Stein Guidance without Gradients: Discretization Uniform Diffusion Inference Time Scaling beyond Particle SMC

Anonymous authors
Paper under double-blind review

Abstract

Diffusion inference can scale through more solver steps or larger trajectory ensembles. These define the numerical depth n and statistical width N. In Sequential Monte Carlo (SMC) with full resampling at every solver step, the resampling depth also equals n. A scalable method should decouple depth and width: the number of trajectories needed for a prescribed sampling accuracy should remain bounded as numerical depth increases. Repeated resampling can violate this principle: we construct a Brownian example with single-output total variation (TV) error $\Theta(n/N)$ for $2 \le n \le N$. This raises a central question: *how can we decouple sequential and statistical complexity, so that increasing n does not force N to grow merely to maintain accuracy?*

We propose Stein Guidance, a family of derivative-free guidance methods based on Gaussian Stein's identity, with whole-path independence Metropolis–Hastings (MH) correction. Each proposal uses value-weighted candidate selection along one trajectory. The value function need not be exact. Approximate values steer trajectories toward promising regions, while a derivative free MH correction applied after generation can asymptotically remove the bias introduced by the approximation. Generating proposal paths independently avoids the shared ancestry caused by resampling across particles. Under uniform value and spatial bounds, our theory guarantees that the number of complete proposals needed for a prescribed finite-grid sampling accuracy stays bounded or even decreases as numerical depth increases. Stein guidance also parallelizes value queries within each trajectory. At a fixed per-proposal query budget, the pool-noise term in the error bound has the same order for every pool size $K \ge 2$.

1 Introduction

Diffusion models (Sohl-Dickstein et al., 2015; Song et al., 2021) learn and implement a progressive denoising process driven by a score function to generate high-fidelity samples. Such a paradigm empowers and transforms broad engineering and scientific domains. However, in these applications, the generated samples need to satisfy complex conditions. For example, in text-to-image generation, the generated images need to comply with the prompt. In robotics and control, the generated action sequence needs to achieve large task reward or high success rates. Inference time scaling—directly applying training-free modifications to the denoising process—is widely adopted to steer the generation toward desired downstream reward (Uehara et al., 2025), explore more candidates (Ma et al., 2025; Skreta et al., 2025; Chen et al., 2025), or correct sampling bias (Singhal et al., 2025; Zhu & Lu, 2026; He et al., 2026; Wang et al., 2025; Ren et al., 2026), and meanwhile to avoid computational overhead from fine-tuning.

A growing body of inference time methods revolves around adding a *guidance* term to the original score function. The well-known classifier guidance computes an external gradient of a classifier logit to satisfy categorical conditions. For broader complex conditions, we assess the quality of generated samples by a reward function ϕ. Thus, steering can be formulated as sampling from a reward tilted law

$$\Pi(\mathrm{d}X_{0:T}) = Z^{-1}\phi(X_T)P(\mathrm{d}X_{0:T}), \qquad Z = \mathbb{E}_P[\phi(X_T)], \tag{1}$$

where P is the path measure of the original diffusion denoising process $X_{0:T}$. In this case, stochastic optimal control provides principled derivations for guidance. One notable example is the Doob's

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

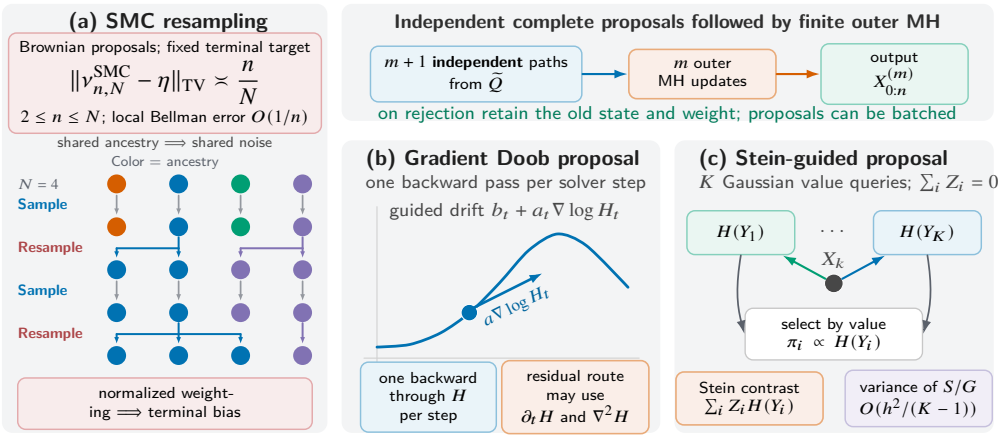

Figure 1: **Guidance depth and efficient correction.** (a) Repeated resampling accumulates population fluctuations; normalized weighting converts them into terminal bias. (b) Direct Doob guidance requires value gradients. (c) Centered Stein pools provide derivative-free guidance and low-variance estimates of local normalizers. The same queries give path weights for outer MH correction with independent complete proposals (top).

h-transform, which defines the guidance term as $\nabla \log h_t(x)$ with $h_t(x) = \mathbb{E}_P[\phi(X_T) \mid X_t = x]$ (Tang & Xu, 2025; Nguyen et al., 2025; Denker et al., 2024; Zhu et al., 2026). Beyond direct gradient, implicit guidance includes other approaches like *importance resampling* and *proposal and rejection*.

Despite the mathematical cleanness, efficiently and accurately computing the guidance term is challenging. The performance of classifier guidance heavily relies on the accuracy of the external classifier. Doob's h-function involves the intractable expectation over the path measure P. Therefore, practical implementations of guidance rely on stochastic approximation, and inevitably lead to sampling biases. These biases distort the generated distribution and are empirically tied to the well-recognized failure mode—reward hacking (Dandapanthula & Boffi, 2026).

In this regard, *correction* is an essential alternative to stay faithful to the reward tilted law. Correction needs extra computation through maintaining a pool of particles or fresh rollouts, of which the size we term as the *statistical width*. Similarly, we term the number of guidance steps we erect on a base diffusion denoising process as *guidance depth*. Roughly speaking, the depth distributes the guidance effort, while the width controls correction accuracy. A popular mechanism for guidance and correction is Sequential Monte Carlo (SMC) (Wu et al., 2023; Cardoso et al., 2024; Skreta et al., 2025; Wang et al., 2026), which simulates a batch of particles, reweights and resamples them by the reward during evolution. There are two critical design choices: 1) the correction batch size N (i.e. statistical width), and 2) the number of reweighting steps n (i.e. guidance depth). A large depth makes guidance easier per step, yet repeated resampling creates shared ancestry and accumulates population fluctuations. Normalized weighting can turn these fluctuations into terminal bias. We exhibit a Brownian example with single-output TV error $\Theta(n/N)$ for $2 \le n \le N$ (Theorem 1). Thus increasing depth can require greater width to maintain accuracy. This motivates our central question:

Can guidance depth be increased without forcing statistical width to grow?

To achieve this goal, increasing the number of guidance steps should not lead to stronger correlations between trajectories. This motivates rejection based sampling: generate proposal paths independently and correct them by acceptance or rejection.

The challenge remains is to retain the benefits of guidance while correcting the bias introduced by an imperfect value function at low computational cost. Explicit guidance requires value gradients, and its correction either need gradients or time derivatives and Hessians. We therefore propose Stein Guidance. Its key idea is simple: **selecting centered candidates by their values approximates gradient guidance in expectation, without evaluating derivatives.** Stein's identity makes this connection precise. At each step, we query a centered pool of Gaussian candidates in parallel and select one with probability proportional to its value. Centering cancels the linear noise in the average

Table 1: Guidance and correction methods (conditions in Section G.4).

Method	Correction with approximate values	Stable with guidance depth?	Parallelizable?
SVDD (Li et al., 2024)	✗	✓	✓
SMC	✓	✗	✓
DOIT, MC (Zhu et al., 2026)	✗	✗†	✓ (M)
Stein ($K = 2$) with outer MH	✓	✓	fixed at 2
Stein with outer MH	✓	✓	✓ (K)

At fixed statistical width, we compare how output-TV bounds depend on guidance depth. For Stein, see Theorem 4; for the SMC counterexample, see Theorem 1. †At fixed M, DOIT's TV bound can worsen as smaller noise scales amplify the rollout error bound (Zhu et al., 2026, Lemma 5.2 and Theorem 5.4).

value. This average becomes more stable as the step size decreases or the pool grows. The same queries also give an explicit path weight for correction with approximate values. We then correct complete paths with another outer MH rejection sampling for m steps. Under uniform value and regularity bounds, the finite-grid TV error decays superpolynomially in m, with constants independent of guidance depth (Theorem 4). The bound separates accumulated Bellman error from finite-pool variance. This is consistent with the intuition behind the Doob h-transform: exact Doob transitions need no correction. In our finite-pool implementation, exact Doob values eliminate the Bellman error term, while finer grids and larger pools reduce the remaining pool variance.

Table 1 isolates three complementary benefits: correction with approximate values, weight stability under numerical refinement, and adjustable parallel query width. Stein pools combine all three under the stated assumptions; two queries retain the first two and provide their finite difference origin.

We also demonstrate our theory through experiments. The first synthetic experiment shows that centered pools with just two queries per step match the sampling efficiency of much larger independent pools. In offline RL experiment, our algorithm reduces critic-Q distribution error most effectively under the same statistical width. Further text-to-image examples illustrate improved alignment with multi-object prompts.

2 The Obstruction: Resampling Bias Grows with Depth

Repeated resampling creates shared ancestry and accumulates fluctuations in the particle population. Subsequent normalized weighting can turn these fluctuations into terminal bias. We work in the $O(1/n)$ local Bellman-error regime of Zhu & Lu (2026), which controls the exponential dependence on depth. Even in this regime, full resampling can require width to grow with depth.

Theorem 1 (Brownian SMC marginal bias)**.** *Let $P_h(x,\cdot) = \mathcal{N}(x,h)$, where $h = 1/n$. There exist a smooth reward $\phi > 0$, a value function $H : [0,1] \times \mathbb{R} \to (0,\infty)$ bounded above and away from zero, and constants $C, c_-, c_+ > 0$, all independent of n, N, such that for every $n \geq 2$, $H_k = H(k/n,\cdot)$ satisfies $H_0 \equiv 1$, $H_n = \phi$, and*

$$\max_{0\leq k<n} \sup_x \max\left\{\frac{P_h H_{k+1}(x)}{H_k(x)}, \frac{H_k(x)}{P_h H_{k+1}(x)}\right\} \leq 1 + \frac{C}{n}. \tag{2}$$

Starting at zero, run N-particle SMC with proposal $P_h(x,\mathrm{d}y)$, incremental weight $H_k(y)/H_{k-1}(x)$, and multinomial resampling after every step. The law $\nu^{\mathrm{SMC}}_{n,N}$ of one terminal particle satisfies

$$c_- \frac{n}{N} \leq \left\|\nu^{\mathrm{SMC}}_{n,N} - \eta\right\|_{\mathrm{TV}} \leq c_+ \frac{n}{N}, \qquad 2 \leq n \leq N, \tag{3}$$

where $\eta(\mathrm{d}x) \propto \phi(x)\mathcal{N}(0,1)(\mathrm{d}x)$ is fixed independently of n, N.

Appendix E gives the construction and proof, relating the accumulated fluctuations to a sum of Jensen gaps.

3 Derivative-Free Guidance via Stein's Identity

Repeated SMC resampling creates shared ancestry and population fluctuations. Normalized weighting can turn these fluctuations into terminal bias. To avoid this dependence, we generate trajectory proposals independently and use acceptance or rejection in place of population resampling. The

proposals use approximate Doob values and finite candidate pools, so they do not generally recover the target path law. An outer correction accounts for the error accumulated along each trajectory. We use independence Metropolis–Hastings updates on complete trajectories. This preserves independent proposal generation without introducing shared ancestry during propagation. The complete algorithm has two parts: *proposal generation* (Algorithm 1) and *correction* (Algorithm 2).

We fix $t_k = kh$, $T = nh$, and $X_0 = x_\star \in \mathbb{R}^{d_x}$. We aim to sample from

$$P_n = \delta_{x_\star} \prod_{k=0}^{n-1} P_k, \qquad \Pi_n = Z_n^{-1} \phi(X_n) P_n, \qquad \text{where } Z_n = \mathbb{E}_{P_n} \phi(X_n) \in (0, \infty). \tag{4}$$

Here $P_k(x, \cdot)$ is the Gaussian Euler reference transition corresponding to the discretized diffusion process. It is the law of $x + hb_k(x) + \sqrt{h}\,\sigma_k(x)Z$, with $Z \sim N(0, I_d)$. The values $H_0, \ldots, H_n$ are finite, strictly positive, and measurable, with $H_n = \phi$. Each $G_k = P_k H_{k+1}$ is finite on reachable states. The target normalizer Z_n is finite and positive, as in (4).

The exact local tilt is $K_k^H(x, \mathrm{d}y) = P_k(x, \mathrm{d}y) H_{k+1}(y)/G_k(x)$, and $\Pi_n^H = \delta_{x_\star} \prod_k K_k^H$. It exactly equals the target Π_n when $H_k = G_k$. At a parent x, freeze $\mu = x + hb_k(x)$, $\sigma = \sigma_k(x) \in \mathbb{R}^{d_x \times d}$, and write $H = H_{k+1}$ and $G = G_k(x)$. The value functions $H_0, \ldots, H_n$ and proposal mechanism stay fixed across complete proposals. We reserve m for the number of outer MH updates. Random initialization requires the additional factor in Proposition 22.

Derivative-free guidance for efficient correction. Value gradients can be expensive or unavailable, so we seek guidance and correction using value evaluations alone. One approach targets the local tilt K_k^H through rejection sampling (Golowich et al., 2026) or local MH; another selects a candidate from a finite pool with probabilities proportional to its value (Li et al., 2024; Zhu & Lu, 2026; Golowich et al., 2026; Jiang et al., 2026). With approximate values, even exact local sampling generally leaves path-level bias. Correcting this bias directly requires the local normalizers $G_k = P_k H_{k+1}$, which are generally intractable.

Our design principle is to improve the guided transition and its correction weight together. The augmented construction for local MH retains the initial pool average in its incremental correction weight (Section G.2). Additional local updates refine the transition but do not reduce the conditional fluctuations of this initial average. A candidate pool makes the value queries serve both purposes: selecting the next state and estimating the normalizer used in an exact augmented correction (Zhu & Lu, 2026). Increasing the pool size therefore directly controls a source of variability in the whole-path correction weight.

This motivates our Stein guidance. Stein's identity gives a derivative-free construction based on K parallel value queries (Section 3.1.2). The centered Gaussian noises sum to zero, so the candidates' linear value changes cancel in the pool average. The remaining fluctuation is governed by curvature. Under Assumption 3, its normalized variance is $O(h^2/(K-1))$, compared with $O(h/K)$ for independent candidates. Section 4 shows how this reduces the finite-pool term in the outer MH output-TV bound. For $K = 2$, the centered pool contains two opposite Gaussian perturbations. Their value difference gives a central finite difference along the sampled direction. **With centering, two value queries already recover the leading guidance drift in expectation** (Section 3.1.1).

3.1 Derivative-Free Guidance via Stein's Identity

3.1.1 Warm up: Derivative-Free Guidance via Two Query Finite Difference

To avoid derivative evaluations while still steering the dynamics toward regions of higher reward, a natural approach is to estimate the guidance using finite differences. We show that this guidance can be implemented by selecting between two symmetric queries based on their value function evaluations.

Instead of using a single diffusion proposal $Y = x + hb(x) + \sqrt{h}\,\sigma Z$ with $Z \sim N(0, I_d)$, we construct two symmetric queries $Y^\pm = x + hb(x) \pm \sqrt{h}\,\sigma Z$. We select between them with probabilities $\Pr(J = \pm 1 \mid Z) = H^\pm/(H^+ + H^-)$, where $H^\pm = H(Y^\pm)$. For positive C^2 values, Taylor expansion and averaging over Z give, as $h \to 0$,

$$\mathbb{E}[Y^J] = x + hb(x) + \sqrt{h}\,\sigma \mathbb{E}\left[Z \frac{H^+ - H^-}{H^+ + H^-} \right] = x + h\left[b(x) + \sigma\sigma^\top \nabla \log H(x) \right] + o(h). \tag{5}$$

Thus value-weighted selection recovers the leading Doob guidance drift in expectation. Selecting between two symmetric queries implements derivative-free guidance based on finite differences.

3.1.2 Stein Guidance

We then build on this two query intuition to construct guidance from a pool of parallel queries. For standard Gaussian Z, integration by parts gives Gaussian Stein's identity $\mathbb{E}[Zf(Z)] = \mathbb{E}[\nabla_Z f(Z)]$. Under its differentiability and integrability conditions, this identity lets us estimate the expected gradient from function values alone. Applying it to $f(z) = H(\mu + \sqrt{h}\sigma z)$ gives

$$\frac{1}{\sqrt{h}}\mathbb{E}[ZH(\mu + \sqrt{h}\sigma Z)] = \sigma^\top \mathbb{E}[\nabla H(\mu + \sqrt{h}\sigma Z)], \qquad s_h = \frac{\sigma^\top \mathbb{E}\nabla H(Y)}{G}. \tag{6}$$

Here $Y = \mu + \sqrt{h}\sigma Z$ is a reference candidate. The ideal local tilt has mean $\mu + h\sigma s_h$. We estimate its guidance using values alone by replacing the numerator and normalizer expectations with sample averages. To cancel the leading fluctuation in the normalizer estimate, we center the Gaussian candidates. Draw $G_i^{\text{noise}} \overset{\text{iid}}{\sim} N(0, I_d)$ and set

$$Z_i = \sqrt{\frac{K}{K-1}}(G_i^{\text{noise}} - \bar{G}^{\text{noise}}), \qquad Y_i = \mu + \sqrt{h}\sigma Z_i, \qquad S(U) = \frac{1}{K}\sum_i H(Y_i), \quad K \geq 2. \tag{7}$$

Here $U = (Y_1, \ldots, Y_K)$ is the candidate pool. Every Z_i is standard Gaussian, $\sum_i Z_i = 0$, and $\mathrm{Cov}(Z_i, Z_j) = -I_d/(K-1)$ for $i \neq j$. Correct marginals preserve Stein's identity despite dependence within the pool. Query all K candidates in parallel and select J proportionally to $H(Y_J)$. Then

$$\widehat{v} = \frac{1}{K\sqrt{h}}\sum_i Z_i H(Y_i), \qquad \mathbb{E}\widehat{v} = Gs_h, \qquad \widehat{s} = \frac{\widehat{v}}{S}, \qquad \mathbb{E}[Y_J \mid U] = \mu + h\sigma\widehat{s}. \tag{8}$$

The numerator is unbiased, while the ratio is generally biased. Centering makes $\widehat{v}$ exactly zero for constant values and removes the linear fluctuation of S. For $K = 2$, $Z_1 = (G_1^{\text{noise}} - G_2^{\text{noise}})/\sqrt{2} = -Z_2$, so averaging (8) recovers the two-query expression (5). Thus Stein guidance extends the finite difference construction to parallel queries, including odd K. It is not restricted to averaging independent pairs. The mean local displacement error is $O(h^2/(K-1))$ under the assumptions of Appendix B.5. Independent candidates instead attenuate leading raw guidance by $1 - 1/K$; see (24).

Remark 2 (Why centering helps). Centering the Gaussian proposals improves the approximation order by canceling the linear noise term. A Taylor expansion of the average value gives

$$S(U) = \frac{1}{K}\sum_i H(\mu + \sqrt{h}\,\sigma Z_i) = H(\mu) + \frac{\sqrt{h}}{K}\nabla H(\mu)^\top \sigma \sum_i Z_i + O_{L^2}(h).$$

Centering the proposals removes the leading term at order $\sqrt{h}$, leaving fluctuations of order h and hence variance of order h^2. The centered Gaussian covariance further gives $\mathrm{Var}(S/G) = O(h^2/(K-1))$, compared with $O(h/K)$ for independent candidates.

Algorithm 1: One complete Stein Guidance proposal

Input: $x_\star, h, n, K \geq 2, \{b_k, \sigma_k, H_{k+1}\}_{k=0}^{n-1}$. **Init:** $X_0 \leftarrow x_\star$, $\ell \leftarrow 0$.
for $k = 0, \ldots, n-1$ **do**
  Draw a fresh $U_k = (Y_1, \ldots, Y_K)$ at X_k using (7).
  $v_i \leftarrow H_{k+1}(Y_i)$ $(i = 1, \ldots, K$, in parallel); $S_k \leftarrow K^{-1}\sum_{i=1}^K v_i$.
  $\ell \leftarrow \ell + \log S_k$; **if** $k \geq 1$ **then** $\ell \leftarrow \ell - \log c$.
  Draw $J \sim \mathrm{Categorical}\big((v_i/(KS_k))_{i=1}^K\big)$; $X_{k+1} \leftarrow Y_J$; $c \leftarrow v_J$.
end for
return $(X_{0:n}, R = e^\ell)$.

3.2 Terminal Metropolis–Hastings Correction

Each complete proposal from Algorithm 1 comes with a computable path importance weight R. The weight uses the values already queried for guidance; it needs neither derivatives nor the local normalizers G_k. Let $\widetilde{Q}$ denote the joint law of the candidate pools, selected indices, and path generated by the algorithm. Theorem 12 in Appendix C.1 gives the weight formula and shows that $W = R/Z_n$ defines a probability law $\widetilde{\Pi} = W\widetilde{Q}$. Under $\widetilde{\Pi}$, the selected path $X_{0:n}$ has the target law Π_n.

Whole-path Metropolis correction. Since R/Z_n is the target-to-proposal density ratio on the augmented space, independence MH accepts a fresh complete proposal with probability $1 \wedge R'/R$. The unknown normalizing constant Z_n cancels.

Algorithm 2: Outer independence MH with m updates

Input: $m \geq 0$; fixed inputs to Algorithm 1. **Init:** $(X_{0:n}, R) \leftarrow$ Algorithm 1.
for $j = 1, \ldots, m$ **do**
  $(X'_{0:n}, R') \leftarrow$ Algorithm 1, $u \sim \mathrm{Uniform}(0, 1)$, independently.
  if $u \leq \min\{1, R'/R\}$ **then** $(X_{0:n}, R) \leftarrow (X'_{0:n}, R')$.
end for
return $X_{0:n}$ (or X_n).

The accepted augmented flow is $\widetilde{Q}(\mathrm{d}\omega)\widetilde{Q}(\mathrm{d}\omega') \min\{W(\omega), W(\omega')\}$, which is symmetric. Hence the MH kernel M preserves $\widetilde{\Pi}$ and its exact finite-grid physical marginal. The finite-run law is $\widetilde{Q}\mathsf{M}^m$; Section 4 quantifies its initialization error. Only the retained path or endpoint and its weight need be stored. All $m + 1$ proposals can be generated in parallel before the ordered accept/retain decisions. Independent outputs require independent repetitions of the entire procedure.

Appendix G extends the proposal with r local pool MH updates and proves its proper weighting identity.

4 Sampling Error and Guidance Depth

We bound the sampling error after m outer independence MH updates. Theorem 4 controls the total variation distance between the output law ν_m and the finite-grid target Π_n defined in (4). The bound separates accumulated Bellman error from local Stein variance.

We use the following regularity conditions to bound the sampling error.

Assumption 3 (Uniform value and spatial bounds)**.** Fix the horizon T. There are constants $0 < c_0 \leq C_0 < \infty$ and $L_2, \Sigma < \infty$ that do not depend on the grid or pool size. For every grid with this horizon, every $k = 0, \ldots, n-1$, and every $x \in \mathbb{R}^{d_x}$,

$$H_{k+1} \in C^2(\mathbb{R}^{d_x}), \quad 0 < c_0 \leq H_{k+1}(x) \leq C_0, \quad \|\nabla^2 H_{k+1}(x)\|_{\mathrm{op}} \leq L_2, \quad \|\sigma_k(x)\|_{\mathrm{op}} \leq \Sigma. \tag{9}$$

In this section, we use the uniform grid $h = T/n$.

Bellman error and the Doob transform. Ideal intermediate value functions H_k^* satisfy the Bellman equation $H_k^* = P_k H_{k+1}^*$ with $H_n^* = \phi$ (Zhao et al., 2024; Zhu & Lu, 2026). Their tilted transitions give the discrete Doob transform and sample Π_n. For approximate values, the ratio G_k/H_k measures the failure of this equation, as in the Bellman error of Zhu & Lu (2026). We use the logarithmic form $\log(P_k H_{k+1}) - \log H_k$ as the local Bellman error. This is also the Bellman residual for stochastic control with terminal reward $\log \phi$ and a KL cost for changing the reference transitions (Todorov, 2006).

We measure this error along the ideal value-tilted path law

$$K_k^H(x, \mathrm{d}y) = \frac{P_k(x, \mathrm{d}y) H_{k+1}(y)}{G_k(x)}, \qquad \Pi_n^H = \delta_{x_\star} \prod_{k<n} K_k^H, \qquad G_k = P_k H_{k+1}. \tag{10}$$

Let B be the accumulated Bellman error along a path, defined below. We measure its variation under Π_n^H by the centered norm $\mathcal{A}_\ell$:

$$\mathcal{A}_\ell = \inf_{b \in \mathbb{R}} \|B - b\|_{L^\ell(\Pi_n^H)}, \qquad B(X_{0:n}) = \sum_{k<n} \log \frac{G_k(X_k)}{H_k(X_k)}, \quad 1 \leq \ell < \infty. \tag{11}$$

For each fixed ℓ, Assumption 3 bounds $\mathcal{A}_\ell$ uniformly in guidance depth (Lemma 14). Subtracting a constant removes the part of B that cancels when the path law is normalized. For a fixed time grid and fixed value functions $H_0, \ldots, H_n$, $\mathcal{A}_\ell$ is independent of K. It vanishes for exact discrete Doob values, or whenever B is constant. The following theorem bounds the sampling error. Its proof appears in Section D.5.

Theorem 4 (Sampling error bound with approximate values)**.** *Use Algorithm 1 with the fixed start, terminal value $H_n = \phi$, and fixed value functions of Section 3. Suppose Assumption 3 holds. For*

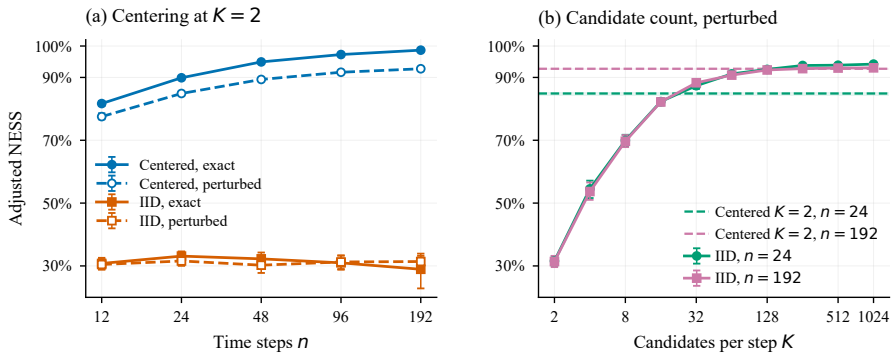

Figure 2: Effect of centering and pool size. (a) Mesh refinement with $K = 2$ for both centered pools and uncentered IID pools, under exact and perturbed guides. (b) Uncentered IID pools require much larger K to reach the adjusted NESS of centered $K = 2$ under the perturbed guide; horizontal bands show the centered references.

every real $p \geq 1$, there is a constant C_p depending only on $p, T, d, c_0, C_0, L_2, \Sigma$. For all $n \geq 1$, $K \geq 2$, and integers $m \geq p$, the physical path law ν_m returned by Algorithm 2 satisfies

$$\|\nu_m - \Pi_n\|_{\mathrm{TV}} \leq \min\Big\{1, \frac{C_p}{m^p} \min\Big\{1, \Big(\underbrace{\mathcal{A}_{2p+2}^2}_{\text{Bellman error}} + \underbrace{\frac{Th}{K-1}}_{\text{local Stein variance}} \Big)^{(p+1)/2}\Big\}\Big\}. \tag{12}$$

The same bound holds for the endpoint marginal law. One output uses $Kn(m+1)$ candidate values.

The theorem indicates a superpolynomial decaying rate for the sampling error with regard to the number of MH correction m, this is faster than the $O(1/N)$ rate of SMC based sampling (Zhu & Lu, 2026), and the $O(1/\sqrt{M})$ rate of DOIT (Zhu et al., 2026), where N, M are the statistical widths for the two types of algorithms respectively.

The coefficient of Theorem 4 separates the Bellman error term from candidate-pool noise. The local Stein variance term accumulates the $O(h^2/(K-1))$ variance of each normalized pool average over $n = T/h$ steps. The gain from K diminishes once the Bellman error term dominates. When B is constant, as for exact discrete Doob values, the Bellman error term vanishes. At fixed m, finer grids and larger pools reduce the remaining term in the error bound.

Correction cost and guidance depth. For any fixed TV accuracy, Theorem 4 bounds the required number of outer MH updates independently of guidance depth. Therefore, this algorithm provides a positive answer to the question we raise in the beginning.

5 Experiments: Centering, Pool Size, and Tilted Distributions

We first isolate the effect of centering and pool size in a controlled SIS benchmark, then compare critic-tilted action distributions under a frozen learned policy. These experiments measure normalizer efficiency and distribution accuracy, respectively.

5.1 Centering and pool size

On a scalar Brownian benchmark with target $\mathcal{N}(1.5, 1)$, we compare centered and IID pools under exact and perturbed value functions. Centered $K = 2$ uses opposite Gaussian innovations; IID pools use independent ones. Adjusted NESS measures normalizer-estimation efficiency; its definition and the protocol appear in Appendix H.1.

With $K = 2$, finer grids improve centered pools, while IID pools remain near 30% adjusted NESS under both guides (Figure 2(a)). Under the perturbed guide at $n = 192$, centered $K = 2$ reaches 92.76% adjusted NESS, versus 31.40% for IID $K = 2$. IID $K = 128$ reaches 92.38% and $K = 256$ reaches 92.81%, using 64 and 128 times as many candidate-value evaluations per path as centered

Method	30%	60%	80%	Method	30%	60%	80%
Stein $K2/L32$	**0.0306**	**0.0697**	**0.1128**	FK $N32$	0.0504	0.1093	0.1816
Stein $K8/L32$	0.0426	0.0942	0.1511	DAS $N32$	0.0307	0.0769	0.1355
DOIT native	0.2418	0.4778	0.6179	TFG	0.1788	0.2808	0.3586
DOIT fixed β	0.2780	0.4774	0.5903	TTS-local	0.1781	0.2767	0.3752
SVDD $M32$	0.1218	0.2276	0.2979				

Table 2: Mean conditional critic-Q KS over eight fixed Hopper observations (lower is better).

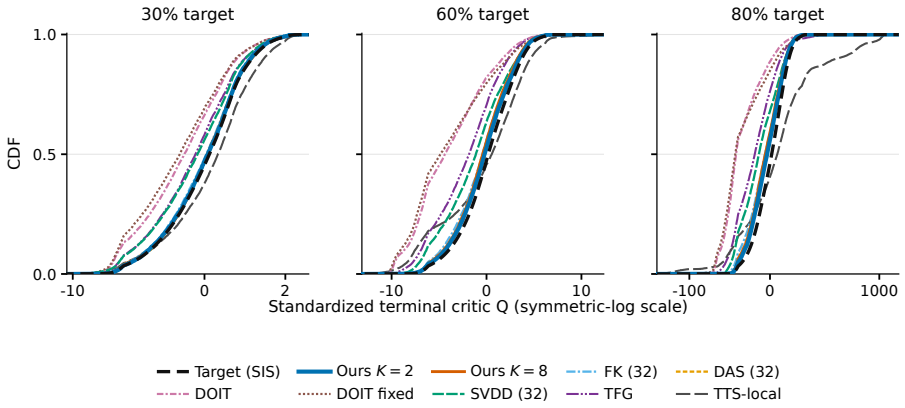

Figure 3: Critic-Q distributions at the three target strengths. Curves equally mix the eight states after standardization by the target reference. The horizontal axis is linear near zero and logarithmic in the tails; black is the numerical SIS target reference.

$K = 2$. Thus a two-point centered pool matches the normalizer efficiency of much larger independent pools in this benchmark.

5.2 Critic-tilted actions in offline RL

On Hopper-medium-v2, we compare sampling from $\pi_{\beta_s}(a|s) \propto p_0(a|s)e^{\beta_s Q(s,a)}$ with a fixed diffusion policy p_0 and critic Q. The three targets assign 30%, 60%, or 80% of their mass to the original policy's highest-Q 10% of actions. We report the Kolmogorov–Smirnov (KS) distance between each method's critic-Q distribution and an independent numerical target reference, averaged over eight fixed states; lower is better.

Stein uses $K = 2, 8$ candidates per step and $L = m + 1 = 32$ complete proposals per output, with Tweedie value estimates and no backpropagation. We compare against FK (Singhal et al., 2025), DAS (Kim et al., 2025), DOIT (Zhu et al., 2026) (native and fixed-temperature), uncorrected SVDD (Li et al., 2024), TFG (Ye et al., 2024), and TTS-local (Zhang et al., 2025). Settings, calibration, and uncertainty estimates appear in Appendix H.2.

The $K = 2$ sampler has the lowest measured KS at all three targets (Table 2). It closely matches DAS at 30% and has lower KS at 60% and 80%. Its KS is also 36–39% lower than FK's. Bootstrap intervals that account for multiple comparisons support its gains over both DOIT variants, SVDD, TFG, and TTS-local at all three targets (15 comparisons); $K = 8$ improves over these baselines in 14 of 15 comparisons. Figure 3 shows how closely Stein's corrected Q distributions follow the target reference.

Whole-path MH correction provides a large part of the improvement. For $K = 2$, KS falls from 0.2174, 0.2804, and 0.4041 before correction to 0.0306, 0.0697, and 0.1128 at $L = 32$. Larger candidate pools do not consistently improve accuracy at this proposal count. This is consistent with Theorem 4: increasing K reduces pool noise but does not remove the Bellman error of fixed approximate values.

Prompt *a photo of ...*	*a book and a laptop.*	*a knife and a zebra.*	*a laptop and a carrot.*	*a dining table and a bear.*
FK-Steering SDv1.5, $N = 4$				
SVDD SDv1.5 $N = 4$				
URGE SDv1.5 $N = 4$				
Stein SDv1.5 $K = 2$				
Stein SDv1.5 $K = 4$				

Table 3: Comparison of four methods under four prompts. In these examples, Stein better preserves both requested objects and their spatial separation, improving prompt alignment.

5.3 Text-to-Image Generation

We apply our algorithm to text-to-image generation with Stable Diffusion v1.5 (Rombach et al., 2022). The pretrained model provides the conditional image distribution $p_{\text{data}}(x \mid c)$, where x is the generated image and c is the text prompt. The pool sizes are $K = 2$ and $K = 4$, with $L = 11$ complete proposals per image: one initial draw followed by 10 whole-path MH updates. FK-Steering (Singhal et al., 2025), SVDD (Li et al., 2024), and URGE (Wang et al., 2026) use $N = 4$ particles on the same backbone. Table 3 shows four two-object prompts.

The Stein row with $K = 4$ closely matches the prompts, keeping both requested objects present and spatially separate. Meanwhile, it improves on the $K = 2$ row, while particle methods often retain only one of the two requested objects or fuse the pair into one implausible object.

Reproducibility Statement

Complete statements, proofs, essential experimental settings, and selected results follow in the appendices.

References

Abbas Abdolmaleki, Jost Tobias Springenberg, Yuval Tassa, Remi Munos, Nicolas Heess, and Martin Riedmiller. Maximum a posteriori policy optimisation. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=S1ANxQW0b.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided diffusion for bayesian linear inverse problems. In *International Conference on Learning Representations*, 2024.

Haoxuan Chen, Yinuo Ren, Martin Renqiang Min, Lexing Ying, and Zachary Izzo. Solving inverse problems via diffusion-based priors: An approximation-free ensemble sampling approach, 2025. AFDPS; arXiv:2506.03979.

Nicolas Chopin, Sumeetpal S. Singh, Tomás Soto, and Matti Vihola. On resampling schemes for particle filters with weakly informative observations. *The Annals of Statistics*, 50(6):3197–3222, 2022.

Sanjit Dandapanthula and Nicholas M Boffi. Are we really tilting? the mechanics of reward guidance in flow and diffusion models. *arXiv preprint arXiv:2606.02884*, 2026.

George Deligiannidis, Pierre E. Jacob, El Mahdi Khribch, and Guanyang Wang. On importance sampling and independent metropolis-hastings with an unbounded weight function. *arXiv preprint arXiv:2411.09514*, 2026.

Alexander Denker, Francisco Vargas, Shreyas Padhy, Kieran Didi, Simon Mathis, Vincent Dutordoir, Riccardo Barbano, Emile Mathieu, Urszula J Komorowska, and Pietro Lio. Deft: Efficient fine-tuning of diffusion models by learning the generalised h-transform. *Advances in Neural Information Processing Systems*, 37:19636–19682, 2024.

Noah Golowich, Fan Chen, Dhruv Rohatgi, Raghav Singhal, Carles Domingo-Enrich, Dylan J. Foster, and Akshay Krishnamurthy. Reject, resample, repeat: Understanding parallel reasoning in language model inference. *arXiv preprint arXiv:2603.07887*, 2026. URL https://arxiv.org/abs/2603.07887.

Jiajun He, José Miguel Hernández Lobato, Yuanqi Du, and Francisco Vargas. Rne: plug-and-play diffusion inference-time control and energy-based training. In *International Conference on Learning Representations*, volume 2026, pp. 75732–75786, 2026.

Jeremy Heng, Adrian N. Bishop, George Deligiannidis, and Arnaud Doucet. Controlled sequential monte carlo. *The Annals of Statistics*, 48(5):2904–2929, 2020.

Jiachen Jiang, Huminhao Zhu, and Zhihui Zhu. Smcevolve: Principled scientific discovery via sequential monte carlo evolution. *arXiv preprint arXiv:2605.15308*, 2026.

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization, 2025. URL https://arxiv.org/abs/2501.05803.

Alexander K. Lew, Marco Cusumano-Towner, and Vikash K. Mansinghka. Recursive Monte Carlo and variational inference with auxiliary variables. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 1096–1106. PMLR, 2022. URL https://proceedings.mlr.press/v180/lew22a.html.

Xiner Li, Yulai Zhao, Chenyu Wang, Gabriele Scalia, Gokcen Eraslan, Surag Nair, Tommaso Biancalani, Shuiwang Ji, Aviv Regev, Sergey Levine, and Masatoshi Uehara. Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding. *arXiv preprint arXiv:2408.08252*, 2024.

Jun S. Liu, Faming Liang, and Wing Hung Wong. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, 2000. doi: 10.1080/01621459.2000.10473908.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, et al. Inference-time scaling for diffusion models beyond scaling denoising steps. *arXiv preprint arXiv:2501.09732*, 2025.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17:527–566, 2017. doi: 10.1007/s10208-015-9296-2.

Toan Nguyen, Kien Do, Duc Kieu, and Thin Nguyen. h-edit: Effective and flexible diffusion-based editing via doob's h-transform. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28490–28501. IEEE, 2025.

Yinuo Ren, Wenhao Gao, Lexing Ying, Grant Rotskoff, and Jiequn Han. Driftlite: Lightweight drift control for inference-time scaling of diffusion models. In *International Conference on Learning Representations*, volume 2026, pp. 76664–76702, 2026.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alán Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman–kac correctors in diffusion: Annealing, guidance, and product of experts. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 55906–55949, 2025. URL https://proceedings.mlr.press/v267/skreta25a.html.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981. doi: 10.1214/aos/1176345632.

WENPIN Tang and RENYUAN Xu. A stochastic analysis approach to conditional diffusion guidance. *Working paper*, 2025.

Emanuel Todorov. Linearly-solvable markov decision problems. In *Advances in Neural Information Processing Systems*, volume 19, 2006. URL https://papers.nips.cc/paper_files/paper/2006/hash/d806ca13ca3449af72a1ea5aedbed26a-Abstract.html.

Masatoshi Uehara, Yulai Zhao, Chenyu Wang, Xiner Li, Aviv Regev, Sergey Levine, and Tommaso Biancalani. Reward-guided controlled generation for inference-time alignment in diffusion models: Tutorial and review. *arXiv preprint arXiv:2501.09685*, 2025.

Chenyang Wang, Weizhong Wang, Yinuo Ren, Jose Blanchet, and Yiping Lu. Simple approximation and derivative free inference-time scaling for diffusion models via sequential monte carlo on path measures. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. arXiv:2605.17850.

Guanghan Wang, Yair Schiff, Subham Sahoo, and Volodymyr Kuleshov. Remasking discrete diffusion models with inference-time scaling. *Advances in Neural Information Processing Systems*, 38: 147282–147339, 2025.

Nick Whiteley and Anthony Lee. Twisted particle filters. *The Annals of Statistics*, 42(1):115–141, 2014.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, 2023.

Xiaodong Yang and Jun S. Liu. Convergence rate of multiple-try metropolis independent sampler. *arXiv preprint arXiv:2111.15084*, 2023.

Haotian Ye, Haowei Lin, Jiaqi Han, Minkai Xu, Sheng Liu, Yitao Liang, Jianzhu Ma, James Zou, and Stefano Ermon. TFG: Unified training-free guidance for diffusion models, 2024. URL https://arxiv.org/abs/2409.15761.

Xiangcheng Zhang, Haowei Lin, Haotian Ye, James Zou, Jianzhu Ma, Yitao Liang, and Yilun Du. Inference-time scaling of diffusion models through classical search, 2025. URL https://arxiv.org/abs/2505.23614.

Stephen Zhao, Rob Brekelmans, Alireza Makhzani, and Roger Baker Grosse. Probabilistic inference in language models via twisted sequential Monte Carlo. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 60704–60748, 2024.

Qijie Zhu, Zeqi Ye, Han Liu, Zhaoran Wang, and Minshuo Chen. Training-free adaptation of diffusion models via Doob's h-transform, 2026. arXiv:2602.16198.

Youheng Zhu and Yiping Lu. On the power of (approximate) reward models for inference-time scaling. *arXiv preprint arXiv:2602.01381*, 2026.

AI Use Disclosure

Generative AI assisted with manuscript restructuring, mathematical exposition, some of the proofs, experiment design and analysis, implementation, and figure preparation. The authors take responsibility for the final content, including all proofs, citations, and numerical results.

Reading the appendices. Appendix A discusses related work and scope. Appendix B provides the Stein identities and local pool moments. Appendices C and D contain the complete-weight analysis and the proof of Theorem 4. Appendix E proves the Brownian resampling obstruction. Appendices F and G give terminal importance resampling and the proper weighting of local refinement. Essential experimental settings and selected supporting results appear in Appendix H.

A Related Work and Scope

This section relates centered candidate pools and whole-path correction to existing guidance methods. Value-based decoding uses future values for derivative-free selection (Li et al., 2024); twisted and controlled particle methods use approximate backward information (Whiteley & Lee, 2014; Heng et al., 2020), and URGE uses path correction (Wang et al., 2026). Gaussian Stein estimators (Stein, 1981; Nesterov & Spokoiny, 2017), multiple try Metropolis (Liu et al., 2000; Yang & Liu, 2023), and independence-MH coupling bounds (Deligiannidis et al., 2026) supply established ingredients. Uniform regularity supplies the moments needed for depth-uniform correction. Variation in the accumulated Bellman error and centered pool moments quantify the two contributions to the sampling error bound. Section G.4 records the comparison conditions for Monte Carlo Doob guidance (Zhu et al., 2026).

The centered norm $\mathcal{A}_r$ measures variation of the accumulated Bellman error under the ideal value-tilted path law. It is independent of pool size and measures near-Doob accuracy at the path level. Multiplying each exact value by a positive time-dependent factor leaves this variation zero. Uniform regularity guarantees finite moments. Constants can depend on dimension, and a continuous target requires a separate discretization bound. Centered pools need not improve every antithetic construction: pairs cancel all odd Hermite components, whereas larger centered pools can retain them. The Brownian SMC obstruction uses full resampling at every step, consistent with the role of resampling intensity for weakly informative potentials (Chopin et al., 2022). The Gaussian experiments measure local-refinement and estimator effects. The learned-policy RL comparison shows that a small, forward-only candidate pool followed by whole-path MH can substantially improve on uncorrected guidance. At $L = 32$, $K = 2$ achieves the lowest observed Q-KS at all three targets, with supported gains over DOIT, SVDD, TFG, and TTS-local and competitive or superior accuracy relative to FK and DAS in observed Q-KS.

B Gaussian Stein Pools: Construction, Moments, and Guidance

The local cancellation in Algorithm 1 supplies the sharper pool moments used by the finite-moment path analysis. We establish those bounds here, then establish the estimator and mean-guidance guarantees for the same candidate construction.

B.1 Targets and properly weighted candidate pools

The pool identity below is the common starting point for local and complete-path correction: weighting a selected candidate by its pool average recovers the unnormalized value-tilted transition.

Fix a time grid $t_k = kh$, $k = 0, \dots, n$, with $T = nh$ and initial state $X_0 = x_\star$. Let P_k be the transition kernels of a reference simulator. The reference path law and its positive terminal reward tilt are

$$P_n = \delta_{x_\star} \prod_{k=0}^{n-1} P_k, \qquad \Pi_n(\mathrm{d}x_{0:n}) = \frac{\phi(x_n)}{Z_n} P_n(\mathrm{d}x_{0:n}), \qquad Z_n = \mathbb{E}_{P_n} \phi(X_n) \in (0, \infty). \tag{13}$$

All statements of exactness below concern this fixed grid target. A discretization error between P_n and a continuous process is a separate quantity.

Let $H_0, \dots, H_n$ be positive value functions with $H_n = \phi$. Define

$$G_k(x) = P_k H_{k+1}(x), \qquad \Gamma_k(x) = \frac{G_k(x)}{H_k(x)}, \qquad K_k^H(x, \mathrm{d}y) = \frac{P_k(x, \mathrm{d}y) H_{k+1}(y)}{G_k(x)}. \tag{14}$$

We assume $0 < G_k(x) < \infty$ on the states under consideration, and write $\Pi_n^H = \delta_{x_\star} \prod_k K_k^H$. These locally normalized kernels define the intermediate target of local correction. When the discrete Bellman equations $H_k = P_k H_{k+1}$ hold, $\Pi_n^H = \Pi_n$. We use $\|\mu - \nu\|_{\mathrm{TV}} = \sup_A |\mu(A) - \nu(A)|$.

At a fixed parent x, abbreviate $H = H_{k+1}$ and $G = G_k(x)$. A candidate pool $U = (Y_1, \dots, Y_K)$ has distribution $q_{\text{pool}} = q_{k,x,K}$, and each of its coordinates has marginal law $P_k(x, \cdot)$. Dependence within the pool is allowed. Set

$$S(U) = \frac{1}{K}\sum_{i=1}^{K} H(Y_i), \qquad w(U) = \frac{S(U)}{G}, \qquad \pi_i(U) = \frac{H(Y_i)}{\sum_{j=1}^{K} H(Y_j)}. \tag{15}$$

Let C_x select Y_J with conditional probabilities $\mathbb{P}(J = i \mid U) = \pi_i(U)$. Thus $C_x f(U) = \sum_i H(Y_i) f(Y_i)/(KS(U))$ is the conditional mean of $f(Y_J)$.

Lemma 5 (Proper weighting of a pool)**.** *For every bounded measurable f,*

$$\mathbb{E}_{q_{\text{pool}}}[S(U) C_x f(U)] = P_k(Hf)(x). \tag{16}$$

Consequently $\mathbb{E}_{q_{\text{pool}}} S = G$, $\mathbb{E}_{q_{\text{pool}}} w = 1$, and the probability measure $p_{\text{pool}} = w q_{\text{pool}}$ satisfies $p_{\text{pool}} C_x = K_k^H(x, \cdot)$. The weight $S(U)$ and selection probabilities require neither derivatives nor an evaluation of G.

Proof of Lemma 5. By the conditional selection probabilities and the correct marginal law of every candidate,

$$\mathbb{E}_{q_{\text{pool}}}[S C_x f] = \frac{1}{K}\sum_i \mathbb{E}_{q_{\text{pool}}}[H(Y_i) f(Y_i)] = P_k(Hf)(x).$$

Taking $f = 1$ gives $\mathbb{E}_{q_{\text{pool}}} S = G$. Dividing the identity by G proves $p_{\text{pool}} C_x = K_k^H(x, \cdot)$. □

This identity requires only correct candidate marginals. The Gaussian Stein identity provides additional structure: it explains the guidance implicit in weighted selection and identifies dependence patterns that reduce fluctuations of S.

B.2 Centered Gaussian candidates and implicit guidance

Centering couples the candidates without changing their Gaussian marginals. It removes constant noise from the Stein estimator and linear noise from the pool average; the identities below also identify the mean guidance implemented by weighted selection.

For the remainder of the local analysis, take an Euler reference kernel with mean $\mu = x + h b_k(x)$ and covariance $h\sigma\sigma^\top$, where $\sigma = \sigma_k(x) \in \mathbb{R}^{d_x \times d}$. Here d_x and d are the state and noise dimensions, and $\gamma_d = N(0, I_d)$. Draw independent $G_1, \dots, G_K \sim N(0, I_d)$ and construct

$$\bar{G} = \frac{1}{K}\sum_{i=1}^{K} G_i, \qquad Z_i = \sqrt{\frac{K}{K-1}}(G_i - \bar{G}), \qquad Y_i = \mu + \sqrt{h}\,\sigma Z_i, \qquad K \geq 2. \tag{17}$$

The resulting vectors satisfy

$$Z_i \sim N(0, I_d), \qquad \sum_i Z_i = 0, \qquad \operatorname{Cov}(Z_i, Z_j) = -\frac{I_d}{K-1} \quad (i \neq j). \tag{18}$$

Thus centering preserves every required marginal after rescaling. The case $K = 2$ is an antithetic pair. The case $K = 3$, for example, can equivalently be generated from independent $G, F \sim N(0, I_d)$ by

$$Z_1 = G, \qquad Z_2 = -G/2 + \sqrt{3}F/2, \qquad Z_3 = -G/2 - \sqrt{3}F/2. \tag{19}$$

The construction therefore works without a decomposition into pairs.

Proposition 6 (Stein estimation and mean guidance)**.** *Suppose $H \in C^1$ is positive and, for $F(z) = H(\mu + \sqrt{h}\sigma z)$, Gaussian integration by parts is valid. A sufficient condition is $F \in W^{1,2}(\gamma_d)$. Define*

$$\widehat{v} = \frac{1}{K\sqrt{h}}\sum_i Z_i H(Y_i), \qquad v_h = \sigma^\top \mathbb{E}\nabla H(\mu + \sqrt{h}\sigma Z), \qquad s_h = \frac{v_h}{G}, \qquad \widehat{s} = \frac{\widehat{v}}{S}, \tag{20}$$

where $Z \sim N(0, I_d)$ and the displayed expectations exist. Then $\mathbb{E}\widehat{v} = v_h$. Moreover, for any finite common baseline B, even one depending on the entire pool,

$$\widehat{v} = \frac{1}{K\sqrt{h}}\sum_i Z_i\{H(Y_i) - B\}. \tag{21}$$

Weighted selection from an uncorrected pool satisfies

$$\mathbb{E}[Y_J \mid U] = \mu + h\sigma\widehat{s}, \tag{22}$$

whereas the exact local tilt has mean

$$\mathbb{E}_{K_k^H(x,\cdot)}Y = \mu + h\sigma s_h = \mu + h\sigma\sigma^\top\nabla_\mu \log \mathbb{E}H(\mu + \sqrt{h}\sigma Z), \tag{23}$$

whenever the last differentiation under the expectation is justified.

Proof of Proposition 6. Direct covariance calculation in (17) gives (18). Gaussian integration by parts for each standard Gaussian marginal yields

$$\mathbb{E}[Z_i H(Y_i)] = \sqrt{h}\,\sigma^\top \mathbb{E}\nabla H(Y_i).$$

All these expectations are the same. Averaging proves $\mathbb{E}\widehat{v} = v_h$. The identity $\sum_i Z_i = 0$ proves (21), including for a random common baseline. Furthermore,

$$\mathbb{E}[Y_J \mid U] = \mu + \sqrt{h}\sigma \sum_i \pi_i Z_i = \mu + h\sigma\frac{\widehat{v}}{S}.$$

Under the exact tilt the mean is $\mu + \sqrt{h}\sigma\mathbb{E}[ZH(Y)]/G$, which equals $\mu + h\sigma v_h/G$ by the same integration by parts identity. Differentiation with respect to μ gives the final expression in (23) when the asserted interchange is justified. □

Equation (23) differentiates with respect to μ while holding σ fixed. It does not take the total derivative of $P_k H_{k+1}(x)$ when the drift and diffusion depend on x. The numerator $\widehat{v}$ is unbiased, while the finite pool ratio $\widehat{s}$ is generally biased. The zero sum identity removes a constant baseline exactly, without querying $H(\mu)$.

The same structure can be read directly from the selection probabilities. Under Assumption 3, or more generally whenever the following integration by parts is valid,

$$\mathbb{E}\sum_i \pi_i Z_i = \sqrt{h}\frac{K}{K-1}\mathbb{E}\sum_i \pi_i(1-\pi_i)\sigma^\top\nabla \log H(Y_i). \tag{24}$$

For fixed K, as $h \to 0$ the factor $K/(K-1)$ cancels the limit $\sum_i \pi_i(1-\pi_i) \to 1 - 1/K$. With independent Gaussian candidates, the factor $K/(K-1)$ is absent, and uncorrected weighted selection attenuates the leading guidance by $1 - 1/K$. This comparison concerns selection from a raw pool.

Proof of the selection identity (24). Let $\rho_{ii} = 1$ and $\rho_{ij} = -1/(K-1)$ for $i \neq j$. Gaussian integration by parts in the independent variables G_i implies, even though the covariance of the Z_i is singular,

$$\mathbb{E}[Z_i\pi_i] = \sum_j \rho_{ij}\mathbb{E}[\nabla_{Z_j}\pi_i].$$

Put $a_j = \sigma^\top\nabla \log H(Y_j)$. Differentiating the quotient that defines π_i gives

$$\nabla_{Z_j}\pi_i = \sqrt{h}\,\pi_i(\mathbf{1}_{i=j}a_i - \pi_j a_j).$$

After summing over i, the coefficient of $\pi_j a_j$ is $1 - \sum_i \rho_{ij}\pi_i$. Since

$$\sum_i \rho_{ij}\pi_i = \frac{K\pi_j - 1}{K-1},$$

this coefficient is $K(1-\pi_j)/(K-1)$, proving the identity. Under Assumption 3, gradients are bounded as shown below, so the differentiation and integration by parts are justified by Gaussian integrability. The same bounds justify dominated convergence as $h \to 0$. For independent candidates, replace ρ by the identity matrix; the resulting coefficient is $1 - \pi_j$, which proves the stated attenuation of the leading guidance. □

B.2.1 The $K = 2$ case

Both the proposal and correction weight have a simple form for Stein ($K = 2$). For $K = 2$, the centered Gaussian construction is exactly

$$Z_1 = (G_1 - G_2)/\sqrt{2} =: Z, \qquad Z_2 = -Z, \qquad Z \sim N(0, I_d).$$

Its candidate law, categorical selection, and query estimator are therefore identical to the antithetic finite difference in (5). The two forms of its path weight also agree:

$$\phi(X_n)\prod_{k=0}^{n-1}\frac{S_{2,k}}{H_{k+1}(X_{k+1})} = \frac{\prod_{k=0}^{n-1}S_{2,k}}{\prod_{k=1}^{n-1}H_k(X_k)} = H_0(x_\star)\prod_{k=0}^{n-1}\frac{S_{2,k}}{H_k(X_k)}.$$

These identities give an exact antithetic finite-difference representation of Algorithm 1 at $K = 2$, including its correction weight.

With a bounded third derivative, the familiar central-difference expansion is

$$\frac{H(\mu+\sqrt{h}\sigma Z) - H(\mu-\sqrt{h}\sigma Z)}{2\sqrt{h}} = (\sigma Z)^\top \nabla H(\mu) + O(h\|\sigma Z\|^3).$$

Its leading ratio is $ZZ^\top\sigma^\top\nabla \log H(\mu)$, whose expectation is $\sigma^\top\nabla \log H(\mu)$. Proposition 6 gives the exact form of Stein's identity behind this Taylor interpretation. Under Assumption 3, Theorem 11 directly gives an $O(h^2)$ error in the one-step mean displacement, without requiring a third derivative. A mean displacement bound is distinct from a second-order weak expansion for every test function.

B.3 Exact spectrum and sharp normalizer variance

The sum constraint cancels the first order term of any local value expansion. A Gaussian Hermite calculation gives a stronger conclusion with no Taylor remainder. Expand $F \in L^2(\gamma_d)$ in the orthonormal multivariate Hermite basis as $F = \sum_\alpha c_\alpha \psi_\alpha$. Define its energy at degree ℓ by $a_\ell = \sum_{|\alpha|=\ell} c_\alpha^2$.

Lemma 7 (Exact Hermite spectrum of the pool average)**.** *For the construction* (17)*,*

$$\mathrm{Var}\left(\frac{1}{K}\sum_i F(Z_i)\right) = \sum_{\ell\geq 1}\lambda_{K,\ell}a_\ell, \qquad \lambda_{K,\ell} = \frac{1+(K-1)(-1/(K-1))^\ell}{K}. \tag{25}$$

In particular $\lambda_{K,1} = 0$*, and* $0 \leq \lambda_{K,\ell} \leq 1/(K-1)$ *for every* $\ell \geq 2$*. Hence*

$$\mathrm{Var}\left(\frac{1}{K}\sum_i F(Z_i)\right) \leq \frac{1}{K-1}\sum_{\ell\geq 2}a_\ell \leq \frac{\mathrm{Var}_{\gamma_d}(F)}{K-1}. \tag{26}$$

The last inequality also holds componentwise for vector valued F*.*

Proof of Lemma 7. If Z, Z' are jointly standard Gaussian and $\mathbb{E}[ZZ'^\top] = \rho I_d$, their exponential generating function obeys

$$\mathbb{E}\left[e^{t^\top Z - \|t\|^2/2}e^{s^\top Z' - \|s\|^2/2}\right] = e^{\rho t^\top s}.$$

Comparing coefficients gives $\mathbb{E}[\psi_\alpha(Z)\psi_\beta(Z')] = \mathbf{1}_{\alpha=\beta}\rho^{|\alpha|}$. Thus, for a finite Hermite expansion,

$$\mathrm{Var}\left(\frac{1}{K}\sum_i F(Z_i)\right) = \frac{1}{K^2}\sum_{i,j}\sum_{\ell\geq 1}\rho_{ij}^\ell a_\ell = \sum_{\ell\geq 1}\frac{1+(K-1)(-1/(K-1))^\ell}{K}a_\ell.$$

Finite expansions are dense in $L^2(\gamma_d)$. The averaging map is continuous into the joint L^2 space, since $\left\|K^{-1}\sum_i F(Z_i)\right\|_2 \leq \|F\|_{L^2(\gamma_d)}$. The formula therefore extends to every $F \in L^2(\gamma_d)$. For $K = 2$, the eigenvalues are zero at odd degrees and one at even degrees. For $K > 2$, the largest even degree coefficient with $\ell \geq 2$ occurs at $\ell = 2$ and is $1/(K-1)$; the odd degree coefficients are between zero and $1/K$. Degree one vanishes in every case. These observations prove (26). □

Theorem 8 (Sharp curvature bound)**.** *Let* $H \in C^2$ *be positive, let* $F(z) = H(\mu+\sqrt{h}\sigma z) \in W^{2,2}(\gamma_d)$*, and let* $G = \mathbb{E}_{\gamma_d}F > 0$*. For the normalized pool weight* $w = S/G$*,*

$$\mathrm{Var}(w) \leq \frac{h^2}{2(K-1)G^2}\mathbb{E}\left\|\sigma^\top\nabla^2 H(\mu+\sqrt{h}\sigma Z)\sigma\right\|_F^2. \tag{27}$$

There is no assumption that H is bounded or that w has a uniform positive lower bound. Equality holds whenever F is a polynomial of degree at most two. In particular, positive quadratic value functions show that the constant $1/2$ cannot be reduced uniformly.

Proof of Theorem 8. The normalized Hermite differentiation identity $\partial_j \psi_\alpha = \sqrt{\alpha_j}\psi_{\alpha-e_j}$ and orthogonality imply

$$\mathbb{E}_{\gamma_d} \left\|\nabla^2 F\right\|_F^2 = \sum_\alpha |\alpha|(|\alpha|-1)c_\alpha^2 = \sum_{\ell\geq 2} \ell(\ell-1)a_\ell.$$

The identity holds first for polynomials and then for $W^{2,2}(\gamma_d)$ by approximation in that Sobolev norm. Because $\ell(\ell-1) \geq 2$ for $\ell \geq 2$, Lemma 7 yields

$$\mathrm{Var}(S) \leq \frac{1}{2(K-1)} \mathbb{E}_{\gamma_d} \left\|\nabla^2 F\right\|_F^2 .$$

The chain rule gives $\nabla^2 F = h\sigma^\top \nabla^2 H(\mu + \sqrt{h}\sigma Z)\sigma$. Division by G^2 proves the result. If F has degree at most two, the only nonconstant degree surviving the pool average is degree two, and $\lambda_{K,2} = 1/(K-1)$, so both inequalities are equalities. A positive nonconstant quadratic with nonzero curvature gives a nonzero equality case. □

For later path bounds, the useful uniform assumption is a relative curvature bound:

$$\sup_{k,x,h} \frac{\mathbb{E}\left\|\sigma_k(x)^\top \nabla^2 H_{k+1}(\mu_k(x) + \sqrt{h}\sigma_k(x)Z)\sigma_k(x)\right\|_F^2}{G_k(x)^2} \leq C_2^2. \tag{28}$$

Together with the regularity in Theorem 8, it yields $\sup_{k,x} \mathrm{Var}(w) \leq C_2^2 h^2/[2(K-1)]$. The Hessian appears only in the analysis; the proposal queries values of H.

B.4 All fixed moments from the same regularity conditions

For each output-TV order p, the path analysis uses several fixed pool-moment orders to separate Bellman error from candidate noise. We derive every fixed real order $\ell \geq 2$ from the bounded-value and bounded-Hessian conditions; higher derivatives are unnecessary. The main theorem combines these local moments with the direct complete-weight bound in Theorem 13. The lower bound obtained along the way is useful for the local-refinement construction.

We use Assumption 3 throughout this subsection. Write $m_\ell = (\mathbb{E}|N(0,1)|^\ell)^{1/\ell}$ and $M_{\ell,d} = (\mathbb{E}\,\|N(0,I_d)\|^\ell)^{1/\ell}$.

Lemma 9 (Gaussian rotation inequality)**.** *Let G be a standard Gaussian vector in any finite dimension. If $\ell \geq 2$ and $f \in W^{1,\ell}$ with respect to its Gaussian measure, then*

$$\|f(G) - \mathbb{E}f(G)\|_{L^\ell} \leq \frac{\pi}{2} m_\ell \left\|\|\nabla f(G)\|\right\|_{L^\ell}. \tag{29}$$

Proof. For smooth f, take an independent copy G' and put $G_t = G\cos t + G'\sin t$ and $V_t = -G\sin t + G'\cos t$. For each t, the vectors G_t, V_t are independent standard Gaussian vectors. Conditional Jensen, the fundamental theorem of calculus and Minkowski give

$$\begin{aligned}
\|f(G) - \mathbb{E}f(G)\|_\ell &\leq \|f(G) - f(G')\|_\ell \\
&\leq \int_0^{\pi/2} \left\|\nabla f(G_t)^\top V_t\right\|_\ell \,\mathrm{d}t \\
&= \frac{\pi}{2} m_\ell \left\|\|\nabla f(G)\|\right\|_\ell .
\end{aligned}$$

The last equality conditions on G_t: the inner product is a scalar Gaussian with conditional standard deviation $\|\nabla f(G_t)\|$. Truncation and smooth approximation in Gaussian $W^{1,\ell}$ extend the inequality to its stated domain. □

Theorem 10 (Uniform bounds for all fixed moments)**.** *Under Assumption 3, for every $K \geq 2$ and real $\ell \geq 2$,*

$$w \geq c_0/C_0, \qquad \|w-1\|_{L^\ell(q_{\mathrm{pool}})} \leq B_\ell \frac{h}{\sqrt{K-1}}, \qquad B_\ell = \frac{\pi m_\ell M_{\ell,d} L_2 \Sigma^2}{2c_0}. \tag{30}$$

In particular,

$$B_\ell \leq \frac{\pi}{2}\frac{L_2\Sigma^2}{c_0}\sqrt{(\ell+2)(d+\ell+2)}. \tag{31}$$

The constants are independent of K and h.

Proof of Theorem 10. Set $c_K = \sqrt{K/(K-1)}$. Regard S as a function of the independent Gaussian input $\mathbf{G} = (G_1, \ldots, G_K)$. Then

$$\nabla_{G_i} S = \frac{c_K\sqrt{h}}{K}\sigma^\top\left(\nabla H(Y_i) - \frac{1}{K}\sum_j \nabla H(Y_j)\right).$$

The average minimizes the sum of squared Euclidean distances. Using this fact and the Hessian bound gives

$$\begin{aligned}\|\nabla_{\mathbf{G}} S\|^2 &\leq \frac{c_K^2 h\Sigma^2}{K^2}\sum_i \|\nabla H(Y_i) - \nabla H(\mu)\|^2 \\ &\leq \frac{c_K^2 h^2 L_2^2 \Sigma^4}{K^2}\sum_i \|Z_i\|^2.\end{aligned}$$

For $\ell \geq 2$, Minkowski in $L^{\ell/2}$ yields

$$\left\|\left(\sum_i \|Z_i\|^2\right)^{1/2}\right\|_\ell \leq \sqrt{K} M_{\ell,d}.$$

Consequently Lemma 9 and $\mathbb{E}_{q_{\mathrm{pool}}} S = G$ imply

$$\|S - G\|_\ell \leq \frac{\pi}{2} m_\ell M_{\ell,d} L_2 \Sigma^2 \frac{h}{\sqrt{K-1}}.$$

All Sobolev requirements hold because S is bounded and the displayed derivative bound has every finite Gaussian moment. Finally $G \geq c_0$ gives the relative moment estimate, and $S \geq c_0$, $G \leq C_0$ give $w \geq c_0/C_0$.

For its explicit dependence on ℓ, Gaussian integration gives

$$m_\ell = \left\{\frac{2^{\ell/2}\Gamma((\ell+1)/2)}{\sqrt{\pi}}\right\}^{1/\ell}, \qquad M_{\ell,d} = \left\{\frac{2^{\ell/2}\Gamma((d+\ell)/2)}{\Gamma(d/2)}\right\}^{1/\ell}.$$

Let $s = \lceil \ell/2 \rceil$. Monotonicity of L^ℓ norms and the even moment identities imply

$$m_\ell \leq m_{2s} = \left\{\prod_{j=1}^{s}(2j-1)\right\}^{1/(2s)} \leq \sqrt{\ell+2}, \quad M_{\ell,d} \leq \left\{\prod_{j=0}^{s-1}(d+2j)\right\}^{1/(2s)} \leq \sqrt{d+\ell+2}.$$

Substitution proves (31). □

The proof differentiates the pool average with respect to its independent Gaussian inputs. Centering leaves differences of gradients, which provide the extra factor $\sqrt{h}$. Every fixed finite moment is covered, including at $K = 2$; there is no restriction $\ell < K$.

Assumption 3 is sufficient for these moment bounds. The generic outer-MH inequalities in Appendix D use moments of the normalized complete weight. The path moment bound permits conditional pool moments to be supplied directly, without a uniform positive lower bound on the normalized pool weight. The weaker relative L^2 curvature bound (28) alone does not provide all higher moments.

B.5 Stein ratio accuracy and mean displacement

The guidance estimator and the corrected output have different error metrics. Here we quantify the estimator's variance and bias, then show how multiplication by the step size reduces its effect on the selected candidate's mean.

The same candidate geometry controls the guidance estimator. Under Assumption 3, let $\|\nabla H\|_\infty \le L_1$. On the full Euclidean state space one may already take $L_1 = \sqrt{2L_2(C_0 - c_0)}$. The bound is written separately to expose the dependence on gradient size.

To see the claimed gradient bound, use the full Euclidean state space. If $L_2 > 0$, the descent inequality at $y = x - \nabla H(x)/L_2$ gives

$$c_0 \le H(y) \le H(x) - \frac{\|\nabla H(x)\|^2}{2L_2} \le C_0 - \frac{\|\nabla H(x)\|^2}{2L_2}.$$

Thus $\|\nabla H(x)\|^2 \le 2L_2(C_0 - c_0)$. If $L_2 = 0$, H is affine and bounded on the full state space, hence constant.

Theorem 11 (Stein variance, ratio bias and mean error)**.** *Let $\nu = K - 1$, $V = L_1^2\Sigma^2 d(d+2)$ and $A_s = dL_2^2\Sigma^4/2$. Under Assumption 3,*

$$\mathbb{E}\,\|\widehat{\nu} - \nu_h\|^2 \le \frac{V}{\nu}, \tag{32}$$

$$\|\mathbb{E}\widehat{s} - s_h\| \le \frac{\sqrt{VA_s}}{c_0^2}\frac{h}{\nu} + \frac{\Sigma L_1 A_s}{c_0^3}\frac{h^2}{\nu}, \tag{33}$$

$$\mathbb{E}\,\|\widehat{s} - s_h\|^2 \le \frac{2V}{c_0^2\nu} + \frac{2\Sigma^2 L_1^2 A_s}{c_0^4}\frac{h^2}{\nu}. \tag{34}$$

Consequently, for $0 < h \le h_0$, uncorrected weighted selection obeys

$$\left\|\mathbb{E}_{q_{\mathrm{pool}}} Y_J - \mathbb{E}_{K_k^H(x,\cdot)} Y\right\| \le C\frac{h^2}{K-1}, \tag{35}$$

where C depends only on $c_0, L_1, L_2, \Sigma, d, h_0$.

Proof of Theorem 11. The baseline identity writes $\widehat{\nu} = K^{-1}\sum_i F_\nu(Z_i)$, where

$$F_\nu(z) = \frac{z}{\sqrt{h}}\{H(\mu + \sqrt{h}\sigma z) - H(\mu)\}, \qquad \|F_\nu(z)\| \le L_1\Sigma\,\|z\|^2.$$

Sum (26) over its coordinates. Since $\mathbb{E}\,\|Z\|^4 = d(d+2)$, this proves

$$\mathbb{E}\,\|\widehat{\nu} - \nu_h\|^2 \le \frac{V}{\nu}.$$

Theorem 8 and $\left\|\sigma^\top\nabla^2 H\sigma\right\|_F^2 \le dL_2^2\Sigma^4$ give $\mathrm{Var}(S) \le A_s h^2/\nu$. Also $\|\nu_h\| \le \Sigma L_1$.

Because $S, G \ge c_0$,

$$\|1/S - 1/G\|_2 \le c_0^{-2}\sqrt{\mathrm{Var}(S)}, \qquad \mathbb{E}\frac{1}{S} - \frac{1}{G} = \mathbb{E}\frac{(S-G)^2}{G^2 S}.$$

The second equality follows by expanding $1/S$ around G and using $\mathbb{E}(S - G) = 0$. Hence

$$\mathbb{E}\widehat{s} - s_h = \mathbb{E}\left[(\widehat{\nu} - \nu_h)(1/S - 1/G)\right] + \nu_h\left(\mathbb{E}(1/S) - 1/G\right),$$

$$\|\mathbb{E}\widehat{s} - s_h\| \le c_0^{-2}\sqrt{\mathbb{E}\,\|\widehat{\nu} - \nu_h\|^2\,\mathrm{Var}(S)} + c_0^{-3}\,\|\nu_h\|\,\mathrm{Var}(S).$$

Substitution proves (33). The decomposition

$$\widehat{s} - s_h = \frac{\widehat{\nu} - \nu_h}{S} + \frac{\nu_h(G - S)}{SG}$$

and $\|a+b\|^2 \le 2\,\|a\|^2 + 2\,\|b\|^2$ prove (34). Finally, subtracting (23) from the expectation of (22) multiplies the ratio bias by $h\sigma$. Using $h \le h_0$ proves (35), for example with $C = \Sigma\sqrt{VA_s}/c_0^2 + h_0\Sigma^2 L_1 A_s/c_0^3$. □

C Complete-Path Weights and Moment Bounds

We first prove the complete-path importance-weight identity. Uniform spatial regularity then supplies all fixed moments of a complete proposal weight. A change of measure separates finite-pool noise from the accumulated Bellman error. The resulting bound on centered moments quantifies the improvement when this error varies little across ideal tilted paths and yields the correction bounds in Appendix D.

C.1 From a weighted pool to the target path

The following theorem justifies using the weight returned by Algorithm 1 in whole-path MH. It shows that reweighting the augmented proposal recovers the target path law. We use the standing assumptions of Section 3, including the fixed start and $H_n = \phi$. The statement also allows nonuniform step sizes h_k and pool sizes K_k. Let $\widetilde{Q}$ denote the joint law of the candidate pools, selected indices, and path generated by Algorithm 1. Each step uses fresh Gaussian inputs and a fresh selection uniform, independent of the preceding history.

Theorem 12 (Exact augmented correction)**.** *Every bounded measurable path function F satisfies*

$$R = H_0(x_\star) \prod_{k=0}^{n-1} \frac{S_k}{H_k(X_k)}, \qquad \mathbb{E}_{\widetilde{Q}}[RF] = \mathbb{E}_{P_n}[\phi(X_n)F]. \tag{36}$$

In particular,

$$\mathbb{E}_{\widetilde{Q}}[R] = Z_n, \qquad W = \frac{R}{Z_n}, \qquad \mathbb{E}_{\widetilde{Q}}[W] = 1. \tag{37}$$

Thus $\widetilde{\Pi} = W\widetilde{Q}$ is a probability law. For every measurable set of paths A,

$$\widetilde{\Pi}(X_{0:n} \in A) = \Pi_n(A).$$

This is the physical marginal: after the candidate pools and selected indices are integrated out, the selected path $X_{0:n}$ has law Π_n.

Proof of Theorem 12. Equation (16) makes the weighted child transition $P_k(H_{k+1}f)/H_k$. Iteration for nonnegative product cylinder functions telescopes the intermediate values, leaving $H_n = \phi$ and cancellation of the initial H_0. Taking $F = 1$ gives finite mass Z_n; a monotone-class argument extends the identity to bounded measurable path functions. In particular, $\mathbb{E}_{\widetilde{Q}}R = Z_n$ and $\widetilde{\Pi} = (R/Z_n)\widetilde{Q}$ has physical marginal Π_n. □

For the moment bounds below, we factor the same weight into pool and Bellman terms. At a parent x, let C_x select a child from a pool with probabilities proportional to $H_{k+1}(Y_i)$. Correct candidate marginals give

$$\mathbb{E}_q[S_k(U)C_x f(U)] = \frac{1}{K_k} \sum_{i=1}^{K_k} \mathbb{E}H_{k+1}(Y_i)f(Y_i) = P_k(H_{k+1}f)(x). \tag{38}$$

Here q denotes the pool law at this parent. Write

$$\Gamma_k = G_k/H_k, \qquad B(X_{0:n}) = \sum_{k=0}^{n-1} \log \Gamma_k(X_k). \tag{39}$$

Here $\log \Gamma_k$ is the local Bellman error and B is the accumulated Bellman error along the path. The centered norm $\mathcal{A}_r$ in (11) measures the variation of B around a constant under Π_n^H. With normalized pool averages $w_k = S_k/G_k(X_k)$, the complete weight factors as

$$R = H_0(x_\star) \prod_{k=0}^{n-1} \frac{S_k}{H_k(X_k)} = H_0(x_\star)e^B \prod_{k=0}^{n-1} w_k. \tag{40}$$

C.2 Complete-weight moments from spatial regularity

Telescoping against the selected values gives a second factorization that does not involve the accumulated Bellman error. It proves the depth-uniform moment bounds used in the main theorem, using only the candidate marginals and their common Gaussian scale.

Theorem 13 (Complete-path moments from spatial regularity)**.** *Under (9), the fixed start and terminal agreement $H_n = \phi$, let $\sum_k h_k = T$ and allow arbitrary pool sizes $K_k \geq 2$. For every real $\ell \geq 2$,*

$$\mathbb{E}_{\widetilde{Q}} W^\ell \leq \mathsf{M}_\ell, \qquad \mathsf{M}_\ell = \left(\frac{C_0}{c_0}\right)^\ell e^{a_\ell T}, \tag{41}$$

where

$$V = \frac{2L_2(C_0 - c_0)\Sigma^2 d}{c_0^2}, \qquad a_\ell = \frac{\ell(\ell-1)}{2}\left(\frac{C_0}{c_0}\right)^{\ell-2} V. \tag{42}$$

These constants are independent of the numerical depth and pool sizes. The result also holds for uncentered pools with the same Gaussian candidate marginals.

Proof. Let $\mathcal{F}_k$ contain the history before pool k is drawn, and write $H_{k,i} = H_{k+1}(Y_{k,i})$, $S_k = K_k^{-1}\sum_i H_{k,i}$. For the selected index J_k, define

$$\eta_k = \frac{S_k}{H_{k,J_k}}, \qquad N_j = \prod_{k<j} \eta_k, \qquad N_0 = 1. \tag{43}$$

Conditional on $\mathcal{F}_k$ and the pool U_k, selection has probability $H_{k,i}/(K_k S_k)$, and hence

$$\mathbb{E}[\eta_k \mid \mathcal{F}_k, U_k] = \sum_i \frac{H_{k,i}}{K_k S_k}\frac{S_k}{H_{k,i}} = 1.$$

Thus N_j is a positive mean-one martingale. Cancellation of the selected values in (40) gives the pathwise identity

$$R = H_n(X_n)N_n = \phi(X_n)N_n. \tag{44}$$

The global Hessian and value bounds imply $\|\nabla H_{k+1}\|^2 \leq L_g^2 := 2L_2(C_0 - c_0)$. Indeed, for $L_2 > 0$ the descent inequality at $x - \nabla H_{k+1}(x)/L_2$ gives $c_0 \leq H_{k+1}(x) - \|\nabla H_{k+1}(x)\|^2/(2L_2)$; if $L_2 = 0$, boundedness makes the affine value constant. Writing μ_k for the pool's common frozen mean, we obtain

$$\begin{aligned}\mathbb{E}[(\eta_k - 1)^2 \mid \mathcal{F}_k, U_k] &= \frac{1}{K_k S_k}\sum_i \frac{(S_k - H_{k,i})^2}{H_{k,i}} \\ &\leq \frac{1}{K_k c_0^2}\sum_i \{H_{k,i} - H_{k+1}(\mu_k)\}^2.\end{aligned}$$

The last inequality uses the minimizing property of the sample mean. Each candidate marginal is $\mu_k + \sqrt{h_k}\sigma_k Z$ with $Z \sim N(0, I_d)$, so the gradient bound gives

$$\mathbb{E}[(\eta_k - 1)^2 \mid \mathcal{F}_k] \leq \frac{L_g^2 h_k}{c_0^2}\mathbb{E}\|\sigma_k Z\|^2 \leq V h_k.$$

No independence or centering within the pool is used. Since $c_0/C_0 \leq \eta_k \leq C_0/c_0$, Taylor's theorem for z^ℓ about 1, together with the conditional mean-one identity, gives

$$\mathbb{E}[\eta_k^\ell \mid \mathcal{F}_k] \leq 1 + a_\ell h_k.$$

Iteration yields $\mathbb{E}N_n^\ell \leq \prod_k (1 + a_\ell h_k) \leq e^{a_\ell T}$. Finally, $c_0 \leq \phi \leq C_0$ implies $c_0 \leq Z_n \leq C_0$. Equation (44) therefore gives $W \leq (C_0/c_0)N_n$ and proves (41). □

This argument controls complete weights directly. It needs neither temporal regularity of the values nor a drift bound: candidates are compared with their common frozen mean. It supplies no mesh-vanishing coefficient by itself. The following change-of-measure argument quantifies such a coefficient through centered moments of the accumulated Bellman error.

C.3 The accumulated Bellman error under the ideal tilted path

The ideal transitions $K_k^H(x, \mathrm{d}y) = H_{k+1}(y)P_k(x, \mathrm{d}y)/G_k(x)$ define $\Pi_n^H = \delta_{x_\star}\prod_k K_k^H$. Telescoping gives

$$L := \frac{\mathrm{d}\Pi_n}{\mathrm{d}\Pi_n^H} = \frac{H_0(x_\star)}{Z_n}e^B = \frac{e^B}{\mathbb{E}_{\Pi_n^H} e^B}, \qquad \Lambda_s := \|L\|_{L^s(\Pi_n^H)}. \tag{45}$$

All centered norms of the accumulated Bellman error in this appendix use this ideal path law. Neither L nor $\mathcal{A}_r$ depends on the pool size when the reference grid and value functions are fixed.

Lemma 14 (Ideal density moments and Bellman error bounds)**.** *Under (9), the fixed start and $H_n = \phi$, let $\rho = C_0/c_0$ and use V, a_s from (42). For every real $s \geq 2$ and $1 \leq r < \infty$,*

$$\Lambda_s \leq \overline{\Lambda}_s := \rho e^{a_s T/s}, \qquad \mathbb{E}_{\Pi_n^H} L^{-1} \leq \rho e^{VT}, \tag{46}$$

$$\mathcal{A}_r \leq \|\log L\|_{L^r(\Pi_n^H)} \leq \frac{r}{e}(1 + \rho e^{VT})^{1/r}. \tag{47}$$

Thus the stated regularity conditions bound every required centered norm of the accumulated Bellman error and every required density moment, with constants independent of n and K.

Proof. Under Π_n^H, put $\eta_k^H = G_k(X_k)/H_{k+1}(X_{k+1})$ and $N_j^H = \prod_{k<j} \eta_k^H$. Relative to the physical path history $\mathcal{F}_k^H = \sigma(X_0, \ldots, X_k)$, the ideal transition gives $\mathbb{E}[\eta_k^H \mid \mathcal{F}_k^H] = 1$. Moreover, $\rho^{-1} \leq \eta_k^H \leq \rho$, and

$$\mathbb{E}[(\eta_k^H - 1)^2 \mid X_k = x] = \frac{1}{G_k(x)} P_k\left(\frac{(G_k(x) - H_{k+1})^2}{H_{k+1}}\right)(x) \leq c_0^{-2} \operatorname{Var}_{P_k(x,\cdot)} H_{k+1} \leq V h_k.$$

The last inequality uses the global gradient bound established in the proof of Theorem 13, comparing the Gaussian candidate with its frozen mean. Taylor expansion of z^s about 1 gives $\mathbb{E}[(\eta_k^H)^s \mid \mathcal{F}_k^H] \leq 1 + a_s h_k$ for $s \geq 2$. Conditional iteration therefore yields $\mathbb{E}(N_n^H)^s \leq e^{a_s T}$. The pathwise identity

$$N_n^H = \frac{H_0(x_\star)}{\phi(X_n)} e^B, \qquad L = \frac{\phi(X_n)}{Z_n} N_n^H$$

and $\phi/Z_n \leq \rho$ prove the positive-moment bound.

For the inverse factor the same kernel gives

$$\mathbb{E}[(\eta_k^H)^{-1} \mid X_k = x] = \frac{P_k(H_{k+1}^2)(x)}{G_k(x)^2} = 1 + \frac{\operatorname{Var}_{P_k(x,\cdot)} H_{k+1}}{G_k(x)^2} \leq 1 + V h_k.$$

Iterating this inequality, without claiming that the inverse product is a mean-one martingale, proves $\mathbb{E}(N_n^H)^{-1} \leq e^{VT}$ and hence the inverse-density bound. Taking $b = \log \mathbb{E}_{\Pi_n^H} e^B$ in the definition of $\mathcal{A}_r$ gives its first bound. Finally, for $z > 0$, $|\log z|^r \leq (r/e)^r (z + z^{-1})$; this follows from $\sup_{u \geq 0} u^r e^{-u} = (r/e)^r$ on each side of $z = 1$. Use $\mathbb{E}L = 1$ and the inverse-density bound to conclude. □

The upper bound in (47) establishes finiteness uniformly in numerical depth. A small $\mathcal{A}_r$ additionally requires that the accumulated Bellman error vary little across paths under Π_n^H.

Proposition 15 (Accumulated Bellman error and path discrepancy)**.** *Whenever Λ_2 and $\mathcal{A}_2$ are finite,*

$$\|\Pi_n^H - \Pi_n\|_{\mathrm{TV}} \leq \min\{1, 2\Lambda_2 \mathcal{A}_2\}. \tag{48}$$

Let $\Pi_{n,k}^H$ be the kth marginal of Π_n^H. For any $r \geq 1$, if

$$\inf_{a \in \mathbb{R}} \|\log \Gamma_k - a\|_{L^r(\Pi_{n,k}^H)} \leq h_k \epsilon_{r,k}, \tag{49}$$

then $\mathcal{A}_r \leq \sum_k h_k \epsilon_{r,k}$. In particular, $\epsilon_{r,k} \leq \epsilon_r$ gives $\mathcal{A}_r \leq T \epsilon_r$.

Proof. Let X, X' be independent ideal paths, with corresponding B, B' and L, L'. Since $\mathbb{E}L' = 1$, conditional Jensen and $|e^u - e^v| \leq (e^u + e^v)|u - v|$ imply

$$\tfrac{1}{2}\mathbb{E}|L - 1| \leq \tfrac{1}{2}\mathbb{E}|L - L'| \leq \tfrac{1}{2}\mathbb{E}[(L + L')|B - B'|] \leq 2\Lambda_2 \mathcal{A}_2.$$

The last step uses Cauchy–Schwarz, $\|L + L'\|_2 \leq 2\Lambda_2$, and $\|B - B'\|_2 \leq 2 \inf_b \|B - b\|_2$. TV is also at most one. For the bound from local residuals, choose deterministic centers a_k arbitrarily close to the displayed infima. Minkowski gives

$$\mathcal{A}_r \leq \left\| \sum_k (\log \Gamma_k(X_k) - a_k) \right\|_r \leq \sum_k \|\log \Gamma_k - a_k\|_{L^r(\Pi_{n,k}^H)}.$$

Letting the centering errors tend to zero proves the claim; no independence across time is used. □

The distribution in (49) is the ideal path marginal. A bound under the reference or finite-pool proposal marginal requires a separate comparison.

C.4 The conditional-moment interface

To make the dependence on the accumulated Bellman error explicit, we use conditional moments of the normalized pool averages. Uniform regularity supplies these at every fixed order, but more general proposals can use the same bounds. They are distinct from the selected-index moments proved above. We write $\ell \geq 2$ for the moment order throughout Appendices B–F; the main decay order p uses $\ell = p + 1$.

Assumption 16 (Conditional moments at a fixed order)**.** Let $\mathcal{F}_k$ contain the proposal history before pool k is drawn. At a specified order $\ell \geq 2$, deterministic finite $u_{\ell,k} \geq 0$ satisfy

$$\mathbb{E}[w_k \mid \mathcal{F}_k] = 1, \qquad \mathbb{E}[|w_k - 1|^\ell \mid \mathcal{F}_k] \leq u_{\ell,k}^\ell.$$

The mean-one identity follows from the candidate marginals. Under (9), Theorem 10 gives $u_{\ell,k} = B_\ell h_k / \sqrt{K_k - 1}$ for every fixed real $\ell \geq 2$, with no derivatives above order two and no restriction $\ell < K_k$. The weaker relative-curvature condition (28) supplies only $u_{2,k}^2 = C_2^2 h_k^2 / [2(K_k - 1)]$. Neither the product argument nor outer MH requires a positive lower bound on w_k. Define the accumulated local fluctuation budget

$$\mathcal{V}_\ell = \sum_{k=0}^{n-1} \log\{1 + (\ell - 1) u_{\ell,k}^2\}. \tag{50}$$

C.5 A centered product-martingale bound

Conditional centering, rather than independence across time, controls the accumulated pool noise. The next inequality keeps the bound centered at one as the proposal approaches its target.

Lemma 17 (L^ℓ quadratic smoothness)**.** *Let $\ell \geq 2$ and $X, Y \in L^\ell$. If X is $\mathcal{F}$-measurable and $\mathbb{E}[Y \mid \mathcal{F}] = 0$, then*

$$\|X + Y\|_\ell^2 \leq \|X\|_\ell^2 + (\ell - 1) \|Y\|_\ell^2 . \tag{51}$$

Proof. For $\ell = 2$, expand the square and use $\mathbb{E}XY = 0$. For $\ell > 2$, set $a(t) = \mathbb{E}|X + tY|^\ell$ and $f_\eta(t) = (a(t) + \eta^\ell)^{2/\ell}$, where $\eta > 0$. Hölder's inequality and domination by powers of $|X| + |Y|$ justify differentiation. The squared-first-derivative term in f_η'' is nonpositive, because $2/\ell - 1 < 0$. Therefore

$$\begin{aligned} f_\eta''(t) &\leq 2(\ell - 1)(a(t) + \eta^\ell)^{2/\ell - 1} \mathbb{E}\{|X + tY|^{\ell - 2} Y^2\} \\ &\leq 2(\ell - 1)(a(t) + \eta^\ell)^{-(\ell-2)/\ell} a(t)^{(\ell-2)/\ell} \|Y\|_\ell^2 \leq 2(\ell - 1) \|Y\|_\ell^2 . \end{aligned}$$

Conditional centering gives $a'(0) = \ell \mathbb{E}[|X|^{\ell - 2} XY] = 0$, so $f_\eta'(0) = 0$. Integrating on $[0, 1]$ gives $f_\eta(1) \leq f_\eta(0) + (\ell - 1) \|Y\|_\ell^2$. Letting $\eta \downarrow 0$ proves the result. □

Theorem 18 (Product moments at one fixed order)**.** *Under Assumption 16 at order $\ell \geq 2$, the product $M_j = \prod_{k<j} w_k$, $M_0 = 1$, is a positive mean-one martingale. Writing $\mathcal{V}_\ell = \sum_k \log\{1 + (\ell - 1) u_{\ell,k}^2\}$,*

$$\|M_n\|_\ell^2 \leq e^{\mathcal{V}_\ell}, \qquad \|M_n - 1\|_\ell^2 \leq e^{\mathcal{V}_\ell} - 1. \tag{52}$$

Proof. The filtration $\mathcal{F}_k$ contains the generated history up to the current parent, before pool k is drawn. Proper weighting gives $\mathbb{E}[w_k \mid \mathcal{F}_k] = 1$. Put $Y_k = M_k(w_k - 1)$. Since M_k is $\mathcal{F}_k$-measurable, the conditional moment assumption gives

$$\mathbb{E}[Y_k \mid \mathcal{F}_k] = 0, \qquad \mathbb{E}|Y_k|^\ell = \mathbb{E}\{M_k^\ell \mathbb{E}[|w_k - 1|^\ell \mid \mathcal{F}_k]\} \leq u_{\ell,k}^\ell \mathbb{E} M_k^\ell.$$

Starting with $M_0 = 1$, induction establishes finiteness of all moments used below. Let $b_k = (\ell - 1) u_{\ell,k}^2$. Applying Lemma 17 to $M_{k+1} = M_k + Y_k$ gives

$$\|M_{k+1}\|_\ell^2 \leq (1 + b_k) \|M_k\|_\ell^2 \leq \prod_{j=0}^{k} (1 + b_j).$$

Also $\mathbb{E}[M_{k+1} \mid \mathcal{F}_k] = M_k$, which proves the mean-one martingale assertion. Applying the same lemma to $M_{k+1} - 1 = (M_k - 1) + Y_k$ and summing gives

$$\|M_n - 1\|_\ell^2 \le \sum_{k=0}^{n-1} b_k \|M_k\|_\ell^2 \le \sum_{k=0}^{n-1} b_k \prod_{j<k}(1 + b_j)$$
$$= \prod_{k=0}^{n-1}(1 + b_k) - 1 = e^{\mathcal{V}_\ell} - 1.$$

This proves both bounds without independent increments or an increasing sequence of moment orders along the path. □

C.6 Change of measure and centered complete-path weights

The pool martingale and the ideal density depend on the same physical path. The following change of measure separates their moments without assuming independence.

Lemma 19 (Finite-moment separation)**.** *Put $M = M_n$ and $\widehat{\widetilde{Q}} = M\widetilde{Q}$. This is a probability law whose physical path marginal is Π_n^H, and*

$$W = ML. \tag{53}$$

For every real $\ell \ge 2$ for which the right-hand side is finite,

$$\|W - 1\|_{L^\ell(\widetilde{Q})} \le \|M - 1\|_{L^\ell(\widetilde{Q})} + 4\Lambda_{4\ell}\|M\|_{L^{4\ell-3}(\widetilde{Q})}^{(4\ell-3)/(4\ell)}\mathcal{A}_{2\ell}. \tag{54}$$

Proof. The product martingale gives $\mathbb{E}_{\widetilde{Q}} M = 1$. Dividing the weighted-pool identity (38) by $G_k(x)$ gives $\mathbb{E}_q[w_k C_x f] = K_k^H f(x)$. Iterating this identity proves that the physical marginal of $M\widetilde{Q}$ is Π_n^H. Equations (40) and (45) then give $W = ML$.

Extend $\widehat{\widetilde{Q}}$ by an independent physical path $X' \sim \Pi_n^H$ and denote its values by B', L'. Since $\mathbb{E}L' = 1$, Jensen conditional on the augmented state and the exponential difference inequality give

$$\mathbb{E}_{\widetilde{Q}}|M(L-1)|^\ell = \mathbb{E}_{\widehat{\widetilde{Q}}} M^{\ell-1}|L-1|^\ell$$
$$\le \mathbb{E}_{\widehat{\widetilde{Q}}\otimes\Pi_n^H} M^{\ell-1}(L + L')^\ell |B - B'|^\ell.$$

Apply Hölder with exponents $4, 4, 2$ and then take the ℓth root. The three factors are bounded respectively by

$$\left(\mathbb{E}_{\widehat{\widetilde{Q}}} M^{4\ell-4}\right)^{1/(4\ell)} = \|M\|_{L^{4\ell-3}(\widetilde{Q})}^{(4\ell-3)/(4\ell)}, \quad 2\Lambda_{4\ell}, \quad 2\mathcal{A}_{2\ell}.$$

For the last factor subtract an arbitrary deterministic center from each of B, B' and take the infimum. The marginal identity just proved justifies both ideal-path norms; it does not make M and L independent. Finally, use $W - 1 = (M - 1) + M(L - 1)$ and Minkowski. □

Theorem 20 (Complete-path moments with Bellman error dependence)**.** *Suppose Assumption 16 holds at orders ℓ and $4\ell - 3$, where $\ell \ge 2$, and $\Lambda_{4\ell}, \mathcal{A}_{2\ell} < \infty$. Then*

$$\|W - 1\|_{L^\ell(\widetilde{Q})} \le \Delta_\ell := \sqrt{e^{\mathcal{V}_\ell} - 1} + 4\Lambda_{4\ell} e^{(4\ell-3)\mathcal{V}_{4\ell-3}/(8\ell)}\mathcal{A}_{2\ell}. \tag{55}$$

In particular $\mathbb{E}_{\widetilde{Q}} W^\ell \le (1 + \Delta_\ell)^\ell$, and at order two

$$D = \chi^2(\widetilde{\Pi}\|\widetilde{Q}) = \mathbb{E}_{\widetilde{Q}}(W - 1)^2 \le D_\star := \Delta_2^2. \tag{56}$$

Independently of the centered moment bound, for any $\ell \ge 2$ the local moment at order $2\ell - 1$ and $\Lambda_{2\ell} < \infty$ give

$$\mathbb{E}_{\widetilde{Q}} W^\ell \le \Lambda_{2\ell}^\ell e^{(2\ell-1)\mathcal{V}_{2\ell-1}/4}. \tag{57}$$

If B is constant, then $L = 1$ and $W = M$. This case needs only the local moment at order ℓ and gives $\|W - 1\|_\ell \le \sqrt{e^{\mathcal{V}_\ell} - 1}$. In particular a conditional variance bound v_k gives $D \le \prod_k(1 + v_k) - 1$. Under the relative-curvature condition on a uniform grid,

$$D \le \exp\left\{\frac{C_2^2 T h}{2(K-1)}\right\} - 1. \tag{58}$$

Proof. Insert the two product bounds from Theorem 18 into Lemma 19. Minkowski gives the uncentered moment, and $\mathbb{E}W = 1$ identifies the variance with the centered second moment. For (57), change measure once and use Cauchy–Schwarz:

$$\mathbb{E}_{\widetilde{Q}} W^{\ell} = \mathbb{E}_{\widehat{\widetilde{Q}}} M^{\ell-1} L^{\ell} \le (\mathbb{E}_{\widetilde{Q}} M^{2\ell-1})^{1/2} (\mathbb{E}_{\Pi_n^H} L^{2\ell})^{1/2} \le e^{(2\ell-1)\mathcal{V}_{2\ell-1}/4} \Lambda_{2\ell}^{\ell}.$$

Constant B makes $L = 1$ by normalization, so the sharper assertions follow directly from the product theorem without higher-order moment conditions on the density or accumulated Bellman error. Finally, take $v_k = C_2^2 h^2/[2(K-1)]$ and use $\log(1+v) \le v$ and $nh^2 = Th$. □

Corollary 21 (Uniform scaling with Bellman error)**.** *Under (9), on a uniform grid $h = T/n$ with pool size $K \ge 2$, all conditions of Theorem 20 hold at every fixed real $\ell \ge 2$. The local moment bounds give*

$$\mathcal{V}_{\ell} \le (\ell-1) B_{\ell}^2 \frac{Th}{K-1} \le \frac{\pi^2 L_2^2 \Sigma^4}{4c_0^2} (\ell-1)(\ell+2)(d+\ell+2) \frac{Th}{K-1}. \tag{59}$$

There is $D_{\ell} < \infty$, depending only on $\ell, T, d, c_0, C_0, L_2, \Sigma$, such that

$$\|W-1\|_{\ell} \le \Delta_{\ell} \le D_{\ell} \left(\mathcal{A}_{2\ell} + \sqrt{\frac{Th}{K-1}} \right). \tag{60}$$

Consequently,

$$D \le D_{\star} \le 2D_2^2 \left(\mathcal{A}_4^2 + \frac{Th}{K-1} \right). \tag{61}$$

Proof. Theorem 10 supplies each required local moment, and Lemma 14 supplies the ideal density moments and centered norms of the accumulated Bellman error. Set $\delta^2 = Th/(K-1) \le T^2$ and $\gamma_s = (s-1)B_s^2$. Summing the local bounds with $\log(1+x) \le x$ proves (59). Since $e^v - 1 \le ve^v$, one admissible constant is

$$D_{\ell} = \max \left\{ \sqrt{\gamma_{\ell}}\, e^{\gamma_{\ell} T^2/2}, \quad 4\overline{\Lambda}_{4\ell} e^{(4\ell-3)\gamma_{4\ell-3} T^2/(8\ell)} \right\}. \tag{62}$$

Substitute into (55), then square and use $(a+b)^2 \le 2(a^2+b^2)$ at order two. □

The constants depend on the fixed moment order and can grow rapidly with it. The local variance scale $h^2/(K-1)$ accumulates to $Th/(K-1)$ despite dependence across time; no conditional centering of $\log w_k$ is needed.

C.7 Initialization and value function requirements

The fixed starting point and terminal value determine which factors appear in the weight. This subsection records the adjustments needed when either is changed.

Proposition 22 (Random initialization)**.** *Let the reference start from μ_0 and the proposal draw $X_0 \sim \nu_0$, where μ_0 and ν_0 are equivalent and $\ell_0 = \mathrm{d}\mu_0/\mathrm{d}\nu_0$ is finite and positive almost surely. For $P_n^{\mu_0} = \mu_0 \prod_k P_k$ and $\Pi_n^{\mu_0} = Z_n^{-1} \phi(X_n) P_n^{\mu_0}$, the proper weight is*

$$R = \ell_0(X_0) H_0(X_0) \prod_{k=0}^{n-1} \frac{S_k(U_k)}{H_k(X_k)}. \tag{63}$$

Moment bounds for this initialization require control of its additional initial density factor.

Proof. Conditioning on $X_0 = x$, the fixed-initial-state proper-weighting identity gives, for bounded physical path functions F,

$$\mathbb{E}_{\widetilde{Q}}[RF] = \int \ell_0(x) \nu_0(\mathrm{d}x) \mathbb{E}_{P_n^x}[\phi(X_n) F] = \mathbb{E}_{P_n^{\mu_0}}[\phi(X_n) F].$$

In particular, $\mathbb{E}R = Z_n$. Including X_0 in the filtration preserves $\mathbb{E}[w_k \mid \mathcal{F}_k] = 1$, so M_n remains mean one. The selected-value telescoping identity becomes $R = \ell_0(X_0)\phi(X_n)N_n$. Thus any moment claim must also account for $\ell_0(X_0)$; proper weighting alone does not supply its moments. The generic outer-MH results continue to apply whenever their weight-moment hypotheses are verified. □

If $\nu_0 = \mu_0$ and the values solve the exact discrete Doob equations, the incremental Bellman sum is zero but $W = H_0(X_0)M_n/Z_n$. The initial value need not be constant, so the fixed-start exact-Doob rate does not automatically extend to this initialization. Sampling instead from $\nu_0(\mathrm{d}x) = H_0(x)\mu_0(\mathrm{d}x)/\mu_0(H_0)$ makes $\ell_0 H_0$ constant, when $0 < \mu_0(H_0) < \infty$; the cost of sampling that initial tilt must be counted separately.

The terminal condition $H_n = \phi$ is also substantive. If the queried terminal value differs, proper weighting requires the factor $\phi(X_n)/H_n(X_n)$. Its moments require separate control. The value functions and complete-proposal mechanism are fixed across outer proposals. Value functions fitted on separate data can be handled conditionally on those data. Adaptation from the proposals being corrected, fresh noisy evaluations of the value functions, or numerical clipping do not inherit the weighting identity without an additional argument.

C.8 An unbounded-value diffusion with all fixed moments

Uniform value bounds are sufficient rather than necessary. This example verifies the conditional moments directly.

Proposition 23 (Brownian exponential tilt)**.** *Take $P_k(x,\cdot) = N(x,h)$, $X_0 = 0$, $T = nh$, and $\phi(x) = e^{\beta x}$ with $\beta \neq 0$. The exact discrete Doob values are $H_k(x) = \exp\{\beta x + \beta^2(T - t_k)/2\}$. For $K = 2$,*

$$w_k = \xi_h(Z_k) = e^{-\beta^2 h/2}\cosh(\beta\sqrt{h}Z_k), \qquad Z_k \overset{\text{iid}}{\sim} N(0,1), \tag{64}$$

and

$$D = \{\cosh(\beta^2 h)\}^n - 1 = \frac{\beta^4 T}{2}h + O(h^2). \tag{65}$$

For every fixed finite $\ell \geq 2$ and $0 < h \leq h_0$, $\|\xi_h - 1\|_\ell \leq C_{\ell,\beta,h_0}h$, uniformly in the parent. Thus this model satisfies the conditional-moment assumptions at every fixed finite order although H has neither a global upper bound nor a global positive lower bound.

Proof. Gaussian integration gives $P_k H_{k+1} = H_k$. Averaging the two values $H_{k+1}(x \pm \sqrt{h}Z_k)$ and dividing by $H_k(x)$ proves (64). The parent cancels, so fresh Gaussian draws make these particular factors independent across time despite the selections. With $u = \beta^2 h$, the Gaussian moment-generating function gives $\mathbb{E}\xi_h = 1$ and $\mathbb{E}\xi_h^2 = \cosh u$. Here $B = 0$ and $W = \prod_k \xi_h(Z_k)$, proving the exact second moment. The expansion follows from $\log\cosh u = u^2/2 + O(u^4)$ and $n = T/h$.

For the fixed-order moment bound, use $\cosh t - 1 \leq t^2\cosh|t|/2$ and $1 - e^{-u/2} \leq u/2$. If $u \leq u_0 = \beta^2 h_0$, then

$$|\xi_h(Z) - 1| \leq \frac{u}{2}\{1 + Z^2\cosh(\sqrt{u_0}Z)\}.$$

The bracket has every finite Gaussian moment, so one may take $C_{\ell,\beta,h_0} = (\beta^2/2)\left\|1 + Z^2\cosh(|\beta|\sqrt{h_0}Z)\right\|_\ell < \infty$. The bound does not depend on the parent, proving the required conditional moment estimate directly. □

In this example, exact values eliminate the accumulated Bellman error but leave nonzero finite-pool noise. Appendix G.3 gives the exact variance formulas for general pool size K.

D Sampling Error Bounds for Outer Metropolis–Hastings

This appendix proves Theorem 4. Generic MH moment bounds give a depth-uniform baseline. A bound on the centered complete weight makes the dependence on the variation of the accumulated Bellman error and on pool size explicit. All required moments follow from the stated spatial regularity. Coupling and moment bounds for independence MH with unbounded weights are established tools; see Deligiannidis et al. (2026). We give the constants used here because their dependence on the centered weight moments determines the near-Doob diffusion rates.

D.1 Generic setting and an ordered coupling

The coupling isolates the initial excess target mass and follows how long that mass can remain uncorrected. Let q be a probability measure on a standard Borel space, let $0 < w < \infty$ q-almost surely, and suppose $\mathbb{E}_q w = 1$. Set $\pi = wq$, and let M be independence MH with proposal q and

acceptance probability $\min\{1, w(y)/w(x)\}$. Throughout this appendix the initial law is q. For an integer $m \geq 0$, define

$$\begin{aligned} d_m &= \|q\mathsf{M}^m - \pi\|_{\mathrm{TV}}, \\ s &= \mathbb{E}_q(w-1)_+ = \mathbb{E}_q(1-w)_+ = \|q-\pi\|_{\mathrm{TV}}, \qquad c = \mathbb{E}_q \min(w,1) = 1 - s > 0. \end{aligned} \tag{66}$$

When finite, write $D = \mathbb{E}_q(w-1)^2$. For a possible weight value $a > 0$, define its rejection probability by

$$r(a) = \mathbb{E}_q(1 - w/a)_+, \qquad 1 - r(a) = \mathbb{E}_q \min(1, w/a) > 0. \tag{67}$$

The variable a is an argument, not a lower bound on the weights.

Theorem 24 (Proposal-initialized coupling bound)**.** *For every integer $m \geq 0$,*

$$d_m \leq \mathbb{E}_q[(w-1)_+ r(w)^m]. \tag{68}$$

For every $a > 1$,

$$1 - r(a) \geq \frac{c}{a}, \qquad r(a) \leq \frac{a-1+s}{a}, \qquad \frac{r(a)}{1-r(a)} \leq \frac{a-1+s}{c}. \tag{69}$$

Proof. If $s = 0$, then $q = \pi$ and every error is zero. Otherwise, couple $X_0 \sim q$ and $Y_0 \sim \pi$ by making them equal with probability c, with common law $c^{-1}\min(1, w)q$. On the remaining event draw their states from the residual laws

$$\frac{(1-w)_+}{s} q \quad \text{and} \quad \frac{(w-1)_+}{s} q,$$

respectively. Their weights then satisfy $w(X_0) < 1 < w(Y_0)$ almost surely.

Use the same independent proposal and uniform random number for both chains at every update. Until the chains meet, the higher-weight chain stays at Y_0. Indeed, if it accepts a proposal Z, the lower-weight chain also accepts because its acceptance threshold is no smaller. If the higher-weight chain rejects, then $w(Z) < w(Y_0)$, so the lower chain's retained weight remains below $w(Y_0)$ whether it accepts or rejects. This proves the ordering and freezing by induction. Once the chains meet, keep them together.

Conditionally on a residual $Y_0 = y$, failure to meet in m updates has probability $r(w(y))^m$. The Y-chain is stationary, and the coupling inequality gives

$$d_m \leq s \int r(w(y))^m \frac{(w(y)-1)_+}{s}\, q(\mathrm{d}y),$$

proving (68). For $a > 1$, $\min(a, w) \geq \min(1, w)$ gives $1 - r(a) \geq c/a$. Also

$$a r(a) = \mathbb{E}_q(a - w)_+ = a - 1 + \mathbb{E}_q(w - a)_+ \leq a - 1 + s.$$

Combining these two estimates proves (69). □

D.2 One update and stationary rejection

One update already converts weight variance into output TV. Stationary rejection is a separate quantity, which the same weights also control.

Proposition 25 (One-step bias and stationary rejection)**.** *If $D < \infty$, then*

$$d_1 \leq \mathbb{E}_q \frac{(w-1)_+^2}{w} \leq D. \tag{70}$$

For independent $U, U' \sim q$, the stationary rejection probability is

$$\rho_{\mathrm{stat}} = \frac{1}{2}\mathbb{E}|w(U) - w(U')| \leq \sqrt{D/2}. \tag{71}$$

Proof. At a point of weight $a > 0$, the density of $q\mathsf{M}$ with respect to q is

$$f_1(a) = r(a) + \mathbb{E}_q \min(1, a/w) = 1 + \mathbb{E}_q \frac{a - w}{\max(a, w)}. \tag{72}$$

The function $v \mapsto (a - v)/\max(a, v)$ is convex: it equals $1 - v/a$ below a and $a/v - 1$ above a, with matching one-sided derivatives. Jensen and $\mathbb{E}_q w = 1$ therefore imply

$$f_1(a) \geq \begin{cases} a, & a \leq 1, \\ 2 - a^{-1}, & a \geq 1. \end{cases}$$

Thus $(a - f_1(a))_+ \le (a-1)_+^2/a$. Integrating this positive density deficit proves the first inequality in (70); the second uses $1/w \le 1$ on $\{w > 1\}$.

At stationarity, the average acceptance probability is

$$\iint \pi(\mathrm{d}x)q(\mathrm{d}y)\min(1, w(y)/w(x)) = \mathbb{E}\min(w(U), w(U')).$$

The identity $\min(a,b) = (a+b-|a-b|)/2$ gives the stated rejection probability. Independence and Cauchy–Schwarz give $\mathbb{E}|w(U) - w(U')| \le \sqrt{2D}$. □

The stationary rejection bound concerns a chain already at its invariant law; it is distinct from the proposal-initialized error after one update.

D.3 A centered second-moment bound for every update count

A variance bound alone gives a bound valid at every positive update count, without any lower bound on the weight.

Theorem 26 (All-m second-moment guarantee)**.** *If $D < \infty$, then for every integer $m \ge 1$,*

$$d_m \le \min\{s, D, 3D/m\}. \tag{73}$$

If $D \le 1$, the last term can be replaced by $2D/m$.

Proof. TV contraction and invariance give $d_m \le d_0 = s$ and $d_m \le d_1 \le D$. For $0 \le r < 1$ and integer $m \ge 1$,

$$mr^m \le \sum_{j=1}^{m} r^j \le \frac{r}{1-r}.$$

Theorem 24 therefore gives

$$md_m \le \mathbb{E}_q\left[(w-1)_+ \frac{r(w)}{1-r(w)}\right]. \tag{74}$$

Let $D_+ = \mathbb{E}_q(w-1)_+^2$ and $D_- = \mathbb{E}_q(1-w)_+^2$. Then $D = D_+ + D_-$, $s^2 \le D_-$, and $s \le \sqrt{D}/2$ because $2s = \mathbb{E}_q|w-1|$. If $D \le 1$, then $c \ge 1/2$, and (69) bounds the expectation in (74) by

$$\frac{D_+ + s^2}{c} \le \frac{D}{c} \le 2D.$$

If $D > 1$, use $\min(1, w/a) \ge w/(a+w)$ and Jensen under π to obtain

$$1 - r(a) \ge \mathbb{E}_q \frac{w}{a+w} = \mathbb{E}_\pi \frac{1}{a+w} \ge \frac{1}{a + \mathbb{E}_\pi w} = \frac{1}{a+1+D}.$$

Hence $r(a)/(1-r(a)) \le a + D$, and the expectation in (74) is at most

$$\mathbb{E}_q[(w-1)_+(w+D)] = D_+ + (1+D)s \le D + (1+D) \le 3D,$$

where $s \le 1$ and $D > 1$ were used. □

The numerator is the centered variance D, rather than $1 + D$. No uniform upper or positive lower bound on w is required.

D.4 Higher centered moments and explicit overlap constants

Higher moments improve the power of the update budget. We use ℓ for moment order, so the corresponding decay power is $\ell - 1$.

Theorem 27 (High-order near-target MH bound)**.** *Let $\ell \ge 2$ and suppose $\mathbb{E}_q|w-1|^\ell < \infty$. For every integer $m \ge \ell - 1$,*

$$d_m \le \left(\frac{2(\ell-1)}{cm}\right)^{\ell-1} \mathbb{E}_q(w-1)_+^\ell. \tag{75}$$

If $D \le 1$, the prefactor can be replaced by $[4(\ell-1)/m]^{\ell-1}$. More generally,

$$c \ge \frac{1}{2+D}. \tag{76}$$

Proof. Put $\alpha = \ell - 1 \geq 1$. For $m \geq \alpha$ and $0 \leq r < 1$,

$$r^m \leq \left(\frac{\alpha}{m}\right)^\alpha \left(\frac{r}{1-r}\right)^\alpha. \tag{77}$$

For $0 < r < 1$, maximize $r^{m-\alpha}(1-r)^\alpha$; its maximum is

$$\left(\frac{m-\alpha}{m}\right)^{m-\alpha} \left(\frac{\alpha}{m}\right)^\alpha \leq \left(\frac{\alpha}{m}\right)^\alpha.$$

The endpoint $m = \alpha$ follows by continuity; $r = 0$ is immediate. Set $Z = (w-1)_+$. The coupling bound and (69) give

$$d_m \leq \left(\frac{\alpha}{cm}\right)^\alpha \mathbb{E}_q[Z(Z+s)^\alpha].$$

By convexity, $(Z+s)^\alpha \leq 2^{\alpha-1}(Z^\alpha + s^\alpha)$. Since $s = \mathbb{E}_q Z$, Jensen gives $s^{\alpha+1} \leq \mathbb{E}_q Z^{\alpha+1}$, whence

$$\mathbb{E}_q[Z(Z+s)^\alpha] \leq 2^{\alpha-1}\{\mathbb{E}_q Z^{\alpha+1} + s^{\alpha+1}\} \leq 2^\alpha \mathbb{E}_q Z^\ell.$$

This proves (75). The small-D simplification uses $c \geq 1/2$. For the general overlap bound, Cauchy–Schwarz applied to

$$1 = \mathbb{E}_q w = \mathbb{E}_q\left[\sqrt{\min(w,1)}\,\frac{w}{\sqrt{\min(w,1)}}\right]$$

gives

$$1 \leq c\,\mathbb{E}_q \frac{w^2}{\min(w,1)} = c\,\mathbb{E}_q[w\mathbf{1}_{\{w\leq 1\}} + w^2\mathbf{1}_{\{w>1\}}] \leq c(2+D).$$

□

Remark 28 (Moment order and update count). The restriction $m \geq \ell - 1$ belongs to the bound, not to existence of the moment. At a fixed update count m, its largest usable order is $\ell = m + 1$. In particular, a fourth centered moment cannot be inserted with $m = 1$ to claim fourth-order centered accuracy after a single update.

D.5 From complete-weight moments to the main theorem

Uniform weight moments supply the baseline. Centered moment bounds then quantify the improvement when the Bellman error term and pool noise are small. Let ν_m and ν_m^T denote the physical path and endpoint laws returned by Algorithm 2; Π_n^T is the target endpoint marginal.

Proof of Theorem 4. Theorem 12 identifies the augmented target as $\widetilde{\Pi} = W\widetilde{Q}$, with physical marginal Π_n. Algorithm 2 starts from $\widetilde{Q}$ and uses independence MH for this target. For every real $\ell \geq 2$, Theorem 13 gives $\mathbb{E}_{\widetilde{Q}} W^\ell \leq \mathsf{M}_\ell$, uniformly in n, K. Since $D = \mathbb{E}_{\widetilde{Q}} W^2 - 1$, equation (76) yields

$$c = \mathbb{E}_{\widetilde{Q}} \min(W,1) \geq \frac{1}{1 + \mathbb{E}_{\widetilde{Q}} W^2} \geq \frac{1}{1+\mathsf{M}_2} > 0. \tag{78}$$

Fix a real $p \geq 1$ and take $\ell = p+1$ in Theorem 27. For integers $m \geq p$,

$$\begin{aligned}\|\widetilde{Q}\mathsf{M}^m - \widetilde{\Pi}\|_{\mathrm{TV}} &\leq \frac{[2p(1+\mathsf{M}_2)]^p}{m^p} \mathbb{E}_{\widetilde{Q}}(W-1)_+^{p+1} \\ &\leq \frac{[2p(1+\mathsf{M}_2)]^p}{m^p} \min\{\mathsf{M}_{p+1}, \mathbb{E}_{\widetilde{Q}}|W-1|^{p+1}\}.\end{aligned}$$

The last step uses $(W-1)_+^{p+1} \leq W^{p+1}$ and $(W-1)_+^{p+1} \leq |W-1|^{p+1}$. As $\mathsf{M}_{p+1} \geq 1$, a common constant for both branches is

$$C_p^{\mathrm{reg}} = [2p(1+\mathsf{M}_2)]^p \mathsf{M}_{p+1}. \tag{79}$$

Projection cannot increase TV, so in particular

$$\|\nu_m - \Pi_n\|_{\mathrm{TV}} \leq \min\left\{1, \frac{C_p^{\mathrm{reg}}}{m^p} \min\{1, \mathbb{E}_{\widetilde{Q}}|W-1|^{p+1}\}\right\}. \tag{80}$$

Thus the baseline $C_p^{\mathrm{reg}} m^{-p}$ holds with no additional assumption on the accumulated Bellman error.

Set $x = Th/(K-1)$ and $\ell = p+1$. Corollary 21, proved from the local moment and ideal-density lemmas independently of this MH argument, gives

$$\mathbb{E}_{\widetilde{Q}}|W-1|^{p+1} \leq D_{p+1}^{p+1}(\mathcal{A}_{2p+2} + \sqrt{x})^{p+1} \leq 2^{(p+1)/2} D_{p+1}^{p+1}(\mathcal{A}_{2p+2}^2 + x)^{(p+1)/2}.$$

For $a, t \geq 0$, $\min\{1, at\} \leq \max\{1, a\} \min\{1, t\}$. Therefore

$$C_p = C_p^{\mathrm{reg}} \max\{1, 2^{(p+1)/2} D_{p+1}^{p+1}\} \tag{81}$$

in (80) gives (12). Its constants depend only on $p, T, d, c_0, C_0, L_2, \Sigma$ and not on n, K, m or the centered norms of the accumulated Bellman error. Lemma 14 supplies the required bounds on these norms and on density moments without an additional hypothesis. Projection to endpoints again cannot increase TV. One initial complete proposal and m fresh proposals use exactly $Kn(m+1)$ candidate values and $n(m+1)$ shared-parent coefficient evaluations, including rejected proposals. □

Equation (80) also holds with the common constant C_p and remains useful when centered moment bounds are available directly. More general proposals can instead combine the moment bounds in Theorem 20 with the generic MH bounds above, checking each required moment order.

D.6 Accumulated Bellman error and fixed correction budgets

Outer MH reduces the contribution of the Bellman error term and pool noise to output error while leaving the value functions and proposal mechanism fixed.

Corollary 29 (Near-Doob correction)**.** *Under the conditions of Theorem 4, for every integer $m \geq 1$,*

$$\|\nu_m - \Pi_n\|_{\mathrm{TV}} = O\left(\frac{\mathcal{A}_4^2 + Th/(K-1)}{m}\right). \tag{82}$$

For any fixed real $p \geq 1$ and integers $m \geq p$,

$$\|\nu_m - \Pi_n\|_{\mathrm{TV}} = O_p\left(\frac{(\mathcal{A}_{2p+2}^2 + Th/(K-1))^{(p+1)/2}}{m^p}\right).$$

The constants are independent of n, K, m. The local Bellman error bound (49) with $r = 4$ and $\epsilon_{4,k} \leq \epsilon_4$ replaces $\mathcal{A}_4^2$ by at most $T^2\epsilon_4^2$ in the first numerator.

Proof. Use Theorem 4 at $p = 1$ and at the specified order, respectively, and Proposition 15 for the bound on the centered error norm from local Bellman errors. Equivalently, the first conclusion follows from (61) and Theorem 26. □

Corollary 30 (Fixed outer budget)**.** *Under the conditions of Theorem 4, each fixed integer $m \geq 1$ satisfies*

$$\|\nu_m - \Pi_n\|_{\mathrm{TV}} = O_m\left(\left[\mathcal{A}_{2m+2}^2 + \frac{Th}{K-1}\right]^{(m+1)/2}\right), \tag{83}$$

with constants independent of n, K. If B is constant, including exact discrete Doob values at the fixed starting state, then every $\mathcal{A}_r = 0$ and the numerator reduces to $(Th/(K-1))^{(m+1)/2}$. More generally, this constant-B conclusion needs only the conditional moment bound $u_{m+1,k} \leq B_{m+1} h/\sqrt{K-1}$ with grid-independent B_{m+1}. A variance bound from relative curvature alone gives

$$\|\nu_m - \Pi_n\|_{\mathrm{TV}} = O\left(\frac{C_2^2 Th}{(K-1)m}\right)$$

when the variance budget remains bounded.

Proof. Take $p = m$ in Theorem 4 and absorb the fixed factor m^{-m} into the constant. For the more general constant-B claim, Theorem 20 gives $W = M$ and $\|W - 1\|_{m+1} \leq \sqrt{e^{\mathcal{V}_{m+1}} - 1} = O_m(\sqrt{Th/(K-1)})$. Also $\mathbb{E}W^2 \leq \|W\|_{m+1}^2 \leq e^{\mathcal{V}_{m+1}}$, so the bounded moment budget supplies a uniform positive overlap in (76). Theorem 27 proves the fixed-budget rate. Theorem 26 and (58) prove the variance-only claim. □

The constants can grow with the order. Fixed-order bounds therefore do not imply a geometric rate in m.

E A Brownian Resampling Obstruction

This appendix proves Theorem 1. We construct a fixed Brownian target and value function, then bound the bias caused by full multinomial resampling. The proof expresses the accumulated population fluctuations as a sum of Jensen gaps.

E.1 The fixed target and value function

Let P_t be the standard Brownian semigroup on $\mathbb{R}$, and fix $c \in (0,1)$. Set

$$\begin{aligned} \phi_c(x) &= 1 + c\sin x, \qquad \eta_c(dx) = \phi_c(x)\mathcal{N}(0,1)(dx), \\ u_t(x) &= e^{-(1-t)/2}\sin x, \quad P_{t-s}u_t = u_s \quad (0 \le s \le t \le 1). \end{aligned} \tag{84}$$

The target is normalized because $\mathbb{E}\sin B_1 = 0$ for Brownian motion started at zero. We use $\phi = \phi_c$ and $\eta = \eta_c$ in Theorem 1. Define constants depending only on c by

$$\begin{aligned} R_c &= \frac{1+c}{1-c}, \qquad q_c = R_c^2, \qquad b_c = \frac{1}{2}q_c(q_c^2 - 1), \\ a_c &= \frac{c(1-c)e^{-1}\sinh(1/8)}{8(1+c)^2}, \quad \rho_c = \min\left\{\frac{1}{8}, \frac{a_c}{4b_c}\right\}. \end{aligned} \tag{85}$$

Choose the fixed schedule

$$\alpha(t) = \frac{(t - 1 + \rho_c)_+}{\rho_c}, \qquad H(t,x) = 1 + c\alpha(t)u_t(x), \qquad 0 \le t \le 1. \tag{86}$$

It is Lipschitz in time and smooth in space, satisfies $1 - c \le H \le 1 + c$, and has $H(0,\cdot) = 1$ and $H(1,\cdot) = \phi_c$. The choice of c fixes the entire schedule before n and N vary.

Write $h = 1/n$, $t_k = k/n$, $H_k = H(t_k,\cdot)$, $u_k = u_{t_k}$, and $\alpha_k = \alpha(t_k)$. By (84),

$$\frac{P_h H_{k+1}(x)}{H_k(x)} = \frac{1 + c\alpha_{k+1}u_k(x)}{1 + c\alpha_k u_k(x)}, \qquad 0 \le \alpha_{k+1} - \alpha_k \le \frac{1}{\rho_c n}. \tag{87}$$

If $u_k(x) \ge 0$, the displayed ratio lies between one and $1 + c(\alpha_{k+1} - \alpha_k)/(1-c)$. If $u_k(x) < 0$, its reciprocal obeys the same bound. Thus

$$\max\left\{\frac{P_h H_{k+1}(x)}{H_k(x)}, \frac{H_k(x)}{P_h H_{k+1}(x)}\right\} \le 1 + \frac{c}{(1-c)\rho_c n}. \tag{88}$$

The exact value is $H_k^\star = 1 + cu_k$. Since both H_k and $H_k^\star$ lie in $[1-c, 1+c]$, either ratio between them is at most R_c.

E.2 Pair representation and the exact terminal target

This representation records the two positions needed for each incremental weight. Let $E_0 = \{\star\}$, $E_p = \mathbb{R}^2$ for $1 \le p \le n+1$, and write $z_p = (x_p, x_{p-1})$ for $p \le n$. Define

$$\begin{aligned} M_1(\star, d(x_1, x_0')) &= P_h(0, dx_1)\delta_0(dx_0'), \\ M_p((x,y), d(z,x')) &= P_h(x, dz)\delta_x(dx'), \quad 2 \le p \le n, \\ M_{n+1} &= \mathrm{Id}, \qquad G_0 = 1, \\ G_k(x_k, x_{k-1}) &= \frac{H_k(x_k)}{H_{k-1}(x_{k-1})}, \quad 1 \le k \le n. \end{aligned} \tag{89}$$

Every physical mutation is a Brownian increment. The terminal identity kernel applies the last edge weight without another mutation. Each edge weight and its reciprocal are bounded by R_c.

Set $Q_p f = G_{p-1}M_p f$ and $Q_{p,q} = Q_{p+1}\cdots Q_q$, with $Q_{q,q} = \mathrm{Id}$. Their normalized maps are

$$\Phi_p(\mu)(f) = \frac{\mu(Q_p f)}{\mu(Q_p 1)}, \qquad \Phi_{p,q}(\mu)(f) = \frac{\mu(Q_{p,q} f)}{\mu(Q_{p,q} 1)}. \tag{90}$$

Start from $\eta_0^N = \delta_\star$. Given the particle history $\mathcal{F}_{p-1}$, sample $Z_p^1, \ldots, Z_p^N$ independently from $\mu_p^N = \Phi_p(\eta_{p-1}^N)$ and let $\eta_p^N = N^{-1}\sum_i \delta_{Z_p^i}$. For $2 \le p \le n$, this means multinomial resampling with G_{p-1} followed by Brownian mutation; $p = n+1$ is terminal resampling with G_n. Equivalently, each of the n Brownian mutations is followed by its edge-weight resampling. Keeping the same ancestors and increments couples this pair system with that implementation exactly.

Along each compatible trajectory, the weights telescope:

$$\prod_{k=1}^{n} G_k(x_k, x_{k-1}) = \frac{H_n(x_n)}{H_0(0)} = \phi_c(x_n). \tag{91}$$

Consequently, the exact terminal pair law is

$$\eta_{n+1}(dx_n, dx_{n-1}) = \phi_c(x_n)\mathcal{N}(0, t_{n-1})(dx_{n-1})P_h(x_{n-1}, dx_n), \tag{92}$$

whose first-coordinate marginal is η_c. With $\mathrm{pr}_1(x, y) = x$, the actual output law is

$$\nu_{n,N}^{\mathrm{SMC}} = \mathbb{E}[\eta_{n+1}^N \circ \mathrm{pr}_1^{-1}]. \tag{93}$$

Exchangeability makes this the law of any single terminal particle.

The remaining weights from stage p onward telescope to $\phi_c(x_n)/H_{p-1}(x_{p-1})$. Future Brownian increments depend only on x_p, so, for $1 \le p \le n$,

$$Q_{p,n+1}(f \circ \mathrm{pr}_1)(x_p, x_{p-1}) = \frac{P_{1-t_p}(\phi_c f)(x_p)}{H_{p-1}(x_{p-1})}. \tag{94}$$

In particular,

$$Q_{p,n+1}1(x_p, x_{p-1}) = \frac{1 + cu_p(x_p)}{H_{p-1}(x_{p-1})} \in [R_c^{-1}, R_c]. \tag{95}$$

E.3 A one-step bound for normalized empirical sampling

The following estimate controls both the upper bound and the possible cancellation during the final part of the schedule. Write $\mathrm{osc}(f) = \sup f - \inf f$.

Lemma 31 (A local Feynman–Kac bias bound)**.** *Consider positive operators $Q_1, \dots, Q_\ell$, the normalized maps in (90), and conditionally independent empirical sampling η_p^N from $\mu_p^N = \Phi_p(\eta_{p-1}^N)$, with deterministic initial law $\eta_0^N = \eta_0$. Put*

$$q_{p,\ell} = \sup_{x,y} \frac{Q_{p,\ell}1(x)}{Q_{p,\ell}1(y)}, \qquad P_{p,\ell}f = \frac{Q_{p,\ell}f}{Q_{p,\ell}1}, \qquad \beta(P) = \sup_{\mathrm{osc}(f)\le 1} \mathrm{osc}(Pf).$$

If every $q_{p,\ell} < \infty$, then for $\mathrm{osc}(f) \le 1$,

$$\left|\mathbb{E}\left[\Phi_{p,\ell}(\eta_p^N)(f) - \Phi_{p,\ell}(\mu_p^N)(f) \mid \mathcal{F}_{p-1}\right]\right| \le \frac{q_{p,\ell}(q_{p,\ell}^2 - 1)}{2N}\beta(P_{p,\ell}). \tag{96}$$

Writing $\eta_\ell = \Phi_{0,\ell}(\eta_0)$ gives

$$\|\mathbb{E}\eta_\ell^N - \eta_\ell\|_{\mathrm{TV}} \le \frac{1}{2N}\sum_{p=1}^{\ell} q_{p,\ell}(q_{p,\ell}^2 - 1)\beta(P_{p,\ell}). \tag{97}$$

Proof. Condition on $\mathcal{F}_{p-1}$ and abbreviate $\mu = \eta_p^N$, $\nu = \mu_p^N$, $w = Q_{p,\ell}1$, $g = w/\nu(w)$, $P = P_{p,\ell}$, and $q = q_{p,\ell}$. Let $a = \Phi_{p,\ell}(\nu)(f)$ and $D = g(Pf - a)$. Then $\nu(g) = 1$, $\nu(D) = 0$, and direct expansion gives

$$\Phi_{p,\ell}(\mu)(f) - \Phi_{p,\ell}(\nu)(f) = (\mu - \nu)(D) - \frac{(\mu - \nu)(g)(\mu - \nu)(D)}{\mu(g)}. \tag{98}$$

The first term has conditional mean zero. The range of w gives $q^{-1} \le g \le q$, $1/\mu(g) \le q$, and $\mathrm{osc}(g) \le q - q^{-1}$. For any bounded v, w_0, conditional independence, Cauchy–Schwarz, and $\mathrm{Var}_\nu(v) \le \mathrm{osc}(v)^2/4$ imply

$$\mathbb{E}\left[|(\mu - \nu)(v)(\mu - \nu)(w_0)| \mid \mathcal{F}_{p-1}\right] \le \frac{\mathrm{osc}(v)\,\mathrm{osc}(w_0)}{4N}.$$

Thus the absolute conditional mean in (98) is at most $(q^2 - 1)\,\mathrm{osc}(D)/(4N)$. Since a is an average of values of Pf, $|Pf - a| \le \mathrm{osc}(Pf) \le \beta(P)$. It follows that $\mathrm{osc}(D) \le 2q\beta(P)$, proving (96).

The semigroup identity $\Phi_{p,\ell} \circ \Phi_p = \Phi_{p-1,\ell}$ gives the pathwise telescoping formula

$$\eta_\ell^N(f) - \eta_\ell(f) = \sum_{p=1}^{\ell}\left[\Phi_{p,\ell}(\eta_p^N)(f) - \Phi_{p,\ell}(\mu_p^N)(f)\right]. \tag{99}$$

Taking expectations, applying (96), and taking the supremum over functions of oscillation at most one proves (97). □

E.4 Neutral fluctuations and their Jensen gaps

Define

$$p_0 = \lfloor (1-\rho_c)n \rfloor + 1, \qquad f_c(x) = \frac{\sin x}{1 + c\sin x}, \qquad \widetilde{f_c} = \frac{1-c^2}{2} f_c. \tag{100}$$

Here $\eta_c(f_c) = 0$ and $\mathrm{osc}(f_c) = 2/(1-c^2)$, so $\mathrm{osc}(\widetilde{f_c}) = 1$. For $p \le p_0$, $H_{p-1} = 1$. The first coordinate therefore follows equal-weight resampling and Brownian mutation throughout this block. Also, (94) gives the scalar identity

$$\Phi_{p,n+1}(\mu)(f_c \circ \mathrm{pr}_1) = F_c\big(\mu(u_p \circ \mathrm{pr}_1)\big), \qquad F_c(z) = \frac{z}{1+cz}, \quad p \le p_0. \tag{101}$$

Indeed, the numerator is $\mu(u_p \circ \mathrm{pr}_1)$ and the denominator is $1 + c\mu(u_p \circ \mathrm{pr}_1)$. On $[-1, 1]$,

$$-F_c''(z) = \frac{2c}{(1+cz)^3} \ge \frac{2c}{(1+c)^3}. \tag{102}$$

Lemma 32 (Variance in the neutral block)**.** *For $1 \le p \le p_0$, write $X_p^i = \mathrm{pr}_1(Z_p^i)$ and set*

$$s_p = e^{-1}\sinh(t_p), \quad V_p = \mathbb{E}\left[\eta_p^N(u_p^2 \circ \mathrm{pr}_1) - \eta_p^N(u_p \circ \mathrm{pr}_1)^2\right], \quad r = 1 - \frac{1}{N},$$

with $s_0 = V_0 = 0$. Then

$$\begin{aligned} \mathbb{E}\,\mathrm{Var}_{\mu_p^N}(u_p \circ \mathrm{pr}_1) &= V_{p-1} + s_p - s_{p-1}, \\ V_p = r(V_{p-1} + s_p - s_{p-1}) &= \sum_{j=1}^{p} r^{p-j+1}(s_j - s_{j-1}) \ge r^p s_p. \end{aligned} \tag{103}$$

Proof. Conditional empirical sampling is unbiased. Since all weights used in this block are one, induction yields $\mathbb{E}\eta_p^N(g \circ \mathrm{pr}_1) = \mathbb{E}g(B_{t_p})$. In particular,

$$\mathbb{E}\eta_p^N(u_p^2 \circ \mathrm{pr}_1) = e^{-(1-t_p)}\frac{1 - e^{-2t_p}}{2} = s_p.$$

By $P_h u_p = u_{p-1}$, the conditional mean of $u_p(X_p^i)$ is $\eta_{p-1}^N(u_{p-1} \circ \mathrm{pr}_1)$ (zero at $p = 1$). Therefore the expected conditional variance is $s_p - (s_{p-1} - V_{p-1})$, proving the first equality in (103). For N independent samples from a law ν, the expected empirical variance is $(1 - 1/N)\,\mathrm{Var}_\nu$. Applying this conditionally gives the recursion for V_p. Its iteration gives the displayed sum. Finally, $s_j - s_{j-1} \ge 0$ and $r^{p-j+1} \ge r^p$, so the sum is at least $r^p s_p$. □

For $1 \le p \le n+1$, define the signed local contributions

$$L_p = \mathbb{E}\left[\Phi_{p,n+1}(\mu_p^N)(\widetilde{f_c} \circ \mathrm{pr}_1) - \Phi_{p,n+1}(\eta_p^N)(\widetilde{f_c} \circ \mathrm{pr}_1)\right]. \tag{104}$$

Equation (99) implies

$$\eta_c(\widetilde{f_c}) - \nu_{n,N}^{\mathrm{SMC}}(\widetilde{f_c}) = \sum_{p=1}^{n+1} L_p. \tag{105}$$

For $p \le p_0$, the function $F_c(z) + cz^2/(1+c)^3$ is concave by (102). Conditional Jensen's inequality and the variance of an empirical mean give

$$L_p \ge \frac{(1-c^2)c}{2(1+c)^3 N}\mathbb{E}\,\mathrm{Var}_{\mu_p^N}(u_p \circ \mathrm{pr}_1) \ge 0. \tag{106}$$

Connection to shared ancestry. The same calculation makes the role of particle dependence explicit. Let $U_p = N^{-1}\sum_i u_p(X_p^i)$ and $U_0 = 0$ in the neutral block. Brownian harmonicity gives

$\mathbb{E}[U_p \mid \mathcal{F}_{p-1}] = U_{p-1}$, hence

$$\begin{aligned}\mathbb{E}U_q^2 &= \frac{1}{N}\sum_{p=1}^{q}\mathbb{E}\,\mathrm{Var}_{\mu_p^N}(u_p \circ \mathrm{pr}_1)\\ &= \frac{s_q}{N} + \frac{N-1}{N}\,\mathrm{Cov}(u_q(X_q^1), u_q(X_q^2)), \qquad q \le p_0.\end{aligned} \tag{107}$$

The first equality sums the conditional variances of the martingale increments. The second follows by expanding the square and using exchangeability and $\mathbb{E}U_q = 0$. Equal-weight ancestor draws merge with probability $1/N$ per round, allowing distinct particles to share Brownian increments. Their covariance contributes to the variance in (107).

More precisely, (101) and the martingale property give $L_p = (1-c^2)\mathbb{E}[F_c(U_{p-1}) - F_c(U_p)]/2$ for $p \le p_0$. Using $F_c(z) = z - cz^2/(1+cz)$ yields the exact identity

$$\sum_{p=1}^{q} L_p = \frac{c(1-c^2)}{2}\mathbb{E}\left[\frac{U_q^2}{1+cU_q}\right], \qquad q \le p_0. \tag{108}$$

Since $|U_q| \le 1$, this sum is between $c(1-c^2)\mathbb{E}U_q^2/[2(1+c)]$ and $c(1-c^2)\mathbb{E}U_q^2/[2(1-c)]$. Thus the accumulated Jensen gap measures the same empirical-mean fluctuations as the covariance formula. Individual marginals in the neutral block remain Brownian; these gaps measure their effect after the future normalized weighting.

E.5 Proof of the terminal TV bounds

The neutral block supplies a positive contribution at order n/N. The fixed width ρ_c makes the remaining stages too few to cancel it.

Proof of Theorem 1. The construction (86), the target identity (91), and (88) establish the value-function and Bellman-ratio claims, with $C = c/[(1-c)\rho_c]$.

For the lower bound, define

$$J_{n,c} = \left\{\left\lceil \frac{n}{8}\right\rceil + 1, \dots, p_0\right\}.$$

Since $\rho_c \le 1/8$, for $n \ge 4$,

$$|J_{n,c}| \ge \left\lfloor \frac{7n}{8}\right\rfloor - \left\lceil \frac{n}{8}\right\rceil + 1 \ge \frac{3n}{4} - 1 \ge \frac{n}{2}.$$

For $n = 2, 3$, the same inequality follows by direct substitution. If $p \in J_{n,c}$, then $t_{p-1} \ge 1/8$ and $p - 1 \le n \le N$. Also $(1-1/N)^N \ge 1/4$ for $N \ge 2$. Thus Lemma 32 gives

$$\mathbb{E}\,\mathrm{Var}_{\mu_p^N}(u_p \circ \mathrm{pr}_1) \ge V_{p-1} \ge r^{p-1}s_{p-1} \ge \frac{1}{4}e^{-1}\sinh(1/8).$$

Substitution into (106) gives $L_p \ge a_c/N$ for every $p \in J_{n,c}$.

To bound the remaining contributions, note from (95) that the ratio $q_{p,n+1}$ in Lemma 31 is at most $R_c^2 = q_c$ for $1 \le p \le n$. Its normalized operator is a Markov kernel, so $\beta(P_{p,n+1}) \le 1$. The local bound therefore gives

$$|L_p| \le \frac{b_c}{N} \quad (1 \le p \le n), \qquad L_{n+1} = 0. \tag{109}$$

The last equality holds because the future map is the identity at $p = n+1$ and conditional empirical sampling is unbiased. There are at most $\rho_c n$ stages with $p_0 < p \le n$, since

$$n - p_0 = n - \lfloor (1-\rho_c)n\rfloor - 1 \le \rho_c n.$$

All terms with $p \le p_0$ are nonnegative by (106). Equations (105) and (109) now yield

$$\begin{aligned}\eta_c(\widetilde{f_c}) - \nu_{n,N}^{\mathrm{SMC}}(\widetilde{f_c}) &\ge \frac{a_c}{N}|J_{n,c}| - \frac{b_c\rho_c n}{N}\\ &\ge \left(\frac{a_c}{2} - b_c\rho_c\right)\frac{n}{N} \ge \frac{a_c}{4}\frac{n}{N}.\end{aligned} \tag{110}$$

Since $\mathrm{osc}(\widetilde{f_c}) = 1$, this proves the TV lower bound with $c_- = a_c/4 > 0$.

For the upper bound, apply (97) to the pair system with $\ell = n + 1$. Each of the first n terms is at most b_c/N, and the last has $q_{n+1,n+1} = 1$ and contributes zero. Projection onto the first coordinate cannot increase TV, so

$$\|\nu_{n,N}^{\mathrm{SMC}} - \eta_c\|_{\mathrm{TV}} \le \|\mathbb{E}\eta_{n+1}^N - \eta_{n+1}\|_{\mathrm{TV}} \le b_c \frac{n}{N}. \tag{111}$$

Taking $c_+ = b_c$ completes the proof. All constants and the schedule depend only on the fixed reward amplitude c. □

F Terminal Importance Resampling

This appendix proves a TV bound for one output selected by terminal importance resampling and applies it to complete Stein proposals.

F.1 Terminal SIR controls an unconditional output law

One final weight-proportional selection produces a probability law whose TV error can be bounded directly. For a generic pair $\pi = wq$, suppose $0 < w < \infty$ q-almost surely and $\mathbb{E}_q w = 1$. Draw $\Omega_1, \dots, \Omega_N$ independently from q and write $W_i = w(\Omega_i)$. The self-normalized importance measure is

$$\widehat{\pi}_N = \frac{\sum_{i=1}^N W_i \delta_{\Omega_i}}{\sum_{i=1}^N W_i}. \tag{112}$$

Terminal sampling–importance–resampling (SIR) additionally selects an index I with conditional probability $W_i / \sum_j W_j$ and returns Ω_I. Its unconditional output law is

$$\overline{\pi}_N = \mathbb{E}[\widehat{\pi}_N]. \tag{113}$$

This is one terminal selection from independent proposals, with no intermediate population resampling. For every bounded f, $\mathbb{E}\widehat{\pi}_N(f) = \overline{\pi}_N(f)$, and hence

$$\sup_{\mathrm{osc}(f)\le 1} |\mathbb{E}\widehat{\pi}_N(f) - \pi(f)| = \|\overline{\pi}_N - \pi\|_{\mathrm{TV}}. \tag{114}$$

The supremum is outside the expectation. In contrast, for a nonatomic target, the random finitely supported measure $\widehat{\pi}_N$ has TV distance one from the target almost surely. Estimator MSE and the unconditional sampling bias in (114) are therefore different quantities.

Theorem 33 (SIR TV and SNIS mean bias)**.** *Suppose $D = \mathbb{E}_q(w-1)^2 < \infty$. For every integer $N \ge 1$,*

$$\|\overline{\pi}_N - \pi\|_{\mathrm{TV}} \le \mathbb{E}_q \frac{w(w-1)_+}{w + N - 1} \tag{115}$$

$$\le \min\left\{s, \frac{D_+ + s}{N}\right\} \le \min\left\{s, \frac{D + \frac{1}{2}\sqrt{D}}{N}\right\}, \tag{116}$$

where $s = \mathbb{E}_q(w-1)_+$ and $D_+ = \mathbb{E}_q(w-1)_+^2$. The same bound controls the supremum of the SNIS mean bias in (114). No uniform positive lower bound or negative moment of w is required.

Proof. Exchangeability gives, for nonnegative bounded f,

$$\overline{\pi}_N(f) = N\mathbb{E}\left[\frac{W_1 f(\Omega_1)}{W_1 + \cdots + W_N}\right].$$

Conditionally on $\Omega_1 = \omega$, the remaining weights are independent copies of w under q. Therefore the density relative to π is

$$g_N(w(\omega)) := \frac{\mathrm{d}\overline{\pi}_N}{\mathrm{d}\pi}(\omega) = N\mathbb{E}\left[\frac{1}{w(\omega) + \sum_{i=2}^N W_i}\right]. \tag{117}$$

For every $a > 0$, this expectation is finite and bounded by N/a. Jensen's inequality gives

$$g_N(a) \ge \frac{N}{a + N - 1}.$$

Both $g_N\pi$ and π have mass one, so

$$\|\overline{\pi}_N - \pi\|_{\mathrm{TV}} = \mathbb{E}_\pi(1 - g_N(w))_+ \le \mathbb{E}_\pi \frac{(w-1)_+}{w + N - 1} = \mathbb{E}_q \frac{w(w-1)_+}{w + N - 1}.$$

The denominator is at least w, and is at least N on $\{w > 1\}$. Thus the last expression is at most s and at most $N^{-1}\mathbb{E}_q[w(w-1)_+] = (D_+ + s)/N$. Finally, $D_+ \le D$ and $2s = \mathbb{E}_q|w-1| \le \sqrt{D}$. □

Corollary 34 (Terminal importance resampling)**.** *Let $\bar{\nu}_N$ be the physical path law obtained by one weight-proportional selection from N independent complete Stein proposals. Under the finite-moment conditions of Theorem 20 at order two, write $D_\star = \Delta_2^2$.*

$$\|\bar{\nu}_N - \Pi_n\|_{\mathrm{TV}} \le \min\left\{1, \frac{D_\star + \frac{1}{2}\sqrt{D_\star}}{N}\right\}. \tag{118}$$

The same bound holds for the endpoint, at cost KnN value queries.

Proof. Set $q = \widetilde{Q}$, $\pi = \widetilde{\Pi}$, and $w = R/\mathbb{E}_{\widetilde{Q}}R$. Normalizing the weights does not change the selection probabilities. Theorem 20 gives $D \le D_\star$. If $\overline{\widetilde{\Pi}}_N$ denotes the resulting augmented output law, Theorem 33 yields

$$\left\|\overline{\widetilde{\Pi}}_N - \widetilde{\Pi}\right\|_{\mathrm{TV}} \le \frac{D_\star + \frac{1}{2}\sqrt{D_\star}}{N}. \tag{119}$$

For any measurable projection T, the preimage of an event is admissible in the definition of TV; hence $\|T_\#\mu - T_\#\nu\|_{\mathrm{TV}} \le \|\mu - \nu\|_{\mathrm{TV}}$. Apply this to the physical path and then to its endpoint. Under (9), the centered moment bound gives $\Delta_2 = O(\mathcal{A}_4 + \sqrt{Th/(K-1)})$, uniformly over the grid and pool size. The same assumptions bound this expression, so $D_\star + \frac{1}{2}\sqrt{D_\star} = O(\mathcal{A}_4 + \sqrt{Th/(K-1)})$. Consequently,

$$\left\|\overline{\pi}_N^{\mathrm{end}} - \Pi_n^{\mathrm{end}}\right\|_{\mathrm{TV}} \lesssim \frac{\mathcal{A}_4 + \sqrt{Th/(K-1)}}{N}. \tag{120}$$

Generating the N complete proposals uses KnN candidate value queries and nN coefficient evaluations. □

G Local Refinement and Exact Weighting

This extension records the exact weight when local pool MH is included in a complete proposal, and verifies the exponential model's pool and path variances for general pool size. Throughout, r counts local pool updates and m counts outer MH updates.

G.1 The local proposal and its query cost

Algorithm 3 defines the local-refinement family. Its $r = 0$ case is Algorithm 1; retaining the initial pool average also makes every finite local run usable in an outer correction.

Algorithm 3: Stein Guidance with pool size K and r local MH updates

Initialize $X_0 = x_\star$ and $\log R = \log H_0(x_\star)$. For $k = 0, \dots, n-1$:

1. Draw the initial pool $U^{(0)}$ by (7) and query its values. Record $S_0 = S(U^{(0)})$ and retain $U = U^{(0)}$.
2. Repeat exactly r times: draw a fresh pool V at the *same* parent and query its values. Replace U by V with probability $\min\{1, S(V)/S(U)\}$; otherwise retain U.
3. Select one candidate from the retained pool proportionally to its value and advance to X_{k+1}. Set $\log R \leftarrow \log R + \log S_0 - \log H_k(X_k)$.

Return $(X_{0:n}, R)$. Launch paths independently for SNIS, or propose fresh complete paths in outer MH with acceptance $\min\{1, R'/R\}$.

All $r+1$ pools can be queried in one batch before the sequential MH decisions. The run uses $K(r+1)n$ candidate values; each retained child value is cached for the next parent. At $K = 2, r = 0$, the weight telescopes to the two-query statistic in (36).

The invariant pool law. At a fixed parent $X_k = x$, let $q_{k,x,K}$ be the centered candidate pool law in (17). Its pool average is $S(U)$, with expectation $G_k(x)$, and its normalized pool average is

$w(U) = S(U)/G_k(x)$. The pool target $\pi_{k,x,K} = wq_{k,x,K}$ followed by selection with probabilities proportional to $H_{k+1}(Y_i)$ gives the exact local tilted kernel K_k^H by Lemma 5.

Algorithm 3 approximates this tilted pool law with exactly r independence MH updates. Each update accepts a fresh pool V from the same parent with probability

$$\alpha(U, V) = \min\{1, S(V)/S(U)\}. \tag{121}$$

The unknown $G_k(x)$ cancels from the acceptance ratio. Candidates within a pool are correlated, whereas the $r + 1$ proposed pools are independent at the same fixed parent. Their value queries can be evaluated in one batch before the sequential acceptance decisions. The algorithm executes a prescribed number of updates and retains rejected states. Stopping at the first acceptance changes its law. Likewise, updating only an index within one fixed pool does not correct the distribution of that pool.

Write $M = M_{k,x,K}$ for this independence-MH pool kernel.

G.2 Exact outer correction after local MH

Local MH can also be retained inside the complete-path proposal. At time k, initialize $U_k^{(0)} \sim q_{k,X_k}$, apply r pool MH updates, and select the next physical state from $U_k^{(r)}$. Retain the initial pool average and use

$$R^{\mathrm{hyb}} = H_0(x_\star) \prod_{k=0}^{n-1} \frac{S(U_k^{(0)})}{H_k(X_k)}. \tag{122}$$

The initial average is essential: weighting the initial pool changes its law from q to the invariant pool target π, after which any finite number of invariant MH updates preserves the weighted target.

Proposition 35 (Proper weighting for any finite local run)**.** *Let $\widetilde{Q}^{\mathrm{hyb}}$ retain the local pool histories and the physical path. For every fixed finite $r \geq 0$ and bounded measurable physical path function F,*

$$\mathbb{E}_{\widetilde{Q}^{\mathrm{hyb}}}\left[R^{\mathrm{hyb}} F(X_{0:n})\right] = \mathbb{E}_{P_n}\left[\phi(X_n) F(X_{0:n})\right]. \tag{123}$$

Therefore $\widetilde{\Pi}^{\mathrm{hyb}} = (R^{\mathrm{hyb}}/Z_n)\widetilde{Q}^{\mathrm{hyb}}$ is a probability law whose physical path marginal is exactly Π_n. The case $r = 0, K = 2$ recovers the path weight of Stein ($K = 2$).

Proof. Fix time k and parent x, suppress these indices on q, π, M, and retain C_x for the terminal child selection kernel. The identities $S(U)q(\mathrm{d}U) = G_k(x)\pi(\mathrm{d}U)$ and $\pi M^r = \pi$ give

$$\begin{aligned}
\mathbb{E}\left[\frac{S(U^{(0)})}{H_k(x)} f(X_{k+1})\right] &= \frac{G_k(x)}{H_k(x)} \pi M^r C_x f \\
&= \frac{G_k(x)}{H_k(x)} \pi C_x f \\
&= \frac{P_k(H_{k+1} f)(x)}{H_k(x)},
\end{aligned} \tag{124}$$

where the last step uses (38). For a cylinder function $F(x_{0:n}) = \prod_{j=0}^n f_j(x_j)$, iterate this weighted transition identity from the final time back to the initial time. Every intermediate factor H_k cancels. The prefactor $H_0(x_\star)$ in (122) cancels the initial denominator, leaving $H_n(X_n) = \phi(X_n)$ and the base transition product P_n. This proves (123) for cylinder functions and hence for bounded measurable path functions by the monotone class theorem. In particular $\mathbb{E}R^{\mathrm{hyb}} = Z_n$, and the normalized tilted augmented law has the stated physical marginal.

For completeness, the standard MH acceptance rule is valid even when the augmented proposal has no convenient Lebesgue density: with respect to the product proposal measure, its accepted probability flow is

$$\frac{1}{Z_n} \min\{R^{\mathrm{hyb}}(u), R^{\mathrm{hyb}}(v)\}\, \widetilde{Q}^{\mathrm{hyb}}(\mathrm{d}u)\widetilde{Q}^{\mathrm{hyb}}(\mathrm{d}v),$$

which is symmetric in u, v. The rejection term is supported on the diagonal. Detailed balance therefore establishes the claimed invariant law, without asserting finite update stationarity. Under an envelope $R^{\mathrm{hyb}} \leq M_\star$, independent proposal acceptance with probability $R^{\mathrm{hyb}}/M_\star$ has conditional accepted law $(R^{\mathrm{hyb}}/Z_n)\widetilde{Q}^{\mathrm{hyb}}$, establishing the rejection sampling claim.

Finally, $w_k^{(0)} = S(U_k^{(0)})/G_k(X_k)$ has conditional mean one before pool k is generated, regardless of the subsequent local MH run. Any conditional pool-moment bounds valid at every reachable hybrid parent therefore give the product bounds of Theorem 18 for $M_{\mathrm{in}} = \prod_k w_k^{(0)}$. The query count is multiplied by $r + 1$ because each parent evaluates one initial pool and r proposed pools. □

Outer MH on the augmented state consequently uses the acceptance probability

$$\alpha = \min\left\{1, \frac{(R^{\mathrm{hyb}})'}{R^{\mathrm{hyb}}}\right\}.$$

On rejection it retains the current augmented state and weight. This gives the exact invariant physical marginal at any fixed grid. A finite number of outer MH updates from a generic initialization still incurs mixing error. If a known finite envelope $R^{\mathrm{hyb}} \le M_\star$ is available, independent rejection sampling with acceptance probability $R^{\mathrm{hyb}}/M_\star$ instead produces exact independent samples at that grid. If P_n is an Euler law, neither construction removes the Euler error relative to a continuous diffusion target.

Replacing $S(U_k^{(0)})$ by $S(U_k^{(r)})$ in (122) is generally invalid. Retaining the initial pool average in the correction weight does not transfer the higher-order marginal accuracy of local MH to that weight.

The conditional distribution of the initial pool average is unchanged by subsequent local updates at the same parent.

G.3 Exact variances for an exponential Brownian tilt

The following formulas describe the initial normalized pool average and its complete-path product at every pool size.

Consider the one dimensional Brownian reference transition $P_h(x, \cdot) = N(x, h)$, with $X_0 = 0, T = nh$, and terminal reward $\phi(x) = e^{\beta x}$, where $\beta \neq 0$. The exact discrete Doob value is

$$H_t(x) = \exp\left(\beta x + \frac{\beta^2(T-t)}{2}\right). \tag{125}$$

For an antithetic pool, its normalized average is independent of the parent:

$$\xi_h(Z) = e^{-\beta^2 h/2}\cosh(\beta\sqrt{h}Z), \qquad Z \sim N(0,1). \tag{126}$$

Consequently, the raw path proposal has normalized complete weight $W_h = \prod_{k=1}^n \xi_h(Z_k)$ for independent Z_k. Although H is unbounded, $\xi_h \ge e^{-\beta^2 h/2}$ and $\|\xi_h - 1\|_{L^p} = O(h)$ for every fixed $p < \infty$. Thus the direct weight assumptions for local MH hold uniformly for $0 < h \le h_0$, without assuming a bounded value function.

This example also admits an exact variance formula for every centered pool size $K \ge 2$. Writing $u = \beta^2 h$, the normalized pool average and its variance are

$$\xi_{h,K} = \frac{e^{-u/2}}{K}\sum_{i=1}^K e^{\beta\sqrt{h}Z_i}, \qquad v_K(h) := \mathrm{Var}(\xi_{h,K}) = \frac{e^u + (K-1)e^{-u/(K-1)}}{K} - 1. \tag{127}$$

By symmetry, the sign of β does not affect this formula. With exact H, the initial pool weights are independent across physical steps even when local MH is used. Consequently the auxiliary proposal and its exactly weighted target satisfy

$$\chi^2(\widetilde{\Pi}_n \| \widetilde{Q}) = (1 + v_K(h))^n - 1. \tag{128}$$

Both identities hold at every $h > 0$ and follow directly from Gaussian moment generating functions. For $K > 2$ the second identity is on the auxiliary space; projection gives a physical path upper bound.

The distribution of this normalized complete weight is exactly independent of the number of local MH updates. By (71), the stationary rejection probability of the outer MH chain is therefore also independent of that number in this model. This statement concerns outer acceptance; the physical proposal law does change with local correction.

Verification of the exact moment formulas. For the centered Gaussian construction, $\mathbb{E}Z_iZ_j = -1/(K-1)$ whenever $i \neq j$. The Gaussian moment generating function gives $\mathbb{E}\xi_{h,K} = 1$, and

expansion of the square in (127) yields

$$\mathbb{E}\xi_{h,K}^2 = \frac{1}{K^2}\left(Ke^u + K(K-1)e^{-u/(K-1)}\right).$$

This proves (127). Exactness of H implies that each incremental factor in the normalized complete weight is its initial normalized pool average. In the exponential example this average is independent of the parent state and depends only on the fresh Gaussian variables of that pool. The normalized complete weight therefore has second moment $(\mathbb{E}\xi_{h,K}^2)^n$, both for raw paths and for paths with local pool MH. The exact auxiliary weighting identity then proves (128).

G.4 Conditions behind the method comparison

Table 1 concerns correction with approximate values, unconditional output TV as numerical depth changes, and parallel candidate queries. For Stein guidance, the primary correction is proposal-initialized outer independence MH. The depth-uniform guarantee uses the fixed start, terminal agreement and uniform regularity conditions of Theorem 4; $K = 2$ is included. The theorem separates finite-moment Bellman error from candidate-pool noise. Each output uses $Kn(m+1)$ value queries. Our Gaussian theory does not cover discrete state diffusion.

At finite candidate count, SVDD gives approximate local sampling even with exact values and has no outer exact correction. The SMC obstruction in Theorem 1 uses Brownian proposals and multinomial resampling at every step. Its single-output TV error is $\Theta(n/N)$ for $2 \le n \le N$, with a fixed target and local Bellman error $O(1/n)$.

Discussion on the DOIT bound. For the Monte Carlo estimator in Zhu et al. (2026, Lemma 5.2 and Theorem 5.4), the local mean square error contains $\sigma_{t_\ell}^{-2}$, and its accumulated contribution depends on

$$\kappa_\sigma = \frac{T}{L}\sum_{\ell=1}^{L}\sigma_{t_\ell}^{-2}. \tag{129}$$

Symbols retain that paper's conventions. Under its stated $O(\log(T/t_1))$ schedule scaling, $t_1 \asymp T/L$ gives $O(\log L)$. This is an upper bound dependence, not observed or necessary variance growth. Increasing M or keeping a positive time cutoff changes the comparison. Algorithms 1 and 2 distinguish full rollouts from the practical surrogate with correction only for $1 < \ell \le \ell^\star$; the latter incurs additional approximation error.

H Experimental details and selected results

This section collects the essential experimental settings, uncertainty estimates, the learned-policy correction-length ablation, and Gaussian diagnostics.

H.1 Synthetic centering protocol

For scalar Brownian motion with $X_0 = 0$, horizon one, and $t_j = j/n$, the terminal potential $g(x) = e^{1.5x}$ gives target $\mathcal{N}(1.5, 1)$. The step size is $h = 1/n$. We compare the exact value $H^*(t,x) = \exp\{1.5x + 1.5^2(1-t)/2\}$ with the perturbed value $H_\delta(t,x) = H^*(t,x)e^{\delta(1-t)\sin x}$, with $\delta = 0.5$. Each step draws K children from the Brownian transition, with opposite innovations for centered $K = 2$ and independent innovations for IID pools. Selection probabilities are proportional to the child values. Paths remain independent, with no path resampling or MH. Writing X_{j+1} for the selected child, each path retains weight

$$R = e^{1.5X_n}\prod_{j=0}^{n-1}\frac{K^{-1}\sum_{k=1}^{K}H_\delta(t_{j+1}, Y_{j,k})}{H_\delta(t_{j+1}, X_{j+1})}.$$

Each independent system averages 16 path weights to obtain $\widehat{Z}$. We report adjusted NESS,

$$\mathrm{NESS}_{\mathrm{adj}} = \left(1 + 16\frac{s^2(\widehat{Z})}{\overline{\widehat{Z}}^2}\right)^{-1},$$

where the sample variance s^2 and mean $\overline{\widehat{Z}}$ are computed across independent systems. The $K = 2$ study uses 4096 systems at each $n = 12, 24, 48, 96, 192$. The larger IID pools,

$K = 4, 8, 16, 32, 64, 128, 256, 512, 1024$, use 1024 systems at $n = 24,192$. The 95% intervals in Figure 2 use 600 bootstrap draws of complete systems.

H.2 RL settings, reference, and uncertainty

The critic-tilted target favors higher-scoring actions while penalizing KL divergence from the original policy (Kim et al., 2025). The penalty limits policy shifts that could exploit critic errors. MPO (Abdolmaleki et al., 2018) uses the same form of target to train an actor. The eight fixed Hopper observations have source indices $0, 2, 3, 4, 6, 12, 14, 15$. Every method uses 15 stochastic DDIM steps with $\eta = 0.6$ and the frozen `behavior_ckpt600.pth` policy and `critic_ckpt100.pth` critic. The score is $Q = \min(Q_1, Q_2)$, evaluated before environment action clipping. The main comparison uses $K = 2, 8$ and $L = m + 1 = 32$ complete proposals per output for Stein: one initial draw followed by 31 whole-path MH updates. FK and DAS use $N = 32$ particles; both DOIT variants and SVDD use $M = 32$ candidates. Native DOIT scales lookahead scores by their sample standard deviation; the fixed-temperature variant uses the declared β_s. SVDD uses the posterior-mean approximation. FK resamples at every nonterminal step; DAS resamples when ESS falls below half its particle count, both multinomially. TFG uses eight recurrences and one clean-prediction update; TTS-local uses four recurrences and four clean-prediction updates.

Reference. A separate pilot of 4×65536 actions per state fixes the highest-decile threshold c_s and the three inverse temperatures. After calibration, four independent blocks of 262144 base-policy actions per state are reserved for validation and reference consistency checks. The remaining twelve blocks, containing 3,145,728 actions per state, define the evaluation reference through weights $w_i \propto e^{\beta_s Q(s,a_i)}$. The weighted CDF uses all these actions directly. Across states and target strengths, reference ESS ranges from 49674 to 1851193, and the largest KS difference between the validation and evaluation references is 0.01584.

Proposal values and selection. Let $D_j(x)$ denote the clean prediction at node j. For $j < 15$, our sampler uses $H_j^{(c)}(x) = \exp\{c\beta_s[Q(s, D_j(x)) - c_s]\}$; at the endpoint it uses $H_{15}(a) = \exp\{\beta_s[Q(s,a) - c_s]\}$. Thus c changes the intermediate guidance while preserving the target. With $S_j = K^{-1}\sum_k H_{j+1}(Y_{j,k})$, the complete-path weight satisfies $\log R = \beta_s Q(s, X_{15}) + \sum_{j=0}^{14}[\log S_j - \log H_{j+1}(X_{j+1})]$. Whole-path MH accepts with probability $\min\{1, R_{\text{new}}/R_{\text{old}}\}$. For each $K = 2, 8, 32$, four validation blocks of 16 chains select $c \in \{0.25, 0.5, 1\}$ by mean KS at $L = 32$. In order of increasing target strength, the selected values are $(0.5, 1, 1)$ for $K = 2$ and $(0.5, 0.5, 0.5)$ for $K = 8$. The follow-up $K = 32$ ablation selects $(1, 0.25, 1)$ using the same rule and fresh validation draws. Tuned baseline settings use four validation blocks of 64 outputs per state. All selections precede the fresh evaluation draws: 16 independent blocks of 64 outputs, totaling 1024 outputs per state and method.

Distribution curves. Figure 3 uses the same main-table settings and fresh evaluation outputs. For each state and target, let μ_s, σ_s be the weighted mean and standard deviation of the numerical target reference. Every method uses $z = (Q - \mu_s)/\sigma_s$, and the displayed CDF equally mixes the eight states. The horizontal axis is linear near zero and logarithmic in the tails. Black denotes the independent numerical SIS target reference.

Statistical comparisons. For each of 10000 bootstrap draws, method blocks are resampled independently and the twelve reference blocks are resampled jointly across methods. The states remain fixed. For method m, the centered CDF-process norm is

$$R_m^* = \frac{1}{8}\sum_s \left\| (F_{m,s}^* - \widehat{F}_{m,s}) - (F_{\text{ref},s}^* - \widehat{F}_{\text{ref},s}) \right\|_\infty .$$

Its $1 - 0.05/27$ bootstrap quantile gives a simultaneous radius for each of the 27 main method–target rows. Subtracting the resulting intervals gives the 42 comparisons of $K = 2, 8$ against seven baselines at three targets. The calculation uses upper CDF-grid envelopes at all method values and 4097 weighted reference quantiles. Table 4 reports the corresponding pointwise intervals.

Method	30%	60%	80%
Stein $K2/L32$	0.0306 [0.0000, 0.0656]	0.0697 [0.0362, 0.1032]	0.1128 [0.0778, 0.1477]
Stein $K8/L32$	0.0426 [0.0086, 0.0766]	0.0942 [0.0574, 0.1310]	0.1511 [0.1152, 0.1870]
DOIT native	0.2418 [0.2073, 0.2763]	0.4778 [0.4419, 0.5138]	0.6179 [0.5828, 0.6531]
DOIT fixed β	0.2780 [0.2436, 0.3123]	0.4774 [0.4414, 0.5134]	0.5903 [0.5545, 0.6262]
SVDD $M32$	0.1218 [0.0853, 0.1583]	0.2276 [0.1929, 0.2623]	0.2979 [0.2604, 0.3354]
FK $N32$	0.0504 [0.0138, 0.0869]	0.1093 [0.0746, 0.1439]	0.1816 [0.1439, 0.2193]
DAS $N32$	0.0307 [0.0000, 0.0662]	0.0769 [0.0401, 0.1136]	0.1355 [0.0974, 0.1737]
TFG	0.1788 [0.1437, 0.2139]	0.2808 [0.2456, 0.3160]	0.3586 [0.3199, 0.3972]
TTS-local	0.1781 [0.1404, 0.2158]	0.2767 [0.2416, 0.3118]	0.3752 [0.3381, 0.4123]

Table 4: Mean KS with pointwise 95% centered-CDF bootstrap bands.

H.3 Whole-path correction with learned values

The configurations in Table 5 use the same proposals at each reported correction length. Thus the L comparisons isolate whole-path MH. The intermediate strength is selected separately for each K and target.

Target	K	c	$L = 1$	$L = 8$	$L = 32$	Accept.	Unmoved
30%	2	0.5	0.2174	0.0463	0.0306	51.73%	0.20%
30%	8	0.5	0.2136	0.0525	0.0426	46.43%	0.26%
30%	32	1	0.1180	0.0544	0.0469	35.85%	0.50%
60%	2	1.0	0.2804	0.1032	0.0697	34.74%	0.57%
60%	8	0.5	0.3617	0.1376	0.0942	29.79%	0.66%
60%	32	0.25	0.4553	0.1602	0.0907	30.24%	0.56%
80%	2	1.0	0.4041	0.1804	0.1128	29.08%	0.63%
80%	8	0.5	0.4713	0.2333	0.1511	24.37%	0.92%
80%	32	1	0.2933	0.1958	0.1436	19.25%	1.40%

Table 5: Fresh-sample ablations. The intermediate strength c is selected separately for each K and target by validation KS at $L = 32$. Acceptance and unmoved fractions refer to the 31 MH decisions.

H.4 Gaussian guidance and proper-weight diagnostics

Figure 4 separates estimator accuracy, raw endpoint moments, and weighted estimation. These runs use local pool refinement and self-normalized importance sampling. Here r counts the local pool MH updates at each transition. For Brownian motion started at zero with horizon one, $\tau^2 = 2$, and $y = \mathbf{1}/2$, the terminal potential and approximate values are

$$\phi(x) = e^{-\|x-y\|^2/(2\tau^2)}, \qquad H_t^{\alpha}(x) = \left(\frac{\tau^2}{v_\alpha(t)}\right)^{d/2} e^{-\|x-y\|^2/(2v_\alpha(t))}, \quad v_\alpha(t) = \tau^2+\alpha(1-t). \tag{130}$$

The value is exact at $\alpha = 1$; changing α preserves the terminal potential. The target is $\mathcal{N}(\mathbf{1}/6, (2/3)I_d)$. The observables $f(x) = d^{-1}\sum_j x_j$ and $q(x) = d^{-1}\|x - \mathbf{1}/6\|^2$ have target means $1/6$ and $2/3$. Each complete-path setting uses 32 independent groups of 128 paths. Mean intervals use t_{31} across group summaries; RMSE intervals use 10000 bootstrap draws of whole groups. In the equal-query experiment, path methods use $n = 16$ and $K(r + 1) = 16$, giving 32768 candidate-value evaluations per group. The SMC comparator uses 2048 particles and 16 mutations at the same query cost, with multinomial resampling when ESS falls below half the particle count. For local refinement, all path weights retain the pool averages recorded before the local MH updates.

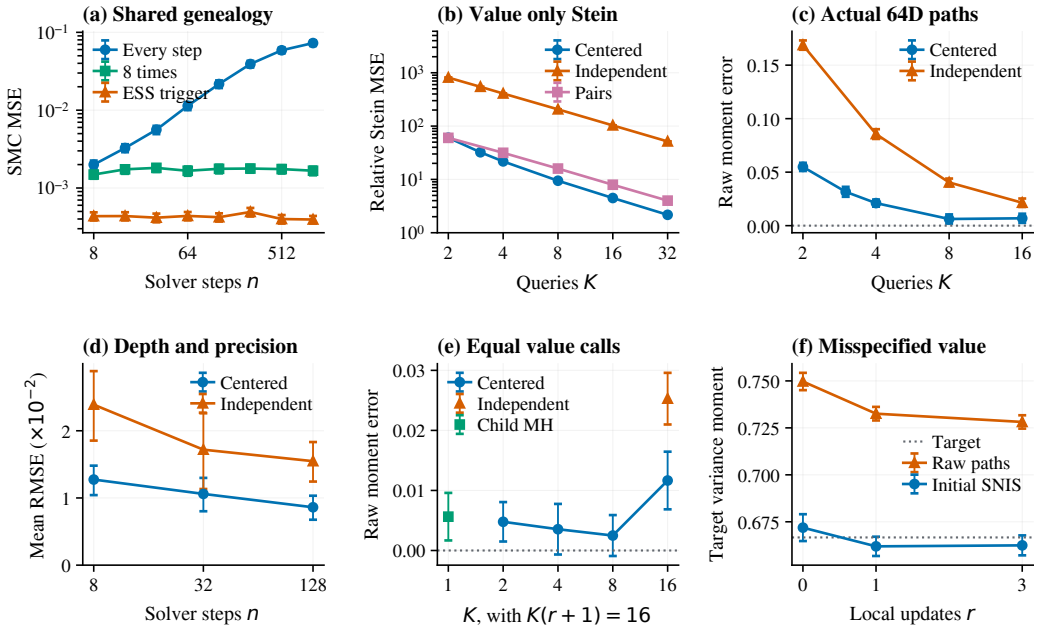

Figure 4: Executed experiments. (a) Neutral SMC, 500 populations of $N = 256$. (b) Relative MSE of $\widehat{s}$, $d = 64$, $h = 1/64$, 8192 pools per point. (c) Raw variance moment error, $d = 64$, $n = 32$. (d) SNIS mean RMSE, $K = 4$. (e) Actual equal cost runs, $n = 16$, $K(r + 1) = 16$. (f) Misspecified values, $\alpha = 4$, $K = 4$, $n = 32$. Panels (c–f) use 32 groups of 128 complete paths. Bars are pointwise 95% intervals; (d) uses a population bootstrap. The dotted lines mark exact targets.