_D. A sixth-moment estimate and a global growth bound control the exterior contribution. This permits a whole-space gradient estimate from local residual evaluations. The unknown invariant mean enters through a scalar that disappears on differentiation.

The finite construction searches over smooth candidate functions. A Galerkin, finite-element or regression method can supply a candidate instead, provided that its global derivative envelopes are certified. The same local coefficient and residual verifier then gives a valid gradient certificate. The coefficient event holds simultaneously for every admissible candidate. The same estimates can therefore be used to select and validate a candidate at the stated confidence level.

In the pure Gaussian model with evaluable coefficients, the paired Euler constructions in Section 10 additionally give a polynomial bound for the total auxiliary simulation and arithmetic count, for fixed dimension and model envelopes. In both access models the sampling cost grows only logarithmically in the inverse failure probability. This matters when the required probability decays exponentially in r^2.

Known structural and moment envelopes remain part of the problem specification. Proposition B.1 gives finite-sample obstructions to learning unrestricted moment bounds or certifying exterior dissipativity from finitely many queries.

11.4 The target and the limits of the theorem

The confidence intervals following Theorem 2.3 concern $\theta_\eta = \pi_\eta h$, the modified diffusion mean. The discrete chain's invariant mean is a different object. The joint limit in (8) is essential: a remaining deterministic weak bias can accumulate if η is held fixed while T tends to infinity. An interval for $\pi_0 h$ would additionally require a quantitative comparison of those invariant means at the desired relative-tail scale.

For a weakly dissipative drift, the deterministic Euler map need not contract at every state. For a state-dependent unbounded innovation factor, the exact transition need not be a location family with the global curvature bound (26). The broader residual certificate and weighted

variance expansion remain valid, while the enlarged CMD proof requires these additional tilted-chain properties. They cannot be deduced from the computational certificate alone.

Two estimates determine the radius in Corollary 2.4. The finite martingale comparison requires $r^4 \log(T+e)/T \to 0$. The weak bias left after the first correction requires $r\eta^2\sqrt{T} \to 0$. Higher Hermite orders improve the variance estimate alone, so the second condition persists.

Lower primitive regularity, general superlinear coefficients, heavier-tailed main-trajectory innovations, and joint limits $d \to \infty$ or $\delta \to 0$ require additional estimates. For example, the residual certificate explicitly contains $\delta^{-1/2}$ and the contraction constants; the covering numbers and tensor norms depend on dimension. Their finiteness for a fixed model is insufficient for either joint limit. The finite-fourth-moment oracle result is a numerical theorem and does not assert a heavy-tail CMD for the main chain.

11.5 Reproducible algebraic checks

Two supplementary scripts evaluate the completed-square identity, the Gaussian Hermite residual, and the paired-gradient formula for the linear model above using 80-decimal-digit arithmetic. For $F(x) = \cos x$ and $Y = a + sG$, the Hermite coefficients and variance are

$$c_n = \frac{s^n}{\sqrt{n!}} e^{-s^2/2} \cos(a + n\pi/2), \qquad \operatorname{Var}(\cos Y) = \tfrac{1}{2}\{1 + e^{-2s^2}\cos(2a)\} - e^{-s^2}\cos^2 a.$$

The variance checks use truncation orders $p = 1, \dots, 4$ and include locations of the form $x = \eta^{-3/4}$, for which ηx^2 grows as η decreases. At the tested points the observed residuals satisfy the factorial upper bound in (87). The paired-gradient checks compare with the exact finite-step formula above.

A Effective contraction and global Poisson estimates

The following estimates also underlie the leading-order normalization in [Sun26]. For the modified diffusion, we bound the contraction rate, invariant second moment, and derivatives of the Poisson solution in terms of the stated envelopes, dimensions, and temperature δ. The diffusion factor may depend on the state and grow linearly. The lower bound on the limiting energy also depends on the choice of observable.

Write

$$B_0 \ge |b(0)|, \qquad S_0 \ge \|\sigma_0(0)\|_{\mathrm{HS}}, \qquad \|D^j b\|_\infty \le B_j, \quad \|D^j \sigma_0\|_\infty \le S_j.$$

For derivatives of $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^{d\times\ell}$, the last norm is the multilinear operator norm with Hilbert–Schmidt output. Decrease a fixed $\eta_0 > 0$, if necessary, so that $\eta_0 S_1^2 \le K_1$, where K_1, K_2 are the weak-dissipativity constants of Assumption 2.1. Throughout this section the weak-dissipativity condition means

$$\langle b(x) - b(y), x - y\rangle \le -K_1|x-y|^2 + K_2, \qquad x, y \in \mathbb{R}^d, \quad K_1 > 0, \quad K_2 \ge 0.$$

The diffusion associated with (2) can be represented as

$$dX_t^\eta = b(X_t^\eta)\,dt + \sqrt{\delta}\,dB_t + \sqrt{\eta}\,\sigma_0(X_t^\eta)\,dW_t, \qquad 0 \le \eta \le \eta_0, \tag{106}$$

with independent Brownian motions of dimensions d and ℓ. Global Lipschitz continuity gives a unique nonexplosive strong solution and a Markov semigroup $\mathcal{P}_t^\eta$.

A.1 A common effective contraction rate

Lemma A.1 (Contraction and invariant moments)**.** *Put*

$$\overline{K} = K_1 - \frac{\eta_0 S_1^2}{2}, \qquad \overline{L} = B_1 + \frac{\eta_0 S_1^2}{2}.$$

If $K_2 > 0$*, choose* $0 < k < \overline{K}$ *and define*

$$R_c^2 = \frac{K_2}{\overline{K} - k}, \qquad a_c = \frac{\overline{L} + k}{4\delta}, \qquad A_c = a_c R_c^2, \qquad C_c = e^{A_c}, \qquad \rho = k e^{-A_c}.$$

If $K_2 = 0$*, take* $C_c = 1$ *and* $\rho = \overline{K}$*. For every* $0 \le \eta \le \eta_0$*, the diffusion has a unique invariant probability* π_η *in the space of probabilities with finite first moment. For all such probabilities* ν, ν'*,*

$$W_1(\nu \mathcal{P}_t^\eta, \nu' \mathcal{P}_t^\eta) \le C_c e^{-\rho t} W_1(\nu, \nu'), \qquad \operatorname{Lip}(\mathcal{P}_t^\eta \varphi) \le C_c e^{-\rho t} \operatorname{Lip}(\varphi). \tag{107}$$

There is also the explicit common bound

$$\begin{aligned} D_* &= 2K_2 + \frac{4B_0^2}{K_1} + \delta d + 5\eta_0 S_0^2, \qquad M_* = \frac{2D_*}{K_1}, \\ \mathbb{E}_x |X_t^\eta|^2 &\le e^{-K_1 t/2} |x|^2 + M_*, \qquad \pi_\eta |x|^2 \le M_*. \end{aligned} \tag{108}$$

Proof. Reflect the additive Brownian motion in the direction joining two solutions and couple their multiplicative Brownian motions synchronously. After the solutions meet, use the same noises for both. Away from the diagonal this construction is defined by locally Lipschitz coefficients and localization; continuity permits the switch at the meeting time. The reflected driving process is Brownian by Lévy's characterization: its quadratic variation is tI_d, and its cross-variation with W is zero. Thus both marginals have the required independent additive and multiplicative drivers. Before meeting, set

$$r = |X - Y|, \qquad e = (X - Y)/r, \qquad \Delta\sigma = \sigma_0(X) - \sigma_0(Y).$$

For an increasing concave radial function F, Itô's formula gives

$$\begin{aligned} \mathcal{A}F(r) = 2\delta F''(r) + F'(r) \Big[\langle e, b(X) - b(Y) \rangle + \frac{\eta}{2r} \{ \|\Delta\sigma\|_{\mathrm{HS}}^2 - |\Delta\sigma^\top e|^2 \} \Big] \\ + \frac{\eta}{2} F''(r) |\Delta\sigma^\top e|^2. \end{aligned}$$

The last term is nonpositive. The expression in square brackets is bounded above by both $\overline{L} r$ and $-\overline{K} r + K_2/r$. Eberle [Ebe16] uses a concave distance with reflection coupling for constant diffusion matrices. Here the two terms involving $\Delta\sigma$ account for the state-dependent noise coupled synchronously.

For $K_2 > 0$, choose $F(0) = 0$ and

$$F'(r) = \exp\{-a_c \min(r, R_c)^2\}.$$

Then $e^{-A_c} r \le F(r) \le r$. For $r < R_c$,

$$\mathcal{A}F(r) \le (-4\delta a_c + \overline{L}) r F'(r) = -k r F'(r) \le -k e^{-A_c} F(r).$$

For $r > R_c$, the second drift bound is at most $-kr$, so the same conclusion holds. The first derivative of F is continuous at R_c, and hence the piecewise second derivative creates no local-time term. Stopping and then removing the stopping times proves exponential contraction for the transportation cost $F(|x-y|)$. Its comparison with Euclidean distance proves (107). If $K_2 = 0$, use $F(r) = r$ directly.

The semigroup maps probabilities with finite first moment into themselves. At a sufficiently large fixed time its W_1 contraction factor is less than one. The contraction mapping theorem on the complete W_1 space gives its unique invariant law. Commutation of the semigroup implies invariance at every time.

For the second moment, weak dissipativity and Young's inequality give

$$2\langle x, b(x)\rangle \le -\frac{7}{4}K_1|x|^2 + 2K_2 + \frac{4B_0^2}{K_1}, \qquad \|\sigma_0(x)\|_{\mathrm{HS}}^2 \le \frac{5}{4}S_1^2|x|^2 + 5S_0^2.$$

Thus $L_\eta|x|^2 \le -K_1|x|^2/2 + D_*$. Stopped Itô and Gronwall prove the finite-time bound in (108). Time averages started at zero converge weakly to the invariant law, by (107). Truncation and the uniform second-moment bound give the invariant bound. □

A.2 An integrable semigroup smoothing estimate

We record a finite recursion because the numerical certificates need effective derivative bounds. The construction uses stochastic integration by parts in the additive Brownian motion; the underlying first-derivative formula is the Bismut–Elworthy–Li formula [EL94]. We give the higher-order cancellation explicitly. Only positive-order derivatives of σ_0, rather than a uniform bound on its values or derivatives of a matrix square root, enter this argument.

Fix an integer $N \ge 2$, and suppose for this subsection that $b, \sigma_0 \in C^N$, with the positive-order envelopes B_j, S_j finite through order N. Set $s_j = \sqrt{\eta_0}S_j$. For $p \ge 2$, define

$$a_p = pB_1 + \frac{p(p-1)}{2}s_1^2, \qquad b_p = pB_1 + (p-1)^2 + p(p-1)s_1^2.$$

The distinction between these constants is that a_p treats a homogeneous linear variation equation, whereas b_p also permits forcing. Indeed, if

$$dY = (A_tY + F_t)\,dt + (\mathcal{B}_tY + G_t)\,dW_t, \quad \|A_t\| \le B_1, \quad \|\mathcal{B}_tv\|_{\mathrm{HS}} \le s_1|v|,$$

then Itô and Young give

$$\frac{d}{dt}\mathbb{E}|Y_t|^p \le b_p\mathbb{E}|Y_t|^p + \mathbb{E}|F_t|^p + 2(p-1)\mathbb{E}\|G_t\|_{\mathrm{HS}}^p. \tag{109}$$

For $F_t = G_t = 0$, the coefficient b_p can be replaced by a_p. These inequalities follow first after stopping, so they do not presume moment estimates for the variations.

Let $\mathcal{P}_n$ be the partitions of $\{1,\dots,n\}$. Define recursively

$$\begin{aligned}
V_{1,p} &= e^{a_p/p},\\
F_{n,p} &= \sum_{\substack{\pi\in\mathcal{P}_n\\ |\pi|\ge 2}} B_{|\pi|}\prod_{A\in\pi} V_{|A|,p|\pi|},\\
H_{n,p} &= \sum_{\substack{\pi\in\mathcal{P}_n\\ |\pi|\ge 2}} s_{|\pi|}\prod_{A\in\pi} V_{|A|,p|\pi|},\\
V_{n,p} &= e^{b_p/p}\{F_{n,p}^p + 2(p-1)H_{n,p}^p\}^{1/p}, \qquad 2 \le n \le N.
\end{aligned} \tag{110}$$

This is a finite recursion: every block on the right has size less than n, although larger moment orders are used.

For initial unit directions, let $J_A(t)$ be the spatial variation indexed by a nonempty set A. Differentiating the SDE gives the homogeneous coefficients Db, $\sqrt{\eta}D\sigma_0$. For $|A| \geq 2$, the drift forcing is

$$\sum_{\substack{\pi \in \mathcal{P}(A) \\ |\pi| \geq 2}} D^{|\pi|} b(X_t)[J_B(t) : B \in \pi],$$

and the stochastic forcing has the analogous derivatives of $\sqrt{\eta}\sigma_0$. Hölder and (109) prove

$$\sup_{x,\, 0 \leq \eta \leq \eta_0,\, 0 < t \leq 1} \|J_A(t)\|_{L^p} \leq V_{|A|,p}. \tag{111}$$

For $|A| \geq 2$, their p-th moments are in fact at most $tV_{|A|,p}^p$, because these variations start at zero. Localization, difference quotients and the same bounds justify the spatial differentiations and their finite-time L^p continuity. Neither the recursion nor the bound uses a uniform bound on the coefficient values.

For $2 \leq n \leq N$ and $0 < t \leq 1$, use the adapted additive-noise control

$$q_s = \frac{J_{\{n\}}(s)}{t\sqrt{\delta}}, \qquad W_n = \int_0^t \langle q_s, dB_s \rangle.$$

Writing D_q^B for the Malliavin directional derivative, the variation equation gives

$$D_q^B X_s = (s/t) J_{\{n\}}(s), \qquad \mathbb{E}W_n^2 \leq \frac{e^{a_2}}{\delta t}. \tag{112}$$

The first identity holds for $0 \leq s \leq t$. It follows by differentiating $(s/t)J_{\{n\}}(s)$: its additional drift is $J_{\{n\}}(s)/t$, exactly the forcing $\sqrt{\delta}\, q_s$ in the Malliavin variation equation. The random control itself is not differentiated in the directional derivative $\langle D^B(\cdot), q \rangle$. As q is adapted and square integrable, duality reads $\mathbb{E}D_q^B H = \mathbb{E}(HW_n)$; no exponential moment of its quadratic energy is needed.

For $i < n$, put

$$R_{in} = J_{\{i,n\}} - D_q^B J_{\{i\}}.$$

For compactness write $J_i = J_{\{i\}}$ in the following equation. Direct subtraction of the two variation equations gives

$$\begin{aligned} dR_{in}(s) = {} & Db(X_s)R_{in}(s)\, ds + \sqrt{\eta} D\sigma_0(X_s)[R_{in}(s)]\, dW_s \\ & + (1 - s/t) D^2 b(X_s)[J_i(s), J_n(s)]\, ds \\ & + \sqrt{\eta}(1 - s/t) D^2 \sigma_0(X_s)[J_i(s), J_n(s)]\, dW_s, \qquad R_{in}(0) = 0. \end{aligned}$$

Thus the homogeneous terms agree with those of the second spatial variation, and each forcing term has the extra factor $1 - s/t$. Equation (109) implies $\|R_{in}(t)\|_{L^p} \leq V_{2,p}$. The stochastic forcing remains present although integration by parts is performed only in B.

In the Faà di Bruno formula for $D^n \mathcal{P}_t^\eta \varphi$, apply duality to its all-singleton term. Each derivative $D_q^B J_{\{i\}}$ then combines with the partition whose only nonsingleton block is $\{i, n\}$. If

$\mathcal{P}_n^\circ$ excludes these $n-1$ partitions and the all-singleton partition, the resulting exact identity is

$$\begin{aligned} D^n\mathcal{P}_t^\eta\varphi = \mathbb{E}\{D^{n-1}\varphi(X_t)[J_{\{1\}},\dots,J_{\{n-1\}}]W_n\} \\ + \sum_{i<n}\mathbb{E}D^{n-1}\varphi(X_t)[J_{\{1\}},\dots,R_{in},\dots,J_{\{n-1\}}] \\ + \sum_{\pi\in\mathcal{P}_n^\circ}\mathbb{E}D^{|\pi|}\varphi(X_t)[J_A : A\in\pi]. \end{aligned} \tag{113}$$

Only Malliavin derivatives of first spatial variations are needed in this identity, for every n.

Set $P = 2N$, and define

$$K_n = \delta^{-1/2}e^{a_2/2}V_{1,P}^{n-1} + (n-1)V_{2,P}V_{1,P}^{n-2} + \sum_{\pi\in\mathcal{P}_n^\circ}\prod_{A\in\pi}V_{|A|,P}.$$

Cauchy–Schwarz for W_n, followed by Hölder for the variations, gives the integrable short-time estimate

$$\|D^n\mathcal{P}_t^\eta\varphi\|_\infty \le K_n t^{-1/2}\sum_{j=1}^{n-1}\|D^j\varphi\|_\infty, \qquad 0<t\le 1. \tag{114}$$

The identity is first proved for smooth inputs by the adapted Malliavin duality just described. For a C^{n-1} input, convolution with a smooth approximate identity preserves every positive-order derivative bound and converges, with those derivatives, uniformly on compact sets. In (113), all input derivatives have order at most $n-1$. The variation bounds at larger moment orders, together with finite-time moments of X_t, make the products uniformly integrable locally uniformly in the initial state. Truncating to a large state ball, then removing this truncation, therefore gives a locally uniform limit of the right-hand side. Its continuity follows from the L^p continuity of the variations. The elementary theorem on locally uniform limits of derivatives then identifies this limit as the n-th derivative of $\mathcal{P}_t^\eta\varphi$. This proves (114) for the stated C^{n-1} inputs. Only linear growth of their values is needed.

We also need smoothing from merely Lipschitz inputs. There is a separate first gain of regularity. For a unit vector v, set $q_s^v = J_v(s)/(t\sqrt{\delta})$, and let W_v be its Itô integral against B. The same duality gives

$$D\mathcal{P}_t^\eta\varphi(x)v = \mathbb{E}_x\{\varphi(X_t)W_v\}.$$

To justify this formula for globally Lipschitz φ, use its smooth convolutions φ_ϵ. Their value errors are uniformly $O(\epsilon\,\mathrm{Lip}(\varphi))$; hence the corresponding right-hand sides converge uniformly by the L^2 bound on W_v. The derivative-limit argument gives $\mathcal{P}_t^\eta\varphi\in C^1$, and the ordinary first variation estimate yields

$$\|D\mathcal{P}_t^\eta\varphi\|_\infty \le V_{1,P}\,\mathrm{Lip}(\varphi), \qquad 0<t\le 1.$$

No derivative of the original Lipschitz input is assumed.

Write

$$\|\varphi\|_{[r]} = \sum_{j=1}^{r}\|D^j\varphi\|_\infty, \qquad A_r = \sum_{j=1}^{r}\sum_{\pi\in\mathcal{P}_j}\prod_{B\in\pi}V_{|B|,P},$$

and define the finite, deliberately conservative constant

$$\mathsf{S}_N = V_{1,P}\prod_{r=1}^{N-1}\{A_r + \sqrt{N}\,K_{r+1}\}.$$

The ordinary spatial variation formula gives $\|\mathcal{P}_t^\eta\varphi\|_{[r]} \le A_r\|\varphi\|_{[r]}$ for $t \le 1$. Split time one into N intervals of length $1/N$. The first interval gives the C^1 estimate above, and each of the remaining intervals gains one derivative using (114). Thus

$$\|\mathcal{P}_1^\eta\varphi\|_{[N]} \le \mathsf{S}_N \operatorname{Lip}(\varphi). \tag{115}$$

A.3 Canonical Poisson solutions and explicit derivative bounds

Lemma A.2 (Global Poisson bounds)**.** *Suppose the weak-dissipativity and ellipticity conditions above hold. For a fixed $N \ge 2$, assume $b, \sigma_0 \in C^N$ with bounded positive-order derivatives through order N. Let $\varphi \in C^{N-1}$, and write $D_j^\varphi = \|D^j\varphi\|_\infty < \infty$, $1 \le j \le N-1$. Its value may grow linearly. There is a unique globally Lipschitz C^N solution, anchored at zero, of*

$$L_\eta F_\eta^\varphi = \varphi - \pi_\eta\varphi.$$

Its absolutely convergent representation is

$$F_\eta^\varphi(x) = -\int_0^\infty \{\mathcal{P}_t^\eta\varphi(x) - \mathcal{P}_t^\eta\varphi(0)\}\,dt. \tag{116}$$

The following computable bounds hold uniformly for $0 \le \eta \le \eta_0$:

$$\begin{aligned} \|DF_\eta^\varphi\|_\infty &\le \mathcal{G}_1(D^\varphi) := (C_c/\rho)D_1^\varphi, \\ \|D^nF_\eta^\varphi\|_\infty &\le \mathcal{G}_n(D^\varphi) := 2K_n\sum_{j=1}^{n-1} D_j^\varphi + (C_c/\rho)\mathsf{S}_N D_1^\varphi, \qquad 2 \le n \le N. \end{aligned} \tag{117}$$

In particular, under Assumption 2.1, take $N = 6$ and the known upper bounds $\|D^jh\|_\infty \le H_j$, $1 \le j \le 5$. Then $f_\eta := F_\eta^h$ has bounds $G_j = \mathcal{G}_j(H)$, $1 \le j \le 6$. Set

$$\theta_\eta = \pi_\eta h, \qquad w_\eta = |\nabla f_\eta|^2, \qquad \mu_\eta = \pi_\eta w_\eta.$$

The second anchored Poisson solution u_η, defined by

$$L_\eta u_\eta = w_\eta - \mu_\eta,$$

also has bounded global derivatives through order six. Explicitly, put

$$W_j = \sum_{a=0}^{j}\binom{j}{a} G_{a+1}G_{j-a+1}, \qquad 1 \le j \le 5.$$

Then $\|D^ju_\eta\|_\infty \le U_j := \mathcal{G}_j(W)$, $1 \le j \le 6$, using the same coefficient constants.

Proof. Contraction and the invariant first moment imply absolute convergence of

$$F_\eta^\circ(x) = -\int_0^\infty (\mathcal{P}_t^\eta\varphi(x) - \pi_\eta\varphi)\,dt.$$

Subtracting its value at zero gives (116). Integrating (107) proves the gradient bound. For derivatives of order $2, \dots, N$, integrate (114) over $(0,1)$; the singular factor has integral two.

On $[1,\infty)$, apply (115) to $\mathcal{P}^\eta_{t-1}\varphi$ and then use its exponentially decreasing Lipschitz constant. This gives exactly (117) and justifies differentiating the resolvent.

Fubini's theorem and the semigroup property give

$$\mathcal{P}^\eta_s F^\circ_\eta - F^\circ_\eta = \int_0^s (\mathcal{P}^\eta_t\varphi - \pi_\eta\varphi)\,dt.$$

After division by s and passage to zero, the right-hand side converges to $\varphi - \pi_\eta\varphi$. Localized Itô's formula for the established C^2 function identifies the left limit as $L_\eta F^\circ_\eta$. Its first two derivatives are bounded, the coefficients grow at most linearly, and finite-time second moments justify removal of the localization. This proves the Poisson equation pointwise. If two globally Lipschitz classical solutions have the same forcing, their difference v satisfies $L_\eta v = 0$. Localized Itô's formula gives $\mathcal{P}^\eta_t v = v$. To remove the localization, use boundedness of ∇v, linear growth of the diffusion factor, and finite-time second moments; the stochastic integral is a true martingale, and the terminal values are uniformly integrable. Therefore

$$|v(x) - v(y)| \le C_c e^{-\rho t}\,\mathrm{Lip}(v)|x-y| \longrightarrow 0.$$

Anchoring removes the constant.

The preceding proof applied to $\varphi = h$ gives f_η. Leibniz's rule gives $\|w_\eta\|_\infty \le G_1^2$ and $\|D^j w_\eta\|_\infty \le W_j$, $1 \le j \le 5$. Apply the same construction to $\varphi = w_\eta$ to obtain the second equation and the U_j bounds. □

In particular, C^4 coefficients with the stated bounded positive-order derivatives and a C^3 forcing give four bounded global Poisson derivatives. Uniform ellipticity, common dissipativity constants and common derivative envelopes make these bounds uniform across a coefficient family. The integrable singularity in (114) gives the fourth derivative from three derivatives of the forcing.

A.4 Energy nondegeneracy and global cutoffs

Lemma A.3 (Positive limiting energy)**.** *For the nonconstant observable in Assumption 2.1,*

$$W_1(\pi_\eta, \pi_0) = O(\sqrt{\eta}), \qquad f_\eta \longrightarrow f_0 \text{ in } C^2_{\mathrm{loc}}, \qquad \mu_\eta \longrightarrow \mu_0 > 0.$$

Consequently there is a fixed-model constant $\mu_* > 0$ such that $\mu_\eta \ge \mu_*$ for all sufficiently small η. Rescaling a fixed nonconstant observable by a factor in $(0,1)$ preserves its upper derivative envelopes and scales the energy by the square of that factor.

Proof. Couple X^η and X^0 from the same initial state using the same additive Brownian motion. For their difference,

$$\frac{d}{dt}\mathbb{E}|X^\eta_t - X^0_t|^2 \le 2B_1\mathbb{E}|X^\eta_t - X^0_t|^2 + 2\eta\{S_0^2 + S_1^2\mathbb{E}|X^\eta_t|^2\}.$$

Choose the initial law π_η. The moment estimate gives

$$W_1(\pi_\eta, \pi_\eta\mathcal{P}^0_t) \le \sqrt{\eta}\,C_t, \qquad C_t^2 = 2(S_0^2 + S_1^2 M_*)\,te^{2B_1 t}.$$

Fix t_* with $C_c e^{-\rho t_*} \le 1/2$. The triangle inequality and contraction of $\mathcal{P}^0_{t_*}$ then give $W_1(\pi_\eta, \pi_0) \le 2C_{t_*}\sqrt{\eta}$. In particular, $\theta_\eta \to \theta_0$.

The anchored functions and their first three derivatives have uniform local bounds. Every sequence $\eta \downarrow 0$ therefore has a subsequence converging in C^2 on compact sets to a globally Lipschitz function $\bar{f}$, anchored at zero. The coefficients of L_η converge locally to those of L_0; hence $L_0\bar{f} = h - \theta_0$. The uniqueness argument in Lemma A.2 identifies $\bar{f} = f_0$, proving convergence of the whole family. Since $w_\eta \le G_1^2$, local gradient convergence and uniform tightness of π_η imply $\mu_\eta \to \pi_0|\nabla f_0|^2 = \mu_0$.

The additive-noise diffusion at $\eta = 0$ has full support at every positive time. Indeed, a smooth controlled path can join the initial point to any nonempty open ball, the additive solution map is continuous in the driving path, and Brownian motion assigns positive probability to every such path neighborhood. Invariance implies that π_0 has full support as well. If $\mu_0 = 0$, continuity forces $\nabla f_0 \equiv 0$, and its Poisson equation makes h constant, contrary to the assumption. □

Lemma A.4 (Cutoffs retaining all derivative bounds)**.** *Let $F(0) = 0$ and $\|D^j F\|_\infty \le C_j$, $1 \le j \le N$. Choose a fixed smooth χ, equal to one on B_2, zero outside B_3, and taking values in $[0,1]$. For $R \ge 1$, set $\chi_R(x) = \chi(x/R)$ and $F_R = \chi_R F$. Then $F_R = F$ on B_{2R}, and all derivatives of orders $1, \dots, N$ are bounded uniformly in R. More precisely, if $c_{\chi,j} = \|D^j \chi\|_\infty$, then*

$$\|D^j F_R\|_\infty \le C_j + \sum_{a=1}^{j-1} \binom{j}{a} c_{\chi,a} C_{j-a} + 3c_{\chi,j} C_1, \qquad 1 \le j \le N. \tag{118}$$

This applies with $N = 6$ to f_η, u_η, and with the corresponding derivative order to other anchored solutions.

Proof. Anchoring and the gradient bound give $|F(x)| \le C_1|x|$. Leibniz's formula bounds the terms with a positive derivative on F by the displayed derivative envelopes, using $R^{-a} \le 1$. The term with all j derivatives on the cutoff is supported in B_{3R} and is at most $3c_{\chi,j}C_1R^{1-j}$. This is bounded for $j \ge 1$. □

These are global cutoff estimates. Conditional expectations of $F_R(X_{k+1})$ can consequently be expanded over the entire one-step innovation law, including jumps outside the localization ball. No one-step overshoot is removed from the transition kernel.

B Limits of finite oracle sampling

The moment and analytic upper envelopes used by the numerical constructors cannot themselves be certified uniformly from an unrestricted finite sampling interface. The following elementary obstructions make this information requirement precise.

Proposition B.1 (Finite-sample information obstructions)**.** *Consider an adaptive algorithm that stops after finitely many calls almost surely under an identically zero, state-independent scalar oracle.*

1. *Suppose the procedure reports an interval for the oracle mean and the interval is bounded almost surely under the zero oracle. For every $0 < p < 1$, some finite-support oracle law gives this interval coverage less than $1 - p$.*

2. *Even with known mean zero and known variance upper bound $S^2 > 0$, no such procedure can estimate variance to accuracy $e < S^2/2$ with uniform success probability $1 - p > 1/2$.*

In addition, without a supplied exterior condition, no procedure using finitely many adaptive drift evaluations can both certify $b(x) = -x$ as globally dissipative with probability at least $1 - p$ and have false-certification probability at most $p < 1/2$ uniformly over smooth drifts. This assertion still holds under a common finite Lipschitz upper bound permitting contracting and expanding linear exterior drifts.

Proof. For the first assertion, fix $q > 0$. Under the zero oracle, choose finite N_0, A so that the algorithm stops by N_0 and reports an interval in $[-A, A]$ with probability at least $1 - q$, including its internal randomness. For $\theta > A$, replace the oracle by $(1-\varepsilon)\delta_0 + \varepsilon\delta_{\theta/\varepsilon}$. Couple the internal randomness and successive observations. With probability at least $1 - N_0\varepsilon$, the first N_0 answers are identical, irrespective of the adaptive query locations. The new mean is θ, so its coverage failure is at least $1 - q - N_0\varepsilon$, which can exceed any prescribed $p < 1$. Each chosen contamination law has finite support.

For the variance assertion, choose $2e < v \leq S^2$ and put mass $1 - \varepsilon$ at zero and mass $\varepsilon/2$ at each of $\pm\sqrt{v/\varepsilon}$. Its variance is v, whereas that of the zero oracle is zero; the two accuracy intervals of radius e are disjoint. Choose N_0 so that the zero-oracle stopping tail is at most q. If baseline accuracy has probability at least $1 - p$, accuracy under contamination is at most $p + q + N_0\varepsilon$, contradicting the same guarantee when $q + N_0\varepsilon < 1 - 2p$.

For the exterior assertion, finitely many query locations under $b_0(x) = -x$ lie in a deterministic ball B_R with probability at least $1-q$, for sufficiently large R. Choose a fixed smooth scalar cutoff χ, zero on $[0, 1]$ and one on $[2, \infty)$, and define $b_R(x) = \{-1 + 2\chi(|x|/R)\}x$. This agrees with $-x$ on B_R, equals $+x$ outside B_{2R}, and has a Lipschitz bound independent of R. Coupled adaptive transcripts agree on the bounded-query event. A baseline certification probability of at least $1 - p$ therefore forces false certification for b_R with probability at least $1 - p - q > p$ when $q < 1 - 2p$. □

References

[DFL26] H. Dai, X. Fan, and J. Lu. "Self-normalized Cramér-type moderate deviation of stochastic gradient Langevin dynamics". In: *Journal of Applied Probability* 63.3 (2026), pp. 1034–1061. DOI: 10.1017/jpr.2025.10065. arXiv: 2410.22047.

[Ebe16] A. Eberle. "Reflection couplings and contraction rates for diffusions". In: *Probability Theory and Related Fields* 166.3–4 (2016), pp. 851–886. DOI: 10.1007/s00440-015-0673-1. arXiv: 1305.1233.

[EL94] K. D. Elworthy and X.-M. Li. "Formulae for the derivatives of heat semigroups". In: *Journal of Functional Analysis* 125.1 (1994), pp. 252–286. DOI: 10.1006/jfan.1994.1124.

[FGL13] X. Fan, I. Grama, and Q. Liu. "Cramér large deviation expansions for martingales under Bernstein's condition". In: *Stochastic Processes and their Applications* 123.11 (2013), pp. 3919–3942. DOI: 10.1016/j.spa.2013.06.010. arXiv: 1210.2198.

[FHX24] X. Fan, H. Hu, and L. Xu. "Normalized and self-normalized Cramér-type moderate deviations for the Euler–Maruyama scheme for the SDE". In: *Science China Mathematics* 67.8 (2024), pp. 1865–1880. DOI: 10.1007/s11425-022-2161-4. arXiv: 2209.06475.

[Hai26] M. Hairer. *Advanced Stochastic Analysis*. Lecture notes, Sections 5.1–5.2. May 19, 2026. URL: https://www.hairer.org/notes/StochasticAnalysisCourse.pdf.

[LTX22] J. Lu, Y. Tan, and L. Xu. "Central limit theorem and self-normalized Cramér-type moderate deviation for Euler–Maruyama scheme". In: *Bernoulli* 28.2 (2022), pp. 937–964. DOI: 10.3150/21-bej1372. arXiv: 2012.04328.

[Sun26] Z. Sun. "Moderate deviations for stochastic gradient Langevin averages with a computed Poisson normalizer". Manuscript, Hexagon identifier 2611.01077. 2026.