hexagon:2611.01077v1

Moderate deviations for stochastic gradient Langevin averages with a computed Poisson normalizer

Zhenyu Sun

November 24, 2026

Abstract

We establish a relative Gaussian tail approximation for time averages of stochastic gradient Langevin dynamics centered at the invariant mean of an associated modified diffusion. Self-normalization is achieved via the trajectory-averaged squared gradient of the corresponding Poisson solution. Under weak dissipativity, conditionally sub-Gaussian gradient noise, and uniform derivative bounds on the drift, covariance factor, and observable, the approximation holds uniformly up to a growing radius $r \to \infty$ whenever

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

where η denotes the step size and $T = m\eta$ is the observation horizon. Here $\gamma = 1$ for the uncorrected empirical average and $\gamma = 2$ after removing the leading weak discretization bias. The proof controls the predictable quadratic variation under an exponential tilt of the full centered one-step Poisson increment. Given coefficient evaluations together with known growth and derivative bounds and an admissible positive dissipativity constant, we further construct finite-budget simulation procedures for both the normalizer and the weak-bias correction. To preserve relative Gaussian tail accuracy for the resulting confidence intervals, it suffices to control the relative error of the empirical energy to $o(r^{-2})$ and the additive error of the corrected average to $o((r\sqrt{T})^{-1})$, with combined failure probability $o(1-\Phi(r))$. The gradient noise is permitted to be asymmetric with a linearly growing covariance factor; in a linear Gaussian model, the excess variance induced by this noise alters the limiting tail ratio at the boundary radius $r = c\eta^{-1/2}$ whenever $\eta T \to \infty$.

Contents

1 Introduction **2**

2 Model and main results **4**
2.1 Dynamics and assumptions 4
2.2 The Poisson normalizer and the first correction 4
2.3 Joint moderate deviations 5

3 Weak expansion, occupation and conditional variance **6**
3.1 Local moment and weak-error bounds 7
3.2 Exponential localization and occupation 8
3.3 The full centered conditional increment 9

4 Exponential tilting of the full increment **10**
4.1 The conditional cumulant estimate . . . 10
4.2 The tilted transition kernel . . . 12
4.3 A bounded-increment normal approximation . . . 15
4.4 The tilted integral at the self-normalized threshold . . . 15
4.5 The martingale moderate-deviation window . . . 18

5 Transfer to the empirical statistic **19**
5.1 A relative-tail perturbation lemma . . . 19
5.2 Proof of the main trajectory theorem . . . 20

6 Certified computation of the normalizer and correction **22**
6.1 Explicit finite-time simulation bounds . . . 22
6.2 Estimating the Poisson gradient . . . 23
6.3 Relative error in the empirical energy . . . 25
6.4 Estimating the weak-bias correction . . . 26
6.5 Tail probabilities and confidence intervals . . . 27

7 Examples and the variance boundary **29**
7.1 A nonconvex model with asymmetric gradient noise . . . 29
7.2 A variance obstruction for the original denominator . . . 30

A Effective contraction and global Poisson estimates **32**
A.1 A common effective contraction rate . . . 32
A.2 An integrable semigroup smoothing estimate . . . 34
A.3 Canonical Poisson solutions and explicit derivative bounds . . . 36
A.4 Energy nondegeneracy and global cutoffs . . . 37

B A Gaussian-mixture check of the local coefficients **38**

1 Introduction

Consider an m-step trajectory of stochastic gradient Langevin dynamics (SGLD) with step size η, whose empirical average is used to estimate the invariant mean of an associated diffusion sharing the same drift. At each state, the instantaneous diffusion covariance coincides with the conditional covariance of a single SGLD increment divided by η. Solving the Poisson equation for this modified diffusion provides a canonical self-normalizer for the empirical average, whose evaluation requires the gradient of the Poisson solution along the observed trajectory. A fundamental question is how accurately these gradients must be approximated so as not to distort a tail probability that decays to zero.

The requisite numerical precision is governed by the tail level. Writing $\overline{\Phi}(x) = 1 - \Phi(x)$ for the upper standard normal tail, a relative error β in the squared normalizer perturbs the argument of $\overline{\Phi}$ by order $x\beta$. At a growing radius r, this induces a relative tail error of order $r^2\beta$ whenever $r^2\beta$ is small. Moreover, a numerical failure event whose probability is comparable to $\overline{\Phi}(r)$ can corrupt the leading-order tail asymptotics. We construct finite-time estimators with explicit budgets that control both sources of error simultaneously.

For the uncorrected self-normalized average, Theorem 2.2 establishes uniform relative Gaussian tail accuracy up to $r = o(\eta^{-1/3}[\log(1/\eta)]^{-1/6})$ along the first observation schedule in Corollary 2.3. Removing the leading weak discretization bias extends this range to $r = o(\eta^{-1/2})$ on the longer second schedule. Both results hold for any fixed dimension and positive additive noise intensity. In gradient models, the mean potential may be nonconvex and the centered, conditionally sub-Gaussian gradient noise may be asymmetric. Its covariance admits a smooth factor of at most linear growth under Assumption 2.1. Individual sampled gradients need only be measurable in the state.

Dai, Fan, and Lu [DFL26] established self-normalized Cramér-type moderate deviations for SGLD by solving the Stein equation of a modified diffusion, and in Remark 3 of their published paper posed the question of how numerical approximation of the Stein derivative affects the resulting statistic. Poisson decompositions have also been employed in the Euler–Maruyama moderate-deviation analyses of Lu, Tan, and Xu [LTX22] and Fan, Hu, and Xu [FHX24]. For an additive, state-independent diffusion matrix, Fan, Hu, and Xu [FHX24] obtain relative Gaussian tail approximations up to $o(\eta^{-1/4})$ at $m = \lfloor \eta^{-2} \rfloor$. At that same observation budget, our sufficient conditions yield $r = o((\eta \log(1/\eta))^{-1/4})$, whereas the larger radii achieved here exploit longer horizons and higher-order smoothness.

To prove the main trajectory approximation, we center the full one-step increment of the Poisson solution and apply an exponential tilt to the resulting martingale. Under a linear tilt, the additive Gaussian component remains Gaussian, so its third cumulant vanishes identically. Incorporating the gradient noise and nonlinear remainder terms yields an absolute third cumulant at state x bounded by $C\eta^2(1 + |x|^2)$ prior to exit from a localization ball, and a quadratic occupation bound controls the cumulative effect of these errors. Because the tilt modifies the transition kernel, we estimate the predictable variance directly under the tilted law. At the self-normalized threshold, completing the square isolates a positive quadratic term in the variance fluctuation, whose tail integral under the tilted measure controls the likelihood ratio; near the threshold itself, a martingale normal approximation due to Fan, Grama, and Liu [FGL13, Theorem 3.1] completes the bound.

For the numerical normalizer, we simulate coupled pairs of Euler paths for the modified diffusion started from nearby perturbations using common Brownian increments, estimating the Poisson gradient via central finite differences of their integrated observables. For the weak-bias correction, we form higher-order forward differences of Poisson values anchored at the origin. Conditional on the observed SGLD trajectory, median-of-means estimators deliver simultaneous error bounds across all visited states. These bounds yield a certified relative-error interval for the empirical energy whenever its lower endpoint is positive and the interval is sufficiently narrow; on the remaining paths, the procedure returns a flagged positive fallback value, and the probability of invoking this fallback is negligible compared with $\overline{\Phi}(r)$ under the hypotheses of Theorem 6.5.

The remainder of the paper is organized as follows. Section 2 introduces the SGLD model, standing assumptions, and main asymptotic results. Sections 3–5 develop the local expansions, exponential tilting estimates, and proof of the trajectory theorem. Section 6 details the certified simulation algorithms and confidence intervals. Section 7 illustrates the theory on a nonconvex potential with asymmetric noise and characterizes the exact limiting tail ratio at the variance boundary in a linear Gaussian model. Appendix A establishes the global contraction and Poisson derivative bounds used in both the probabilistic and numerical analyses, and Appendix B checks the local expansion coefficients for an explicit Gaussian mixture.

2 Model and main results

2.1 Dynamics and assumptions

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

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

Here $(\xi_k)_{k\geq 1}$ is a sequence of independent standard d-dimensional Gaussian vectors, $(\zeta_k)_{k\geq 1}$ is an independent and identically distributed sequence independent of the entire Gaussian sequence, and Z is jointly measurable. We write P_η for the Markov transition kernel of (1). Throughout, vector norms are Euclidean and derivative norms are multilinear operator norms, equipped with the Hilbert–Schmidt norm on the codomain for matrix-valued derivatives.

Assumption 2.1. *Throughout the asymptotic regime $\eta \downarrow 0$, the following objects and constants remain fixed:*

(i) The drift $b : \mathbb{R}^d \to \mathbb{R}^d$ belongs to C^6 and satisfies

$$\|D^j b\|_\infty \leq B_j, \quad 1 \leq j \leq 6, \qquad \langle b(x) - b(y), x - y\rangle \leq -K_1|x-y|^2 + K_2$$

for some constants $K_1 > 0$ and $K_2 \geq 0$.

(ii) The centered gradient noise satisfies, for all $x, t \in \mathbb{R}^d$,

$$\mathbb{E}Z(x,\zeta) = 0, \qquad \mathbb{E}e^{\langle t, Z(x,\zeta)\rangle} \leq \exp\{K^2(1+|x|^2)|t|^2/2\}. \tag{2}$$

(iii) Its covariance $\Sigma(x) = \mathbb{E}[Z(x,\zeta)Z(x,\zeta)^\top]$ admits a factorization $\Sigma = \sigma_0\sigma_0^\top$, where $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^{d\times\ell}$ belongs to C^6 and satisfies $\|D^j\sigma_0\|_\infty \leq S_j$ for $1 \leq j \leq 6$.

(iv) The observable $h : \mathbb{R}^d \to \mathbb{R}$ is nonconstant, belongs to C^5, and satisfies $\|D^j h\|_\infty \leq H_j$ for $1 \leq j \leq 5$.

Under Assumption 2.1, the functions b, σ_0, and h may grow linearly at infinity. No symmetry is imposed on the law of Z, nor is spatial differentiability required of individual sample realizations $x \mapsto Z(x,\zeta)$. We fix a small upper step-size threshold $\eta_0 > 0$ satisfying $\eta_0 S_1^2 \leq K_1$, which may be decreased finitely many times to accommodate the explicit small-step conditions in the proofs.

For the numerical procedures in Section 6, we assume oracle access to point evaluations of b, σ_0, and h, together with a known admissible dissipativity constant $K_1 > 0$ (any known positive lower bound on K_1 may be used in its place) and known upper envelopes for $K_2, K, B_j, S_j, H_j, |b(0)|$, and $\|\sigma_0(0)\|_{\mathrm{HS}}$. Computational complexity is measured in terms of real arithmetic operations, coefficient evaluations, and independent standard Gaussian draws. The target parameter of inference is the invariant mean θ_η defined below.

2.2 The Poisson normalizer and the first correction

Let $\mathcal{P}_t^\eta$ denote the Markov semigroup associated with the modified diffusion

$$d\mathsf{X}_t = b(\mathsf{X}_t)\,dt + \sqrt{\delta}\,dB_t + \sqrt{\eta}\,\sigma_0(\mathsf{X}_t)\,dW_t,$$

whose invariant probability measure is denoted by π_η and whose infinitesimal generator is given by

$$L_\eta = b \cdot \nabla + \frac{1}{2} A_\eta : D^2, \qquad A_\eta = \delta I + \eta \Sigma. \tag{3}$$

The empirical statistics below are centered at the stationary mean $\theta_\eta = \pi_\eta h$ under this modified diffusion law.

Appendix A establishes the existence and uniqueness of π_η, uniform Wasserstein contraction of $\mathcal{P}_t^\eta$, and the well-posedness of the anchored, globally Lipschitz Poisson solutions

$$\begin{aligned} &L_\eta f_\eta = h - \theta_\eta, \quad f_\eta(0) = 0, \quad \theta_\eta = \pi_\eta h, \\ &L_\eta u_\eta = w_\eta - \mu_\eta, \quad u_\eta(0) = 0, \quad w_\eta = |\nabla f_\eta|^2, \quad \mu_\eta = \pi_\eta w_\eta. \end{aligned} \tag{4}$$

Their derivatives of orders one through six are bounded uniformly in $\eta \in [0, \eta_0]$. Moreover, for each fixed model, $\mu_\eta \to \mu_0 > 0$ as $\eta \downarrow 0$, although our numerical procedures require no explicit positive lower bound on μ_η.

Define the frozen-coefficient second-order weak-error operator

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

The term involving $\eta\Sigma$ is already included in L_η. The operator $\mathcal{A}_2$ gives the coefficient of η^2 in the frozen one-step expansion after subtracting $F + \eta L_\eta F$.

For any measurable function $v : \mathbb{R}^d \to \mathbb{R}$, write $\bar{v}_m = m^{-1} \sum_{k<m} v(X_k)$ for its empirical average, and define

$$Y_m = \bar{w}_{\eta,m}, \qquad W_{m,\eta}^{(j)} = \frac{\sqrt{T}\{\bar{h}_m + j\eta \bar{a}_{\eta,m} - \theta_\eta\}}{\sqrt{\delta Y_m}}, \qquad j \in \{0, 1\}, \tag{6}$$

with the convention $W_{m,\eta}^{(j)} = 0$ whenever $Y_m = 0$. The index $j = 0$ corresponds to the uncorrected self-normalized average, whereas $j = 1$ incorporates the explicit first-order weak-bias correction; the positive sign on $j\eta\bar{a}_{\eta,m}$ reflects our sign convention $L_\eta f_\eta = h - \theta_\eta$.

2.3 Joint moderate deviations

Theorem 2.2 (Relative Gaussian tails)**.** *Suppose Assumption 2.1 holds, fix $j \in \{0, 1\}$, and set $\gamma = j + 1$. If*

$$\eta \downarrow 0, \quad T = m\eta \to \infty, \quad r \to \infty, \qquad r^2 \sqrt{\frac{\log(T + e)}{T}} \to 0, \quad r^2 \eta \to 0, \quad r\eta^\gamma \sqrt{T} \to 0, \tag{7}$$

then, for both signs $s \in \{-1, 1\}$,

$$\sup_{0 \le x \le r} \left| \frac{\mathbb{P}(sW_{m,\eta}^{(j)} > x)}{\overline{\Phi}(x)} - 1 \right| \longrightarrow 0. \tag{8}$$

The same limit holds when the strict inequality is replaced by a weak inequality.

The clock condition in (7) is equivalent to $r^4 \log(T+e)/T \to 0$ and emerges from the exponential tilt of the full one-step Poisson increment. The second and third conditions in (7) respectively control the discrepancy between the predictable quadratic variation of the martingale and $\delta T Y_m$, and the residual weak discretization bias. Implicit constants may depend on the fixed model parameters as well as d, ℓ, δ, x_0, but are independent of m, η, r.

Corollary 2.3 (Explicit observation schedules)**.** *Write $L = \log(1/\eta)$. For the uncorrected statistic ($j = 0$), Theorem 2.2 holds along the schedule*

$$T \asymp \eta^{-4/3} L^{1/3}, \qquad m \asymp \eta^{-7/3} L^{1/3}, \qquad r = o(\eta^{-1/3} L^{-1/6}).$$

For the bias-corrected statistic ($j = 1$), Theorem 2.2 holds along the longer schedule

$$T \asymp \eta^{-2} L, \qquad m \asymp \eta^{-3} L, \qquad r = o(\eta^{-1/2}).$$

Proof. Direct substitution verifies (7). Under the first schedule, both the clock and weak-bias conditions reduce to $r\eta^{1/3} L^{1/6} \to 0$ upon taking the square root of the clock condition. Under the second schedule, the clock and variance conditions both reduce to $r^2 \eta \to 0$, while the bias condition becomes $r\eta\sqrt{L} = o(\eta^{1/2}\sqrt{L}) \to 0$. □

Under power-law scalings $T \asymp \eta^{-a}$ and $r = \eta^{-c}$, the sufficient conditions for the uncorrected statistic require $c < \min\{a/4, 1 - a/2\}$; the two upper bounds coincide at $a = 4/3$. Their common value $1/3$ is the supremum of the exponents guaranteed by (7) for uncorrected averages on power-law schedules. Proposition 7.2 exhibits a variance obstruction at the corrected step-size radius $r \asymp \eta^{-1/2}$ on longer time scales when the leading-order denominator δY_m is retained without variance correction.

In Section 6, we construct a strictly positive numerical normalizer $\widehat{Y}_m$ and, when $j = 1$, numerical correction values $\widehat{a}_k$. Let β denote the target relative tolerance for the empirical energy, e_a and e_c denote absolute tolerances for the weak-bias correction and an optional numerical center, and p_g, p_a, p_c bound the respective failure probabilities of the gradient, correction, and center estimates. Theorem 6.5 guarantees that the statistic with numerical normalization satisfies (8) whenever

$$r^2 \beta \to 0, \qquad r\sqrt{T}(\eta e_a + e_c) \to 0, \qquad (p_g + p_a + p_c)/\overline{\Phi}(r) \to 0.$$

The construction requires no known positive lower bound on μ_η. The simulation guarantees hold conditionally on the observed trajectory. The transfer uses these guarantees and allows dependence between the resulting estimates and the empirical average.

3 Weak expansion, occupation and conditional variance

All constants in this section are uniform over $0 < \eta \le \eta_0$, after decreasing $\eta_0 > 0$ if necessary. For $R \ge 1 + |x_0|$, define the exit time, pre-exit indicator, and quadratic weight

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

We impose the local step-size restriction $\eta(1 + 9R^2) \le c_0$, where $c_0 > 0$ is a fixed, sufficiently small constant. Given a globally Lipschitz anchored Poisson solution F, we work with its smooth truncation $F_R = \chi_R F$ constructed in Lemma A.4, which agrees with F on B_{2R} and has uniformly bounded positive-order derivatives through all required orders. Every conditional expectation below integrates over the complete one-step transition law, including jumps that overshoot B_R or the support of χ_R. The arguments also apply directly to an uncut test function possessing the stated global derivative bounds, in which case its value at the origin is immaterial.

3.1 Local moment and weak-error bounds

At a fixed pre-step state $x \in \mathbb{R}^d$, set

$$y = x + \eta b(x), \quad W = \eta Z(x,\zeta) + \sqrt{\eta\delta}\,\xi, \quad \Delta = X_1 - x = \eta\{b(x) + Z(x,\zeta)\} + \sqrt{\eta\delta}\,\xi.$$

The centered innovation W has directional sub-Gaussian variance proxy

$$a_x = \eta\delta + \eta^2 K^2 q_x \le C\eta, \qquad |x| \le R.$$

Introducing an independent standard Gaussian vector G and applying auxiliary Gaussian integration yields

$$\mathbb{E}e^{|W|^2/(4a_x)} \le 2^{d/2}, \qquad \mathbb{E}e^{\theta|y+W|^2} \le (1-2\theta a_x)^{-d/2} \exp\left\{\frac{\theta|y|^2}{1-2\theta a_x}\right\}, \tag{9}$$

where the second bound holds for $0 \le 2\theta a_x < 1$; both follow from the identity $e^{\theta|z|^2} = \mathbb{E}_G e^{\sqrt{2\theta}\langle G,z\rangle}$ combined with (2). In particular, for every fixed $p \ge 2$,

$$\mathbb{E}|b(x) + Z(x,\zeta)|^p \le C_p q_x^{p/2}, \qquad \mathbb{E}|\Delta|^p \le C_p \eta^{p/2}, \qquad |x| \le R. \tag{10}$$

Lemma 3.1 (Weak expansion with Gaussian cancellation)**.** *Let $F(0) = 0$ have uniformly bounded derivatives of orders one through six. Then, for its cutoff F_R, at every $|x| \le R$,*

$$\begin{aligned} P_\eta F_R(x) - F_R(x) &= \eta L_\eta F_R(x) + \eta^2 \mathcal{A}_2 F_R(x) + \mathcal{R}_\eta^F(x), \\ |\mathcal{R}_\eta^F(x)| &\le C\eta^3 q_x^2, \qquad |\mathcal{A}_2 F_R(x)| \le C q_x. \end{aligned} \tag{11}$$

Consequently, the uncorrected weak defect is bounded by $C\eta^2 q_x$, and this latter bound remains valid assuming only four bounded positive-order derivatives.

Proof. Set $B = b(x) + Z(x,\zeta)$ and expand $F_R(x+\Delta)$ around x through degree five. By (10), the sixth-order Taylor remainder has expectation at most $C\mathbb{E}|\Delta|^6 \le C\eta^3$. For the first two degrees,

$$\mathbb{E}\Delta = \eta b, \qquad \mathbb{E}\Delta^{\otimes 2} = \eta\delta I + \eta^2\Sigma + \eta^2 b^{\otimes 2},$$

which generate $\eta L_\eta F_R$ and $\eta^2 D^2 F_R[b,b]/2$. Independence and centering give

$$\mathbb{E}\Delta^{\otimes 3} = \eta^2\delta\ \mathrm{Sym}(b \otimes I) + \eta^3 \mathbb{E}B^{\otimes 3},$$

where Sym denotes summation over the three tensor positions. The first term produces the third-order component of $\mathcal{A}_2 F_R$, while the second contributes at most $C\eta^3 q_x^{3/2}$. At degree four, the pure Gaussian moment supplies the final term in (5), the mixed Gaussian–B terms contribute $O(\eta^3 q_x)$, and the pure B term is $O(\eta^4 q_x^2)$. At degree five, the pure Gaussian moment vanishes by symmetry, and the remaining cross terms are bounded by

$$C\{\eta^3 q_x^{1/2} + \eta^4 q_x^{3/2} + \eta^5 q_x^{5/2}\}.$$

The local condition $\eta q_x \le c_0$ absorbs the excess powers of ηq_x, establishing (11). The bound $|\mathcal{A}_2 F_R(x)| \le C q_x$ follows directly from the linear growth of b. Expanding through degree three with a fourth-order remainder yields the uncorrected bound directly. Nowhere in the proof is symmetry of Z required. □

3.2 Exponential localization and occupation

Lemma 3.2 (Stopped quadratic and quartic occupation)**.** *There exist fixed constants $\theta, c, C > 0$ such that*

$$\mathbb{P}(\tau_R \le m) \le (e^{\theta|x_0|^2} + CT)e^{-\theta R^2}, \tag{12}$$

$$\mathbb{P}\left\{\eta \sum_{k<m} I_k q_{X_k} > C(T + t + 1)\right\} \le e^{-t}, \qquad t \ge 1, \tag{13}$$

$$\mathbb{P}\left\{\eta \sum_{k<m} I_k q^2_{X_k} > C(T + R^2 t + 1)\right\} \le e^{-t}, \qquad t \ge 1. \tag{14}$$

Proof. Applying the dissipativity condition in Assumption 2.1(i) at $y = 0$ together with Young's inequality gives

$$x \cdot b(x) \le -\kappa|x|^2 + A, \qquad \kappa = K_1/2, \quad A = K_2 + |b(0)|^2/(2K_1).$$

Consequently, for sufficiently small $\eta_0 > 0$,

$$|x + \eta b(x)|^2 \le (1 - \kappa\eta)|x|^2 + C\eta. \tag{15}$$

Combining (9), $a_x \le C\eta$, and a sufficiently small fixed $\theta > 0$ yields

$$\mathbb{E}_x e^{\theta|X_1|^2} \le e^{\theta|x|^2} \exp\{\eta(C - c|x|^2)\}, \qquad |x| \le R. \tag{16}$$

Explicitly, choosing a local proxy bound A_N with $a_x \le \eta A_N$ and taking $2\theta A_N \le \kappa/2$ and $2\theta A_N \eta_0 \le 1/2$, the elementary inequality $-\log(1 - u) \le 2u$ yields fixed constants $c, C > 0$.

The right-hand side of (16) is bounded above by $e^{\theta|x|^2} + C\eta$: outside a fixed compact ball the exponential factor is at most one, while inside that ball its excess is $O(\eta)$. Freezing the chain at τ_R (while retaining its exit state) and iterating gives $\mathbb{E}e^{\theta|X_{m\wedge\tau_R}|^2} \le e^{\theta|x_0|^2} + CT$, which implies (12). Moreover, multiplying (16) along the predictably stopped trajectory produces a supermartingale with exponent

$$\theta|X_{m\wedge\tau_R}|^2 - \theta|x_0|^2 + c\eta \sum_{k<m} I_k|X_k|^2 - C\eta \sum_{k<m} I_k.$$

Discarding the nonnegative terminal term and applying Markov's inequality yields (13).

For the quartic occupation estimate, fix a smooth, increasing, convex function $\phi : [0, \infty) \to [0, \infty)$ satisfying $\phi(z) = z^2$ for $z \le 9$, affine for $z \ge 16$, and having bounded first and second derivatives. Define

$$\psi_R(x) = R^2\phi(q_x/R^2),$$

so that $\psi_R(x) = q_x^2/R^2$ on B_{2R} and ψ_R grows only quadratically outside B_{4R}. Its Hessian is positive semidefinite and satisfies

$$\|D^2\psi_R\|_\infty \le C, \qquad |D^2\psi_R(z)| \le Cq_z/R^2.$$

For sufficiently small η_0, linear growth of b guarantees that $y = x + \eta b(x) \in B_{2R}$ whenever $|x| \le R$, so the entire line segment joining x and y lies within the region where $\psi_R = q^2/R^2$. A deterministic Taylor expansion together with the radial drift inequality gives

$$\psi_R(y) - \psi_R(x) \le \eta R^{-2}(C - cq_x^2),$$

since the first-order term is $4\eta q_x(x \cdot b(x))/R^2$ and the second-order remainder is at most $C\eta^2 q_x^2/R^2$, which is absorbed by decreasing η_0.

Next, decompose

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

Convexity and the Hessian bounds imply

$$0 \le Q \le C|W|^2, \qquad Q \le CR^{-2}(q_y|W|^2 + |W|^4), \qquad |\nabla\psi_R(y)|^2 \le Cq_x^3/R^4.$$

Applying the Cauchy–Schwarz inequality, the directional sub-Gaussian bound for ℓ, and $e^{2\theta Q} - 1 \le 2\theta Q e^{2\theta C|W|^2}$, we obtain

$$\log \mathbb{E} e^{\theta(\ell+Q)} \le C\theta^2\eta q_x^3/R^4 + C\theta\eta q_x/R^2,$$

where the exponentially weighted second and fourth moments of W are bounded by $C\eta$ and $C\eta^2$ via (9). Since $q_x/R^2 \le 2$ on the pre-exit ball, choosing $\theta > 0$ sufficiently small absorbs these terms into the negative quadratic drift term, yielding

$$\mathbb{E}_x e^{\theta\psi_R(X_1)} \le e^{\theta\psi_R(x)} \exp\{\eta R^{-2}(C - cq_x^2)\}.$$

The associated stopped supermartingale establishes (14), with initial contribution $R^2\psi_R(x_0) = q_{x_0}^2$; the quadratic growth of ψ_R at infinity ensures integrability of the one-step overshoot. □

For later reference, if $s = d + r^2 + \log(T+e)$, $R^2 = C_J s$, and $s/T \to 0$, then choosing C_J sufficiently large ensures that the exit and quadratic-occupation failure probabilities are each bounded by $C_J(T+e)^{-J}e^{-Jr^2}$. Setting the quartic confidence level t to a sufficiently large multiple of s likewise gives, on the same probability scale,

$$\eta \sum_{k<m} q_{X_k}^2 \le C_J(T + s^2)$$

whenever $\tau_R > m$.

3.3 The full centered conditional increment

Lemma 3.3 (Conditional variance and its first correction)**.** *Let $F(0) = 0$ have uniformly bounded derivatives of orders one through three. With $d^F = F_R(X_1) - P_\eta F_R(x)$ and $|x| \le R$,*

$$\left|\operatorname{Var}_x F_R(X_1) - \eta\delta|\nabla F_R(x)|^2\right| \le C\eta^2 q_x, \qquad \mathbb{E}_x e^{c(d^F)^2/\eta} \le C. \tag{17}$$

If F has uniformly bounded derivatives through order six, then uniformly on every fixed compact subset of the interior of B_{2R}, with $g = \nabla F$,

$$\eta^{-1}\operatorname{Var}_x F_R(X_1) = \delta|g|^2 + \eta v_1^F(x) + O(\eta^2), \tag{18}$$

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

$$\mathbb{E}_x (d^F)^3 = 3\eta^2\delta^2 D^2F[g,g] + O(\eta^3). \tag{19}$$

Proof. Global Lipschitz continuity of F_R implies $|F_R(y+W)-F_R(y)| \le C|W|$. Combining the square-exponential bound in (9) with centering yields the second inequality in (17).

Next, expanding $F_R(x+\Delta)$ through degree two and centering gives

$$d^F = L + \eta g\cdot Z + (Q_2 - \mathbb{E}Q_2) + (R_3 - \mathbb{E}R_3), \quad L = \sqrt{\eta\delta}\, g\cdot\xi, \quad Q_2 = \tfrac{1}{2}D^2F_R(x)[\Delta,\Delta],$$

where $\mathbb{E}R_3^2 \le C\eta^3$. By independence and odd Gaussian moment cancellation,

$$\mathrm{Cov}(L, \eta g\cdot Z) = 0, \qquad \mathrm{Cov}(L, Q_2) = \eta^2\delta D^2F_R(x)[b,g].$$

Furthermore, $\mathrm{Var}\, Q_2 \le C\eta^2$, $\mathrm{Var}(\eta g\cdot Z) \le C\eta^2 q_x$, and all remaining cross covariances are bounded in magnitude by $C\eta^2 q_x$, establishing the first estimate in (17).

For the higher-order expansions (18)–(19), put $z = \sqrt{\eta}$ and $u = \sqrt{\delta}\,\xi$. Collecting terms through degree four in z in the centered Taylor expansion gives $zA_1 + z^2A_2 + z^3A_3 + z^4A_4$, where A_1, A_3 are odd in u, A_2, A_4 are even in u, and the L^3 remainder is $O(z^5)$ uniformly on compact sets. Consequently, all apparent odd powers z^3 and z^5 vanish upon taking expectations in the variance and third-moment expansions. In the coefficient of η^2 in the variance, the centered quadratic Gaussian term contributes $\delta^2\|D^2F\|_{\mathrm{HS}}^2/2$, the drift–Gaussian cross term contributes $2\delta D^2F[b,g]$, twice the linear–cubic Gaussian covariance contributes $\delta^2\sum_i D^3F[g,e_i,e_i]$, and the gradient noise contributes $g^\top\Sigma g$. For the third centered moment, the coefficient of η^2 is $3\mathbb{E}(A_1^2A_2) = 3\delta^2D^2F[g,g]$. The moment bounds in (10) control all terms of order η^3 and higher; in particular, any skewness of the gradient innovation Z enters only at order η^3. □

4 Exponential tilting of the full increment

Throughout this section, P_η denotes the transition kernel of the SGLD chain (1), and the estimates apply to either choice $F_R = \chi_R f_\eta$ or $F_R = -\chi_R f_\eta$. Recall that the cutoff χ_R from Lemma A.4 equals one on B_{2R}, vanishes outside B_{3R}, and ensures uniform bounds on D^jF_R for $1 \le j \le 6$ across $R \ge 1$. Set

$$g_R = \nabla F_R, \qquad w_\eta = |\nabla f_\eta|^2, \qquad \tau_R = \inf\{k \ge 0 : |X_k| > R\}, \qquad I_k = \mathbf{1}_{\{k<\tau_R\}}.$$

All constants below are independent of m, η, R and of the tilt parameter λ within its stated range, depending only on the fixed model data in Assumption 2.1. Every conditional expectation is taken with respect to the full one-step transition law, including steps that exit the localization ball.

4.1 The conditional cumulant estimate

Lemma 4.1 (Cumulants of the full centered increment)**.** *Let $F \in C^2(\mathbb{R}^d)$ have uniformly bounded gradient and Hessian. At a fixed state $x \in \mathbb{R}^d$, define*

$$d_x = F(X_1) - P_\eta F(x), \qquad \psi_x(\lambda) = \log \mathbb{E}_x e^{\lambda d_x}, \qquad v_x = \mathbb{E}_x d_x^2, \qquad q_x = 1 + |x|^2.$$

For any fixed $\lambda_0, c_0 > 0$, there exist sufficiently small $\eta_0 > 0$ and a constant $C < \infty$ such that

$$\sup_{|\lambda|\le\lambda_0} |\psi_x'''(\lambda)| \le C\eta^2 q_x, \qquad 0 < \eta \le \eta_0, \quad \eta q_x \le c_0. \tag{20}$$

Consequently,

$$|\psi_x(\lambda) - \lambda^2 v_x/2| \le C|\lambda|^3\eta^2 q_x,$$
$$|\psi_x'(\lambda) - \lambda v_x| \le C\lambda^2\eta^2 q_x, \qquad |\psi_x''(\lambda) - v_x| \le C|\lambda|\eta^2 q_x. \tag{21}$$

The constants require no spatial derivatives of individual realizations $x \mapsto Z(x,\zeta)$.

Proof. Set $y = x + \eta b(x)$ and

$$W = \sqrt{\eta\delta}\,\xi + \eta Z(x,\zeta), \qquad a = \eta\delta + \eta^2 K^2 q_x \le C\eta.$$

The directional sub-Gaussian bound (2) and independence yield $\mathbb{E}e^{t\cdot W} \le e^{a|t|^2/2}$, which upon integration against an independent Gaussian vector gives

$$\mathbb{E}e^{|W|^2/(4a)} \le 2^{d/2}. \tag{22}$$

Expanding around the drifted point y, write

$$F(y+W) - F(y) = L + Q, \qquad L = u\cdot W, \quad u = \nabla F(y), \quad |Q| \le \tfrac{1}{2}\|D^2F\|_\infty |W|^2.$$

We first tilt by the linear part $e^{\lambda L}/\mathbb{E}e^{\lambda L}$ and denote this auxiliary probability measure by $\widetilde{\mathbb{Q}}_\lambda$. Under $\widetilde{\mathbb{Q}}_\lambda$, the innovations ξ and ζ remain independent, and the Gaussian component undergoes a pure mean shift; hence its third cumulant vanishes and

$$\operatorname{Cum}_{3,\widetilde{\mathbb{Q}}_\lambda}(L) = \eta^3 \operatorname{Cum}_{3,\widetilde{\mathbb{Q}}_\lambda}(u\cdot Z).$$

The scalar variable $u\cdot Z$ has sub-Gaussian variance proxy Cq_x and tilt parameter $\lambda\eta$. Because its uncentered normalizing constant $\mathbb{E}e^{\lambda\eta u\cdot Z} \ge 1$ by Jensen's inequality, Cauchy–Schwarz and the sixth-moment bound in (10) give

$$\mathbb{E}_{\widetilde{\mathbb{Q}}_\lambda}|u\cdot Z|^3 \le Cq_x^{3/2} e^{C\lambda^2\eta^2 q_x} \le Cq_x^{3/2},$$

so that $|\operatorname{Cum}_{3,\widetilde{\mathbb{Q}}_\lambda}(L)| \le C\eta^3 q_x^{3/2}$.

Moreover, for any fixed $c \ge 0$ and $0 \le p \le 6$, the linear-tilted measure satisfies

$$\mathbb{E}_{\widetilde{\mathbb{Q}}_\lambda}[|W|^p e^{c|W|^2}] \le C_p a^{p/2}. \tag{23}$$

Indeed, its normalizing constant is at least one, while $\lambda u\cdot W \le |W|^2/(16a) + 4a\lambda_0^2|u|^2$; for small η_0, absorbing $c|W|^2$ and the polynomial factor into (22) yields (23).

The full tilt of $F(X_1)$ reweights $\widetilde{\mathbb{Q}}_\lambda$ by $e^{\lambda Q}$; write $\mathbb{E}_\lambda$ for expectation under this law. Its relative normalizing constant $\mathbb{E}_{\widetilde{\mathbb{Q}}_\lambda} e^{\lambda Q}$ equals $1 + O(|\lambda|a)$ and is bounded below by $e^{-C\lambda_0 a}$. Setting $A = L - \mathbb{E}_{\widetilde{\mathbb{Q}}_\lambda} L$, reweighting gives, for $j = 1,2,3$,

$$\left|\mathbb{E}_\lambda A^j - \mathbb{E}_{\widetilde{\mathbb{Q}}_\lambda} A^j\right| \le C|\lambda| a^{j/2+1},$$

which follows by subtracting $\mathbb{E}_{\widetilde{\mathbb{Q}}_\lambda} A^j$ in the numerator and applying $|e^{\lambda Q} - 1| \le |\lambda Q| e^{|\lambda Q|}$ together with (23). The resulting perturbation to the centered third moment of L is therefore $O(|\lambda| a^{5/2})$. Under the full tilt,

$$\|L - \mathbb{E}_\lambda L\|_3 \le Ca^{1/2}, \qquad \|Q - \mathbb{E}_\lambda Q\|_3 \le Ca,$$

so the three remaining cross and cubic terms in the expansion of $\mathbb{E}_\lambda[(L + Q - \mathbb{E}_\lambda(L+Q))^3]$ are bounded by Ca^2, $Ca^{5/2}$, and Ca^3. Combining these estimates yields

$$|\operatorname{Cum}_{3,\lambda}(L+Q)| \le C\{a^2 + \eta^3 q_x^{3/2}\} \le C\eta^2 q_x,$$

where the final step uses $a \le C\eta$, $q_x \ge 1$, and $\eta\sqrt{q_x} \le \sqrt{c_0\eta}$. Global Lipschitz continuity and (22) justify differentiating $\psi_x(\lambda)$ under the expectation on compact tilt intervals, and identifying $\psi_x'''(\lambda) = \operatorname{Cum}_{3,\lambda}(L+Q)$ followed by integration from $\lambda = 0$ establishes (21). □

4.2 The tilted transition kernel

The one-step exponential tilt associated with $F_R(X_1) - P_\eta F_R(x)$ is the Markov kernel

$$P_\eta^{\lambda,R}(x, dy) = \frac{e^{\lambda F_R(y)} P_\eta(x, dy)}{P_\eta e^{\lambda F_R}(x)}. \tag{24}$$

We write $\mathbb{P}_{\lambda,R}$ for the path measure generated by (24) from x_0, and $\mathbb{E}_x^\lambda$ for its one-step expectation starting at x, suppressing R in the notation.

Lemma 4.2 (Diffusion associated with the tilted kernel)**.** *There exist fixed positive constants c_0, λ_0, η_0 such that whenever $\eta(1 + 9R^2) \le c_0$ and $|\lambda| \le \lambda_0$, the diffusion with covariance $A_\eta = \delta I + \eta\Sigma$ and tilted drift*

$$b_{\lambda,R} = b + \lambda\{\delta g_R + \eta\chi_R\Sigma g_R\} \tag{25}$$

satisfies uniform Wasserstein contraction and second-moment bounds. Let $\pi_{\lambda,R}$ denote its unique invariant law and set $\mu_{\lambda,R} = \pi_{\lambda,R} w_\eta$. Then

$$0 < \mu_*/2 \le \mu_{\lambda,R} \le C, \qquad |\mu_{\lambda,R} - \mu_{\lambda',R}| \le C|\lambda - \lambda'|.$$

Furthermore, the anchored Poisson solution to

$$L_{\lambda,R} u_{\lambda,R} = w_\eta - \mu_{\lambda,R}$$

has uniformly bounded derivatives of orders one through four, and the same bounds hold for its cutoff $U_R = \chi_R u_{\lambda,R}$.

Proof. Lemma A.2 (with $N = 6$) and Lemma A.4 yield uniform bounds on $D^j F_R$ for $1 \le j \le 6$. By the Leibniz rule, linear growth of σ_0, and the support and derivative properties of χ_R,

$$\|\delta g_R + \eta\chi_R\Sigma g_R\|_{C_b^4} \le C(1 + \eta R^2),$$

where the C_b^4 norm includes the supremum norm of the perturbation itself. Choosing $\lambda_0 > 0$ sufficiently small ensures that the Lipschitz constant of the drift perturbation is at most $K_1/2$, preserving pairwise weak dissipativity with constant $K_1/2$. Decreasing η_0 if necessary, Lemma A.1 yields the uniform contraction and moment bounds.

Because w_η has uniformly bounded derivatives through order five, applying Lemma A.2 (with $N = 4$) to the drift $b_{\lambda,R}$ and forcing w_η gives uniform derivative bounds through order four for $u_{\lambda,R}$, which transfer to U_R via Lemma A.4. For any two tilts $\lambda, \lambda' \in [-\lambda_0, \lambda_0]$, stationarity and the Poisson equation yield

$$\mu_{\lambda,R} - \mu_{\lambda',R} = -\pi_{\lambda,R}[(b_{\lambda,R} - b_{\lambda',R}) \cdot \nabla u_{\lambda',R}],$$

justified by stopped Itô's formula since $\nabla u_{\lambda',R}$ is bounded, the diffusion coefficient grows at most linearly, and the invariant second moments are uniformly bounded. The right-hand side is bounded in magnitude by $C|\lambda - \lambda'|$. Finally, at $\lambda = 0$, $\mu_{0,R} = \mu_\eta \ge \mu_*$ by Lemma A.3, so decreasing λ_0 once more secures the uniform lower bound $\mu_{\lambda,R} \ge \mu_*/2$. □

Lemma 4.3 (Local estimates for the quotient kernel)**.** *Under the hypotheses of Lemma 4.2, for every $|x| \le R$,*

$$\left|\mathbb{E}_x^\lambda[U_R(X_1) - U_R(x)] - \eta L_{\lambda,R} U_R(x)\right| \le C\eta^2(1+|x|^2), \tag{26}$$

$$\mathbb{E}_x^\lambda e^{\theta|X_1|^2} \le e^{\theta|x|^2} e^{\eta(C-c|x|^2)}, \tag{27}$$

$$\log \mathbb{E}_x^\lambda e^{z\{U_R(X_1) - \mathbb{E}_x^\lambda U_R(X_1)\}} \le Cz^2\eta, \qquad z \in \mathbb{R}, \tag{28}$$

$$\left|\mathrm{Var}_x^\lambda F_R(X_1) - \eta\delta w_\eta(x)\right| \le C\eta^2(1+|x|^2), \tag{29}$$

where $\theta, c > 0$ are fixed constants.

Proof. Set $y = x + \eta b(x)$, $X_1 = y + W$, and $a_x = \eta\delta + \eta^2 K^2(1+|x|^2) \le C\eta$. Write

$$F_R(y+W) - F_R(y) = a \cdot W + Q(W), \qquad a = \nabla F_R(y), \quad |Q(W)| \le C|W|^2.$$

Canceling $e^{\lambda F_R(y)}$ in (24), the normalizing constant is $Z_y = \mathbb{E}e^{\lambda(a\cdot W + Q(W))}$. Jensen's inequality gives $Z_y \ge e^{-C\eta}$, whereas auxiliary Gaussian integration using the directional sub-Gaussian bound gives $Z_y \le e^{C\eta}$. With $\Delta = X_1 - x$ and $V_x = F_R(x+\Delta) - F_R(x)$, since $|F_R(y) - F_R(x)| \le C\eta(1+|x|)$, it follows that

$$Z_x := \mathbb{E}e^{\lambda V_x} = 1 + O(\eta(1+|x|)), \qquad Z_x \ge 1/2. \tag{30}$$

To prove (26), expand

$$H_x(z) = [U_R(z) - U_R(x)]e^{\lambda[F_R(z) - F_R(x)]}$$

around x through degree three. At $z = x$,

$$DH_x = DU_R, \qquad D^2H_x = D^2U_R + 2\lambda\,\mathrm{sym}(DU_R \otimes g_R),$$

the third derivative is uniformly bounded, and the fourth derivative along the increment is bounded by $Ce^{C|\Delta|}(1+|\Delta|)$. Using the moment identities $\mathbb{E}\Delta = \eta b$, $\mathbb{E}\Delta^{\otimes 2} = \eta A_\eta + \eta^2 b^{\otimes 2}$, and

$$\mathbb{E}\Delta^{\otimes 3} = \eta^2\delta\,\mathrm{Sym}(b \otimes I) + \eta^3\mathbb{E}(b+Z)^{\otimes 3},$$

whose norm is bounded by $C\eta^2(1+|x|)$ since $\eta(1+|x|^2) \le c_0$, together with $\mathbb{E}[e^{C|\Delta|}|\Delta|^j] \le C_j\eta^{j/2}$, we obtain

$$\mathbb{E}H_x(X_1) = \eta\{L_\eta U_R + \lambda A_\eta g_R \cdot DU_R\}(x) + O(\eta^2(1+|x|^2)).$$

On B_R, the operator in braces coincides with $L_{\lambda,R}U_R(x)$. Dividing by (30) preserves the error order since $|L_{\lambda,R}U_R(x)| \le C(1+|x|)$, establishing (26).

For the Lyapunov bound (27), set $\gamma = \theta + C|\lambda|$. The lower bound on Z_y and auxiliary Gaussian integration imply

$$\begin{aligned}\log \mathbb{E}_x^\lambda e^{\theta|X_1|^2} &\le \theta|y|^2 + C\eta - \tfrac{d}{2}\log(1-2\gamma a_x) + \frac{a_x|2\theta y + \lambda a|^2}{2(1-2\gamma a_x)} \\ &\le \theta|y|^2 + C\theta^2\eta|y|^2 + C\eta,\end{aligned}$$

provided $2\gamma a_x \le 1/2$ for sufficiently small η_0. Combining this with $|y|^2 \le (1-\kappa\eta)|x|^2 + C\eta$ and choosing $\theta > 0$ sufficiently small yields (27).

Next, the tilted density of W relative to its original law is bounded by $e^{C\eta + C|W|}$, so $\mathbb{E}_x^\lambda e^{c|W|^2/\eta} \le C$. By global Lipschitz continuity of U_R, the same square-exponential integrability holds for $U_R(X_1) - \mathbb{E}_x^\lambda U_R(X_1)$, which implies the sub-Gaussian bound (28).

Finally, $\mathrm{Var}_x^\lambda F_R(X_1) = \psi_x''(\lambda)$. By (21), $|\psi_x''(\lambda) - \psi_x''(0)| \le C|\lambda|\eta^2 q_x$; since $|g_R|^2 = w_\eta$ on B_R, applying Lemma 3.3 at zero tilt yields (29). □

Theorem 4.4 (Concentration of the tilted clock)**.** *Suppose the chain starts at x_0, evolves under (24), and satisfies the hypotheses of Lemma 4.2 with $T = m\eta \geq 1$ and $R \geq 1 + |x_0|$. There exists a fixed constant $B < \infty$ such that*

$$H_m = \eta \sum_{k<m} I_k(1 + |X_k|^2), \qquad G = \{\tau_R > m,\ H_m \leq BT\}$$

satisfy, uniformly over all admissible tilts,

$$\mathbb{P}_{\lambda,R}(G^c) \leq C(T+1)e^{-\theta R^2} + e^{-cT}. \tag{31}$$

Set $\nu_\lambda = \delta\mu_{\lambda,R}$. On the event G, all three normalized functionals

$$\frac{1}{m\mu_{\lambda,R}} \sum_{k<m} w_\eta(X_k), \qquad \frac{1}{T\nu_\lambda} \sum_{k<m} \mathrm{Var}^\lambda_{X_k} F_R(X_{k+1}), \qquad \frac{1}{T\nu_\lambda} \sum_{k<m} \mathrm{Var}^0_{X_k} F_R(X_{k+1})$$

concentrate around one: for every $t \geq 1$, the probability that G occurs while the maximum absolute deviation of these three quantities from one exceeds

$$C\left[\sqrt{t/T} + \eta + R/T\right] \tag{32}$$

is at most $2e^{-t}$, where the superscript 0 denotes the conditional variance under the original one-step kernel evaluated along the tilted trajectory. In particular, if

$$s = d + r^2 + \log(T + e), \qquad R^2 = C_J s, \qquad s/T \to 0,$$

then for any prescribed fixed J, the outer failure probability $\mathbb{P}_{\lambda,R}(G^c)$ is eventually at most $p_J = C_J(T+e)^{-J}e^{-Jr^2}$.

Proof. Before exit, (27) implies $\mathbb{E}^\lambda_x e^{\theta|X_1|^2} \leq e^{\theta|x|^2} + C\eta$. Freezing the chain at τ_R (retaining its overshoot) and applying optional stopping and Markov's inequality yields

$$\mathbb{P}_{\lambda,R}(\tau_R \leq m) \leq (e^{\theta|x_0|^2} + CT)e^{-\theta R^2}.$$

Multiplying (27) along the stopped trajectory produces a nonnegative supermartingale with exponent

$$\theta|X_{m\wedge\tau_R}|^2 - \theta|x_0|^2 + c\eta \sum_{k<m} I_k|X_k|^2 - C\eta \sum_{k<m} I_k.$$

Discarding the nonnegative terminal term gives $\mathbb{P}_{\lambda,R}\{H_m > C(T+u)\} \leq e^{-u}$ after adjusting fixed constants; setting $u = T$ proves (31).

By (28), the stopped martingale

$$M^U_m = \sum_{k<m} I_k\{U_R(X_{k+1}) - \mathbb{E}^\lambda_k U_R(X_{k+1})\}$$

satisfies $\mathbb{P}_{\lambda,R}(|M^U_m| > C\sqrt{Tt}) \leq 2e^{-t}$. On the event G, summing (26) decomposes $\eta\sum_{k<m}(w_\eta(X_k) - \mu_{\lambda,R})$ into a boundary difference of magnitude at most CR, the martingale M^U_m, and a remainder of magnitude at most $C\eta H_m \leq C\eta T$. Dividing by $T\mu_{\lambda,R}$, whose denominator is uniformly bounded away from zero, yields the concentration bound for the first functional. By (29) applied at λ and at 0, the second and third normalized functionals differ from the first on G by at most $C\eta H_m/T \leq C\eta$, so the same martingale tail bound controls all three simultaneously. The final assertion follows immediately from (31), $R^2 = C_J s$, and $T/s \to \infty$. □

Note that for the predictably stopped full-increment martingale, the exponential tilt uses $P_\eta^{\lambda,R}$ prior to τ_R and reverts to the original kernel P_η from τ_R onward. Since the event G, its complement, and all functionals restricted to G depend solely on the path up to exit, Theorem 4.4 directly supplies the required inputs for the stopped exponential tilt without conditioning an original-law concentration inequality.

4.3 A bounded-increment normal approximation

Lemma 4.5 (Normal approximation with a random bracket)**.** *Let (e_k) be martingale differences satisfying $|e_k| \le \varepsilon$ with $0 < \varepsilon \le 1/2$, and suppose that*

$$\mathbb{P}\left(\left|\sum_k \mathbb{E}_{k-1} e_k^2 - 1\right| > D\right) \le p, \qquad 0 < D \le 1/4.$$

Then, on an enlarged probability space, there exists a martingale whose increments are bounded by ε, whose predictable quadratic variation lies in $[1-D, 1+D]$ almost surely, and whose terminal value coincides with $\sum_k e_k$ outside an event of probability at most p. Moreover,

$$\sup_z \left|\mathbb{P}\left(\sum_k e_k \le z\right) - \Phi(z)\right| \le C\{\varepsilon|\log\varepsilon| + \sqrt{D}\} + p. \tag{33}$$

Proof. Retain the original increments e_k until the first step at which adding $\mathbb{E}_{k-1}e_k^2$ would cause the accumulated predictable bracket to exceed $1+D$, and set all increments from that step onward to zero; denote the resulting terminal bracket by $V_* \le 1+D$. Append $K = \lceil \varepsilon^{-2} \rceil$ conditionally independent Rademacher increments of conditional amplitude $\sqrt{(1-D-V_*)_+/K}$. Each appended increment has magnitude at most ε, and the terminal bracket of the modified martingale equals $\max(V_*, 1-D) \in [1-D, 1+D]$. On the event where the original bracket lies in $[1-D, 1+D]$, truncation is never triggered and all appended amplitudes vanish, ensuring exact agreement of the terminal values.

Because $|e_k| \le \varepsilon$, the increments satisfy $\mathbb{E}_{k-1}|e_k|^j \le \varepsilon^{j-2}\mathbb{E}_{k-1}e_k^2$ for all $j \ge 3$, verifying Bernstein's conditional moment condition. Applying the Kolmogorov bound of Fan, Grama, and Liu [FGL13, Theorem 3.1, equation (24)] to the modified martingale yields $C(\varepsilon|\log\varepsilon| + \sqrt{D})$, and accounting for the exceptional event adds at most p. This construction applies under any chosen base probability measure. □

4.4 The tilted integral at the self-normalized threshold

We isolate the core probabilistic transfer argument so that its dynamical inputs are completely transparent. Let $(d_{k+1})_{0 \le k < m}$ be a martingale difference sequence under $\mathbb{P}$, and set

$$S = \sum_{k<m} d_{k+1}, \quad v_k = \mathbb{E}_k d_{k+1}^2, \quad V = \sum_{k<m} v_k, \quad \psi_k(\lambda) = \log \mathbb{E}_k e^{\lambda d_{k+1}}, \quad \Psi(\lambda) = \sum_{k<m} \psi_k(\lambda).$$

Define the exponentially tilted probability measure $\mathbb{Q}_\lambda$ by

$$\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} = e^{\lambda S - \Psi(\lambda)}, \tag{34}$$

whose Radon–Nikodym derivative is the terminal value of a positive product martingale. Assume the following uniform estimates hold for $|\lambda| \le \lambda_0$:

$$\mathbb{E}_k e^{cd_{k+1}^2/\eta} \le C, \qquad \sup_{|\theta|\le\lambda_0} |\psi_k'''(\theta)| \le C\eta^2 a_k, \quad 0 \le a_k \le 1 + R^2; \tag{35}$$

$$A := \frac{\eta}{T}\sum_{k<m} a_k \le B \text{ on an event } G, \qquad \mathbb{Q}_\lambda(G^c) \le p;$$

$$\mathbb{Q}_\lambda\left(\left\{\left|\frac{V}{T\nu_\lambda} - 1\right| > C[\sqrt{t/T} + \eta + R/T]\right\} \cap G\right) \le Ce^{-t}, \quad t \ge 1, \tag{36}$$

where G is independent of t, and ν_λ is deterministic, continuous in λ, and bounded between two fixed positive constants ν_* and ν^*. For the stopped SGLD increment, Lemma 4.1 and Theorem 4.4 verify (35)–(36) with $a_k = I_k(1 + |X_k|^2)$ and $\nu_\lambda = \delta\mu_{\lambda,R}$.

Proposition 4.6 (Tail estimate at the self-normalized threshold)**.** *Assume (35)–(36). Fix $J \ge 2$, $T = m\eta \ge 2$, $0 < \eta \le 1$, and $r \ge 1$. For a sufficiently large fixed constant C_J, define*

$$a = \eta + R/T, \qquad D = C_J\{\sqrt{\log(T+e)/T} + a\}, \qquad \varepsilon = C_J\sqrt{\eta\log(m+e)/T},$$
$$B_r = \eta r^3/\sqrt{T}, \qquad E_r = (1+r)\sqrt{D} + (1+r)^2 D + (1+r)\varepsilon|\log\varepsilon| + B_r. \tag{37}$$

If

$$r/\sqrt{T\nu_*} \le \lambda_0, \qquad D \le 1/4, \qquad \varepsilon \le e^{-1}, \qquad r^2 \le cT$$

for a sufficiently small fixed constant $c > 0$, then

$$\sup_{0\le x\le r}\left|\frac{\mathbb{P}(S > x\sqrt{V})}{\overline{\Phi}(x)} - 1\right| \le CE_r e^{CE_r} + C_J(1+r)(T+e)^{-J}e^{CE_r} + C_J(1+r)e^{Cr^2+CB_r(1+R^2)}p. \tag{38}$$

Here the constants in the exponent multiplying p depend solely on the fixed input constants and not on J.

Proof. Fix $x \in [0, r]$. By continuity and positivity of ν_λ, the map $\lambda \mapsto \lambda\sqrt{T\nu_\lambda}$ vanishes at $\lambda = 0$ and is at least r at $\lambda = r/\sqrt{T\nu_*}$; hence we may choose a tilt parameter λ satisfying

$$\lambda \in [0, r/\sqrt{T\nu_*}], \qquad \lambda\sqrt{T\nu_\lambda} = x, \tag{39}$$

without requiring uniqueness. Set $\sigma = \sqrt{T\nu_\lambda}$ and $t_\lambda = V/\sigma^2$. The square-exponential integrability in (35) implies $V \le CT$ almost surely, so $0 \le t_\lambda \le M$ for a fixed constant M.

Cumulants and the tilted centered martingale. Define the signed remainder $R_\psi = \Psi(\lambda) - \lambda^2 V/2$. Integrating the third-cumulant bound in (35) from zero gives

$$|R_\psi| \le C\eta|\lambda|^3 TA,$$
$$\left|\sum_k \psi_k'(\lambda) - \lambda V\right| \le C\eta\lambda^2 TA,$$
$$\left|\sum_k \psi_k''(\lambda) - V\right| \le C\eta|\lambda|TA. \tag{40}$$

In particular, $|R_\psi| \le CB_r$ on G and $|R_\psi| \le CB_r(1+R^2)$ everywhere. Under $\mathbb{Q}_\lambda$, the conditional mean and variance of d_{k+1} are $\psi_k'(\lambda)$ and $\psi_k''(\lambda)$, respectively, so

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

is the terminal value of a centered $\mathbb{Q}_\lambda$-martingale with predictable bracket $\sigma^{-2}\sum_k \psi_k''(\lambda)$. On G, this bracket differs from $\sigma^{-2}V$ by at most $C\eta|\lambda| \le C\eta$.

Kolmogorov approximation under $\mathbb{Q}_\lambda$. Since $\psi_k(\lambda) \ge 0$, combining the conditional tilt density $e^{\lambda d - \psi_k(\lambda)}$ with Young's inequality and (35) gives

$$\mathbb{E}_k^{\mathbb{Q}_\lambda} e^{c_1 d_{k+1}^2/\eta} \le C_1.$$

Consequently, $|\psi_k'(\lambda)| \le C\sqrt{\eta}$, and the centered difference $d_{k+1} - \psi_k'(\lambda)$ satisfies the same type of square-exponential bound under $\mathbb{Q}_\lambda$. Truncating this centered difference at level $C_J\sqrt{\eta\log(m+e)}$ and recentering conditionally under $\mathbb{Q}_\lambda$ produces tail first and second moments bounded by $C\sqrt{\eta}(m+e)^{-c_J}$ and $C\eta(m+e)^{-c_J}$, respectively. Choosing the clipping constant sufficiently large ensures that the total truncation probability, the normalized recentering shift on untruncated paths, and the normalized bracket perturbation are each bounded by $C_J(T+e)^{-J-2}$ (for the recentering shift, the prefactor is $m\sqrt{\eta}/\sqrt{T} = \sqrt{m}$, and $m \ge T$).

Invoking (36) at a large fixed multiple of $\log(T+e)$ together with (40), the clipped normalized martingale has increments bounded by ε and predictable bracket within D of one outside an event of $\mathbb{Q}_\lambda$-probability at most $p + C_J(T+e)^{-J}$. Applying Lemma 4.5 under $\mathbb{Q}_\lambda$ and absorbing the small recentering shifts via the bounded standard normal density yields

$$\sup_z |\mathbb{Q}_\lambda(U_\lambda \le z) - \Phi(z)| \le C_J\{\sqrt{D} + \varepsilon|\log\varepsilon| + (T+e)^{-J} + p\}. \tag{41}$$

Notice that only a polynomially small truncation probability is needed here because (41) is an absolute approximation under $\mathbb{Q}_\lambda$, so the clipping level involves $\log(m+e)$ rather than r^2.

The self-normalized threshold. Define $Z_\lambda = (S - x\sqrt{V})/\sigma$. Completing the square in (34) yields the exact identity

$$\begin{aligned} \mathbb{P}(S > x\sqrt{V}) &= e^{-x^2/2}\mathbb{E}^{\mathbb{Q}_\lambda}[e^{-xZ_\lambda}\mathbf{1}_{\{Z_\lambda>0\}}W_\lambda], \\ W_\lambda &= \exp\{R_\psi + \tfrac{1}{2}x^2(\sqrt{t_\lambda}-1)^2\}, \end{aligned} \tag{42}$$

where the quadratic clock fluctuation enters with a positive sign, while the first-order clock effect is absorbed into

$$Z_\lambda = U_\lambda + x(t_\lambda - \sqrt{t_\lambda}) + \frac{\sum_k \psi_k'(\lambda) - \lambda V}{\sigma}. \tag{43}$$

For a fixed input-dependent constant C_4, set

$$h_J = C_J\sqrt{\log(T+e)/T}, \qquad d_J = C_4 a + h_J \le D, \qquad H = G \cap \{|t_\lambda - 1| \le d_J\},$$

enlarging C_J in (37) if necessary. Then $\mathbb{Q}_\lambda(H^c) \le p + C_J(T+e)^{-J}$, and on H we have $|Z_\lambda - U_\lambda| \le C\{xD + \eta x^2/\sqrt{T}\}$. Combining this displacement bound with (41) gives a Kolmogorov distance bound

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

for Z_λ. Because $z \mapsto e^{-xz}\mathbf{1}_{\{z>0\}}$ has total variation at most two for every $x \ge 0$, Stieltjes integration by parts yields

$$\left|\mathbb{E}^{\mathbb{Q}_\lambda} e^{-xZ_\lambda}\mathbf{1}_{\{Z_\lambda>0\}} - I(x)\right| \le 2\kappa_x, \qquad I(x) = e^{x^2/2}\overline{\Phi}(x) \asymp (1+x)^{-1}. \tag{44}$$

Integration of the clock error. Put $U = |t_\lambda - 1| \le M + 1$. By (36), there exist fixed constants $C_4, c_4 > 0$ such that

$$\mathbb{Q}_\lambda(U > C_4 a + v,\ G) \le C_4 e^{-c_4 T v^2}, \qquad v \ge 0,$$

where the bound is trivial for $v = O(T^{-1/2})$ upon enlarging C_4 and follows from setting t to a constant multiple of Tv^2 otherwise. Defining $U_0 = (U - C_4 a)_+$, we have $G \setminus H = \{U_0 > h_J\} \cap G$. For any $0 \le \gamma \le c_4 T/4$, layer-cake integration yields

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}_\lambda}[e^{\gamma U_0^2} \mathbf{1}_{\{U_0 > h_J\}} \mathbf{1}_G] &\le C_4 e^{-(c_4 T - \gamma) h_J^2} + 2C_4 \gamma \int_{h_J}^\infty v e^{-(c_4 T - \gamma) v^2}\, dv \\ &\le C e^{-c_4 T h_J^2/2}.\end{aligned}$$

Using $U^2 \le 2C_4^2 a^2 + 2U_0^2$, setting $\gamma = Cr^2$, and taking $r^2 \le cT$ with $c > 0$ small, increasing the constant in h_J gives

$$\mathbb{E}^{\mathbb{Q}_\lambda}[e^{x^2 U^2/2} \mathbf{1}_{G \setminus H}] \le C_J e^{Cr^2 a^2} (T + e)^{-J}, \qquad 0 \le x \le r. \tag{45}$$

Crucially, integrating the clock tail directly avoids a crude worst-case bound of order $e^{Cr^2}(T + e)^{-J}$.

Put $g_x(z) = e^{-xz} \mathbf{1}_{\{z > 0\}}$ and $b = C(B_r + r^2 D^2)$. On H we have $|\log W_\lambda| \le b$, so (44) gives

$$\left|\mathbb{E}^{\mathbb{Q}_\lambda}[g_x(Z_\lambda)(W_\lambda - 1)\mathbf{1}_H]\right| \le (e^b - 1)\{I(x) + 2\kappa_x\}.$$

On $G \setminus H$, combining $|\sqrt{t_\lambda} - 1| \le U$, (45) and $|R_\psi| \le CB_r$ bounds the weighted contribution by $C_J e^{C(B_r + r^2 a^2)}(T + e)^{-J}$. On G^c the deterministic bounds $t_\lambda \le M$ and $|R_\psi| \le CB_r(1 + R^2)$ give $Ce^{Cr^2 + CB_r(1+R^2)} p$. For the unweighted contributions, use $0 \le g_x \le 1$ and the respective event probabilities $C_J(T + e)^{-J}$ and p. Substitution into (42) and division by $I(x) \ge c/(1 + x)$ yield (38). Here $a \le D \le 1/4$, and $B_r + r^2 D^2 \le B_r + r^2$ bounds the exponent multiplying p independently of J. □

4.5 The martingale moderate-deviation window

Theorem 4.7 (Moderate deviations for the stopped martingale)**.** *Suppose the hypotheses of Proposition 4.6 hold uniformly as $\eta \downarrow 0$, $T = m\eta \to \infty$, and $r \to \infty$. Choose a fixed, sufficiently large J and set*

$$R^2 = C_J[d + r^2 + \log(T + e)], \qquad p = C_J(T + e)^{-J} e^{-Jr^2},$$

with fixed d. If

$$r^2 \sqrt{\frac{\log(T + e)}{T}} \to 0, \qquad r^2 \eta \to 0, \tag{46}$$

then

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

with the convention $S/\sqrt{V} = 0$ when $V = 0$. Both the opposite tail and non-strict tail versions hold under the same conditions when the inputs apply to both signs of the tilt. For the stopped SGLD martingale, the local restriction $\eta(1 + 9R^2) \le c_0$ is also retained.

Proof. The first condition in (46), namely $r^4 \log(T+e)/T \to 0$, implies $r = o(T^{1/4})$, $r/\sqrt{T} \to 0$, and

$$r^2 R/T \le C_J\{r^3/T + r^2\sqrt{\log(T+e)}/T + r^2/T\} \to 0.$$

Hence $r^2 D \to 0$, $r\sqrt{D} \to 0$, and $B_r = (r^2\eta)(r/\sqrt{T}) \to 0$. To verify that $r\varepsilon|\log\varepsilon| \to 0$ without restricting T to a polynomial schedule in η, write $L = \log(1/\eta)$ and $H = \log(T+e)$. Then $\log(m+e) \le C(1+L+H)$, $\varepsilon \to 0$, and $|\log\varepsilon| \le C(1+L+H)$. When $L \le H$, we have

$$r\varepsilon|\log\varepsilon| \le C(r/\sqrt{T})H^{3/2} \to 0$$

since $r = o(T^{1/4})$; when $L > H$, using $r/\sqrt{T} \le 1$ eventually yields $r\varepsilon|\log\varepsilon| \le C\sqrt{\eta}(1+L)^{3/2} \to 0$. Consequently, $E_r \to 0$.

For the final term in (38), note that

$$\frac{B_r(1+R^2)}{r^2} \le C_J\left\{\frac{\eta r^3}{\sqrt{T}} + \frac{\eta r\log(T+e)}{\sqrt{T}} + \frac{\eta r}{\sqrt{T}}\right\} \to 0.$$

Choosing J strictly larger than the fixed exponential constant multiplying r^2 in Proposition 4.6, the exceptional contribution becomes

$$C_J(1+r)(T+e)^{-J}e^{-(J-C-o(1))r^2} \to 0,$$

while all finite-range hypotheses of Proposition 4.6 are eventually satisfied. On the event $\{V = 0\}$, all increments d_{k+1} vanish almost surely, so the convention $S/\sqrt{V} = 0$ agrees with the event $\{S > x\sqrt{V}\}$ for all $x \ge 0$. Replacing S by $-S$ proves the opposite tail. Finally, for $0 < x \le r$, letting $y \uparrow x$ in the strict-tail bounds for $\mathbb{P}(S/\sqrt{V} > y)$ and invoking continuity of $\overline{\Phi}$ gives the non-strict tail bound at x, while $x = 0$ follows by taking the complement of the opposite strict tail. □

Theorem 4.7 governs the centered full-increment martingale normalized by its exact predictable quadratic variation. In Theorem 2.2, we pass to the empirical SGLD statistic by replacing the bracket with $\delta T Y_m$ and accounting separately for the weak discretization bias.

5 Transfer to the empirical statistic

5.1 A relative-tail perturbation lemma

The following elementary comparison lemma accommodates simultaneous, arbitrarily dependent perturbations in both the numerator and the denominator of a self-normalized statistic.

Lemma 5.1 (Dependent numerator and denominator errors)**.** *Let $W = U/\sqrt{V}$, where $V > 0$ on an event G, and let $\widehat{W} = (U+\Delta)/\sqrt{\widehat{V}}$, assigning arbitrary finite values outside G whenever a quotient is undefined. Suppose that*

$$\mathbb{P}(G^c) \le p, \qquad \left|\frac{\widehat{V}}{V} - 1\right| \le \beta < \frac{1}{2}, \qquad |\Delta| \le a\sqrt{V} \quad \text{on } G.$$

Fix $r \geq 1$, and assume that both the strict and non-strict tails of W and $-W$ admit relative standard normal tail errors of at most ϵ_0 uniformly on $[0, r+1]$. If $A = (1+r)a + (1+r)^2\beta \leq c$ for a sufficiently small universal constant $c > 0$, then

$$\sup_{0 \leq x \leq r} \left| \frac{\mathbb{P}(\widehat{W} > x)}{\overline{\Phi}(x)} - 1 \right| \leq C(\epsilon_0 + A)e^{CA} + \frac{p}{\overline{\Phi}(r)}. \tag{47}$$

The same bound holds for the opposite and non-strict tails, without any independence assumption among $U, V, \Delta, \widehat{V}$, and G.

Proof. For any $x \geq 0$, pointwise comparison on the event G gives

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

The shifted thresholds differ from x by at most $C(x\beta + a)$ and lie in $[-1, r+1]$ for $A \leq c$. Because the standard normal hazard rate satisfies $\varphi(z)/\overline{\Phi}(z) \leq C(1+z)$ for $z \geq 0$ and is bounded on $[-1, 0]$, integrating $(\log \overline{\Phi})' = -\varphi/\overline{\Phi}$ yields

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

for each such threshold shift v. At nonnegative shifted thresholds, the assumed tail approximation applies directly. If the lower threshold $z = x\sqrt{1-\beta} - a$ is negative, then $x \leq 2a$ (so $\overline{\Phi}(x)$ is bounded away from zero) and writing $\mathbb{P}(W > z) = 1 - \mathbb{P}(-W \geq -z)$ reduces the estimate to the opposite-tail bound at $-z > 0$, which again contributes at most $C\epsilon_0$. Combining these inequalities establishes (47), and the remaining tail assertions follow by replacing strict inequalities with weak ones or negating U and Δ. □

5.2 Proof of the main trajectory theorem

Proof of Theorem 2.2. Put $\gamma = j + 1$ and apply Theorem 4.7 at radius $r + 1$. Set

$$s = d + (r+1)^2 + \log(T+e), \qquad R^2 = C_J s, \qquad p_J = C_J(T+e)^{-J}e^{-J(r+1)^2},$$

where J is a fixed, sufficiently large constant and C_J is chosen large enough that $R \geq 1 + |x_0|$. Because $r \geq 1$, the bias condition $r\eta^\gamma\sqrt{T} \to 0$ implies $T \leq \eta^{-2\gamma}$ eventually. Combining this bound with $r^2\eta \to 0$ gives

$$\eta \log(T+e) \longrightarrow 0, \qquad \eta(1 + 9R^2) \longrightarrow 0.$$

This verifies the small-step localization condition throughout Sections 3 and 4. Moreover, the clock condition $r^4 \log(T+e)/T \to 0$ implies

$$\frac{s}{T} \to 0, \qquad \frac{s^2}{T} \to 0, \qquad \frac{rR}{\sqrt{T}} \to 0, \tag{48}$$

since $s^2 \leq C\{1 + r^4 + \log^2(T+e)\}$ and $rR/\sqrt{T} \leq C_J\{r^2/\sqrt{T} + r\sqrt{\log(T+e)/T} + r/\sqrt{T}\}$.

Choose $F_R = -\chi_R f_\eta$ and form the predictably stopped full-increment martingale and its predictable quadratic variation:

$$d_{k+1} = I_k\{F_R(X_{k+1}) - P_\eta F_R(X_k)\}, \qquad S_R = \sum_{k<m} d_{k+1}, \qquad V_R = \sum_{k<m} I_k \operatorname{Var}_{X_k} F_R(X_{k+1}).$$

Lemmas 3.3 and 4.1 verify (35), while Theorem 4.4 verifies (36) with

$$a_k = I_k(1 + |X_k|^2), \qquad \nu_\lambda = \delta\mu_{\lambda,R}, \qquad \mathbb{Q}_\lambda(G^c) \le p_J,$$

where ν_λ is continuous and uniformly positive by Lemma 4.2. Since the Markov chain under (24) coincides with the stopped exponential tilt up to τ_R, Theorem 4.7 applies and establishes relative Gaussian tail approximations for both signs of $S_R/\sqrt{V_R}$ uniformly on $[0, r+1]$.

We now compare $S_R/\sqrt{V_R}$ with $W^{(j)}_{m,\eta}$ under the original law $\mathbb{P}$. Intersecting the zero-tilt concentration event of Theorem 4.4 (at confidence level t equal to a sufficiently large multiple of s) with the exit, quadratic, and quartic occupation events of Lemma 3.2 produces an event $\mathcal{G}_J$ satisfying

$$\mathbb{P}(\mathcal{G}_J^c) \le C_J(T+e)^{-J} e^{-Jr^2} = o(\overline{\Phi}(r)), \tag{49}$$

on which, for all sufficiently small η,

$$\begin{aligned} &\tau_R > m, \qquad H_2 := \eta \sum_{k<m} (1 + |X_k|^2) \le C_J T, \\ &H_4 := \eta \sum_{k<m} (1 + |X_k|^2)^2 \le C_J(T + s^2), \qquad Y_m \ge \mu_*/2 > 0. \end{aligned} \tag{50}$$

Indeed, the clock deviation on $\mathcal{G}_J$ is bounded by $C_J\{\sqrt{s/T} + \eta + R/T\} = o(1)$ while $\mu_\eta \ge \mu_*$ by Lemma A.3; note that μ_* is used purely in this proof and never enters the numerical algorithms.

By Lemma 3.3, on $\mathcal{G}_J$ we have

$$|V_R - \delta T Y_m| \le C\eta H_2 \le C_J \eta T, \qquad \left|\frac{\delta T Y_m}{V_R} - 1\right| \le C_J\eta, \tag{51}$$

and in particular $V_R \ge cT$ on $\mathcal{G}_J$.

For the numerator, set $H = \chi_R f_\eta = -F_R$. On $\mathcal{G}_J$, every visited state X_k lies in B_R, so the differential operators in (11) applied to H coincide with those applied to f_η, while the one-step conditional expectation $P_\eta H(X_k)$ integrates over the entire transition law. Telescoping the increments of H yields

$$\eta \sum_{k<m} \{h(X_k) - \theta_\eta + j\eta a_\eta(X_k)\} = S_R + Q_j \quad \text{on } \mathcal{G}_J, \tag{52}$$

where

$$Q_j = H(X_m) - H(x_0) - (1-j)\eta^2 \sum_{k<m} a_\eta(X_k) - \sum_{k<m} \mathcal{R}_\eta^{f_\eta}(X_k).$$

On $\mathcal{G}_J$, the boundary difference $|H(X_m) - H(x_0)|$ is bounded by CR. When $j = 0$, the uncorrected weak defect sums to at most $C\eta H_2$; when $j = 1$, the leading weak-error term cancels and the remaining remainder sums to at most $C\eta^2 H_4$. Dividing by $\sqrt{V_R} \ge \sqrt{cT}$ gives

$$\frac{|Q_j|}{\sqrt{V_R}} \le A_j, \quad A_0 = C_J\left(\frac{R}{\sqrt{T}} + \eta\sqrt{T}\right), \quad A_1 = C_J\left\{\frac{R}{\sqrt{T}} + \eta^2\sqrt{T}\left(1 + \frac{s^2}{T}\right)\right\}. \tag{53}$$

Applying Lemma 5.1 with $U = S_R$, $V = V_R$, $\widehat{V} = \delta T Y_m$, and $\Delta = Q_j$, equations (48) and (7) ensure that

$$(1+r)A_j + (1+r)^2 C_J \eta \longrightarrow 0.$$

Since $\mathbb{P}(\mathcal{G}_j^c) = o(\overline{\Phi}(r))$ by (49) and the relative tail error of $S_R/\sqrt{V_R}$ vanishes uniformly on $[0, r+1]$, we conclude (8) for both signs and both inequality conventions. □

The finite-sample bounds in the preceding proof separate three sources of error: the tilted martingale approximation contributes the error in (38) with outer exception term $(1+r)e^{Cr^2+C\eta r^3(1+R^2)/\sqrt{T}}p_J$; the trajectory transfer introduces $C\{(1+r)A_j + (1+r)^2\eta\}$ along with $\mathbb{P}(\mathcal{G}_j^c)/\overline{\Phi}(r)$; and in Section 6, the probabilities of numerical failure enter through p_g, p_a, p_c divided by $\overline{\Phi}(r)$. Because these exceptional events live under distinct probability measures, they are kept separate until the final joint asymptotic limit is taken.

6 Certified computation of the normalizer and correction

Given an observed SGLD trajectory $(X_k)_{k<m}$, we estimate the Poisson gradient $g_\eta = \nabla f_\eta$ and, in the corrected case $j = 1$, the weak-bias correction $a_\eta = \mathcal{A}_2 f_\eta$ at every visited state. The algorithm takes as inputs the trajectory states, point evaluations of b, σ_0, h, and the effective envelopes specified after Assumption 2.1. Within each finite-difference pair, two auxiliary Euler paths are driven by identical Gaussian increments, while independent Monte Carlo replicates are generated conditionally on the entire observed trajectory. This structure yields simultaneous confidence bounds for any deterministic array of query states and, by conditioning, for the random trajectory $(X_k)_{k<m}$.

6.1 Explicit finite-time simulation bounds

Let $\mathcal{P}_t^\eta$ denote the semigroup of the modified diffusion (3), and let $C_c, \rho > 0$ be the explicit contraction constants from Lemma A.1. Introduce the combined diffusion matrix $\sigma_\eta(x) = (\sqrt{\delta} I_d, \sqrt{\eta}\,\sigma_0(x))$ and define the uniform envelopes

$$L_b = B_1, \quad B_0 = |b(0)|, \quad s_0^2 = \delta d + \eta_0 \|\sigma_0(0)\|_{\mathrm{HS}}^2, \quad s_1 = \sqrt{\eta_0} S_1, \quad s_2 = \sqrt{\eta_0} S_2,$$

valid across $0 \le \eta \le \eta_0$. Set

$$\begin{aligned} \alpha_2 &= 2L_b + s_1^2, \qquad \alpha_4 = 4L_b + 6s_1^2, \qquad c_K = 2L_b + 1 + 2s_1^2, \\ D_K &= B_2^2 + 2s_2^2, \qquad C_H = \max\{\alpha_2, \tfrac{1}{2}\max(c_K, \alpha_4)\}, \\ H(S) &= S\{H_2 + H_1\sqrt{D_K S}\}e^{C_H S}, \qquad S \ge 1, \end{aligned} \tag{54}$$

noting that $H_1 > 0$ since h is nonconstant. For the finite-horizon resolvent approximation

$$f_{\eta,S} = -\int_0^S \mathcal{P}_t^\eta h \, dt,$$

we have the explicit bounds

$$\|\nabla f_\eta - \nabla f_{\eta,S}\|_\infty \le C_c H_1 e^{-\rho S}/\rho, \qquad \|D^2 f_{\eta,S}\|_\infty \le H(S). \tag{55}$$

The first inequality follows by integrating the Lipschitz contraction in (79). For the second, let J_u and K_{uv} denote the first and mixed second spatial variation processes of the diffusion along unit initial directions u, v; Itô's formula gives

$$\mathbb{E}|J_u(t)|^2 \le e^{\alpha_2 t}, \quad \mathbb{E}|J_u(t)|^4 \le e^{\alpha_4 t}, \quad \frac{d}{dt}\mathbb{E}|K_{uv}(t)|^2 \le c_K \mathbb{E}|K_{uv}(t)|^2 + D_K e^{\alpha_4 t}.$$

Since $K_{uv}(0) = 0$, Gronwall's inequality yields $\mathbb{E}|K_{uv}(t)| \le \sqrt{D_K t}\, e^{\max(c_K, \alpha_4)t/2}$, and differentiating $\mathcal{P}_t^\eta h$ twice under the integral sign gives (54). Every constant in (54) is an explicit known envelope.

We next record an explicit strong Euler discretization bound. For $Q \ge 1$ and $S \ge 1$, define

$$\begin{aligned} k_0 &= 1 + 4L_b^2 + 2s_1^2, \qquad c_0 = 4B_0^2 + 2s_0^2, \\ C_e &= 3L_b + 2s_1^2, \qquad C_z = L_b + 2s_1^2, \qquad C_D = 2L_b + L_b^2 + s_1^2, \\ M_Q(S) &= (Q^2 + c_0 S)e^{k_0 S}, \\ D_Q(S) &= 2\{B_0^2 + s_0^2 + (L_b^2 + s_1^2)M_Q(S)\}, \\ E_Q(S) &= \sqrt{D_Q(S)}\{1 + \sqrt{C_z S}\, e^{C_e S/2}\}. \end{aligned} \tag{56}$$

Let $(Y_j^y)_{0 \le j \le K}$ be the Euler–Maruyama discretization of the modified diffusion with step size $\tau = S/K \le 1$ and initial condition $|y| \le Q$. Coupling Y_j^y synchronously to the continuous diffusion X_t^y via the same driving Brownian motions gives

$$\sup_{0 \le t \le S} \mathbb{E}|X_t^y - Y_{\lfloor t/\tau \rfloor}^y| \le E_Q(S)\sqrt{\tau}, \tag{57}$$

where $Y_{\lfloor S/\tau \rfloor}^y = Y_K^y$. The discrete recursion $\mathbb{E}|Y_{j+1}^y|^2 \le (1 + k_0\tau)\mathbb{E}|Y_j^y|^2 + c_0\tau$ gives the moment bound $M_Q(S)$. Write $\mathcal{B} = (B, W)$ for the combined Brownian motion. For $j\tau \le t \le (j+1)\tau$ use the adapted continuous Euler interpolation

$$\overline{Y}_t = Y_j^y + (t - j\tau)b(Y_j^y) + \sigma_\eta(Y_j^y)(\mathcal{B}_t - \mathcal{B}_{j\tau}).$$

Its within-step displacement $z_t = \overline{Y}_t - Y_{\lfloor t/\tau \rfloor}^y$ satisfies $\mathbb{E}|z_t|^2 \le D_Q(S)\tau$. For $e_t = X_t^y - \overline{Y}_t$, Itô's formula gives $(\mathbb{E}|e_t|^2)' \le C_e \mathbb{E}|e_t|^2 + C_z \mathbb{E}|z_t|^2$ between grid times. Integration and the triangle inequality yield (57).

6.2 Estimating the Poisson gradient

For a query state $x \in \mathbb{R}^d$, coordinate $i \in \{1, \dots, d\}$, and finite-difference spacing $0 < u \le 1$, simulate a coupled pair of Euler paths starting from $x + ue_i$ and $x - ue_i$ using identical Brownian increments, and form the central-difference estimator

$$V_i(x) = -\frac{\tau}{2u} \sum_{j=0}^{K-1} \{h(Y_j^{x+ue_i}) - h(Y_j^{x-ue_i})\}. \tag{58}$$

By the synchronous coupling bound $\mathbb{E}|Y_j^{x+ue_i} - Y_j^{x-ue_i}|^2 \le 4u^2 e^{C_D j\tau}$ and the Cauchy–Schwarz inequality,

$$\mathbb{E}V_i(x)^2 \le V(S) := H_1^2 S^2 e^{C_D S}. \tag{59}$$

Crucially, this second-moment bound does not blow up as $u \downarrow 0$ and is uniform over all $x \in \mathbb{R}^d$.

Theorem 6.1 (Simultaneous gradient estimates)**.** *Suppose Assumption 2.1 holds, and let* $x_0, \dots, x_{m-1}$ *be any finite array of states with* $\max_{k<m} |x_k| \le R$ *and* $R \ge 1$*. Given tolerances* $0 < \epsilon \le 1$ *and* $0 < p < 1$*, set*

$$\begin{aligned} S &= \max\left\{1, \rho^{-1} \log_+ \frac{3C_c H_1}{\rho\epsilon}\right\}, \\ u &= \min\{1, \epsilon/[3\sqrt{d}(1 + H(S))]\}, \\ K &= \left\lceil S \max\left\{1, \left(\frac{3\sqrt{d} H_1 S E_{R+1}(S)}{\epsilon u}\right)^2\right\}\right\rceil, \qquad \tau = S/K, \\ n &= \left\lceil \max\{1, 576 d V(S)/\epsilon^2\}\right\rceil, \end{aligned} \tag{60}$$

where $\log_+ z = \max(0, \log z)$*. Let* B *be the smallest positive odd integer satisfying* $2^{-B} \le p/(md)$*. For each state* x_k *and coordinate* i*, let* $v_{k,i}$ *be the median of* B *independent group means of (58), each computed from* n *independent coupled pairs. Then the resulting vectors* $v_k = (v_{k,1}, \dots, v_{k,d})^\top$ *satisfy*

$$\mathbb{P}_{\mathrm{aux}}\left\{\max_{k<m} |v_k - g_\eta(x_k)| > \epsilon \,\middle|\, (x_k)_{k<m}\right\} \le p, \tag{61}$$

and the procedure requires at most $2mdnBK$ *Euler step updates.*

Proof. By (55), truncating the resolvent integral at horizon S incurs a gradient bias of Euclidean norm at most $\epsilon/3$. A central difference of $f_{\eta,S}$ with spacing u has coordinate-wise bias at most $H(S)u/2$, contributing at most $\epsilon/6$ in Euclidean norm. Because $x_k \pm u e_i \in B_{R+1}$, integrating (57) over $[0, S]$ bounds the coordinate-wise Euler discretization bias of (58) by

$$H_1 S E_{R+1}(S)\sqrt{\tau}/u \le \epsilon/(3\sqrt{d}),$$

so the Euclidean Euler bias is at most $\epsilon/3$.

For each group mean of n samples, (59) and Chebyshev's inequality imply that the probability of exceeding tolerance $\epsilon/(6\sqrt{d})$ is at most $1/16$. The median of B groups fails this tolerance only if at least $(B+1)/2$ groups fail, which by a union bound over subsets of size $(B+1)/2$ has probability at most

$$2^B (1/16)^{B/2} = 2^{-B} \le \frac{p}{md}.$$

Taking a union bound over all md state-coordinate pairs ensures that the Monte Carlo error is at most $\epsilon/6$ in Euclidean norm simultaneously for all $k < m$ with probability at least $1 - p$. Summing the four error budgets ($\epsilon/3 + \epsilon/6 + \epsilon/3 + \epsilon/6 = \epsilon$) establishes (61). □

For fixed dimension, noise intensity, and model envelopes, $S = O(1 + \log(\epsilon^{-1}))$, so every exponential factor in (60) grows at most polynomially in ϵ^{-1}. Consequently, there exists a model-dependent polynomial P_g such that the total number of Euler updates is bounded by

$$Cm\{1 + \log(m/p)\}\, P_g(1 + R, \epsilon^{-1}), \tag{62}$$

where the polynomial degree depends on ratios of the growth envelopes to ρ. In a dense matrix implementation, each Euler update requires $O(d(d + \ell))$ arithmetic operations in addition to evaluating b, σ_0, and h; no uniform complexity bound is claimed in the degenerate regimes $\rho \downarrow 0$ or $d \to \infty$.

6.3 Relative error in the empirical energy

Given gradient estimates v_k and error radii $e_k \geq 0$, define the empirical energy enclosure endpoints

$$L_m = \frac{1}{m}\sum_{k<m}(|v_k| - e_k)_+^2, \qquad U_m = \frac{1}{m}\sum_{k<m}(|v_k| + e_k)^2. \tag{63}$$

Whenever $|v_k - g_\eta(x_k)| \leq e_k$ for all $k < m$, the true empirical energy $Y_m = m^{-1}\sum_{k<m}|g_\eta(x_k)|^2$ lies in $[L_m, U_m]$.

Lemma 6.2 (Minimax relative error on an interval)**.** *Suppose $0 < L_m \leq U_m$, and define the harmonic midpoint and relative half-width*

$$\widehat{Y}_m^H = \frac{2L_mU_m}{L_m + U_m}, \qquad b_m = \frac{U_m - L_m}{U_m + L_m}. \tag{64}$$

Then $|\widehat{Y}_m^H/y - 1| \leq b_m$ for every $y \in [L_m, U_m]$, and $\widehat{Y}_m^H$ uniquely minimizes $\sup_{y\in[L_m,U_m]}|z/y-1|$ over $z > 0$. Moreover, if $e_k \leq \epsilon$ and $|v_k - g_\eta(x_k)| \leq e_k$ for all $k < m$, then

$$\sqrt{L_m} \geq (\sqrt{Y_m} - 2\epsilon)_+, \qquad \sqrt{U_m} \leq \sqrt{Y_m} + 2\epsilon. \tag{65}$$

In particular, whenever $Y_m > 0$ and $\epsilon \leq \beta\sqrt{Y_m}/4$ with $0 < \beta < 1$, we have $L_m > 0$ and $b_m \leq \beta$.

Proof. The worst-case relative error $\sup_{y\in[L_m,U_m]}|z/y - 1|$ is attained at one of the endpoints; equating $z/L_m - 1 = 1 - z/U_m$ yields $z = \widehat{Y}_m^H$ with error b_m, while perturbing z strictly increases one endpoint error. Next, the reverse triangle inequality in the empirical Euclidean norm gives (65). When $\sqrt{Y_m} > 2\epsilon$, monotonicity of $(U - L)/(U + L)$ in L and U implies

$$b_m \leq \frac{4\epsilon\sqrt{Y_m}}{Y_m + 4\epsilon^2} \leq \frac{4\epsilon}{\sqrt{Y_m}}.$$

Under the stated condition $\epsilon \leq \beta\sqrt{Y_m}/4$, the last bound is at most β and the lower endpoint is positive. □

To obtain a well-defined algorithm on every finite trajectory, we prescribe the following fixed-budget procedure. Given $0 < \beta < 1/2$ and $0 < p_g < 1$, set

$$R = \max\{1, \max_{k<m}|x_k|\}, \qquad e_g = \beta^2.$$

Invoke Theorem 6.1 with $(\epsilon, p) = (e_g, p_g)$ and compute (63) with $e_k = e_g$. If $L_m > 0$ and $b_m \leq \beta$, return $\widehat{Y}_m = \widehat{Y}_m^H$ together with an acceptance flag; otherwise, return $\widehat{Y}_m = 1$ with a rejection flag. This positive fallback ensures that the self-normalized statistic is well-defined on all paths while explicitly signaling whether the certified interval criterion was met.

Proposition 6.3 (A positive numerical normalizer)**.** *The preceding procedure terminates in finitely many operations for every finite input array and outputs a strictly positive value $\widehat{Y}_m$. Interpreting the relative error as $+\infty$ when $Y_m = 0$, we have*

$$\mathbb{P}_{\text{aux}}\left\{|\widehat{Y}_m/Y_m - 1| > \beta \,\middle|\, (x_k)_{k<m}\right\} \leq p_g + \mathbf{1}_{\{Y_m < 16\beta^2\}}, \tag{66}$$

and the same upper bound controls the conditional probability of returning the rejection flag.

Proof. Every budget parameter in (60) is finite for any finite input array. On the event where all gradient enclosures hold and $Y_m \geq 16\beta^2$, we have $e_g = \beta^2 \leq \beta\sqrt{Y_m}/4$, so Lemma 6.2 guarantees both acceptance and $|\widehat{Y}_m/Y_m - 1| \leq \beta$. When $Y_m < 16\beta^2$, the right-hand side of (66) is at least one, while on $\{Y_m \geq 16\beta^2\}$ the conditional failure probability is bounded by p_g. □

Notably, this construction uses neither μ_η nor the nondegeneracy constant μ_* from Lemma A.3; under the hypotheses of Theorem 2.2, the event $\{Y_m < 16\beta^2\}$ automatically has probability $o(\overline{\Phi}(r))$ as $\beta \to 0$.

6.4 Estimating the weak-bias correction

Evaluating $a_\eta = \mathcal{A}_2 f_\eta$ in (5) requires partial derivatives of f_η through order four, which we approximate using finite differences of anchored Poisson values. Let $\|D^j f_\eta\|_\infty \leq G_j$ $(1 \leq j \leq 6)$ denote the explicit global derivative bounds provided by Lemma A.2; only G_3, G_4, G_5 are required below.

Theorem 6.4 (Simultaneous correction estimates)**.** *Given an input array $(x_k)_{k<m}$ contained in B_R with $R \geq 1$, a tolerance $0 < e_a \leq 1$, and a failure budget $0 < p_a < 1$, define*

$$\begin{aligned}
M_b(R) &= B_0 + L_b R, \qquad C_A = \tfrac{d}{2} M_b(R)^2 + \tfrac{\delta}{2} d^{3/2} M_b(R) + \tfrac{\delta^2}{8} d^2, \\
\vartheta &= \frac{e_a}{1 + C_A}, \qquad G = \max(G_3, G_4, G_5), \qquad u_A = \frac{\vartheta}{12(1+G)}, \qquad a_{\mathrm{v}} = \frac{\vartheta u_A^4}{64}, \\
Q_A &= R + 4, \qquad N_A = 28md^2, \\
S_A &= \max\left\{1, \rho^{-1} \log_+ \frac{3C_c H_1 Q_A}{\rho a_{\mathrm{v}}}\right\}, \\
K_A &= \left\lceil S_A \max\left\{1, \left(\frac{6H_1 S_A E_{Q_A}(S_A)}{a_{\mathrm{v}}}\right)^2\right\}\right\rceil, \qquad \tau_A = S_A/K_A, \\
V_A &= H_1^2 S_A^2 Q_A^2 e^{C_D S_A}, \qquad n_A = \left\lceil \max\{1, 144 V_A / a_{\mathrm{v}}^2\}\right\rceil,
\end{aligned} \tag{67}$$

and let B_A be the smallest positive odd integer satisfying $2^{-B_A} \leq p_a/N_A$.

At each required query point $y \in \mathbb{R}^d$, estimate $f_\eta(y)$ by the median of B_A independent group means, each formed from n_A independent samples of

$$V_y = -\tau_A \sum_{j=0}^{K_A - 1} \{h(Y_j^y) - h(Y_j^0)\}, \tag{68}$$

where Y_j^y and Y_j^0 share identical Brownian increments within each sample. Substitute these value estimates into the forward-difference stencils

$$\mathcal{D}_{\mathbf{i}, u_A} f_\eta(x) = u_A^{-j} \sum_{\epsilon \in \{0,1\}^j} (-1)^{j - |\epsilon|} f_\eta\left(x + u_A \sum_{\ell=1}^{j} \epsilon_\ell e_{i_\ell}\right) \tag{69}$$

for all multi-indices $\mathbf{i}$ of the forms (i,j), (i,j,j), and (i,i,j,j) with $1 \leq i, j \leq d$, and insert the resulting derivative approximations into (5) with drift $b(x_k)$. Then the resulting estimates $\widehat{a}_k$ satisfy

$$\mathbb{P}_{\mathrm{aux}}\left\{\max_{k<m} |\widehat{a}_k - a_\eta(x_k)| > e_a \,\middle|\, (x_k)_{k<m}\right\} \leq p_a, \tag{70}$$

using at most $2N_A n_A B_A K_A$ Euler updates.

Proof. Because $f_\eta(0) = 0$, (88) gives the representation

$$f_\eta(y) = -\int_0^\infty \{\mathcal{P}_t^\eta h(y) - \mathcal{P}_t^\eta h(0)\}\, dt.$$

Every stencil point satisfies $|y| \le R + 4u_A \le Q_A$, so truncating the integral at S_A incurs a bias of at most $C_c H_1 Q_A e^{-\rho S_A}/\rho \le a_{\mathrm{v}}/3$. By (57), replacing the truncated integral by $\mathbb{E}V_y$ introduces an additional discretization bias bounded by

$$2H_1 S_A E_{Q_A}(S_A)\sqrt{\tau_A} \le a_{\mathrm{v}}/3.$$

The synchronous coupling bound $\mathbb{E}|Y_j^y - Y_j^0|^2 \le |y|^2 e^{C_D j \tau_A}$ yields $\mathbb{E}V_y^2 \le V_A$. Hence Chebyshev's inequality bounds the probability that a group mean deviates from $\mathbb{E}V_y$ by more than $a_{\mathrm{v}}/3$ by $1/16$, and the median-of-means argument bounds the failure probability at a single query point by $2^{-B_A} \le p_a/N_A$.

Across all m states, the total number of stencil point occurrences is at most $m(4d^2 + 8d^2 + 16d^2) = N_A$. By a union bound, every value estimate is accurate within a_{v} simultaneously with probability at least $1 - p_a$. Repeated coordinate indices in (69) require no special modification: the sum of absolute stencil weights is $2^j u_A^{-j}$, and the integral representation of forward differences together with the mean-value theorem gives

$$|\mathcal{D}_{\mathbf{i},u_A} f_\eta(x) - D_{\mathbf{i}} f_\eta(x)| \le j u_A G_{j+1}, \qquad j = 2, 3, 4.$$

Combining the discretization and value-estimation errors shows that every approximated partial derivative entry has error at most

$$j u_A G_{j+1} + 2^j a_{\mathrm{v}}/u_A^j \le \vartheta/3 + \vartheta/4 < \vartheta.$$

Finally, the sum of absolute coefficients multiplying the derivative entries in $\mathcal{A}_2 f_\eta(x_k)$ is bounded by C_A: indeed, $\|b(x_k)\|_1^2/2 \le d M_b(R)^2/2$ for the Hessian contraction, $(\delta/2)d\|b(x_k)\|_1 \le (\delta/2)d^{3/2} M_b(R)$ for the third-order contraction, and $(\delta^2/8)d^2$ for the fourth-order contraction. Multiplying by ϑ yields $|\widehat{a}_k - a_\eta(x_k)| \le C_A \vartheta < e_a$, proving (70). □

For fixed dimension and model envelopes, $a_{\mathrm{v}}^{-1} \le C(1+R^2)^5 e_a^{-5}$ and $S_A = O(1 + \log(1+R) + \log(e_a^{-1}))$, so all exponential terms in (67) grow polynomially in $1 + R$ and e_a^{-1}. Consequently, for a model-dependent polynomial P_a, the total number of Euler updates is bounded by

$$C m\{1 + \log(m/p_a)\}\, P_a(1 + R, e_a^{-1}). \tag{71}$$

Differencing against the trajectory started at the origin eliminates the unknown invariant mean θ_η from every simulation, while the derivative envelopes G_j enter solely through the deterministic error certificate rather than through higher-order pathwise tangent equations.

6.5 Tail probabilities and confidence intervals

For $j \in \{0, 1\}$, define the statistic with numerical normalization

$$\widehat{W}_{m,\eta}^{(j)} = \frac{\sqrt{T}\left\{\bar{h}_m + j\eta\, m^{-1} \sum_{k<m} \widehat{a}_k - \widehat{\theta}_\eta\right\}}{\sqrt{\delta \widehat{Y}_m}}, \tag{72}$$

where $\widehat{Y}_m$ is the strictly positive output of Proposition 6.3, and the correction simulations are omitted when $j = 0$. When constructing a pivot for inference on the unknown invariant mean, one sets $\widehat{\theta}_\eta = \theta_\eta$; if instead a numerical approximation $\widehat{\theta}_\eta$ is substituted, its accuracy is supplied as an external condition

$$\mathbb{P}\{|\widehat{\theta}_\eta - \theta_\eta| > e_c\} \le p_c. \tag{73}$$

The numerical center may use the SGLD trajectory, auxiliary randomness, or both. To construct confidence intervals for θ_η, take $e_c = p_c = 0$ and invert the pivot as in Corollary 6.6.

Theorem 6.5 (Relative tails with numerical normalization)**.** *Suppose Assumption 2.1 and the joint asymptotic conditions (7) of Theorem 2.2 hold for $j \in \{0, 1\}$ with $\gamma = j + 1$. Construct $\widehat{Y}_m$ via Theorem 6.1 and Proposition 6.3 with target relative tolerance $\beta \in (0, 1/2)$, gradient tolerance $e_g = \beta^2$, and failure budget p_g. For $j = 1$, construct $(\widehat{a}_k)_{k<m}$ via Theorem 6.4 with tolerance $0 < e_a \le 1$ and budget p_a; for $j = 0$, set $e_a = p_a = 0$. Assume (73) when a numerical center is used. If*

$$r^2\beta \to 0, \qquad r\sqrt{T}(\eta e_a + e_c) \to 0, \qquad \frac{p_g + p_a + p_c}{\overline{\Phi}(r)} \to 0, \tag{74}$$

then the computed statistic (72) satisfies both relative Gaussian tail approximations of Theorem 2.2 uniformly on $0 \le x \le r$, including for non-strict inequalities. Moreover, all gradient and correction routines terminate in finitely many operations on every finite observed trajectory.

Proof. By the proof of Theorem 2.2 (applied on $[0, r + 1]$), there exists a trajectory event $\mathcal{G}_J$ with $\mathbb{P}(\mathcal{G}_J^c) = o(\overline{\Phi}(r))$ on which $Y_m \ge \mu_*/2 > 0$, where $\mu_* > 0$ is guaranteed by Lemma A.3. Since $r \to \infty$, the first condition in (74) implies $\beta \to 0$, so $16\beta^2 \le \mu_*/2$ eventually.

Intersect $\mathcal{G}_J$ with the gradient accuracy event and, when applicable, the correction and center accuracy events. Integrating the conditional gradient and correction bounds over the path, then using (73) and the union bound, gives a total failure probability of at most $o(\overline{\Phi}(r)) + p_g + p_a + p_c$. On this intersection, Proposition 6.3 ensures acceptance of $\widehat{Y}_m^H$ with $|\widehat{Y}_m/Y_m - 1| \le \beta$, while the additive numerator perturbation divided by $\sqrt{\delta Y_m}$ is bounded in magnitude by $C\sqrt{T}(\eta e_a + e_c)$. Applying Lemma 5.1 to $W_{m,\eta}^{(j)}$ on $[0, r + 1]$ shows that all additional relative tail errors vanish under (74). The opposite and non-strict tails follow identically, and finite termination on every finite trajectory holds by construction. □

The group counts grow logarithmically in m/p_g and, when a correction is used, in m/p_a. On $\mathcal{G}_J$ the observed radius satisfies $R = O(\sqrt{d + r^2 + \log(T + e)})$. Equations (62) and (71) then bound the update counts in terms of this radius, the requested tolerances and the failure budgets. For the specific schedule and tolerances below, choosing each auxiliary failure budget proportional to e^{-2r^2} gives a polynomial bound in η^{-1}.

For instance, in the bias-corrected regime ($j = 1$), setting $L = \log(1/\eta)$ and choosing

$$T \asymp \eta^{-2}L, \qquad r = \eta^{-2/5}, \qquad \beta = \eta, \qquad e_a = \eta, \qquad e_c = p_c = 0,$$

with combined failure budget $p_g + p_a \le e^{-2r^2}$, gives

$$r^2\beta = \eta^{1/5} \to 0, \qquad r\sqrt{T}(\eta e_a + e_c) = O(\eta^{3/5}\sqrt{L}) \to 0,$$

so Theorem 6.5 applies with explicit gradient tolerance $e_g = \eta^2$. Inverting the resulting pivot yields fully observable confidence intervals for θ_η.

Corollary 6.6 (Confidence intervals for the diffusion mean)**.** *Under the hypotheses of Theorem 6.5 with $e_c = p_c = 0$, define the computable point estimator and closed confidence interval*

$$\widehat{c}_m^{(j)} = \bar{h}_m + j\eta\, m^{-1} \sum_{k<m} \widehat{a}_k, \qquad I_m^{(j)}(x) = \left[\widehat{c}_m^{(j)} - x\sqrt{\frac{\delta \widehat{Y}_m}{T}}, \widehat{c}_m^{(j)} + x\sqrt{\frac{\delta \widehat{Y}_m}{T}} \right].$$

Then

$$\sup_{0\le x\le r} \left| \frac{\mathbb{P}\{\theta_\eta \notin I_m^{(j)}(x)\}}{2\overline{\Phi}(x)} - 1 \right| \longrightarrow 0,$$

and constructing $I_m^{(j)}(x)$ requires no evaluation of $\theta_\eta = \pi_\eta h$.

Proof. Setting $\widehat{\theta}_\eta = \theta_\eta$ in (72) and noting that $\widehat{Y}_m > 0$ everywhere, the non-coverage event decomposes into the disjoint union $\{\widehat{W}_{m,\eta}^{(j)} > x\} \cup \{\widehat{W}_{m,\eta}^{(j)} < -x\}$. Summing the two strict-tail estimates of Theorem 6.5 yields the uniform limit, including at $x = 0$. □

7 Examples and the variance boundary

7.1 A nonconvex model with asymmetric gradient noise

In one dimension, a double-well potential can satisfy the weak-dissipativity condition in Assumption 2.1 while failing global strong monotonicity. Below we construct an explicit model of this type featuring a linearly growing noise scale and a skewed, unbounded innovation distribution.

Proposition 7.1 (A nonconvex potential with linear noise amplitude)**.** *Fix $a > \kappa > 0$, $\alpha \neq 0$, and $\delta > 0$. In dimension $d = 1$, consider*

$$b(x) = -\kappa x + a\tanh x, \qquad \sigma_0(x) = \alpha x, \qquad h(x) = x + \tanh x + \cos x.$$

Let ζ_0 take values $\sqrt{2}$ and $-1/\sqrt{2}$ with probabilities $1/3$ and $2/3$, respectively. For an independent standard normal random variable G, set

$$\zeta = (\zeta_0 + G)/\sqrt{2}, \qquad Z(x,\zeta) = \alpha x \zeta.$$

Then

$$\mathbb{E}\zeta = 0, \qquad \mathbb{E}\zeta^2 = 1, \qquad \mathbb{E}\zeta^3 = \frac{1}{4}, \qquad \mathbb{E}e^{t\zeta} \le \exp\left(\frac{17t^2}{32}\right) \quad (t \in \mathbb{R}), \tag{75}$$

and Assumption 2.1 holds with

$$K_1 = \kappa/2, \qquad K_2 = 2a^2/\kappa, \qquad K^2 = 17\alpha^2/16, \qquad \Sigma(x) = \alpha^2 x^2.$$

This model therefore satisfies the hypotheses on the coefficients and innovations in Theorem 2.2. Given coefficient evaluations and the envelopes below, Theorem 6.5 also applies under its stated observation, tolerance and failure-probability conditions.

Proof. Direct calculation gives $\mathbb{E}\zeta_0 = 0$, $\mathbb{E}\zeta_0^2 = 1$, and $\mathbb{E}\zeta_0^3 = 1/\sqrt{2}$; independence, Gaussian symmetry, and scaling immediately yield the first three identities in (75).

Recall that any centered random variable U supported on a compact interval of length D satisfies $\log \mathbb{E}e^{tU} \le D^2t^2/8$: differentiating $\log \mathbb{E}e^{tU}$ twice expresses its second derivative as the

variance under the exponentially tilted law, which is at most $D^2/4$ (by shifting to the midpoint of the interval), and integrating twice from $t = 0$ proves the bound. Since ζ_0 is supported on an interval of length $3/\sqrt{2}$, we obtain

$$\mathbb{E}e^{t\zeta} = e^{t^2/4}\mathbb{E}e^{t\zeta_0/\sqrt{2}} \le e^{t^2/4+9t^2/32} = e^{17t^2/32}.$$

Thus ζ has sub-Gaussian variance proxy $17/16$, whence

$$\mathbb{E}e^{tZ(x,\zeta)} \le e^{17\alpha^2x^2t^2/32} \le e^{K^2(1+x^2)t^2/2}, \qquad \mathbb{E}Z(x,\zeta)^2 = \alpha^2x^2.$$

The nonzero third moment $\mathbb{E}\zeta^3 = 1/4$ confirms that the innovation is asymmetric and non-Gaussian, while the Gaussian component G renders its support unbounded.

For the drift, writing $r = x - y$ and applying Young's inequality gives

$$\begin{aligned} r\{b(x) - b(y)\} &= -\kappa r^2 + ar(\tanh x - \tanh y) \\ &\le -\kappa r^2 + 2a|r| \le -\frac{\kappa}{2}r^2 + \frac{2a^2}{\kappa}. \end{aligned}$$

To construct explicit derivative envelopes, define polynomials recursively by $T_0(z) = z$ and $T_{j+1}(z) = (1 - z^2)T_j'(z)$, so that $(d/dx)^j \tanh x = T_j(\tanh x)$ for every $j \ge 1$. Letting C_j denote the sum of the absolute values of the coefficients of T_j, admissible upper envelopes are given by

$$B_j = \kappa\mathbf{1}_{\{j=1\}} + aC_j, \qquad S_j = |\alpha|\mathbf{1}_{\{j=1\}}, \qquad H_j = \mathbf{1}_{\{j=1\}} + C_j + 1$$

through the required derivative orders, while $b(0) = \sigma_0(0) = 0$ and h is nonconstant. □

The mean drift $b = -U'$ corresponds to the double-well potential $U(x) = \kappa x^2/2 - a\log\cosh x$, which satisfies $U''(0) = \kappa - a < 0$ and therefore violates global one-sided Lipschitz contractivity. Moreover, $b + Z = -\partial_x\psi(\cdot,\zeta)$ with random potential $\psi(x,\zeta) = \frac{1}{2}(\kappa - \alpha\zeta)x^2 - a\log\cosh x$ satisfying $\mathbb{E}\psi(x,\zeta) = U(x)$. Finally, the presence of the nonconstant even term $\cos x$ in h precludes any linear relation $h = \theta + cb$ with constants θ, c, ensuring that the canonical Poisson solution f_η is genuinely nonlinear.

7.2 A variance obstruction for the original denominator

The linear model below isolates the role of $r^2\eta \to 0$. Here the gradient noise adds $\eta\sigma^2$ to the innovation variance per unit physical time, the weak-bias correction vanishes, and the normalizer is known exactly.

Proposition 7.2 (Gaussian tail ratio at the variance boundary)**.** *Fix $\kappa, \delta, \sigma > 0$, set $X_0 = 0$, and consider the linear recursion*

$$X_{k+1} = (1 - \kappa\eta)X_k + \sqrt{\eta\delta}\,\xi_{k+1} + \eta\sigma\zeta_{k+1},$$

where $(\xi_k)_{k\ge1}$ and $(\zeta_k)_{k\ge1}$ are independent sequences of independent standard normal random variables. For the linear observable $h(x) = -\kappa x$, the modified diffusion satisfies

$$\theta_\eta = 0, \qquad f_\eta(x) = x, \qquad \mathcal{A}_2 f_\eta = 0, \qquad Y_m = 1.$$

Consequently, the uncorrected and bias-corrected statistics coincide and follow an exact centered Gaussian distribution with variance

$$v_{\eta,m} = \frac{1+\eta\sigma^2/\delta}{m}\sum_{l=1}^{m-1}\{1-(1-\kappa\eta)^l\}^2 = 1+\frac{\eta\sigma^2}{\delta}+O(T^{-1}), \qquad T=m\eta, \tag{76}$$

uniformly as $\eta \downarrow 0$ and $T \to \infty$. If, in addition, $\eta T \to \infty$, then for every fixed $c > 0$ and boundary radius $r = c\eta^{-1/2}$,

$$\frac{\mathbb{P}(W^{(j)}_{m,\eta} > r)}{\overline{\Phi}(r)} \longrightarrow \exp\left(\frac{c^2\sigma^2}{2\delta}\right) > 1, \qquad j = 0, 1. \tag{77}$$

Proof. The associated modified diffusion is an Ornstein–Uhlenbeck process with drift $-\kappa x$ and diffusion variance $\delta+\eta\sigma^2$, so $\theta_\eta = 0$. Since $L_\eta x = -\kappa x$ and x is anchored and globally Lipschitz, uniqueness yields $f_\eta(x) = x$, whence $D^2 f_\eta \equiv 0$, $\mathcal{A}_2 f_\eta \equiv 0$, and $Y_m \equiv 1$.

Write $\varrho_\eta = 1-\kappa\eta$ and $\epsilon_k = \sqrt{\eta\delta}\,\xi_k + \eta\sigma\zeta_k$. For sufficiently small η, $0 < \varrho_\eta < 1$, and summing the exact recursion gives

$$X_k = \sum_{i=1}^{k}\varrho_\eta^{k-i}\epsilon_i, \qquad \sum_{k=0}^{m-1}X_k = \frac{1}{\kappa\eta}\sum_{i=1}^{m-1}(1-\varrho_\eta^{m-i})\epsilon_i.$$

Both statistics $W^{(0)}_{m,\eta}$ and $W^{(1)}_{m,\eta}$ therefore equal

$$-\frac{1}{\sqrt{\delta T}}\sum_{i=1}^{m-1}(1-\varrho_\eta^{m-i})\epsilon_i,$$

which is centered Gaussian with variance

$$v_{\eta,m} = \frac{\delta+\eta\sigma^2}{\delta m}\sum_{l=1}^{m-1}(1-\varrho_\eta^l)^2.$$

Moreover,

$$0 \le 1-\frac{1}{m}\sum_{l=1}^{m-1}(1-\varrho_\eta^l)^2 \le \frac{1}{m}+\frac{2}{m(1-\varrho_\eta)} = \frac{1}{m}+\frac{2}{\kappa T},$$

establishing (76).

When $\eta T \to \infty$, we have $v_{\eta,m} = 1+\eta\sigma^2/\delta+o(\eta)$. Using the standard Gaussian tail asymptotic $\overline{\Phi}(z) = e^{-z^2/2}(1+O(z^{-2}))/(\sqrt{2\pi}z)$ as $z \to \infty$, along with $r \to \infty$ and $v_{\eta,m} \to 1$, we find

$$\frac{\mathbb{P}(W^{(j)}_{m,\eta} > r)}{\overline{\Phi}(r)} = \frac{\overline{\Phi}(r/\sqrt{v_{\eta,m}})}{\overline{\Phi}(r)} = \sqrt{v_{\eta,m}}\,\exp\left\{\frac{r^2}{2}(1-v_{\eta,m}^{-1})\right\}(1+o(1)).$$

Substituting $r = c\eta^{-1/2}$ yields (77). □

Proposition 7.2 shows that relative Gaussian tail accuracy fails at the scale $r = c\eta^{-1/2}$ even when both the clock and weak-bias conditions for the corrected statistic are satisfied. For example, when $T \asymp \eta^{-5/2}$, we have $r^2\sqrt{\log(T+e)/T} = O(\eta^{1/4}\sqrt{\log(1/\eta)}) \to 0$ and $r\eta^2\sqrt{T} = O(\eta^{1/4}) \to 0$, whereas $r^2\eta = c^2$ and $\eta T \to \infty$, so the limiting tail ratio in (77) strictly exceeds

one. (On shorter horizons $T \asymp \eta^{-1}$, the $O(T^{-1})$ edge effect is comparable in magnitude to the $O(\eta)$ excess variance.)

In this linear model, replacing the denominator δY_m by the complete variance proxy $(\delta + \eta\sigma^2)Y_m$ reduces the variance of the statistic to $m^{-1}\sum_{l=1}^{m-1}(1-\varrho_\eta^l)^2 = 1+O(T^{-1})$. At $r = c\eta^{-1/2}$, one then has $r^2/T = c^2/(\eta T) \to 0$, restoring a limiting tail ratio equal to one. For general nonlinear observables, the corresponding first-order variance correction is given by v_1^F in (18); extending the moderate-deviation radius beyond $o(\eta^{-1/2})$ with such a corrected denominator would require controlling the second-order variance remainder, higher-order weak bias, and tilted clock fluctuations at the larger scale.

A Effective contraction and global Poisson estimates

Coupling the additive Brownian motion by reflection while coupling the multiplicative noise synchronously yields an explicit exponential Wasserstein contraction rate for the modified diffusion, which we use to truncate the Poisson resolvent integral in Section 6. In addition, short-time derivative bounds derived via stochastic integration by parts control finite-difference errors and extend directly to the tilted comparison diffusions. All contraction rates and derivative envelopes below are expressed explicitly in terms of the model constants, dimensions, and noise intensities.

Throughout this appendix, write

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

where the norm on $D^j\sigma_0$ is the multilinear operator norm with Hilbert–Schmidt output. Decrease $\eta_0 > 0$ if necessary so that $\eta_0 S_1^2 \le K_1$, where K_1, K_2 are the weak-dissipativity constants from Assumption 2.1. We realize the diffusion associated with (3) via the stochastic differential equation

$$dX_t^\eta = b(X_t^\eta)\,dt + \sqrt{\delta}\,dB_t + \sqrt{\eta}\,\sigma_0(X_t^\eta)\,dW_t, \qquad 0 \le \eta \le \eta_0, \tag{78}$$

driven by independent standard Brownian motions B and W in $\mathbb{R}^d$ and $\mathbb{R}^\ell$, respectively. Global Lipschitz continuity guarantees a unique nonexplosive strong solution and defines the Markov semigroup $\mathcal{P}_t^\eta$.

A.1 A common effective contraction rate

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

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

If $K_2 > 0$, fix any $0 < k < \overline{K}$ and set

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

if $K_2 = 0$, set $C_c = 1$ and $\rho = \overline{K}$. Then for every $0 \le \eta \le \eta_0$, the semigroup $\mathcal{P}_t^\eta$ admits a unique invariant probability measure π_η on $\mathbb{R}^d$, with finite first moment, and for all probability measures with finite first moments ν, ν' and all globally Lipschitz functions φ,

$$W_1(\nu\mathcal{P}_t^\eta, \nu'\mathcal{P}_t^\eta) \le C_c e^{-\rho t} W_1(\nu, \nu'), \qquad \mathrm{Lip}(\mathcal{P}_t^\eta\varphi) \le C_c e^{-\rho t}\,\mathrm{Lip}(\varphi). \tag{79}$$

Moreover, the second moments satisfy the uniform bounds

$$D_* = 2K_2 + \frac{4B_0^2}{K_1} + \delta d + 5\eta_0 S_0^2, \qquad M_* = \frac{2D_*}{K_1}, \tag{80}$$
$$\mathbb{E}_x|X_t^\eta|^2 \le e^{-K_1 t/2}|x|^2 + M_*, \qquad \pi_\eta|x|^2 \le M_*.$$

Proof. Couple two solutions X and Y of (78) by reflecting the additive Brownian motion B across the hyperplane orthogonal to $X - Y$ prior to their coupling time while driving both with the same multiplicative Brownian motion W; once the two processes meet, let them evolve together with identical noises. Away from the diagonal $\{X = Y\}$, this coupling is governed by locally Lipschitz coefficients, and path continuity allows gluing at the meeting time. Prior to meeting, set

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

For any C^2 increasing concave function $F : [0, \infty) \to [0, \infty)$, Itô's formula yields the radial generator action

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

Because $F'' \le 0$, the final term is nonpositive, while the bracketed expression is bounded above by both $\overline{L}r$ and $-\overline{K}r + K_2/r$. This follows the reflection coupling method of Eberle [Ebe16], tailored to make the constants explicit in our split-noise setting.

When $K_2 > 0$, define F by $F(0) = 0$ and $F'(r) = \exp\{-a_c \min(r, R_c)^2\}$, so that $e^{-A_c}r \le F(r) \le r$. For $r < R_c$, we have

$$\mathcal{A}F(r) \le (-4\delta a_c + \overline{L})rF'(r) = -krF'(r) \le -ke^{-A_c}F(r),$$

and for $r > R_c$, the second drift bound is at most $-kr$, yielding the same inequality. Since F' is continuous at R_c, Tanaka's formula introduces no local-time contribution at R_c. Localization and passing to the limit establish exponential contraction in the transportation metric $F(|x-y|)$, and equivalence with Euclidean distance gives (79); when $K_2 = 0$, taking $F(r) = r$ directly gives the claim.

Since $\mathcal{P}_t^\eta$ maps probability measures with finite first moment into themselves and is a strict W_1-contraction for sufficiently large t, Banach's fixed-point theorem on the complete metric space $(\mathcal{P}_1(\mathbb{R}^d), W_1)$ yields a unique invariant probability measure π_η, while the semigroup property implies invariance for all $t \ge 0$. For every bounded Lipschitz φ, contraction gives $\mathcal{P}_t^\eta\varphi(x) \to \pi_\eta\varphi$ at each x. Integrating against any invariant probability and using bounded convergence identifies that probability with π_η.

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

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

whence $L_\eta|x|^2 \le -K_1|x|^2/2 + D_*$. Stopped Itô's formula and Gronwall's lemma prove the finite-time bound in (80). Finally, since $\delta_0\mathcal{P}_t^\eta \to \pi_\eta$ weakly as $t \to \infty$ by (79), truncating $|x|^2$ and passing to the limit establishes $\pi_\eta|x|^2 \le M_*$. □

A.2 An integrable semigroup smoothing estimate

To supply explicit derivative envelopes for the numerical certificates, we record a finite recursion based on stochastic integration by parts in the additive Brownian motion, following Elworthy and Li [EL94].

Fix an integer $N \geq 2$ and assume throughout this subsection that $b, \sigma_0 \in C^N$ with finite positive-order envelopes B_j, S_j for $1 \leq j \leq N$. Put $s_j = \sqrt{\eta_0} S_j$, and for $p \geq 2$ define

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

Here a_p governs homogeneous linear variation equations, whereas b_p accommodates inhomogeneous forcing terms: if

$$dY = (A_t Y + F_t)\, dt + (\mathcal{B}_t Y + G_t)\, dW_t, \quad \|A_t\| \leq B_1, \quad \|\mathcal{B}_t v\|_{\mathrm{HS}} \leq s_1 |v|,$$

then Itô's formula and Young's inequality give

$$\frac{d}{dt}\mathbb{E}|Y_t|^p \leq b_p \mathbb{E}|Y_t|^p + \mathbb{E}|F_t|^p + 2(p-1)\mathbb{E}\|G_t\|_{\mathrm{HS}}^p, \tag{81}$$

and b_p may be replaced by a_p when $F_t = G_t = 0$ (derived first via stopping so that integrability is not assumed a priori).

Let $\mathcal{P}_n$ denote the set of all partitions of $\{1, \dots, n\}$, and define constants recursively for $2 \leq n \leq N$ and $p \geq 2$ by

$$\begin{aligned} V_{1,p} &= e^{a_p/p}, \\ F_{n,p} &= \sum_{\substack{\pi \in \mathcal{P}_n \\ |\pi| \geq 2}} B_{|\pi|} \prod_{A \in \pi} V_{|A|, p|\pi|}, \\ H_{n,p} &= \sum_{\substack{\pi \in \mathcal{P}_n \\ |\pi| \geq 2}} s_{|\pi|} \prod_{A \in \pi} V_{|A|, p|\pi|}, \\ V_{n,p} &= e^{b_p/p} \{F_{n,p}^p + 2(p-1) H_{n,p}^p\}^{1/p}. \end{aligned} \tag{82}$$

Because every block $A \in \pi$ on the right-hand side has cardinality strictly less than n, this recursion terminates in finitely many steps.

Given fixed initial unit directions, let $J_A(t)$ denote the spatial flow variation indexed by a nonempty subset $A \subseteq \{1, \dots, N\}$. Differentiating (78) yields linear coefficients Db and $\sqrt{\eta}\, D\sigma_0$; for $|A| \geq 2$, the drift forcing is

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

with an analogous stochastic forcing involving $D^{|\pi|}(\sqrt{\eta}\, \sigma_0)$. Applying Hölder's inequality and (81) inductively yields

$$\sup_{x,\, 0 \leq \eta \leq \eta_0,\, 0 < t \leq 1} \|J_A(t)\|_{L^p} \leq V_{|A|,p}. \tag{83}$$

For $|A| \geq 2$, since $J_A(0) = 0$, its p-th moment is in fact bounded by $tV_{|A|,p}^p$. Localization and difference quotients justify spatial differentiation and finite-time L^p continuity without requiring uniform bounds on b or σ_0 themselves.

For $2 \le n \le N$ and $0 < t \le 1$, define 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 along q with respect to B, differentiating the flow equation shows that

$$D_q^B X_s = (s/t) J_{\{n\}}(s), \qquad \mathbb{E} W_n^2 \le \frac{e^{a_2}}{\delta t}, \tag{84}$$

since $(s/t)J_{\{n\}}(s)$ satisfies the first variation equation with an additional drift term $J_{\{n\}}(s)/t$. Because q is adapted and square-integrable, Malliavin duality gives $\mathbb{E} D_q^B H = \mathbb{E}(H W_n)$ without requiring exponential integrability of $\int_0^t |q_s|^2 \, ds$.

For each $1 \le i < n$, define the residual process

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

Its linear evolution shares the homogeneous coefficients of $J_{\{i,n\}}$, while both its drift and stochastic forcing terms equal those of $J_{\{i,n\}}$ scaled by $1 - s/t \in [0,1]$; hence $\|R_{in}(t)\|_{L^p} \le V_{2,p}$.

Applying Malliavin integration by parts to the all-singleton term in the Faà di Bruno expansion of $D^n \mathcal{P}_t^\eta \varphi$ pairs each derivative $D_q^B J_{\{i\}}$ with the partition whose unique nonsingleton block is $\{i, n\}$. Letting $\mathcal{P}_n^\circ$ denote the collection of partitions in $\mathcal{P}_n$ excluding the all-singleton partition and these $n-1$ pair partitions, we obtain the exact identity

$$\begin{aligned} D^n \mathcal{P}_t^\eta \varphi = \mathbb{E}\{ & D^{n-1} \varphi(X_t)[J_{\{1\}}, \dots, J_{\{n-1\}}] W_n \} \\ & + \sum_{i<n} \mathbb{E} D^{n-1} \varphi(X_t)[J_{\{1\}}, \dots, R_{in}, \dots, J_{\{n-1\}}] \\ & + \sum_{\pi \in \mathcal{P}_n^\circ} \mathbb{E} D^{|\pi|} \varphi(X_t)[J_A : A \in \pi], \end{aligned} \tag{85}$$

which involves Malliavin derivatives of first-order spatial variations only.

Set $P = 2N$ and define

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

Applying Cauchy–Schwarz to W_n and Hölder's inequality to the variation factors in (85) yields the integrable short-time smoothing bound

$$\|D^n \mathcal{P}_t^\eta \varphi\|_\infty \le K_n t^{-1/2} \sum_{j=1}^{n-1} \|D^j \varphi\|_\infty, \qquad 0 < t \le 1. \tag{86}$$

For a C^{n-1} input with bounded positive-order derivatives, apply (85) to smooth mollifications. The derivatives of the mollifications converge locally uniformly through order $n-1$. The flow moment bounds (83), the weight bound (84) and continuity of the variations in the initial state justify passage to the limit on compact sets and give continuous derivatives. Equation (80) controls the linearly growing function values. This extends (86) to the stated class of inputs.

For completeness, we also record an explicit time-one smoothing constant for globally Lipschitz inputs. Write

$$\|\varphi\|_{[r]} = \sum_{j=1}^{r} \|D^j\varphi\|_\infty, \qquad A_r = \sum_{j=1}^{r} \sum_{\pi\in\mathcal{P}_j} \prod_{B\in\pi} V_{|B|,P}, \qquad \mathsf{S}_N = \prod_{r=1}^{N-1} \{A_r + \sqrt{N-1}\, K_{r+1}\}.$$

The classical variation formula implies $\|\mathcal{P}_t^\eta\varphi\|_{[r]} \le A_r\|\varphi\|_{[r]}$ for $0 < t \le 1$. Partitioning $[0,1]$ into $N-1$ subintervals of length $(N-1)^{-1}$ and gaining one order of differentiability per subinterval via (86) yields, for any globally Lipschitz C^1 function φ,

$$\|\mathcal{P}_1^\eta\varphi\|_{[N]} \le \mathsf{S}_N \,\mathrm{Lip}(\varphi). \tag{87}$$

For a unit direction v, apply the additive-noise control (84) at time s, using the first variation in direction v, and denote its stochastic integral by W_v. Integration by parts gives $D_v\mathcal{P}_s^\eta\varphi(x) = \mathbb{E}_x[\varphi(X_s)W_v]$ for smooth inputs. Mollifications of a Lipschitz function converge uniformly to it. The uniform L^2 bound for W_v therefore gives uniform convergence of the first derivatives at each fixed $s > 0$, so $\mathcal{P}_s^\eta\varphi$ is C^1 for $0 < s \le 1$. Take $s < 1$ and apply (86) on $N-1$ subintervals of the remaining time to obtain $\mathcal{P}_1^\eta\varphi \in C^N$. To retain the constant in (87), apply the original equal-interval estimate to the mollifications, whose Lipschitz constants are at most $\mathrm{Lip}(\varphi)$, and pass to the limit in distributions.

A.3 Canonical Poisson solutions and explicit derivative bounds

Lemma A.2 (Global Poisson bounds)**.** *Under the weak-dissipativity and ellipticity conditions above, fix $N \ge 2$ and suppose $b, \sigma_0 \in C^N$ with bounded positive-order derivatives through order N. Let $\varphi \in C^{N-1}$ have finite derivative envelopes $D_j^\varphi = \|D^j\varphi\|_\infty$ for $1 \le j \le N-1$ (allowing linear growth of φ). Then there exists a unique globally Lipschitz C^N solution, anchored at zero, to the Poisson equation*

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

given by the absolutely convergent integral representation

$$F_\eta^\varphi(x) = -\int_0^\infty \{\mathcal{P}_t^\eta\varphi(x) - \mathcal{P}_t^\eta\varphi(0)\}\, dt. \tag{88}$$

Moreover, the following explicit bounds hold uniformly for $0 \le \eta \le \eta_0$:

$$\begin{aligned} \|DF_\eta^\varphi\|_\infty &\le \mathcal{G}_1(D^\varphi) := (C_c/\rho)D_1^\varphi, \\ \|D^nF_\eta^\varphi\|_\infty &\le \mathcal{G}_n(D^\varphi) := 2K_n \sum_{j=1}^{n-1} D_j^\varphi + (C_c/\rho)\mathsf{S}_N D_1^\varphi, \qquad 2 \le n \le N. \end{aligned} \tag{89}$$

In particular, under Assumption 2.1, taking $N = 6$ and using $D_j^h \le H_j$ $(1 \le j \le 5)$ in the monotone bounds (89) gives $\|D^j f_\eta\|_\infty \le G_j := \mathcal{G}_j(H)$ for $1 \le j \le 6$. Setting $\theta_\eta = \pi_\eta h$, $w_\eta = |\nabla f_\eta|^2$, $\mu_\eta = \pi_\eta w_\eta$, and

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

the second anchored Poisson solution u_η solving $L_\eta u_\eta = w_\eta - \mu_\eta$ likewise satisfies $\|D^j u_\eta\|_\infty \le U_j := \mathcal{G}_j(W)$ for $1 \le j \le 6$.

Proof. Exponential contraction (79) and the finite invariant first moment ensure absolute convergence of $F_\eta^\circ(x) = -\int_0^\infty (\mathcal{P}_t^\eta \varphi(x) - \pi_\eta \varphi)\, dt$, and subtracting $F_\eta^\circ(0)$ gives (88). Integrating (79) yields the first-derivative bound in (89). For $2 \le n \le N$, integrating (86) over $t \in (0,1)$ contributes $2K_n \sum_{j=1}^{n-1} D_j^\varphi$, while on $[1,\infty)$ applying (87) to $\mathcal{P}_{t-1}^\eta \varphi$ followed by (79) contributes $(C_c/\rho)\mathsf{S}_N D_1^\varphi$, justifying differentiation under the integral sign and proving (89).

The semigroup resolvent identity shows that F_η^φ solves the Poisson equation classically. If v is the difference of two globally Lipschitz classical solutions, then $L_\eta v = 0$; by localized Itô's formula, boundedness of ∇v, linear growth of σ_0, and (80), we have $\mathcal{P}_t^\eta v = v$ for all $t \ge 0$, so $|v(x) - v(y)| \le C_c e^{-\rho t} \operatorname{Lip}(v)|x - y| \to 0$ as $t \to \infty$, and $v(0) = 0$ forces $v \equiv 0$. Applying these results first to $\varphi = h$ and then to $\varphi = w_\eta$ (using the Leibniz rule $\|D^j w_\eta\|_\infty \le W_j$) yields the bounds for f_η and u_η. □

Note that Lemma A.2 applies with $N = 4$ whenever $b, \sigma_0 \in C^4$ and $\varphi \in C^3$ have bounded positive-order derivatives; thus an auxiliary tilted drift need not preserve all six derivatives of the original model in order to admit uniform fourth-order Poisson bounds.

A.4 Energy nondegeneracy and global cutoffs

Lemma A.3 (Positive limiting energy)**.** *For the nonconstant observable h 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 exists a model-dependent constant $\mu_* > 0$ such that $\mu_\eta \ge \mu_*$ for all sufficiently small η (no explicit numerical lower bound in terms of upper derivative envelopes alone is claimed).

Proof. Coupling X^η and X^0 from the same initial state with the same additive Brownian motion B gives

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

Starting at the invariant law π_η and applying (80) yields $W_1(\pi_\eta, \pi_\eta \mathcal{P}_t^0) \le C_t\sqrt{\eta}$ with $C_t^2 = 2(S_0^2 + S_1^2 M_*) t e^{2B_1 t}$. Fixing $t_* > 0$ so that $C_c e^{-\rho t_*} \le 1/2$, the triangle inequality and contraction of $\mathcal{P}_{t_*}^0$ give $W_1(\pi_\eta, \pi_0) \le 2C_{t_*}\sqrt{\eta}$, and hence $\theta_\eta \to \theta_0$.

Because f_η and its derivatives through order three are uniformly bounded on compact sets, every sequence $\eta_k \downarrow 0$ admits a subsequence along which f_{η_k} converges in C^2_{loc} to a globally Lipschitz function $\bar{f}$ anchored at zero. Passing to the limit in $L_{\eta_k} f_{\eta_k} = h - \theta_{\eta_k}$ shows that $L_0 \bar{f} = h - \theta_0$, so uniqueness in Lemma A.2 forces $\bar{f} = f_0$ and establishes convergence of the entire family as $\eta \downarrow 0$. Since $w_\eta = |\nabla f_\eta|^2 \le G_1^2$ converges locally uniformly to $|\nabla f_0|^2$ and (π_η) is tight, we obtain $\mu_\eta \to \pi_0 |\nabla f_0|^2 = \mu_0$.

At $\eta = 0$, fix any $t > 0$ and a smooth path from the initial point into a prescribed open ball. Subtracting its initial point and integrated drift and dividing by $\sqrt{\delta}$ gives a continuous driving path starting at zero. Brownian motion assigns positive probability to every uniform neighborhood of this path. The solution map is continuous in the driving path by Gronwall's inequality, so the transition law, and hence the invariant law π_0, has full support. If $\mu_0 = 0$, continuity of ∇f_0 would force $\nabla f_0 \equiv 0$ everywhere, whence $h - \theta_0 = L_0 f_0 \equiv 0$, contradicting the assumption that h is nonconstant. Thus $\mu_0 > 0$. □

Lemma A.4 (Cutoffs retaining all derivative bounds)**.** *Let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy $F(0) = 0$ and $\|D^j F\|_\infty \le C_j$ for $1 \le j \le N$. Fix a smooth cutoff function $\chi : \mathbb{R}^d \to [0,1]$ equal to 1 on B_2*

and 0 *outside* B_3. *For* $R \geq 1$, *define* $\chi_R(x) = \chi(x/R)$ *and* $F_R = \chi_R F$. *Then* $F_R = F$ *on* B_{2R}, *and all derivatives of orders* $1, \ldots, N$ *are bounded uniformly in* $R \geq 1$: *writing* $c_{\chi,j} = \|D^j \chi\|_\infty$,

$$\|D^j F_R\|_\infty \leq C_j + \sum_{a=1}^{j-1} \binom{j}{a} c_{\chi,a} C_{j-a} + 3 c_{\chi,j} C_1, \qquad 1 \leq j \leq N. \tag{90}$$

In particular, this applies with $N = 6$ *to* f_η *and* u_η, *and with the corresponding derivative order to other anchored Poisson solutions.*

Proof. Anchoring and the gradient bound imply $|F(x)| \leq C_1 |x|$. Applying the Leibniz rule to $D^j(\chi_R F)$ and using $R^{-a} \leq 1$ bounds every term with at least one derivative falling on F by the first two terms in (90), while the term with all j derivatives falling on χ_R is supported in B_{3R} and bounded by $3 c_{\chi,j} C_1 R^{1-j} \leq 3 c_{\chi,j} C_1$ for $j \geq 1$. □

Because Lemma A.4 provides global derivative bounds across all of $\mathbb{R}^d$, conditional expectations of $F_R(X_{k+1})$ can be expanded over the complete one-step transition law without truncating jumps that exit the localization ball.

B A Gaussian-mixture check of the local coefficients

For the following one-dimensional test model, all one-step transition moments can be computed in closed form, giving a check, within this model, of the weak-expansion, variance, and third-moment coefficients in Lemmas 3.1 and 3.3.

In dimension $d = 1$, consider

$$F(x) = \cos x, \qquad b(x) = -x, \qquad Z(x, \zeta) = x\zeta, \qquad \mathbb{P}(\zeta = -1) = \frac{2}{3}, \quad \mathbb{P}(\zeta = 2) = \frac{1}{3}.$$

Then $\mathbb{E}\zeta = 0$, $\mathbb{E}\zeta^2 = 2$, and $\mathbb{E}\zeta^3 = 2$. Given a pre-step state $x \in \mathbb{R}$, the next state X' is distributed as the two-component Gaussian mixture

$$X' = \begin{cases} x(1 - 2\eta) + \sqrt{\eta\delta}\, \xi, & \text{with probability } 2/3, \\ x(1 + \eta) + \sqrt{\eta\delta}\, \xi, & \text{with probability } 1/3, \end{cases}$$

where $\xi \sim \mathcal{N}(0, 1)$. In particular, $\Sigma(x) = 2x^2$ and

$$L_\eta F(x) = x \sin x - \tfrac{1}{2}(\delta + 2\eta x^2) \cos x.$$

Here we work directly with the uncut test function F, which possesses globally bounded derivatives of all orders. The forcing $L_\eta F$ depends on η and has unbounded derivatives. This example tests the one-step expansions.

For each integer $k \geq 1$, define

$$C_k(\eta) = e^{-k^2 \delta \eta / 2} \frac{2 \cos\{kx(1 - 2\eta)\} + \cos\{kx(1 + \eta)\}}{3}. \tag{91}$$

Using trigonometric identities and Gaussian characteristic functions, the first three raw moments of $F(X') = \cos X'$ are given exactly by

$$M_1 = C_1, \qquad M_2 = (1 + C_2)/2, \qquad M_3 = (3C_1 + C_3)/4, \tag{92}$$

from which the conditional variance and third central moment follow via $\operatorname{Var} F(X') = M_2 - M_1^2$ and $\mathbb{E}[(F(X') - M_1)^3] = M_3 - 3M_1M_2 + 2M_1^3$. Setting

$$\lambda_{k,-} = -k^2\delta/2 - 2ikx, \qquad \lambda_{k,+} = -k^2\delta/2 + ikx,$$

the coefficient of η^n in the power-series expansion of $C_k(\eta)$ is explicitly

$$\frac{1}{3n!} \operatorname{Re}\left[e^{ikx}\{2\lambda_{k,-}^n + \lambda_{k,+}^n\}\right],$$

and finite Cauchy products of these series yield the exact Taylor coefficients of the centered moments without numerical differentiation.

Residual orders and numerical values

Define $B_\eta = M_1 - F(x) - \eta L_\eta F(x)$ and $d = F(X') - \mathbb{E}F(X')$. Across the geometric step-size grid $\eta = 10^{-1}, 10^{-2}, \ldots, 10^{-12}$, consider the three asymptotic residuals

$$\begin{aligned} R_b(\eta) &= B_\eta - \eta^2 \mathcal{A}_2 F(x), \\ R_v(\eta) &= \eta^{-1} \operatorname{Var} F(X') - \delta(F'(x))^2 - \eta v_1^F(x), \\ R_3(\eta) &= \eta^{-2}\mathbb{E}d^3 - 3\delta^2 F''(x)(F'(x))^2, \end{aligned}$$

whose theoretical convergence rates in η are 3, 2, and 1, respectively. At $x = \delta = 1$, the analytical formulas give

$$\begin{aligned} \mathcal{A}_2 F(1) &= -0.62334885710450064735\ldots, \\ v_1^F(1) &= -0.05526071768889609865\ldots, \\ 3F''(1)(F'(1))^2 &= -1.14772110185143888100\ldots. \end{aligned}$$

We evaluated the theoretical coefficients and the characteristic-function expansions at $x = \delta = 1$ with 80 decimal digits of working precision. The largest discrepancy was less than 2.11×10^{-81}. The supplementary script `check_local_expansions.py` uses `mpmath` to evaluate (91)–(92). The observed orders from the last two step sizes,

$$\frac{\log|R(10^{-12})/R(10^{-11})|}{\log(10^{-12}/10^{-11})}$$

were 2.99999999999, 1.99999999999, and 0.99999999998, respectively. In a separate calculation, the script compares the scaled weak and variance residuals with the next coefficients of their Taylor expansions.

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.

[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.