THEORETICAL GUARANTEES FOR SMC-GUIDED DIFFUSION SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Post-hoc conditioning of pretrained diffusion models can be addressed using Sequential Monte Carlo (SMC) methods. By evolving an interacting particle system, SMC-guided diffusion samplers combine unconditional reverse-diffusion dynamics with sequential reweighting to approximate conditional distributions. Nevertheless, even in the infinite-particle limit, the implemented sampler may differ from the ideal conditional target because of errors in the diffusion model, its numerical implementation, and the guidance mechanism. We characterize how these local errors propagate through forward-smoothing kernels, which jointly account for the reverse dynamics and the remaining conditioning information. This yields non-asymptotic error bounds that capture both finite-particle fluctuations and approximation errors arising from initialization, numerical integration, score approximation, and potential design. In doing so, we extend stability guarantees for diffusion models to the conditional setting. Finally, we apply our framework to several state-of-the-art SMC-guided diffusion algorithms, providing a unified theoretical perspective on their approximation mechanisms and sources of error.

1 INTRODUCTION

Score-based Generative Models (SGMs) also called diffusion models provide a flexible way of learning complex high-dimensional data distributions together with a tractable sequential generative structure (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020; Song et al., 2021). While these models are primarily trained for unconditional generation, many applications require controlling the generated distribution through additional observations, constraints, or rewards. Conditional diffusion methods broadly follow two paradigms: conditioning may be incorporated during training through the model architecture or objective (Ho & Salimans, 2021; Rombach et al., 2022; Saharia et al., 2022), or imposed *post hoc* at inference time by steering a pretrained unconditional model. We focus on the latter setting, which arises for instance in Bayesian inverse problems, missing-data imputation, reward guidance, and conditional generation from partial observations. How to turn a pretrained diffusion model into an *accurate guided sampler* without retraining it?

Sequential Monte Carlo (SMC) provides a natural mechanism for inference-time guidance. It approximates a sequence of distributions using interacting particles that alternate between *mutation* through proposal kernels and *selection* through weighting and resampling (Doucet et al., 2001; Del Moral, 2004; Del Moral et al., 2006; Chopin & Papaspiliopoulos, 2020). Diffusion SMC samplers build their proposals from the pretrained reverse-diffusion dynamics and use intermediate potential functions to incorporate the guidance information (Trippe et al., 2023; Wu et al., 2023a; Cardoso et al., 2024; Dou & Song, 2024; Singhal et al., 2025; Nazemi et al., 2024). With appropriate proposals and potentials, they aim to approximate a tilted target of the form $\pi^\star(\mathrm{d}x) \propto G(x)\pi_{\mathrm{data}}(\mathrm{d}x)$, where $G(x)$ assigns greater weight to samples with the desired attributes.

A natural next question is whether the particles accurately approximate $\pi^\star$. Classical SMC theory quantifies the error due to using finitely many particles for a specified Feynman–Kac flow (Del Moral, 2004; Whiteley, 2012; Douc et al., 2014). In practical diffusion samplers, that flow is constructed using simulable approximations: a learned score and discretized reverse dynamics, and sometimes approximate intermediate potentials. Particle convergence therefore establishes accuracy for the implemented flow, which need not coincide with $\pi^\star$. This motivates an analysis that

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.

combines classical SMC error bounds with the errors introduced when score-based approximations are used to construct the twisted flow.

The resulting error therefore has two qualitatively different sources: finite-particle Monte Carlo error and deterministic approximation bias. The latter already appears in unconditional diffusion sampling, where convergence analyses typically separate initialization mismatch, score-approximation error, and numerical discretization error (Chen et al., 2023; Benton et al., 2024; Conforti et al., 2025b; Strasman et al., 2025a;b; Gentiloni Silveri & Ocello, 2025). In the guided setting, however, local errors in the diffusion dynamics are propagated through the remaining selection–mutation steps, altering their effect on the terminal distribution. This interaction is the missing link between diffusion-model convergence and SMC stability.

Contributions. Our main contributions can be summarized as follows.

- **Joint particle and model-approximation guarantees.** To the best of our knowledge, we derive the first non-asymptotic guarantee for diffusion-based SMC that jointly captures finite-particle error and approximation bias. The bound separates the usual $M^{-1/2}$ Monte Carlo contribution from propagated errors due to initialization, numerical discretization, score approximation, and guidance-potential mismatch. It therefore distinguishes errors that vanish by increasing the particle count from those that require improving the underlying diffusion model or guidance design.
- **Guided forgetting on noncompact state spaces.** We establish a forgetting mechanism for the time-inhomogeneous forward-smoothing dynamics induced by sequential guidance. Under structural conditions on the learned score and guidance potentials, Lyapunov drift and local minorization of the implemented diffusion proposals yield contraction in guidance-adapted weighted total-variation norms. Building on the Harris argument of Hairer & Mattingly (2008) and extending the unconditional analysis of Strasman et al. (2026), the resulting stability coefficients quantify how local particle and approximation errors are attenuated by the subsequent reverse dynamics and remaining guidance.
- **Connection with practical diffusion SMC.** We develop a flexible framework covering several existing diffusion-based SMC samplers, including Wu et al. (2023a); Cardoso et al. (2024), and recast them as particle approximations of an approximate Feynman–Kac flow whose terminal distribution may remain biased with respect to the desired target. By mapping their concrete proposal and weighting mechanisms onto our assumptions, the theory provides tools to analyze their stability, trace how the different sources of error propagate through the sampling trajectory, and better understand their strengths and potential failure modes.

2 Conditional diffusion sampling with SMC

Let $\pi_{\mathrm{data}} \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the data distribution for which a pretrained diffusion model is available. Assigning preferences to samples from π_{data} can be formulated as sampling from the tilted distribution

$$\pi^\star(\mathrm{d}x) \propto G(x)\, \pi_{\mathrm{data}}(\mathrm{d}x)\,,$$

where G is a nonnegative function encoding the desired preference and the normalizing constant is assumed to be finite and positive. For example, Bayesian posterior sampling can be recasted in this setting defining $G(x) = p(y \mid x)$ (Wu et al., 2023a; Cardoso et al., 2024), for a fixed observation y, as well as reward-guided generation, with $G(x) = \exp\{\lambda r(x)\}$ for a reward function r and guidance strength $\lambda > 0$ (Wu et al., 2023b; Black et al., 2024; Fan et al., 2023; Marion et al., 2025). We next show that the sequential structure of diffusion models naturally fits SMC: SGM transitions define the mutation kernels, while intermediate potentials progressively incorporate the guidance G.

2.1 Score-based generative models as priors

We first recall how score-based generative models transform noise into samples, establishing the reverse dynamics used in the guided sampler.

Forward-backward Diffusion processes. Fix a time horizon $T > 0$. Fix $\alpha \geq 0$ and $t \in [0, T] \ni t \mapsto \beta_t > 0$ a continuous noise schedule. The *forward diffusion* is defined as

$$\mathrm{d}\overrightarrow{\mathbf{X}}_t = -\alpha\beta(t)\overrightarrow{\mathbf{X}}_t \mathrm{d}t + \sqrt{2\beta(t)}\mathrm{d}B_t\,, \qquad \overrightarrow{\mathbf{X}}_0 \sim \pi_{\mathrm{data}}\,, \tag{1}$$

where $(B_t)_{t\in[0,T]}$ is a d-dimensional Brownian motion. Appropriate choices of $(\alpha, \beta_\cdot)$ recover the VE and VP diffusions (Song & Ermon, 2019; Sohl-Dickstein et al., 2015; Ho et al., 2020), we refer to Strasman et al. (2026) for more details. We write p_t to refer to both the density and, by abuse of notation, the law of $\overrightarrow{\mathbf{X}}_t$. Under minimal assumptions on π_{data}, the forward dynamics (1) admits a time-reversed process (Haussmann & Pardoux, 1986; Cattiaux et al., 2023), also known as *backward process*, given by

$$\mathrm{d}\overleftarrow{\mathbf{X}}_t = \left(\alpha\bar{\beta}_t \overleftarrow{\mathbf{X}}_t + 2\bar{\beta}_t \nabla \log p_{T-t}(\overleftarrow{\mathbf{X}}_t)\right)\mathrm{d}t + \sqrt{2\bar{\beta}_t}\,\mathrm{d}B_t\,, \qquad \overleftarrow{\mathbf{X}}_0 \sim p_T\,, \tag{2}$$

with $\bar{\beta}_t := \beta(T-t)$. To connect the reverse process with the SMC construction below, we fix the times at which particles will be weighted and resampled. For simplicity, let $N \in \mathbb{N}^*$ and choose the uniform grid

$$\mathcal{T}^{\mathrm{sel}} := \{t_k\}_{k=0}^N\,, \qquad t_k := \frac{kT}{N}\,, \quad k = 0, \dots, N\,.$$

For $k = 0, \dots, N-1$, let Q_{k+1} be the exact transition kernel of (2) over $[t_k, t_{k+1}]$:

$$\mathsf{Q}_{k+1}(\mathbf{x}, \mathrm{A}) := \mathbb{P}\Big(\overleftarrow{\mathbf{X}}_{t_{k+1}} \in \mathrm{A} \,\Big|\, \overleftarrow{\mathbf{X}}_{t_k} = \mathbf{x}\Big)\,, \qquad \mathrm{A} \in \mathcal{B}(\mathbb{R}^d)\,.$$

By time reversal, the unguided composition satisfies $p_T\mathsf{Q}_1 \cdots \mathsf{Q}_N = \pi_{\mathrm{data}}$. These kernels thus define an ideal reference for the mutation steps; the practical, simulable kernels introduced next approximate them.

Practical Implementation. On a practical level, three ingredients are required to implement the reverse dynamics (2): (i) a tractable reference distribution π_∞ approximating p_T, (ii) a learned score network $\mathrm{s}_\theta(\cdot, t)$ approximating $\nabla \log p_t$, and (iii) a numerical scheme to discretize the SDE. We therefore obtain a viable approximation by initializing the reverse process from the tractable reference distribution π_∞ in place of p_T, substituting the learned score $\mathrm{s}_\theta(\cdot, t)$ for the unavailable score $\nabla \log p_t$, and applying a numerical discretization to the resulting reverse SDE. Together, these three approximations turn the ideal reverse dynamics into a simulable Markov chain.

Two distinct grids on $[0, T]$ enter the construction. The grid $\mathcal{T}^{\mathrm{sel}} = \{t_k\}_{k=0}^N$ specifies the times at which particles are weighted and, when required, resampled. These times are determined by the SMC construction, for instance by the times at which guidance information is introduced. By contrast, numerical simulation relies on a possibly finer discretization grid

$$\mathcal{T}^{\mathrm{EM}} := \{s_n\}_{n=0}^{NJ}\,, \qquad \text{with} \quad s_n := \frac{nT}{NJ} \quad \text{and} \quad h := \frac{T}{NJ}\,,$$

where $J \geq 1$ is the number of numerical propagation steps performed between two consecutive SMC times. The grids are aligned through $t_k = s_{kJ}$, but serve different purposes: a single SMC mutation from t_k to t_{k+1} may comprise several Euler–Maruyama updates.

Defining $\Delta_n := \int_{s_n}^{s_{n+1}} \bar{\beta}_u \,\mathrm{d}u$, the corresponding time-changed Euler–Maruyama scheme is

$$\bar{\mathbf{X}}^\theta_{s_{n+1}} = \bar{\mathbf{X}}^\theta_{s_n} + \Delta_n\left(\alpha\bar{\mathbf{X}}^\theta_{s_n} + 2\,\mathrm{s}_\theta\left(\bar{\mathbf{X}}^\theta_{s_n}, T - s_n\right)\right) + \sqrt{2\Delta_n}\,\xi_n\,, \qquad \bar{\mathbf{X}}^\theta_0 \sim \pi_\infty\,, \tag{3}$$

where $(\xi_n)_{n=0}^{NJ-1}$ are i.i.d. $\mathcal{N}(0, \mathbf{I}_d)$ random variables, independent of $\bar{\mathbf{X}}^\theta_0$. When $J = 1$, the numerical grid coincides with the SMC selection grid. For $k = 0, \dots, N-1$, we denote by $\mathsf{Q}^{\theta,h}_{k+1}$ the transition kernel obtained by composing the J Euler–Maruyama updates in (3) over $[t_k, t_{k+1}]$. It is a simulable approximation of the exact reverse kernel Q_{k+1}, incorporating both score-approximation and time-discretization errors. Hence, in the absence of guidance, $\pi_\infty\,\mathsf{Q}^{\theta,h}_1 \cdots \mathsf{Q}^{\theta,h}_N \approx \pi_{\mathrm{data}}$.

2.2 Guided targets and Feynman–Kac representation

Diffusion models provide an unconditional generative model for π_{data}, whereas our target is $\pi^\star(\mathrm{d}x) \propto G(x)\pi_{\mathrm{data}}(\mathrm{d}x)$. A naive approach would propagate particles according to the unconditional reverse dynamics and apply G only at the terminal time. This amounts to a terminal importance-weighting correction: it does not steer trajectories toward regions where G is large and can therefore lead to severe weight degeneracy and high Monte Carlo variance, especially when the desired attributes are rare.

Ideal Feynman–Kac model. Diffusion-guided SMC instead introduces guidance progressively along the reverse-time trajectory through a sequence of positive potentials $(g_k)_{k=0}^{N-1}$. To make the selection step explicit, we introduce the *normalized selection operator*. For a probability measure ν and a positive measurable function g satisfying $0 < \nu[g] < \infty$, where $\nu[f] := \int f \, \mathrm{d}\nu$, define

$$\tau_g(\nu)(\mathrm{d}x) := \frac{g(x)\nu(\mathrm{d}x)}{\nu[g]}.$$

The measure $\tau_g(\nu)$ is the selected distribution associated with ν: states are favored according to their relative potential g, and the denominator ensures that the resulting measure has unit mass. The ideal Feynman–Kac flow alternates this *selection step* with *mutation* through the exact reverse kernels. Given $(g_k, \mathsf{Q}_{k+1})_{k=0}^{N-1}$, the flow is defined recursively by

$$\phi_0 = p_T, \qquad \phi_{k+1} = \tau_{g_k}(\phi_k)\mathsf{Q}_{k+1}, \qquad k = 0, \dots, N-1. \tag{4}$$

We consider potential sequences defining an ideal Feynman–Kac representation of the target, namely

$$\phi_N = \pi^\star \,. \tag{5}$$

The corresponding Feynman–Kac path measure is obtained by tilting the law of the reverse chain $(\overleftarrow{\mathbf{X}}_{t_k})_{k=0}^{N}$ by the accumulated potential $\prod_{k=0}^{N-1} g_k(\overleftarrow{\mathbf{X}}_{t_k})$.

Approximate Feynman–Kac model. The ideal Feynman–Kac model cannot generally be implemented because its initialization, mutation kernels, and potentials may involve intractable quantities. We therefore use the computable approximations introduced above. First, the unavailable initial law p_T is replaced by the tractable reference distribution π_∞. Second, each exact reverse kernel Q_{k+1} is replaced by the simulable kernel $\mathsf{Q}_{k+1}^{\theta,h}$, obtained by composing the Euler–Maruyama updates in (3); the parameter θ records the replacement of the exact score by the learned score, while h records the numerical discretization. Finally, when the ideal potential g_k cannot be evaluated, it is replaced by a positive computable approximation $\hat{g}_k$. These substitutions make selection and mutation implementable, but respectively introduce initialization, score-approximation, time-discretization, and potential-approximation errors.

Starting from $\phi_0^\theta = \pi_\infty$, the resulting approximate guided flow associated with $(\hat{g}_k, \mathsf{Q}_{k+1}^{\theta,h})_{k=0}^{N-1}$ is defined by

$$\phi_{k+1}^\theta = \tau_{\hat{g}_k}(\phi_k^\theta)\mathsf{Q}_{k+1}^{\theta,h}\,, \qquad \text{for } k = 0, \dots, N-1\,. \tag{6}$$

Equivalently, consider the simulable numerical reverse chain $(\bar{\mathbf{X}}_{s_n}^\theta)_{n=0}^{NJ}$ defined in (3) and observe it at the SMC times, $\bar{\mathbf{X}}_{t_k}^\theta = \bar{\mathbf{X}}_{s_{kJ}}^\theta$. The resulting chain has transitions $(\mathsf{Q}_k^{\theta,h})_{k=1}^{N}$, and its Feynman–Kac path measure is obtained by tilting its law by $\prod_{k=0}^{N-1} \hat{g}_k(\bar{\mathbf{X}}_{t_k}^\theta)$. Thus, ϕ_N^θ is the terminal distribution associated with the implemented Feynman–Kac model, while its discrepancy from the desired target $\phi_N = \pi^\star$ captures the price paid for using these approximations.

Particle-based approximation. To implement the recursion defining the flow $(\phi_k^\theta)_{0\le k\le N}$, we resort to the standard SMC population approximation. At each selection time t_k, the distribution ϕ_k^θ is represented by a population of M interacting particles $\{\mathbf{X}_k^{\theta,(i)}\}_{i=1}^M$ and its associated empirical measure

$$\phi_k^{\theta,M} := \frac{1}{M}\sum_{i=1}^{M} \delta_{\mathbf{X}_k^{\theta,(i)}}\,,$$

where $\delta_{\mathbf{x}}$ denotes the Dirac measure at $\mathbf{x}$. Applying the *selection–mutation* update to $\phi_k^{\theta,M}$ gives the following finite mixture

$$\tau_{\hat{g}_k}(\phi_k^{\theta,M})\mathsf{Q}_{k+1}^{\theta,h}(\mathrm{d}\mathbf{x}) = \sum_{i=1}^{M} w_k^{(i)} \mathsf{Q}_{k+1}^{\theta,h}(\mathbf{X}_k^{\theta,(i)}, \mathrm{d}\mathbf{x})\,, \qquad w_k^{(i)} := \frac{\hat{g}_k(\mathbf{X}_k^{\theta,(i)})}{\sum_{j=1}^{M}\hat{g}_k(\mathbf{X}_k^{\theta,(j)})}\,. \tag{7}$$

The resulting empirical measure $\phi_{k+1}^{\theta,M}$ provides an approximation ϕ_{k+1}^θ in (6). Algorithm 1 summarizes this particle implementation. The formulation is sufficiently flexible to recover concrete diffusion-SMC constructions. In particular, Appendix B.4 gives the explicit correspondence with MCGdiff (Cardoso et al., 2024) and SMCDiff (Trippe et al., 2023) for noiseless inpainting.

Algorithm 1 Diffusion-based SMC with multinomial resampling

Input: $M, N, J, \phi_0^\theta, (\hat{g}_k, \mathsf{Q}_{k+1}^{\theta,h})_{k=0}^{N-1}$.
Sample $\mathbf{X}_0^{\theta,(j)} \overset{\text{iid}}{\sim} \phi_0^\theta$, $j = 1, \dots, M$.
for $k = 0, \dots, N-1$ **do**
  Weight: $w_k^{(j)} \leftarrow \frac{\hat{g}_k(\mathbf{X}_k^{\theta,(j)})}{\sum_{m=1}^M \hat{g}_k(\mathbf{X}_k^{\theta,(m)})}$, $j = 1, \dots, M$.
  Select: sample independently $\widetilde{\mathbf{X}}_k^{\theta,(j)} \sim \sum_{m=1}^M w_k^{(m)} \delta_{\mathbf{X}_k^{\theta,(m)}}$, $j = 1, \dots, M$.
  Mutate: $\mathbf{X}_{k+1}^{\theta,(j)} \sim \mathsf{Q}_{k+1}^{\theta,h}(\widetilde{\mathbf{X}}_k^{\theta,(j)}, \mathrm{d}\mathbf{x})$, $j = 1, \dots, M$.
end for
Output: $\phi_N^{\theta,M} = M^{-1} \sum_{j=1}^M \delta_{\mathbf{X}_N^{\theta,(j)}}$.

3 Forgetting and error propagation in diffusion-based SMC

The three constructions introduced in Section 2 naturally decompose the error with respect to the desired target. The ideal flow terminates at $\phi_N = \pi^\star$, the implemented infinite-particle flow terminates at ϕ_N^θ, and the particle output $\phi_N^{\theta,M}$ approximates ϕ_N^θ. Hence, for any admissible test function ψ,

$$\phi_N^{\theta,M}[\psi] - \phi_N[\psi] = \underbrace{\phi_N^{\theta,M}[\psi] - \phi_N^\theta[\psi]}_{\text{Particle error}} + \underbrace{\phi_N^\theta[\psi] - \phi_N[\psi]}_{\text{Approximation bias}} . \tag{8}$$

The first term is the finite-particle Monte Carlo error, while the second measures the discrepancy between the implemented and ideal guided flows. We control both through the forgetting properties governing their propagation.

3.1 Guided forgetting

Stability of the learned proposals. The analysis of SMC methods relies on two closely related tools: *forgetting estimates*, which quantify how the dependence on initial conditions is attenuated along the dynamics, and *stability arguments*, which control how local perturbations propagate through successive selection–mutation steps (Hairer & Mattingly, 2008; Del Moral, 2004; Whiteley, 2012; Douc et al., 2014). For an adaptation of these ideas to the unguided SGM setting, see Strasman et al. (2026), who establish forgetting properties for the exact reverse-time dynamics under structural assumptions on the data distribution.

In diffusion-based SMC, mutation relies on approximate SGM kernels $\mathsf{Q}_k^{\theta,h}$ built from a learned score and numerical discretization. The classical forgetting and stability tools must therefore be adapted to account for these approximations. We show that Harris stability of the SGM kernels can be established from structural properties of s_θ.

To formulate stability on the unbounded state space $\mathbb{R}^d$, we use the quadratic Lyapunov function $V(\mathbf{x}) := \|\mathbf{x}\|^2$, naturally matched to the dissipativity and growth conditions imposed below and controls the second moments of the chain. The associated V-weighted metric records both discrepancies between measures and their behavior in the tails, making it well suited to Markov kernels satisfying Lyapunov drift conditions. Accordingly, for measurable test functions $\psi : \mathbb{R}^d \to \mathbb{R}$ and finite signed measures μ with $|\mu|[V] < \infty$, we define the weighted supremum norm and the V-weighted total-variation norm by

$$\|\psi\|_{V,\infty} := \sup_{\mathbf{x} \in \mathbb{R}^d} \frac{|\psi(\mathbf{x})|}{1 + V(\mathbf{x})}, \qquad \|\mu\|_V := \int_{\mathbb{R}^d} (1 + V)\, \mathrm{d}|\mu| . \tag{9}$$

Assumption 3.1 (Learned score stability)**.** For every $k = 0, \dots, N-1$, there exist $\gamma_k > \alpha/2$, $\kappa_k \geq 0$, and $L_k \geq 0$ such that, for every $u \in [t_k, t_{k+1})$ and $\mathbf{x} \in \mathbb{R}^d$,

$$\langle \mathbf{x}, \mathrm{s}_\theta(\mathbf{x}, T-u) \rangle \leq -\gamma_k V(\mathbf{x}) + \kappa_k , \qquad \|\mathrm{s}_\theta(\mathbf{x}, T-u)\|^2 \leq L_k(1 + V(\mathbf{x})) . \tag{10}$$

These are standard dissipativity and linear-growth conditions for the reverse drift. The first condition, together with $\gamma_k > \alpha/2$, makes the approximate reverse drift dissipative in the tails. The

second is implied, for instance, by a stronger global Lipschitz regularity of the network, as satisfied by standard feed-forward architectures with Lipschitz activations (Virmaux & Scaman, 2018). They are natural learned analogues of the estimates satisfied by the exact score under comparable structural assumptions on π_{data}; see Strasman et al. (2026). Thus, when the score approximation preserves these tail properties, Assumption 3.1 is expected to hold for the learned reverse dynamics.

Under Assumption 3.1, the approximate SGM kernels satisfy the two ingredients of Harris stability given in the following proposition, proved in Appendix C.2.

Proposition 3.2 (Stability of the approximate proposal kernels)**.** *Suppose that Assumption 3.1 hold. Assume moreover that the numerical increments within $[t_k, t_{k+1}]$ satisfy*

$$\Delta_n \le \frac{2\gamma_k - \alpha}{\alpha^2 - 4\alpha\gamma_k + 4L_k}\,, \qquad n = kJ, \dots, (k+1)J - 1\,. \tag{11}$$

Then, we have

$$\mathsf{Q}_{k+1}^{\theta,h} V \le \lambda_{k+1}^{Q} V + K_{k+1}^{Q}\,. \tag{12}$$

with $\lambda_{k+1}^{Q} = \exp\left(-(2\gamma_k - \alpha)\int_{t_k}^{t_{k+1}} \bar\beta_u \,\mathrm{d}u\right)$, $K_{k+1}^{Q} = C_k \int_{t_k}^{t_{k+1}} \bar\beta_u \,\mathrm{d}u$, and C_k defined in (31)*, independent of J and h.*

Moreover, for every $R > 0$, there exist $\varepsilon_{k+1}^{Q}(R) > 0$ and a probability measure $\nu_{k+1,R}^{Q}$ such that

$$\mathsf{Q}_{k+1}^{\theta,h}(\mathbf{x}, A) \ge \varepsilon_{k+1}^{Q}(R)\, \nu_{k+1,R}^{Q}(A)\,, \qquad V(\mathbf{x}) \le R\,, \qquad A \in \mathcal{B}(\mathbb{R}^d)\,. \tag{13}$$

From proposal stability to guided forgetting. Since the potentials enter the Feynman–Kac flow at each selection step, stability of the mutation kernels must be complemented by regularity conditions on the guidance potentials. Following standard assumptions from SMC stability theory (Del Moral, 2004; Whiteley, 2012; Douc et al., 2014), we impose the following condition.

Assumption 3.3 (Guidance potentials)**.** For every $k = 0, \dots, N-1$, the ideal and implemented potentials are strictly positive and bounded:

$$0 < g_k(\mathbf{x}) \le \|g_k\|_\infty < \infty, \qquad 0 < \hat g_k(\mathbf{x}) \le \|\hat g_k\|_\infty < \infty, \qquad \mathbf{x} \in \mathbb{R}^d\,.$$

Unlike classical strong-mixing assumptions, Assumption 3.3 does not require the potentials to be uniformly bounded away from zero; in particular, $\inf_{\mathbf{x}} \hat g_k(\mathbf{x})$ may vanish. This point accommodates bounded likelihood and transition factors that decay to zero, as used for motif conditioning and inverse problems in SMCDiff, MCGDiff, Filtering Posterior Sampling, and particle-filtering latent diffusion (Trippe et al., 2023; Cardoso et al., 2024; Dou & Song, 2024; Nazemi et al., 2024). It also covers the positive twisting and reward potentials used by TDS and Feynman–Kac steering whenever the corresponding scores are bounded above, or are clipped in practice (Wu et al., 2023a; Singhal et al., 2025). Our analysis can therefore compare these methods at the level of their Feynman–Kac flows while allowing them to use different proposals and potential constructions.

To track how a perturbation at stage k propagates to the terminal distribution, we aggregate the remaining guidance weights conditionally on the current state. This defines the normalization underlying the forward-smoothing kernels. Let $\mathbb{E}_{k,\mathbf{x}}^{\theta}$ denote expectation for the unweighted reverse chain (3) started from $\mathbf{x}$ at time t_k. We define the *continuation function* as

$$\widetilde\beta_{k|N}^{\theta}(\mathbf{x}) := \mathbb{E}_{k,\mathbf{x}}^{\theta}\left[\prod_{j=k}^{N-1} \frac{\hat g_j(\bar{\mathbf{X}}_{t_j}^{\theta})}{\|\hat g_j\|_\infty}\right], \qquad \widetilde\beta_{N|N}^{\theta} = \mathbb{1}\,. \tag{14}$$

It measures the expected remaining guidance weight of a trajectory starting from $\mathbf{x}$. The associated terminal *forward-smoothing kernel* is defined, for an appropriate test function ψ, by

$$\mathsf{G}_{k:N}^{\theta}\psi(\mathbf{x}) := \frac{\mathbb{E}_{k,\mathbf{x}}^{\theta}\left[\psi(\bar{\mathbf{X}}_{t_N}^{\theta}) \prod_{j=k}^{N-1} \frac{\hat g_j(\bar{\mathbf{X}}_{t_j}^{\theta})}{\|\hat g_j\|_\infty}\right]}{\widetilde\beta_{k|N}^{\theta}(\mathbf{x})}\,. \tag{15}$$

It is the law of the terminal state obtained by propagating from $\mathbf{x}$ while incorporating all future guidance. In particular, $\phi_N^{\theta} = \tau_{\widetilde\beta_{k|N}^{\theta}}(\phi_k^{\theta})\mathsf{G}_{k:N}^{\theta}$ (see Corollary C.3).

To transfer proposal stability to the forward-smoothing kernels, we introduce the guidance-adapted Lyapunov function $V^\theta_{\ell|N}(\mathbf{x}) := \frac{V(\mathbf{x})}{\widetilde{\beta}^\theta_{\ell|N}(\mathbf{x})}$ for $0 \le \ell \le N$.

Proposition 3.4 (Forward-smoothing contraction)**.** *Under Assumptions 3.1 3.3 and step-size condition* (11)*, there exist $b_N \in (0,1]$ and $\alpha_{k,N} \in (0,1)$, $1 \le k \le N$, such that, for every $0 \le \ell \le N$ and probability measures μ_1, μ_2 with finite $V^\theta_{\ell|N}$-moments,*

$$\left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_V \le \Gamma_{\ell,N} \left\|\mu_1 - \mu_2\right\|_{V^\theta_{\ell|N}} , \tag{16}$$

where $\Gamma_{\ell,N} := \frac{1}{b_N} \prod_{k=\ell+1}^N \alpha_{k,N}$ and $\Gamma_{N,N} := 1$.

The factors $\alpha_{k,N}$ quantify contraction at successive forward-smoothing steps, while b_N^{-1} accounts for conversion to the terminal V-norm. The explicit constants and proof are given in Appendix C. The learned-score and step-size conditions may alternatively be replaced by Assumption C.4. This result provides the forgetting mechanism through which local particle and approximation errors are attenuated sequentially along the guided dynamics, yielding a unified stability view on both errors.

3.2 Unified particle and approximation errors

We now leverage the forgetting property of Proposition 3.4 to combine particle and approximation errors into a single terminal guarantee. The following mild regularity condition is used to control the diffusion-model approximation terms and is standard in KL convergence analyses of SGMs (Conforti et al., 2025b).

Assumption 3.5 (Training-data regularity)**.** The data distribution admits a strictly positive density $p_0 \in C^2(\mathbb{R}^d)$ with finite Fisher information and finite fourth moment.

Alternative sufficient conditions on the data distribution and their relation to the learned-score assumptions, are discussed in Appendix A.1.

Theorem 3.6. *Let $p \ge 1$ and fix the selection grid $\mathcal{T}^{\mathrm{sel}}$. Suppose that Assumptions3.1, 3.3, D.1 (with $q = 2p$), D.2, 3.5 together with the step-size condition* (11) *hold. Then, for every ψ such that $\|\psi\|_{V,\infty} < \infty$, there exist $C_\ell^{\mathrm{MC}}, C_k^{\mathrm{discr}}, C_k^{\mathrm{net}}, C_k^{\mathrm{pot}}, \Lambda_{k,N}$ finite such that*

$$\left\|\phi_N^{\theta,M}[\psi] - \pi^\star[\psi]\right\|_{L_p} \le \|\psi\|_{V,\infty} \Bigg[\frac{1}{\sqrt{M}} \sum_{\ell=0}^N \Gamma_{\ell,N} C_\ell^{\mathrm{MC}} + \Gamma_{0,N}\Lambda_{0,N} \left\|p_T - \pi_\infty\right\|_V + \sum_{k=1}^N \Gamma_{k,N}\Lambda_{k,N} \left(C_k^{\mathrm{discr}} \sqrt{h} + C_k^{\mathrm{net}} \varepsilon_k^{\mathrm{score}}(\theta,h) + C_k^{\mathrm{pot}} \varepsilon_k^{\mathrm{pot}} \right) \Bigg] .$$

where $\varepsilon_k^{\mathrm{score}}(\theta,h)^2 := \sum_{n=(k-1)J}^{kJ-1} \Delta_n \, \mathbb{E}_k[\|\nabla \log p_{T-s_n}(\overleftarrow{\mathbf{X}}_{s_n}) - \mathrm{s}_\theta\left(\overleftarrow{\mathbf{X}}_{s_n}, T - s_n\right)\|^2]$ and $\mathbb{E}_k$ denotes expectation under the exact reverse dynamics (2) *initialized from $\tau_{g_{k-1}}(\phi_{k-1})$ at time t_{k-1}. $\varepsilon_k^{\mathrm{pot}}$ denotes the normalized potential mismatch defined in* (18)*. For $k < N$, $\Gamma_{k,N} = b_N^{-1} \prod_{j=k+1}^N \alpha_{j,N}$ with $\alpha_{j,N} \in (0,1)$, and $\Gamma_{N,N} = 1$ is the forgetting factor.*

All constants in Theorem 3.6 are quantitatively controlled by the corresponding model, stability, and moment parameters; their definitions and derivations are given in the Appendices C, D and E.

Sketch of proof. The bound results from propagating two types of local perturbations through the same forward-smoothing dynamics. Note that,

$$\left\|\phi_N^{\theta,M}[\psi] - \pi^\star[\psi]\right\|_{L_p} \le \left\|\phi_N^{\theta,M}[\psi] - \phi_N^\theta[\psi]\right\|_{L_p} + \|\psi\|_{V,\infty} \left\|\phi_N^\theta - \phi_N\right\|_V .$$

Particle approximation of the approximate model. The particle error is a finite sum of local sampling fluctuations, each of order $M^{-1/2}$ under multinomial resampling by conditional Marcinkiewicz–Zygmund inequalities (Del Moral, 2004). On the unbounded state space, propagating these errors additionally requires Lyapunov moment control and integrability of the random normalization factors induced by future guidance. Such moment and normalization conditions are standard in particle analyses on non-compact spaces (Whiteley, 2012). Under the particle-integrability

conditions of Assumptions D.1 (with $q = 2p$) and D.2, together with the forward-smoothing stability estimate of Proposition 3.4, for any measurable ψ such that $\|\psi\|_{V,\infty} < \infty$,

$$\left\|\phi_N^{\theta,M}[\psi] - \phi_N^{\theta}[\psi]\right\|_{L_p} \leq \frac{\|\psi\|_{V,\infty}}{\sqrt{M}} \sum_{\ell=0}^{N} \Gamma_{\ell,N} \mathcal{C}_\ell^{\mathrm{MC}} \, \mathcal{A}_{\ell,N} \, . \tag{17}$$

See Lemma D.4 for the proof. $\mathcal{C}_\ell^{\mathrm{MC}}$ controls the local empirical fluctuation, while $\mathcal{A}_{\ell,N}$ captures the inverse normalization induced by the remaining guidance; see Appendix D.1 for their definitions.

Approximation bias. For $1 \leq k \leq N$, comparing the ideal and implemented updates from the same input law ϕ_{k-1} gives the local discrepancy. A telescoping argument and forward-smoothing stability of Proposition 3.4 yields (see Lemma E.3)

$$\left\|\phi_N^{\theta} - \phi_N\right\|_V \leq \sum_{k=0}^{N} \Gamma_{k,N} \Lambda_{k,N} \mathcal{B}_k^{\mathrm{bias}}.$$

where $\mathcal{B}_k^{\mathrm{bias}} = \left\|\phi_k - \tau_{\hat{g}_{k-1}}(\phi_{k-1}) \mathsf{Q}_k^{\theta,h}\right\|_V$ and $\mathcal{B}_0^{\mathrm{bias}} = \left\|\phi_0 - \phi_0^{\theta}\right\|_V$. The local discrepancies have three sources: the initialization mismatch $\mathcal{B}_0^{\mathrm{bias}}$, the SGM approximation error $\mathcal{B}_k^{\mathrm{SGM}}$ arising from score estimation and numerical discretization, and the guidance-potential mismatch $\mathcal{B}_k^{\mathrm{pot}}$. Then, under Assumptions 3.1 3.3 3.5 and the step-size condition (11) the following hold:

(i) *Initialization* (Section H.1 in Strasman et al., 2026). $\phi_0 = p_T$ and $\phi_0^{\theta} = \pi_\infty$, hence $\mathcal{B}_0^{\mathrm{bias}} = \|p_T - \pi_\infty\|_V$. In particular, when $\alpha > 0$, $\mathcal{B}_0^{\mathrm{bias}} \lesssim \exp\left\{-\alpha \int_0^T \beta(s)\,\mathrm{d}s\right\}$;

(ii) *SGM kernel approximation* (Lemma E.4). For every $1 \leq k \leq N$, there exist $C_k^{\mathrm{discr}}, C_k^{\mathrm{net}} < \infty$ independent of h such that $\mathcal{B}_k^{\mathrm{SGM}} \leq C_k^{\mathrm{discr}} \sqrt{h} + C_k^{\mathrm{net}} \varepsilon_k^{\mathrm{score}}(\theta, h)$.

(iii) *Potential mismatch* (Proposition E.5). For $1 \leq k \leq N$,

$$\mathcal{B}_k^{\mathrm{pot}} \leq \varepsilon_k^{\mathrm{pot}} \left\{\phi_{k-1}\left[(1 + K_k^Q + \lambda_k^Q V)^2\right]\right\}^{1/2} := \varepsilon_k^{\mathrm{pot}} C_k^{\mathrm{pot}} \, , \tag{18}$$

with $(\varepsilon_k^{\mathrm{pot}})^2 := \phi_{k-1}\left[\left(\frac{\hat{g}_{k-1}}{\phi_{k-1}[\hat{g}_{k-1}]} - \frac{g_{k-1}}{\phi_{k-1}[g_{k-1}]}\right)^2\right]$.

Combining these local estimates and using that $\mathcal{B}_k^{\mathrm{bias}} \leq \mathcal{B}_k^{\mathrm{pot}} + \mathcal{B}_k^{\mathrm{SGM}}$ finishes the proof. □

3.3 Interpretation and controlled setting illustration

Theorem 3.6 separates finite-particle fluctuations from the approximation bias, while making explicit how both are propagated by the guided dynamics. In contrast with many classical SMC error bounds formulated for bounded test functions, Theorem 3.6 applies to test functions with at most quadratic growth. As a consequence, it yields non-asymptotic guarantees for estimating quantities such as the mean and second moments of the target distribution.

The Monte Carlo contribution decreases at the standard rate $M^{-1/2}$, whereas initialization, score approximation, numerical discretization, and potential mismatch persist even in the infinite-particle limit. Increasing M therefore improves the approximation of the implemented flow ϕ_N^{θ}, but does not by itself bring ϕ_N^{θ} closer to the desired target $\pi^\star$. These terms must instead be reduced by improving the reference initialization, score model, numerical discretization, or guidance construction. In particular, the bound yields a discretization contribution of order $\sqrt{h}$ and, in the Ornstein–Uhlenbeck setting with $\alpha > 0$, an initialization error decaying at rate of order $\mathrm{e}^{-\alpha T}$.

Moreover, a perturbation introduced at stage k contributes to the terminal error through the forgetting factor $\Gamma_{k,N}$. Hence, local errors of comparable magnitude may have different terminal effects depending on where they occur along the SMC trajectory. In this regard, the implemented potentials $\hat{g}_k$ play a dual role: they determine the local mismatch $\varepsilon_k^{\mathrm{pot}}$ but also affect the stability through $\Gamma_{k,N}$.

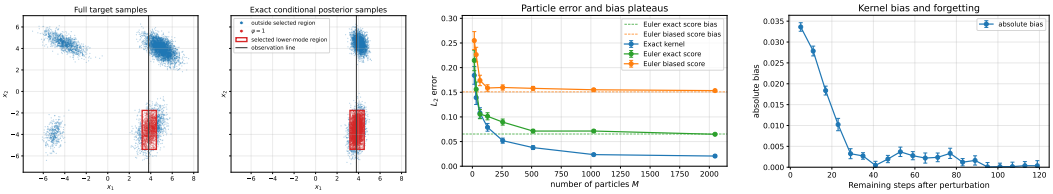


Figure 1: Controlled Gaussian-mixture illustration. Left: prior and conditional posterior, with the red rectangle defining the observable $\psi_{\mathcal{R}}$. Middle: empirical L_2 error versus particle count M, illustrating Monte Carlo decay and approximation-bias plateaus. Right: terminal effect of a single local kernel perturbation versus the number of remaining reverse steps; larger values correspond to earlier perturbations. Error bars show standard errors across independent runs.

Finally, the approximation terms are conditioning-dependent. The score error $\varepsilon_k^{\mathrm{score}}(\theta, h)$ is evaluated along reverse trajectories initialized from the ideal selected law, while $\varepsilon_k^{\mathrm{pot}}$ compares normalized ideal and implemented potentials under the true filtering distribution. The bound therefore measures approximation quality in the regions emphasized by the guidance. In other words, even though the proposal is built from an unconditional diffusion model, its contribution to the guarantee is determined by its accuracy along the trajectories explored by the conditional SMC flow.

Numerical illustration. We illustrate these mechanisms in Gaussian-mixture toy problem, chosen because the scores, reverse kernels, and targets are available in closed form. The purpose is not to reassess the practical performance of diffusion-based SMC, which has been studied extensively in the previously cited literature, but to isolate the error components appearing in the theorem. We use a four-component anisotropic mixture conditioned on a noisy observation of its first coordinate (see details in Appendix F). The left panel of Figure 1 shows the prior and posterior, with $\psi_{\mathcal{R}} = \mathbf{1}_{\mathcal{R}}$ (the red rectangle) indicating a region around one posterior mode; its exact posterior probability is computed from the closed-form Gaussian mixture.

At a fixed reverse-time grid, we compare SMC using exact reverse kernels, Euler kernels driven by the exact score, and Euler kernels with a controlled score perturbation. The exact kernels remove score and discretization errors from the proposals; the two Euler variants introduce discretization error and, in the latter case, an additional score error. The middle panel reports the empirical L_2 error over independent runs as the particle count M increases. It illustrates Monte Carlo decay compatible with the $M^{-1/2}$ scale and error plateaus when approximation bias dominates: more particles improve the approximation of the implemented flow without removing its bias.

The right panel examines local error propagation. We replace only the ℓ-th exact reverse kernel by a perturbed-score Euler kernel, keeping all other kernels exact, and vary ℓ. Plotting the resulting terminal change in the modal-region probability against $N - \ell$ shows that perturbations introduced earlier in the SMC flow, and therefore followed by more guided transitions, are more strongly attenuated. This illustrates the guided forgetting mechanism quantified by the propagation factors $\Gamma_{\ell,N}$. Full specifications, additional diagnostics, and a dimension-50 benchmark are given in Appendix F.

4 Discussion

Our analysis applies to diffusion-based SMC samplers including SMCDiff, TDS, MCGDiff, filtering posterior sampling, and particle-filtering latent diffusion (Trippe et al., 2023; Wu et al., 2023a; Cardoso et al., 2024; Dou & Song, 2024; Nazemi et al., 2024). It separates finite-particle error from approximation bias due to learned dynamics, discretization, and guidance, while propagating both through the same guided forgetting mechanism.

In this work, the choice $V(\mathbf{x}) = \|\mathbf{x}\|^2$ is natural for our learned-score assumptions, but the abstract Feynman–Kac analysis allows geometry-adapted Lyapunov functions, potentially yielding sharper stability bounds or accommodating different observables. Finally, the implemented potentials affect both local approximation errors and the stability factors governing their propagation. Turning this interaction into practical criteria for guidance design is a natural direction for future work.

AI USE STATEMENT

In this work, we used generative AI tools to assist for manuscript drafting and editing. We take responsibility for all the content of this work.

REPRODUCIBILITY STATEMENT

Assumptions used in the theoretical results are stated explicitly in the main text, and complete proofs and intermediate technical results are provided in the Appendix. The numerical appendix specifies the Gaussian-mixture models, observation mechanisms, exact scores and reverse kernels, controlled score perturbations, evaluation criteria, and experimental settings used in the reported experiments. These closed-form benchmarks make the reference posterior quantities directly reproducible. Experimental code is provided in the supplementary material.

ETHICS STATEMENT

This work is primarily theoretical and uses synthetic data in its numerical experiments. It does not involve human subjects, personal or sensitive data, or deployment on real users. As with conditional generative modeling more broadly, particular downstream applications may raise domain-specific ethical considerations that are outside the scope of this work.

REFERENCES

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*. Springer, 2014.

Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=r5njV3BsuD.

Andreas Bergmeister, Stefanie Jegelka, Nikolas Nüsken, Carles Domingo-Enrich, and Jakiw Pidstrigach. Reinforce adjoint matching: Scaling rl post-training of diffusion and flow-matching models. *arXiv preprint arXiv:2605.10759*, 2026. URL https://arxiv.org/abs/2605.10759.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *International Conference on Learning Representations*, volume 2024, pp. 4965–4987, 2024.

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

Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 59(4):1844–1881, 2023.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=zyLVMgsZ0U_.

Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Springer, 2020.

Giovanni Conforti, Alain Durmus, Le-Tuyet-Nhi Pham, and Gael Raoul. Non-asymptotic convergence of discrete diffusion models: Masked and random walk dynamics. *arXiv preprint arXiv:2512.00580*, 2025a.

Giovanni Conforti, Alain Durmus, and Marta Silveri Gentiloni. Kl convergence guarantees for score diffusion models under minimal data assumptions. *SIAM Journal on Mathematics of Data Science*, 7(1):86–109, 2025b. doi: 10.1137/23M1613670. URL https://doi.org/10.1137/23M1613670.

Pierre Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Probability and Its Applications. Springer, 2004.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436, 2006.

Carles Domingo-Enrich and Jiequn Han. Adjoint matching through the lens of the stochastic maximum principle in optimal control. *arXiv preprint arXiv:2604.08580*, 2026. URL https://arxiv.org/abs/2604.08580.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations*, pp. 53791–53846, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/852f50969a9e523ec41d26f2f68bd456-Paper-Conference.pdf.

Carles Domingo-Enrich, Yuanqi Du, and Michael S. Albergo. A unified perspective on fine-tuning and sampling with diffusion and flow models. *arXiv preprint arXiv:2605.00229*, 2026. URL https://arxiv.org/abs/2605.00229.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *International Conference on Learning Representations*, volume 2024, pp. 12069–12095, 2024.

R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov Chains*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018.

Randal Douc, Éric Moulines, and Jimmy Olsson. Long-term stability of sequential monte carlo methods under verifiable conditions. *The Annals of Applied Probability*, 24(5):1767–1802, 2014.

Arnaud Doucet, Nando De Freitas, and Neil James Gordon. *Sequential Monte Carlo methods in practice*. Springer, 2001.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in neural information processing systems*, 36:79858–79885, 2023.

Marta Gentiloni Silveri and Antonio Ocello. Beyond log-concavity and score regularity: Improved convergence bounds for score-based generative models in w2-distance. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55631–55656. PMLR, 2025. URL https://proceedings.mlr.press/v267/silveri25a.html.

Pierre Gloaguen, Sylvain Le Corff, and Jimmy Olsson. A pseudo-marginal sequential monte carlo online smoothing algorithm. *Bernoulli*, 28(4):2606–2633, 2022.

Martin Hairer and Jonathan C. Mattingly. Yet another look at harris' ergodic theorem for markov chains. *arXiv preprint arXiv:0810.2777*, 2008. URL https://arxiv.org/abs/0810.2777.

U. G. Haussmann and E. Pardoux. Time reversal of diffusions. *The Annals of Probability*, 14(4):1188–1205, 1986.

Aaron Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Brandon Wood, Daniel Levine, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. *arXiv preprint arXiv:2504.11713*, 2025. URL https://arxiv.org/abs/2504.11713.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. URL https://openreview.net/forum?id=qw8AKxfYbI.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 6840–6851, 2020.

Zhen Liu, Tim Xiao, Carles Domingo-Enrich, Weiyang Liu, and Dinghuai Zhang. Value gradient guidance for flow matching alignment. In *Advances in Neural Information Processing Systems*, volume 38, pp. 11372–11404. Curran Associates, Inc., 2025. doi: 10.52202/085713-0380. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/10b7e27c8eb9571fbbd2ae6a9f8c3855-Paper-Conference.pdf.

Pierre Marion, Anna Korba, Peter Bartlett, Mathieu Blondel, Valentin De Bortoli, Arnaud Doucet, Felipe Llinares-López, Courtney Paquette, and Quentin Berthet. Implicit diffusion: Efficient optimization through stochastic sampling. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1999–2007. PMLR, 2025. URL https://proceedings.mlr.press/v258/marion25a.html.

Amir Nazemi, Mohammad Hadi Sepanj, Nicholas Pellegrino, Chris Czarnecki, and Paul Fieguth. Particle-filtering-based latent diffusion for inverse problems. *arXiv preprint arXiv:2408.13868*, 2024.

Le-Tuyet-Nhi Pham, Dario Shariatian, Antonio Ocello, Giovanni Conforti, and Alain Oliviero Durmus. Discrete markov probabilistic models: An improved discrete score-based framework with sharp convergence bounds under minimal assumptions. In *Forty-second International Conference on Machine Learning*, 2025.

Jakiw Pidstrigach, Elizabeth Baker, Carles Domingo-Enrich, George Deligiannidis, and Nikolas Nüsken. Conditioning diffusions using malliavin calculus. *arXiv preprint arXiv:2504.03461*, 2025. URL https://arxiv.org/abs/2504.03461.

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 (CVPR)*, pp. 10684–10695, June 2022.

W. Rudin. *Real and Complex Analysis*. McGraw-Hill, Singapore, 1987.

Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, New York, NY, USA, 2022. Association for Computing Machinery. doi: 10.1145/3528233.3530757.

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*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55810–55827. PMLR, 2025. URL https://proceedings.mlr.press/v267/singhal25b.html.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 11918–11930, 2019.

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.

Stanislas Strasman, Antonio Ocello, Claire Boyer, Sylvain Le Corff, and Vincent Lemaire. An analysis of the noise schedule for score-based generative models. *Transactions on Machine Learning Research*, 2025a. ISSN 2835-8856. URL https://openreview.net/forum?id=BlYIPa0Fx1.

Stanislas Strasman, Sobihan Surendran, Claire Boyer, Sylvain Le Corff, Vincent Lemaire, and Antonio Ocello. Wasserstein convergence of critically damped langevin diffusions. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=a7bisLzORM.

Stanislas Strasman, Gabriel Cardoso, Sylvain Le Corff, Vincent Lemaire, and Antonio Ocello. On forgetting and stability of score-based generative models. *arXiv preprint arXiv:2601.21868*, 2026. doi: 10.48550/arXiv.2601.21868.

Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi S. Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. In *The Eleventh International Conference on Learning Representations*, 2023.

Aladin Virmaux and Kevin Scaman. Lipschitz regularity of deep neural networks: analysis and efficient estimation. *Advances in neural information processing systems*, 31, 2018.

Nick Whiteley. Sequential monte carlo samplers: error bounds and insensitivity to initial conditions. *Stochastic Analysis and Applications*, 30(5):774–798, 2012.

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. *Advances in Neural Information Processing Systems*, 36:31372–31403, 2023a.

Xiaoshi Wu, Keqiang Sun, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score: Better aligning text-to-image models with human preference. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2096–2105. IEEE, 2023b.

APPENDIX

A Additional discussion **14**

A.1 Alternative data regularity conditions 14

A.2 Related work 15

B Notation and general Feynman–Kac framework **16**

B.1 Notation summary 16

B.2 Measures, kernels, and weighted norms 16

B.3 Feynman–Kac framework 17

B.4 Correspondence with existing methods: noiseless inpainting 19

C Forward-smoothing kernels: properties and forgetting **19**

C.1 Kernel notation and tail factorization 19

C.2 Proof of Proposition 3.2 22

C.3 Forgetting of the forward smoothing kernel 25

D Monte Carlo Error **30**

D.1 Detailed assumptions 30

D.2 Proof 31

E Approximation bias **36**

E.1 Local discrepancies and moment control 36

E.2 Propagation of local approximation errors 37

E.3 Local error controls 40

F Numerical illustration **42**

F.1 Gaussian-mixture target and closed form diffusion model 42

F.2 Conditional sampling model: noisy inpainting. 44

F.3 Two-dimensional Gaussian mixture 46

F.4 A 25-component Gaussian mixture in dimension 50 48

A ADDITIONAL DISCUSSION

A.1 ALTERNATIVE DATA REGULARITY CONDITIONS

Assumption 3.5 is one possible sufficient condition for convergence bounds in SGMs (Conforti et al., 2025b). An alternative path is taken in Strasman et al. (2026), where π_{data} is assumed to be in $C^2(\mathbb{R}^d)$ such that, there exist $\gamma_{\mathrm{data}} > \alpha/2$, $\kappa_{\mathrm{data}} \geq 0$, $C_{\mathrm{data}} < \infty$, and $r \geq 1$,

$$\langle \mathbf{x}, \nabla \log p_0(\mathbf{x}) \rangle \leq -\gamma_{\mathrm{data}} \|\mathbf{x}\|^2 + \kappa_{\mathrm{data}} \,,$$
$$\left\| \nabla^2 \log p_0(\mathbf{x}) \right\|_{\mathrm{F}} \leq C_{\mathrm{data}}(1 + \|\mathbf{x}\|^r) \,.$$

These conditions, as proven in Strasman et al. (2026), propagate to the forward marginals p_t and their scores, and therefore guarantee that the ideal reverse dynamics satisfy (10). They also impose stability directly on the true data score, while Assumption 3.1 asks the learned score to preserve

analogous tail dissipativity and growth properties. In this regard, the alternative assumption allows us to interpret Assumption 3.1 as requiring the network approximation to preserve the dissipativity structure of the underlying diffusion.

A.2 Related work

Diffusion-based SMC. Several diffusion-based SMC methods provide principled constructions for post-hoc conditional sampling. In particular, Wu et al. (2023a) establish asymptotic exactness, as the number of particles increases, for the conditional distribution induced by a prescribed diffusion model, while Cardoso et al. (2024) develop an SMC construction for Bayesian linear inverse problems. These guarantees concern a fixed diffusion model and therefore do not, by themselves, quantify the discrepancy between the implemented sampler and a target distribution defined from the true data distribution. Our analysis is complementary: it *separates the finite-particle error from the bias* introduced by initialization, score approximation, numerical discretization, and approximate guidance potentials.

Reward fine-tuning of diffusion and flow models. A related line of work modifies the parameters of a pretrained generative model so that its terminal distribution favors high-reward samples. Existing approaches include supervised fine-tuning on reward-filtered datasets (Wu et al., 2023b), reinforcement-learning objectives (Fan et al., 2023; Black et al., 2024), and differentiation through the iterative sampling procedure; see Marion et al. (2025) and the references therein. The latter formulation views fine-tuning as an optimization problem whose inner variable is the distribution generated by an iterative sampler. Implicit Diffusion develops a single-loop method for this problem, avoiding the memory cost of differentiating through a complete sampling trajectory (Marion et al., 2025).

More recent approaches formulate reward fine-tuning and sampling from exponentially tilted distributions as stochastic optimal-control problems. Adjoint Matching turns this control problem into a memoryless regression objective for diffusion and flow models (Domingo-Enrich et al., 2025), while Adjoint Sampling extends this viewpoint to scalable amortized sampling from unnormalized densities (Havens et al., 2025). Subsequent works study reinforcement-based adjoint matching (Bergmeister et al., 2026), provide a unified view of score-matching, adjoint-based, and nonequilibrium objectives (Domingo-Enrich et al., 2026), and derive adjoint matching from the stochastic maximum principle (Domingo-Enrich & Han, 2026). Value-gradient guidance similarly learns the correction to a pretrained flow by matching it to the gradient of a value function (Liu et al., 2025).

Our setting differs in that the pretrained score network is kept fixed. Guidance is introduced at inference time through particle reweighting, resampling, and approximate mutation kernels. Consequently, our goal is not to analyze convergence of a parameter-optimization procedure, but to quantify the sampling error of the resulting guided particle system. The two perspectives are nevertheless related through their common tilted target: fine-tuning seeks to amortize the tilted distribution into new model parameters, whereas diffusion SMC approximates it directly at inference time.

Doob transforms and guided dynamics. For a reference Markov process and a positive terminal reward, the ideal guided process can be expressed through a Doob h-transform, where h is the conditional expectation of the remaining reward. In stochastic control formulations, a useful point of view to analyse SGM convergence bounds (Conforti et al., 2025b; Pham et al., 2025; Conforti et al., 2025a; Gentiloni Silveri & Ocello, 2025), $\nabla \log h$ determines the optimal drift correction; in our Feynman–Kac formulation, the analogous continuation function determines the forward-smoothing kernels and the propagation of local errors. Adjoint- and value-based methods learn or approximate this control field (Domingo-Enrich et al., 2025; Havens et al., 2025; Liu et al., 2025; Domingo-Enrich et al., 2026; Domingo-Enrich & Han, 2026), while Malliavin-based conditioning provides an alternative way to construct controls when terminal rewards are nonsmooth or singular (Pidstrigach et al., 2025). Diffusion SMC instead approximates the same change of measure through a population of weighted particles. Approximate potentials may therefore be interpreted as tractable surrogates for the ideal continuation function, while the coefficients $\Gamma_{k,N}$ in our analysis quantify the stability of the resulting approximate transform.

Feynman–Kac stability and perturbation theory. At the level of general Feynman–Kac theory, Whiteley (2012) establishes finite-particle stability and insensitivity to initialization for prescribed SMC models on noncompact state spaces. Related perturbation questions arise in Gloaguen et al. (2022), who analyze online smoothing when unnormalized transition densities are replaced by possibly biased estimators. In our setting, the perturbations have a specific generative origin: they arise from a learned score, numerical approximation of the reverse diffusion, and approximate guidance potentials. Our results describe how these errors are transported jointly by the guided Feynman–Kac flow.

B Notation and general Feynman–Kac framework

B.1 Notation summary

B.2 Measures, kernels, and weighted norms

Let $\mathcal{M}_+\left(\mathbb{R}^d\right)$ (resp. $\mathcal{P}\left(\mathbb{R}^d\right)$) denote the set of finite (positive) measures (resp. probability measures) on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We define $\mathcal{F}(\mathbb{R}^d)$ as the set of Borel measurable functions from $\mathbb{R}^d \to \mathbb{R}$. For every $(\mu, h) \in \mathcal{M}_+\left(\mathbb{R}^d\right) \times \mathcal{F}(\mathbb{R}^d)$ with $\mu[|h|] < \infty$ we define $\mu[h] := \int h(x)\mu(\mathrm{d}x) \in \mathbb{R}$. We also define, for a set of *particles* $\xi_{0:M-1} \in \mathbb{R}^{Md}$, the occupation (or empirical) measure supported on these particles as

$$\mu_{(\xi_{0:M-1})}\left(\mathrm{d}\mathbf{x}\right) := M^{-1} \sum_{i=0}^{M-1} \delta_{\xi_i}\left(\mathrm{d}\mathbf{x}\right) ,$$

where, for $\mathbf{x} \in \mathbb{R}^d$,

$$\delta_{\mathbf{x}} : \mathcal{B}(\mathbb{R}^d) \ni \mathrm{A} \to \begin{cases} 1 & \mathbf{x} \in \mathrm{A} , \\ 0 & \mathbf{x} \notin \mathrm{A} . \end{cases}$$

We call a function $\mathsf{K} : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}_+$ a finite kernel if, for each $\mathbf{x} \in \mathbb{R}^d$, the map $\mathrm{A} \mapsto \mathsf{K}(\mathbf{x}, \mathrm{A})$ is a measure on $\mathcal{B}(\mathbb{R}^d)$, and for each $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$, the map $\mathbf{x} \mapsto \mathsf{K}(\mathbf{x}, \mathrm{A})$ is a Borel measurable function from $\mathbb{R}^d$ to $[0, \infty)$. Such kernels induce an operator on real-valued measurable functions defined as:

$$\mathsf{K}h : \mathbb{R}^d \ni \mathbf{x} \mapsto \int h(y)\mathsf{K}(\mathbf{x}, \mathrm{d}y) ,$$

and an operator on measures $\mu \in \mathcal{M}_+\left(\mathbb{R}^d\right)$:

$$\mu\mathsf{K} : \mathcal{B}(\mathbb{R}^d) \ni \mathrm{A} \mapsto \int \mu(\mathrm{d}\mathbf{x})\mathsf{K}(\mathbf{x}, \mathrm{A}) .$$

A positive function $h \in \mathcal{F}(\mathbb{R}^d)$ may also act on kernels K by defining a h-weighted kernels

$$h\mathsf{K} : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \ni (\mathbf{x}, \mathrm{A}) \to h(\mathbf{x})\mathsf{K}(\mathbf{x}, \mathrm{A}) \in \mathbb{R}_+ .$$

Denote also by $\mathbb{1}$ the constant function on $\mathbb{R}^d$, which is equal to one. For any positive measurable function g and any probability measure ν such that $0 < \nu[g] < \infty$, we define, the tilting map by

$$\tau_g(\nu)[\psi] := \tfrac{\nu[g\psi]}{\nu[g]} ,$$

for every measurable test function ψ such that $\nu[|g\psi|] < \infty$. We also say that a function $\mu : \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}$ is a signed-measure if there exists μ_1, μ_2 in $\mathcal{M}_+\left(\mathbb{R}^d\right)$ such that $\mu = \mu_1 - \mu_2$. We define the set of signed measures as $\mathcal{M}_s\left(\mathbb{R}^d\right)$. For every $\mu \in \mathcal{M}_s\left(\mathbb{R}^d\right)$ we define the total variation measure $|\mu|$ as the smallest positive measure λ satisfying $\lambda(\mathrm{A}) \geq |\mu(\mathrm{A})|$ for all $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$. We do not prove the existence of such a measure and refer the reader to (Rudin, 1987, Chapter 6) for the proof. For any Borel measurable $V : \mathbb{R}^d \to [0, \infty)$, we define

$$\|\|_V : \mathcal{M}_s\left(\mathbb{R}^d\right) \ni \mu \to |\mu|(1 + V) = \int_{\mathbb{R}^d} (1 + V(\mathbf{x}))\, |\mu|(\mathrm{d}\mathbf{x}) \in [0, \infty] . \tag{19}$$

If we define the set $\mathcal{M}_V\left(\mathbb{R}^d\right) := \{\mu \in \mathcal{M}_s\left(\mathbb{R}^d\right) \mid \|\mu\| < \infty\}$, it is possible to show (Douc et al., 2018, Proposition D.3.3) that $\left(\mathcal{M}_V\left(\mathbb{R}^d\right), \|\|_V\right)$ is a Banach space.

Notation	Meaning
$\mathcal{T}^{\mathrm{EM}}$	Numerical Euler–Maruyama grid $\{s_n\}_{n=0}^{NJ}$, with $t_k = s_{kJ}$
h	Numerical time step, $h = T/(NJ)$
Δ_n	Integrated noise increment, $\Delta_n = \int_{s_n}^{s_{n+1}} \bar{\beta}_u \, \mathrm{d}u$
Q_k	Ideal mutation kernel from time t_{k-1} to time t_k
$\mathsf{Q}_k^{\theta,h}$	Approximate mutation kernel from time t_{k-1} to time t_k
g_k	Ideal potential at time t_k
$\hat{g}_k$	Implemented potential at time t_k
$\tilde{g}_k$	Normalized potential, $\tilde{g}_k = \hat{g}_k / \|\hat{g}_k\|_\infty$
τ_g	Tilting map by g, $\tau_g(\nu)[\psi] = \nu[g\psi]/\nu[g]$
K_k	Ideal unnormalized selection–mutation kernel, $\mathsf{K}_k = g_{k-1}\mathsf{Q}_k$
K_k^θ	Approximate unnormalized selection–mutation kernel, $\mathsf{K}_k^\theta = \hat{g}_{k-1}\mathsf{Q}_k^{\theta,h}$
$\mathbb{F}_k$	Ideal normalized Feynman–Kac transform, $\mathbb{F}_k(\nu) = \nu\mathsf{K}_k/(\nu\mathsf{K}_k\mathbb{1})$
$\mathbb{F}_k^\theta$	Approximate normalized Feynman–Kac transform, $\mathbb{F}_k^\theta(\nu) = \nu\mathsf{K}_k^\theta/(\nu\mathsf{K}_k^\theta\mathbb{1})$
ϕ_k	Ideal Feynman–Kac flow at time t_k
ϕ_k^θ	Approximate Feynman–Kac flow at time t_k
$\phi_k^{\theta,M}$	Empirical particle approximation of ϕ_k^θ
$\widetilde{\phi}_k^{\theta,M}$	Selected empirical measure after reweighting/resampling at time t_k
$\widetilde{\beta}_{k\|\ell}^\theta$	Normalized continuation function from time k to terminal time ℓ
$\mathsf{G}_{k\|\ell}^\theta$	Forward-smoothing kernel from time $k-1$ to time k, associated with terminal time ℓ
$\mathsf{G}_{k:\ell}^\theta$	Product of forward-smoothing kernels from time k to time ℓ
$V_{k\|N}^\theta$	Twisted Lyapunov function, $V/\widetilde{\beta}_{k\|N}^\theta$
$\Gamma_{\ell,N}$	Forward-smoothing stability coefficient from time ℓ to terminal time N
$\Lambda_{\ell,N}$	Normalization factor appearing in the deterministic kernel-bias bound
$\mathcal{B}_k^{\mathrm{bias}}$	Local discrepancy between the ideal and implemented Feynman–Kac updates initialized from the same ideal law ϕ_{k-1}
$\mathcal{B}_k^{\mathrm{SGM}}$	Local SGM approximation error due to score approximation and numerical discretization
$\mathcal{B}_k^{\mathrm{pot}}$	Local discrepancy due to mismatch between the ideal and implemented potentials
$\eta_\ell^{M,\theta}$	One-step predictive law before sampling at time ℓ in the particle system
$\mathcal{A}_{\ell,N}^{(p)}$	Monte Carlo normalization moment coefficient
$\mathcal{C}_\ell^{\mathrm{MC}}$	Local Monte Carlo fluctuation constant
$w_k^{(i)}$	Normalized particle weight at time t_k
$\mathbf{X}_k^{\theta,(i)}$	Particle i at time t_k
$\widetilde{\mathbf{X}}_k^{\theta,(i)}$	Selected particle i after resampling at time t_k
$\mathcal{F}_k^\theta$	Particle filtration up to time t_k
$\widetilde{\mathcal{F}}_k^\theta$	Filtration after selection/resampling at time t_k

Table 1: Summary of the main notation.

B.3 Feynman–Kac framework

We start by describing the Feynman–Kac framework underlying our analysis, and later introduce its specialization to conditional sampling with diffusion models.

Idealized Feynman–Kac model. Let $\mathbb{R}^d$ be the state space. We consider a sequence of positive potentials $(g_k)_{0\le k\le N-1}$ and Markov kernels $(\mathsf{Q}_k)_{1\le k\le N}$. A pair (g_{k-1}, Q_k) induces the following unnormalized predictive kernels

$$\mathsf{K}_k := g_{k-1}\mathsf{Q}_k\,, \qquad \mathsf{K}_k\psi(\mathbf{x}) = g_{k-1}(\mathbf{x})\mathsf{Q}_k\psi(\mathbf{x})\,. \tag{20}$$

The corresponding normalized Feynman–Kac transformation is the nonlinear map

$$\mathbb{F}_k : \mathcal{M}_+\left(\mathbb{R}^d\right) \to \mathcal{P}\left(\mathbb{R}^d\right), \qquad \mathbb{F}_k(\nu) := \frac{\nu \mathsf{K}_k}{\nu \mathsf{K}_k \mathbb{1}}, \tag{21}$$

whenever $0 < \nu \mathsf{K}_k \mathbb{1} < \infty$. Equivalently, for every bounded measurable test function ψ,

$$\mathbb{F}_k(\nu)[\psi] = \frac{\nu[g_{k-1} \mathsf{Q}_k \psi]}{\nu[g_{k-1}]} = \left(\tau_{g_{k-1}}(\nu) \mathsf{Q}_k\right)[\psi] .$$

With this map, we define the selection-*before*-propagation Feynman–Kac flow by the recursion

$$\phi_k = \mathbb{F}_k(\phi_{k-1}), \qquad k = 1, \ldots, N .$$

Approximate Feynman–Kac model. We consider a sequence of strictly positive measurable approximate potentials $(\hat{g}_k)_{0 \le k \le N-1}$ and approximate Markov kernels $(\mathsf{Q}_k^{\theta,h})_{1 \le k \le N}$. For $1 \le k \le N$, define the approximate unnormalized selection–mutation kernel

$$\mathsf{K}_k^\theta := \hat{g}_{k-1} \mathsf{Q}_k^{\theta,h}, \qquad \mathsf{K}_k^\theta \psi(\mathbf{x}) = \hat{g}_{k-1}(\mathbf{x}) \mathsf{Q}_k^{\theta,h} \psi(\mathbf{x}) . \tag{22}$$

The corresponding normalized Feynman–Kac transformation is

$$\mathbb{F}_k^\theta(\nu) := \frac{\nu \mathsf{K}_k^\theta}{\nu \mathsf{K}_k^\theta \mathbb{1}}, \tag{23}$$

whenever $0 < \nu \mathsf{K}_k^\theta \mathbb{1} < \infty$. Equivalently, for every bounded measurable test function ψ,

$$\mathbb{F}_k^\theta(\nu)[\psi] = \frac{\nu[\hat{g}_{k-1} \mathsf{Q}_k^{\theta,h} \psi]}{\nu[\hat{g}_{k-1}]} = \left(\tau_{\hat{g}_{k-1}}(\nu) \mathsf{Q}_k^{\theta,h}\right)[\psi] .$$

Starting from an approximate initial distribution ϕ_0^θ, we define the approximate Feynman–Kac flow recursively by

$$\phi_k^\theta = \mathbb{F}_k^\theta(\phi_{k-1}^\theta), \qquad k = 1, \ldots, N .$$

Particle-based approximation. Although the approximate flow $(\phi_k^\theta)_{0 \le k \le N}$ is well defined recursively, it is generally not available in closed form. SMC approximates it by empirical measures carried by M particles. Suppose that, at time k, we have particles $\mathbf{X}_k^{\theta,(1)}, \ldots, \mathbf{X}_k^{\theta,(M)}$ and the empirical measure

$$\phi_k^{\theta,M} := \frac{1}{M} \sum_{i=1}^M \delta_{\mathbf{X}_k^{\theta,(i)}} .$$

Applying the approximate Feynman–Kac transformation $\mathbb{F}_{k+1}^\theta$ to $\phi_k^{\theta,M}$ gives the finite mixture

$$\mathbb{F}_{k+1}^\theta(\phi_k^{\theta,M})(\mathrm{d}\mathbf{x}) = \sum_{i=1}^M w_k^{(i)} \mathsf{Q}_{k+1}^{\theta,h}(\mathbf{X}_k^{\theta,(i)}, \mathrm{d}\mathbf{x}),$$

where

$$w_k^{(i)} := \frac{\hat{g}_k(\mathbf{X}_k^{\theta,(i)})}{\sum_{j=1}^M \hat{g}_k(\mathbf{X}_k^{\theta,(j)})} .$$

Equivalently, particles are first selected according to the weighted empirical measure

$$\tau_{\hat{g}_k}(\phi_k^{\theta,M}) = \sum_{i=1}^M w_k^{(i)} \delta_{\mathbf{X}_k^{\theta,(i)}},$$

and then propagated through the approximate mutation kernel $\mathsf{Q}_{k+1}^{\theta,h}$. Conditionally on the current particle system, sampling M particles independently from the resulting mixture produces the empirical approximation $\phi_{k+1}^{\theta,M}$.

B.4 Correspondence with existing methods: noiseless inpainting

We illustrate the correspondence of Algorithm 1 with MCGdiff (Cardoso et al., 2024) and SMCDiff (Trippe et al., 2023). For clarity, we consider their common specialization to noiseless inpainting. Write $\mathbf{x} = (\mathbf{x}_{\mathrm{o}}, \mathbf{x}_{\mathrm{u}})$ for the observed and unobserved coordinates, respectively, and suppose that $\mathbf{x}_{\mathrm{o}} = y$.

(i) **MCGdiff** defines for noiseless inpainting, $h_k(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}_{\mathrm{o}}; \mathrm{m}_{T-t_k|0}\, y, \sigma^2_{T-t_k|0}\mathbf{I}_{d_{\mathrm{y}}}\right)$, with

$$\hat{g}_k(\mathbf{x}) = \frac{\int h_{k+1}(z)\, \mathrm{Q}^{\theta}_{t_{k+1}|t_k}(\mathbf{x}, \mathrm{d}z)}{h_k(\mathbf{x})}\,, \qquad \mathrm{Q}^{\theta}_{k+1}(\mathbf{x}, \mathrm{d}z) = \frac{h_{k+1}(z)\, \mathrm{Q}^{\theta}_{t_{k+1}|t_k}(\mathbf{x}, \mathrm{d}z)}{\int h_{k+1}(u)\, \mathrm{Q}^{\theta}_{t_{k+1}|t_k}(\mathbf{x}, \mathrm{d}u)}\,.$$

(ii) **SMCDiff** generates a forward-diffused trajectory $(y_k)_{k=0}^N$ of the observed coordinates, with $y_N = y$. Conditional on this trajectory, define $\mathbf{u}_k := (y_k, \mathbf{x}_{\mathrm{u}})$ and $\mathbf{v}_{k+1} := (y_{k+1}, z_{\mathrm{u}})$. Then

$$\hat{g}_k(\mathbf{x}_{\mathrm{u}}) = \int \mathrm{Q}^{\theta}_{t_{k+1}|t_k}\left(\mathbf{u}_k, \mathbf{v}_{k+1}\right)\, \mathrm{d}z_{\mathrm{u}}\,, \qquad \mathrm{Q}^{\theta}_{k+1}(\mathbf{x}_{\mathrm{u}}, \mathrm{d}z_{\mathrm{u}}) = \int_{\mathbb{R}^{d_{\mathrm{y}}}} \mathrm{Q}^{\theta}_{t_{k+1}|t_k}\left(\mathbf{u}_k, \mathrm{d}z_{\mathrm{o}}\, \mathrm{d}z_{\mathrm{u}}\right)\,.$$

C Forward-smoothing kernels: properties and forgetting

C.1 Kernel notation and tail factorization

In the main paper, we introduced the continuation function associated with the approximate Feynman–Kac model for terminal time N through its probabilistic representation in (14). Here, we extend the notation to an arbitrary terminal index $0 \le k \le \ell \le N$ and give its kernel representation. For $0 \le k < \ell \le N$, the normalized continuation function is

$$\widetilde{\beta}^{\theta}_{k|\ell} := \frac{\mathsf{K}^{\theta}_{k+1} \cdots \mathsf{K}^{\theta}_{\ell} \mathbb{1}}{\prod_{j=k}^{\ell-1} \|\hat{g}_j\|_\infty}\,, \qquad \widetilde{\beta}^{\theta}_{\ell|\ell} := \mathbb{1}\,. \tag{24}$$

Since $\mathsf{K}^{\theta}_{k+1} = \hat{g}_k \mathsf{Q}^{\theta,h}_{k+1}$, the continuation functions satisfy

$$\widetilde{\beta}^{\theta}_{k|\ell} = \frac{\hat{g}_k}{\|\hat{g}_k\|_\infty} \mathsf{Q}^{\theta,h}_{k+1} \widetilde{\beta}^{\theta}_{k+1|\ell}\,, \qquad 0 \le k < \ell \le N\,. \tag{25}$$

We define *forward smoothing kernels*, obtained by tilting the proposal kernel with the future continuation weight. For $1 \le k \le \ell \le N$, define

$$\mathsf{G}^{\theta}_{k|\ell}(\mathbf{x}_{k-1}, \mathrm{d}\mathbf{x}_k) := \frac{\mathsf{Q}^{\theta,h}_k(\mathbf{x}_{k-1}, \mathrm{d}\mathbf{x}_k)\, \widetilde{\beta}^{\theta}_{k|\ell}(\mathbf{x}_k)}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|\ell}(\mathbf{x}_{k-1})}\,. \tag{26}$$

The kernel $\mathsf{G}^{\theta}_{k|\ell}$ describes the transition from time $k-1$ to time k after accounting for the remaining Feynman–Kac weights up to terminal time ℓ. We write

$$\mathsf{G}^{\theta}_{k:\ell} := \mathsf{G}^{\theta}_{k+1|\ell} \cdots \mathsf{G}^{\theta}_{\ell|\ell}\,, \qquad \mathsf{G}^{\theta}_{\ell:\ell} := \mathrm{Id}\,,$$

so that $\mathsf{G}^{\theta}_{k:\ell}$ propagates a perturbation from time k to time ℓ under the guided dynamics.

The above kernel notation coincides with the probabilistic representation used in the main text in (15). Indeed, for $k < r \le \ell$, by (25),

$$\mathsf{Q}^{\theta,h}_r \widetilde{\beta}^{\theta}_{r|\ell}(\mathbf{x}_{r-1}) = \frac{\widetilde{\beta}^{\theta}_{r-1|\ell}(\mathbf{x}_{r-1})}{\hat{g}_{r-1}(\mathbf{x}_{r-1}) / \|\hat{g}_{r-1}\|_\infty}\,.$$

Hence,

$$\mathsf{G}^{\theta}_{r|\ell}(\mathbf{x}_{r-1}, \mathrm{d}\mathbf{x}_r) = \mathsf{Q}^{\theta,h}_r(\mathbf{x}_{r-1}, \mathrm{d}\mathbf{x}_r)\, \frac{\hat{g}_{r-1}(\mathbf{x}_{r-1})}{\|\hat{g}_{r-1}\|_\infty} \frac{\widetilde{\beta}^{\theta}_{r|\ell}(\mathbf{x}_r)}{\widetilde{\beta}^{\theta}_{r-1|\ell}(\mathbf{x}_{r-1})}\,.$$

Therefore, multiplying the forward-smoothing kernels $\mathsf{G}^\theta_{k+1|\ell}, \dots, \mathsf{G}^\theta_{\ell|\ell}$ causes the continuation functions to telescope. Consequently, for every bounded measurable ψ,

$$\mathsf{G}^\theta_{k:\ell}\psi(\mathbf{x}) = \frac{\mathsf{K}^\theta_{k+1}\cdots\mathsf{K}^\theta_{\ell}\psi(\mathbf{x})}{\mathsf{K}^\theta_{k+1}\cdots\mathsf{K}^\theta_{\ell}\mathbb{1}(\mathbf{x})}\,, \qquad 0 \le k < \ell \le N\,. \tag{27}$$

Thus, $\mathsf{G}^\theta_{k:\ell}$ is the normalized tail Feynman–Kac kernel from time k to time ℓ. Taking $\ell = N$ and using the Markov-chain representation of the above recovers the probabilistic representation used in the main text in (15).

Proposition C.1 (Tail factorization)**.** *Let $0 \le k \le m \le N$. For every probability measure $\nu \in \mathcal{P}(\mathbb{R}^d)$ such that $0 < \nu(\widetilde{\beta}^\theta_{k|m}) < \infty$, with $\widetilde{\beta}^\theta_{k|m}$ defined as in* (14)*, we have*

$$\left(\mathbb{F}^\theta_m \circ \cdots \circ \mathbb{F}^\theta_{k+1}\right)(\nu) = \nu^\theta_{k|m}\mathsf{G}^\theta_{k+1|m}\cdots\mathsf{G}^\theta_{m|m}\,, \qquad \textit{with } \nu^\theta_{k|m} := \frac{\widetilde{\beta}^\theta_{k|m}\nu}{\nu[\widetilde{\beta}^\theta_{k|m}]}\,. \tag{28}$$

Proof. For $0 \le k < m \le N$, set

$$C_{k:m} := \prod_{i=k}^{m-1} \|\hat{g}_i\|_\infty\,.$$

Recall that

$$\left(\mathbb{F}^\theta_m \circ \cdots \circ \mathbb{F}^\theta_{k+1}\right)(\nu) = \frac{\nu\mathsf{K}^\theta_{k+1}\cdots\mathsf{K}^\theta_{m}}{\nu\mathsf{K}^\theta_{k+1}\cdots\mathsf{K}^\theta_{m}\mathbb{1}}\,,$$

with the unnormalized kernels K^θ_ℓ, for $k+1 \le \ell \le m$, as in (20). We first show by induction on $m-k$ that, for every bounded measurable test function ψ,

$$\mathsf{K}^\theta_{k+1}\cdots\mathsf{K}^\theta_m\psi = C_{k:m}\widetilde{\beta}^\theta_{k|m}\mathsf{G}^\theta_{k+1|m}\cdots\mathsf{G}^\theta_{m|m}\psi\,. \tag{29}$$

First note that when $m = k+1$,

$$\widetilde{\beta}^\theta_{k|k+1}(\mathbf{x}_k) = \tilde{g}_k(\mathbf{x}_k)\,, \qquad \mathsf{G}^\theta_{k+1|k+1} = \mathsf{Q}^{\theta,h}_{k+1}\,,$$

and therefore,

$$\mathsf{K}^\theta_{k+1}\psi(\mathbf{x}_k) = \hat{g}_k(\mathbf{x}_k)\mathsf{Q}^{\theta,h}_{k+1}\psi(\mathbf{x}_k) = \|\hat{g}_k\|_\infty\,\widetilde{\beta}^\theta_{k|k+1}(\mathbf{x}_k)\mathsf{G}^\theta_{k+1|k+1}\psi(\mathbf{x}_k)\,.$$

Then, assume that (29) holds for the pair $(k+1, m)$, with $k+1 < m$, namely

$$\mathsf{K}^\theta_{k+2}\cdots\mathsf{K}^\theta_m\psi = C_{k+1:m}\widetilde{\beta}^\theta_{k+1|m}\mathsf{G}^\theta_{k+2|m}\cdots\mathsf{G}^\theta_{m|m}\psi\,.$$

Then, for any bounded measurable ψ,

$$\begin{aligned}\mathsf{K}^\theta_{k+1}\mathsf{K}^\theta_{k+2}\cdots\mathsf{K}^\theta_m\psi(\mathbf{x}_k) &= C_{k+1:m}\,\hat{g}_k(\mathbf{x}_k)\mathsf{Q}^{\theta,h}_{k+1}\left(\widetilde{\beta}^\theta_{k+1|m}\mathsf{G}^\theta_{k+2|m}\cdots\mathsf{G}^\theta_{m|m}\psi\right)(\mathbf{x}_k)\\ &= C_{k:m}\,\tilde{g}_k(\mathbf{x}_k)\mathsf{Q}^{\theta,h}_{k+1}\left(\widetilde{\beta}^\theta_{k+1|m}\mathsf{G}^\theta_{k+2|m}\cdots\mathsf{G}^\theta_{m|m}\psi\right)(\mathbf{x}_k)\,.\end{aligned}$$

By definition, the normalized continuation functions satisfy the recursion

$$\widetilde{\beta}^\theta_{k|m}(\mathbf{x}_k) = \tilde{g}_k(\mathbf{x}_k)\mathsf{Q}^{\theta,h}_{k+1}\widetilde{\beta}^\theta_{k+1|m}(\mathbf{x}_k)\,,$$

so we obtain

$$\mathsf{K}^\theta_{k+1}\cdots\mathsf{K}^\theta_m\psi(\mathbf{x}_k) = C_{k:m}\widetilde{\beta}^\theta_{k|m}(\mathbf{x}_k)\frac{\mathsf{Q}^{\theta,h}_{k+1}\left(\widetilde{\beta}^\theta_{k+1|m}\mathsf{G}^\theta_{k+2|m}\cdots\mathsf{G}^\theta_{m|m}\psi\right)(\mathbf{x}_k)}{\mathsf{Q}^{\theta,h}_{k+1}\widetilde{\beta}^\theta_{k+1|m}(\mathbf{x}_k)}\,.$$

By definition of the forward smoothing kernel (26),

$$\begin{aligned}
&\frac{\mathsf{Q}_{k+1}^{\theta,h}\left(\widetilde{\beta}_{k+1|m}^{\theta}\mathsf{G}_{k+2|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi\right)(\mathbf{x}_k)}{\mathsf{Q}_{k+1}^{\theta,h}\widetilde{\beta}_{k+1|m}^{\theta}(\mathbf{x}_k)}\\
&=\frac{\int\widetilde{\beta}_{k+1|m}^{\theta}(\mathbf{x}_{k+1})\left(\mathsf{G}_{k+2|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi\right)(\mathbf{x}_{k+1})\mathsf{Q}_{k+1}^{\theta,h}(\mathbf{x}_k,\mathrm{d}\mathbf{x}_{k+1})}{\mathsf{Q}_{k+1}^{\theta,h}\widetilde{\beta}_{k+1|m}^{\theta}(\mathbf{x}_k)}\\
&=\int\left(\mathsf{G}_{k+2|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi\right)(\mathbf{x}_{k+1})\mathsf{G}_{k+1|m}^{\theta}(\mathbf{x}_k,\mathrm{d}\mathbf{x}_{k+1})\\
&=\mathsf{G}_{k+1|m}^{\theta}\mathsf{G}_{k+2|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi(\mathbf{x}_k)\,.
\end{aligned}$$

Therefore,

$$\mathsf{K}_{k+1}^{\theta}\cdots\mathsf{K}_{m}^{\theta}\psi(\mathbf{x}_k)=C_{k:m}\widetilde{\beta}_{k|m}^{\theta}(\mathbf{x}_k)\mathsf{G}_{k+1|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi(\mathbf{x}_k)\,,$$

which proves (29). Finally, applying (29) with ψ and with $\mathbb{1}$, and using that each $\mathsf{G}_{\ell|m}^{\theta}$ is a Markov kernel, for $k+1\le\ell\le m$, so that

$$\mathsf{G}_{k+1|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\mathbb{1}=\mathbb{1}\,,$$

we get

$$\begin{aligned}
\left(\mathbb{F}_{m}^{\theta}\circ\cdots\circ\mathbb{F}_{k+1}^{\theta}\right)(\nu)[\psi]&=\frac{\nu\mathsf{K}_{k+1}^{\theta}\cdots\mathsf{K}_{m}^{\theta}\psi}{\nu\mathsf{K}_{k+1}^{\theta}\cdots\mathsf{K}_{m}^{\theta}\mathbb{1}}\\
&=\frac{C_{k:m}\nu\left[\widetilde{\beta}_{k|m}^{\theta}\mathsf{G}_{k+1|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi\right]}{C_{k:m}\nu\left[\widetilde{\beta}_{k|m}^{\theta}\right]}\\
&=\frac{\nu\left[\widetilde{\beta}_{k|m}^{\theta}\,\mathsf{G}_{k+1|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\psi\right]}{\nu\left[\widetilde{\beta}_{k|m}^{\theta}\right]}\\
&=\left(\nu_{k|m}^{\theta}\mathsf{G}_{k+1|m}^{\theta}\cdots\mathsf{G}_{m|m}^{\theta}\right)[\psi]\,.
\end{aligned}$$

which holds for every bounded measurable ψ, and therefore proves (28). □

Remark C.2. This factorization is purely algebraic and is not specific to the approximate Feynman–Kac model. It holds for any sequence of Markov kernels and their associated continuation functions and forward-smoothing kernels.

Corollary C.3. *For every $0\le k\le N$, we have*

$$\phi_N^{\theta}=\tau_{\widetilde{\beta}_{k|N}^{\theta}}(\phi_k^{\theta})\mathsf{G}_{k:N}^{\theta}\,. \tag{30}$$

Proof. The case $k=N$ is immediate. For $k<N$, by the recursive definition of the approximate Feynman–Kac flow,

$$\phi_N^{\theta}=\left(\mathbb{F}_{N}^{\theta}\circ\cdots\circ\mathbb{F}_{k+1}^{\theta}\right)(\phi_k^{\theta})\,.$$

Applying Proposition C.1 with $m=N$ and $\nu=\phi_k^{\theta}$ yields

$$\phi_N^{\theta}=\nu_{k|N}^{\theta}\mathsf{G}_{k:N}^{\theta}\,.$$

Since

$$\nu_{k|N}^{\theta}=\tau_{\widetilde{\beta}_{k|N}^{\theta}}(\phi_k^{\theta})\,,$$

the result follows. □

C.2 Proof of Proposition 3.2

We keep the SMC selection grid $\mathcal{T}^{\mathrm{sel}} = (t_k)_{k=0}^N$ fixed and use the possibly varying numerical grid $s_n = nh$, where $h = T/(NJ)$ and $t_k = s_{kJ}$. Let $V(\mathbf{x}) := \|\mathbf{x}\|^2$. For $n = 0, \dots, NJ-1$, define

$$\Delta_n := \int_{s_n}^{s_{n+1}} \bar{\beta}_u \, \mathrm{d}u \, .$$

Note that the total noise variance over a SMC step is

$$\int_{t_k}^{t_{k+1}} \bar{\beta}_u \, \mathrm{d}u = \sum_{n=kJ}^{(k+1)J-1} \Delta_n \, .$$

Denote the Euler kernel associated with the numerical scheme by

$$\mathsf{R}_n^\theta(\mathbf{x}, \cdot) := \mathcal{N}\left(\mathbf{x} + \Delta_n b_n^\theta(\mathbf{x}), 2\Delta_n \mathbf{I}_d\right) \, , \quad \text{with} \quad b_n^\theta(\mathbf{x}) := \alpha\mathbf{x} + 2\, \mathrm{s}_\theta\left(\mathbf{x}, T - s_n\right) \, .$$

The implemented SMC mutation is the composition kernel

$$\mathsf{Q}_{k+1}^{\theta,h} = \mathsf{R}_{kJ}^\theta \cdots \mathsf{R}_{(k+1)J-1}^\theta \, .$$

Lyapunov drift condition. For $n = kJ, \dots, (k+1)J-1$, expanding the time-changed Euler step gives

$$\mathsf{R}_n^\theta V(\mathbf{x}) = V(\mathbf{x}) + 2\Delta_n \left\langle \mathbf{x}, \alpha\mathbf{x} + 2\, \mathrm{s}_\theta\left(\mathbf{x}, T - s_n\right) \right\rangle + \Delta_n^2 \left\| \alpha\mathbf{x} + 2\, \mathrm{s}_\theta\left(\mathbf{x}, T - s_n\right) \right\|^2 + 2d\Delta_n \, .$$

By Assumption 3.1,

$$\left\langle \mathbf{x}, \alpha\mathbf{x} + 2\, \mathrm{s}_\theta\left(\mathbf{x}, T - s_n\right) \right\rangle \leq (\alpha - 2\gamma_k) V(\mathbf{x}) + 2\kappa_k$$

and

$$\left\| \alpha\mathbf{x} + 2\, \mathrm{s}_\theta\left(\mathbf{x}, T - s_n\right) \right\|^2 \leq (\alpha^2 - 4\alpha\gamma_k + 4L_k) V(\mathbf{x}) + 4\alpha\kappa_k + 4L_k \, .$$

Hence, using (11),

$$\mathsf{R}_n^\theta V \leq \left[1 - (2\gamma_k - \alpha)\Delta_n\right] V + C_k \Delta_n \, ,$$

with

$$C_k := 4\kappa_k + 2d + (4\alpha\kappa_k + 4L_k) \frac{2\gamma_k - \alpha}{\alpha^2 - 4\alpha\gamma_k + 4L_k} \, . \tag{31}$$

Iterating this inequality over the J numerical steps in $[t_k, t_{k+1}]$ yields

$$\mathsf{Q}_{k+1}^{\theta,h} V \leq \prod_{n=kJ}^{(k+1)J-1} \left[1 - (2\gamma_k - \alpha)\Delta_n\right] V + C_k \sum_{n=kJ}^{(k+1)J-1} \Delta_n \prod_{m=n+1}^{(k+1)J-1} \left[1 - (2\gamma_k - \alpha)\Delta_m\right] \, .$$

Hence,

$$\mathsf{Q}_{k+1}^{\theta,h} V \leq \prod_{n=kJ}^{(k+1)J-1} \left[1 - (2\gamma_k - \alpha)\Delta_n\right] V + C_k \sum_{n=kJ}^{(k+1)J-1} \Delta_n \, .$$

Using $1 - x \leq e^{-x}$ and $\sum_{n=kJ}^{(k+1)J-1} \Delta_n = \int_{t_k}^{t_{k+1}} \bar{\beta}_u \, \mathrm{d}u$, we obtain

$$\mathsf{Q}_{k+1}^{\theta,h} V \leq \exp\left(-(2\gamma_k - \alpha) \int_{t_k}^{t_{k+1}} \bar{\beta}_u \, \mathrm{d}u\right) V + C_k \int_{t_k}^{t_{k+1}} \bar{\beta}_u \, \mathrm{d}u \, .$$

This proves (12), with constants independent of the numerical refinement.

Minorization. Recall that

$$\mathsf{R}_n^\theta(\mathbf{x},\cdot) = \mathcal{N}\left(\mathbf{x} + \Delta_n b_n^\theta(\mathbf{x}), 2\Delta_n \mathbf{I}_d\right) , \qquad b_n^\theta(\mathbf{x}) = \alpha\mathbf{x} + 2\,\mathrm{s}_\theta\left(\mathbf{x}, T - s_n\right) ,$$

and denotes its associated transition density by

$$r_n^\theta(z_{n+1} \mid z_n) = \frac{1}{(4\pi\Delta_n)^{d/2}} \exp\left\{ -\frac{\|z_{n+1} - z_n - \Delta_n b_n^\theta(z_n)\|^2}{4\Delta_n} \right\} .$$

Expanding the square gives

$$\|z_{n+1} - z_n - \Delta_n b_n^\theta(z_n)\|^2 = \|z_{n+1} - z_n\|^2 - 2\Delta_n \left\langle b_n^\theta(z_n), z_{n+1} - z_n \right\rangle + \Delta_n^2 \|b_n^\theta(z_n)\|^2 ,$$

and therefore

$$r_n^\theta(z_{n+1} \mid z_n) = \mathcal{N}\left(z_{n+1}; z_n, 2\Delta_n \mathbf{I}_d\right) \exp\left\{ \frac{1}{2}\left\langle b_n^\theta(z_n), z_{n+1} - z_n \right\rangle - \frac{\Delta_n}{4}\|b_n^\theta(z_n)\|^2 \right\} . \quad (32)$$

Let $q_{k+1}^{\theta,h}(y \mid \mathbf{x})$ denote the density of the composite kernel $\mathsf{Q}_{k+1}^{\theta,h}$. Setting the endpoints of the numerical path over $[t_k, t_{k+1}]$ to $z_{kJ} = \mathbf{x}$ and $z_{(k+1)J} = y$, composing the transition kernels from the time-changed Euler scheme (3) and (32) give

$$\begin{aligned}
&q_{k+1}^{\theta,h}(y \mid \mathbf{x}) \\
&= \int \prod_{n=kJ}^{(k+1)J-1} r_n^\theta(z_{n+1} \mid z_n) \mathrm{d}z_{kJ+1} \ldots \mathrm{d}z_{(k+1)J-1} \\
&= \int q_{1:J-1}(z_{kJ+1:(k+1)J-1} | x, y) \mathcal{N}\left(y; x, 2\Delta_{kJ:(k+1)J}\mathbf{I}_d\right) \mathrm{e}^{\mathcal{L}_k} \mathrm{d}z_{kJ+1} \ldots \mathrm{d}z_{(k+1)J-1}
\end{aligned}$$

where

$$\mathcal{L}_k := \sum_{n=kJ}^{(k+1)J-1} \left[\frac{1}{2}\left\langle b_n^\theta(z_n), z_{n+1} - z_n \right\rangle - \frac{\Delta_n}{4}\|b_n^\theta(z_n)\|^2 \right] ,$$

and

$$q_{1:J-1}(z_{1:J-1}|x,y) := \frac{\mathcal{N}\left(z_1; x, 2\Delta_{kJ}\mathbf{I}_d\right) \prod_{n=1}^{J-2} \mathcal{N}\left(z_{n+1}; z_n, 2\Delta_{kJ+n}\mathbf{I}_d\right) \mathcal{N}\left(y; z_{kJ-1}, 2\Delta_{kJ}\mathbf{I}_d\right)}{\mathcal{N}\left(y; z_{kJ-1}, 2\Delta_{kJ}\mathbf{I}_d\right)}$$

By Bayes law, we have that

$$q_{1:J-1}(z_{1:J-1}|x,y) := q_{J-1|J,0}(z_{J-1}|y,x) \prod_{n=1}^{J-2} q_{n|n+1,0}(z_n|z_{n+1},x) \quad (33)$$

where

$$q_{n|n+1,0}(z_n|z_{n+1},z_0) := \mathcal{N}\left(z_n; z_0 + \frac{u_n}{u_{n+1}}(z_{n+1} - z_0), 2\frac{u_n}{u_{n+1}}(u_{n+1} - u_n)\mathbf{I} \right) ,$$

where $u_n := \sum_{m=0}^{n-1} \Delta_{kJ+m}$. The law of appearing above, coincides with the discrete Markov Chain satisfying the following recursion

$$Z_{kJ} := y \,, \text{ For } n \in \{J-1, \cdots, 1\} \,, \; Z_n \sim q_{n|n+1,0}(\cdot|Z_{n+1}, x) \,,$$

where $q_{n|n+1,0}(\cdot|z_n, x)$ is the probability law $\mathrm{A} \to \int 1_{\mathrm{A}}(z)\, q_{n|n+1,0}(z|z_n,x)\mathrm{d}z$.

As,

$$q_{k+1}^{\theta,h}(y \mid \mathbf{x}) = \mathcal{N}\left(y; \mathbf{x}, 2u_J\mathbf{I}_d\right) \mathbb{E}\left[\exp\left\{\mathcal{L}_k\left(Z_1, \cdots Z_{J-1}\right)\right\}\right] ,$$

by Jensen's inequality,

$$q_{k+1}^{\theta,h}(y \mid \mathbf{x}) \geq \mathcal{N}\left(y; \mathbf{x}, 2u_J\mathbf{I}_d\right) \exp\left[\mathbb{E}\left\{\mathcal{L}_k\left(Z_1, \cdots Z_{J-1}\right)\right\}\right]$$

It therefore remains to lower bound

$$\exp\left[\mathbb{E}\left\{\mathcal{L}_k\left(Z_1,\cdots Z_{J-1}\right)\right\}\right] = \frac{1}{2}\sum_{n=1}^{J-1}\mathbb{E}\left[\left\langle b^\theta_{n+kJ}(Z_n), Z_{n+1}-Z_n\right\rangle\right] - \frac{1}{4}\sum_{n=1}^{J-1}\Delta_n\mathbb{E}\left[\|b^\theta_{n+kJ}(Z_n)\|^2\right] .$$

We start by controling the quadratic term. Since $b^\theta_n(z) = \alpha z + 2\,\mathrm{s}_\theta\left(z, T-s_n\right)$, the linear-growth condition in Assumption 3.1 implies

$$\|b^\theta_n(z)\|^2 \leq 8L_k + (2\alpha^2+8L_k)(\|z\|^2) \leq C_k(1+\|z\|^2)\,,$$

with $2\alpha^2+8L_k$. Using (33)

$$Z_n \sim \mathcal{N}\left(\frac{u_J-u_n}{u_J}\mathbf{x}+\frac{u_n}{u_J}y, 2\frac{u_n}{u_J}(u_J-u_n)\mathbf{I}_d\right) . \tag{34}$$

Hence, using that if $u \leq a$ then $u(a-u) \leq a^2/4$, for $V(\mathbf{x}) \leq R$ and $\|y\| \leq r$ we have that

$$\mathbb{E}\left[\|Z_n\|^2\right] = \left\|\frac{u_J-u_n}{u_J}\mathbf{x}+\frac{u_n}{u_J}y\right\|^2 + d\frac{2u_n(u_J-u_n)}{u_J} \leq (\sqrt{R}+r)^2 + \frac{du_J}{2}\,.$$

The right-hand side is uniform in J. Therefore there exists M, independent of the numerical refinement, such that

$$\mathbb{E}\left[\|b^\theta_n(Z_n)\|^2\right] \leq M\,.$$

Hence,

$$-\frac{1}{4}\sum_{n=1}^{J-1}\Delta_n\mathbb{E}\left[\|b^\theta_{kJ+n}(Z_n)\|^2\right] \geq -\frac{M}{4}u_J = -\frac{M}{4}\int_{t_k}^{t_{k+1}}\bar{\beta}_u\,\mathrm{d}u\,.$$

For the other term, note that

$$\frac{1}{2}\sum_{n=1}^{J-1}\mathbb{E}\left[\left\langle b^\theta_{kJ+n}(Z_n), Z_{n+1}-Z_n\right\rangle\right] \geq -\frac{1}{2}\sum_{n=1}^{J-1}\mathbb{E}\left[\left\|b^\theta_{kJ+n}(Z_n)\right\|^2\right]^{1/2}\mathbb{E}\left[\|Z_{n+1}-Z_n\|^2\right]^{1/2} . \tag{35}$$

For the term $\mathbb{E}\left[\|Z_{n+1}-Z_n\|^2\right]^{1/2}$, we have that

$$Z_n - Z_{n+1} = x\left(1-\frac{u_n}{u_{n+1}}\right) - Z_n(1-\frac{u_n}{u_{n+1}}) + \left(2\frac{u_n}{u_{n+1}}(u_{n+1}-u_n)\right)^{1/2}\xi_{n-1}\,,$$

where $\xi \sim \mathcal{N}\left(0,\mathbf{I}\right)$ and is independent of Z_n. But by (34)

$$\mathbb{E}\left[Z_{n+1}\right] = x + \frac{u_{n+1}}{u_J}\left(y-x\right)\,,$$

and then $\mathbb{E}\left[Z_n - Z_{n+1}\right] = \left(x-y\right)\left(u_{n+1}/u_J\right)$. For the variance term, we have that

$$\mathbb{V}\left[Z_n - Z_{n+1}\right] = 2(u_{n+1}-u_n)\left(1-\frac{u_n}{u_J}\right)\mathbf{I}\,.$$

Therefore, we have that

$$\mathbb{E}\left[\|Z_n - Z_{n+1}\|^2\right] = \|x-y\|^2\left(\frac{u_{n+1}-u_n}{u_J}\right)^2 + 2(u_{n+1}-u_n)\left(1-\frac{(u_{n+1}-u_n)}{u_J}\right)d\,,$$

and finally

$$\sum_{n=1}^{J-1}\mathbb{E}\left[\|Z_n - Z_{n+1}\|^2\right] = \|x-y\|^2\frac{\sum_{n=1}^{J-1}\Delta_n^2}{u_J^2} + 2\frac{d}{u_J}\left(u_J^2 - \sum_{n=1}^{J-1}\Delta_n^2\right) \leq \|x-y\|^2 + 2du_J\,.$$

Combining the two yields

$$\exp\left[\mathbb{E}\left\{\mathcal{L}_k\left(Z_1, \cdots Z_{J-1}\right)\right\}\right] \geq M' .$$

Therefore,

$$q_{k+1}^{\theta,h}(y \mid \mathbf{x}) \geq \mathcal{N}\left(y; \mathbf{x}, 2a_k \mathbf{I}_d\right) \mathrm{e}^{-M'} .$$

For $V(\mathbf{x}) \leq R$ and $\|y\| \leq r$,

$$\mathcal{N}\left(y; \mathbf{x}, 2a_k \mathbf{I}_d\right) \geq (4\pi a_k)^{-d/2} \exp\left\{-\frac{(\sqrt{R}+r)^2}{4a_k}\right\} := \underline{q}_{k,R,r}$$

Hence, for $V(\mathbf{x}) \leq R$, $\|y\| \leq r$

$$q_{k+1}^{\theta,h}(y \mid \mathbf{x}) \geq \underline{q}_{k,R,r} \mathrm{e}^{-M'}$$

Let $\nu_{k+1,R}^{Q}$ be the uniform probability measure on $\mathrm{B}_r(0)$, namely

$$\nu_{k+1,R}^{Q}(A) := \frac{\mathrm{Leb}(A \cap \mathrm{B}_r(0))}{\mathrm{Leb}(\mathrm{B}_r(0))} .$$

Then, for every $A \in \mathcal{B}(\mathbb{R}^d)$ and $V(\mathbf{x}) \leq R$,

$$\begin{aligned} \mathsf{Q}_{k+1}^{\theta,h}(\mathbf{x}, A) &= \int_A q_{k+1}^{\theta,h}(y \mid \mathbf{x}) \, \mathrm{d}y \\ &\geq \underline{q}_{k,R,r} \mathrm{e}^{-M'} \, \mathrm{Leb}\left(A \cap \mathrm{B}_r(0)\right) \\ &= \varepsilon_{k+1}^{Q}(R) \, \nu_{k+1,R}^{Q}(A) , \end{aligned}$$

where $\varepsilon_{k+1}^{Q}(R) := \underline{q}_{k,R,r} \, \mathrm{Leb}(\mathrm{B}_r(0)) > 0$. This proves the local minorization.

C.3 Forgetting of the forward smoothing kernel

We now establish forgetting for the forward-smoothing kernels under proposal-level stability conditions. The following assumption is stated at an abstract kernel level for generality, but introduces no additional requirement for Algorithm 1: by Proposition 3.2, it follows directly from Assumption 3.1 and the step-size condition (11). Thus, throughout the remainder of the Appendix, any result stated under Assumption C.4 applies to Algorithm 1 under Assumption 3.1 and (11). This formulation also allows the subsequent analysis to apply to other proposal mechanisms satisfying the same drift and minorization properties.

Assumption C.4 (Proposal drift and local minorization)**.** The approximate proposal kernels $(\mathsf{Q}_k^{\theta})_{1 \leq k \leq N}$ satisfy the following conditions:

(I) **Drift.** There exists a measurable function $V : \mathbb{R}^d \to [0, \infty)$ such that, for every $1 \leq k \leq N$, there exist constants $\lambda_k^Q \in (0,1)$ and $K_k^Q < \infty$ satisfying, for any $\mathbf{x} \in \mathbb{R}^d$,

$$\mathsf{Q}_k^{\theta,h} V(\mathbf{x}) \leq \lambda_k^Q V(\mathbf{x}) + K_k^Q . \tag{36}$$

(II) **Minorization.** For every $R > 0$ and every $1 \leq k \leq N$, there exist $\varepsilon_k^Q(R) > 0$ and a probability measure $\nu_{k,R}^Q$ such that, for every $\mathbf{x} \in C_R := \{\mathbf{x} \in \mathbb{R}^d : V(\mathbf{x}) \leq R\}$, and every $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathsf{Q}_k^{\theta,h}(\mathbf{x}, A) \geq \varepsilon_k^Q(R) \nu_{k,R}^Q(A) . \tag{37}$$

Throughout the following results, N and the implemented model are fixed. By Assumption 3.3,

$$0 < \widetilde{\beta}_{k|N}^{\theta} \leq 1.$$

Consequently, for every proposal minorizing measure,

$$c_{k,R}^Q := \nu_{k,R}^Q[\widetilde{\beta}_{k|N}^{\theta}] > 0. \tag{38}$$

Lemma C.5 (Forgetting transfer conditions)**.** *Suppose that Assumption C.4 and Assumption 3.3 hold and define*

$$V^{\theta}_{k|N}(\mathbf{x}) := \frac{V(\mathbf{x})}{\widetilde{\beta}^{\theta}_{k|N}(\mathbf{x})}\,, \qquad 0 \le k \le N\,.$$

Then, for every $R > 0$, every $1 \le k \le N$, every $\mathbf{x}_{k-1} \in C_R$, and every $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathsf{G}^{\theta}_{k|N}(\mathbf{x}_{k-1}, A) \ge \varepsilon^{Q}_{k}(R) c^{Q}_{k,R} \nu^{\mathsf{G}}_{k,R}(A)\,, \tag{39}$$

where

$$\nu^{\mathsf{G}}_{k,R}(A) := \frac{\nu^{Q}_{k,R}\left[\widetilde{\beta}^{\theta}_{k|N} \mathbb{1}_A\right]}{\nu^{Q}_{k,R}\left[\widetilde{\beta}^{\theta}_{k|N}\right]}\,.$$

Here the dependence of $\nu^{\mathsf{G}}_{k,R}$ on N and θ is dropped. Moreover, for every $1 \le k \le N$, choose $R_k > 0$ such that

$$R_k > \frac{K^{Q}_{k}}{1 - \lambda^{Q}_{k}},$$

and define

$$\lambda^{\mathsf{G}}_{k} := \lambda^{Q}_{k} + \frac{K^{Q}_{k}}{R_k} \in (0,1)\,, \qquad K^{\mathsf{G}}_{k} := \frac{K^{Q}_{k}}{\varepsilon^{Q}_{k}(R_k) c^{Q}_{k,R_k}}\,.$$

Then, for every $1 \le k \le N$ and every $\mathbf{x}_{k-1} \in \mathbb{R}^d$,

$$\mathsf{G}^{\theta}_{k|N} V^{\theta}_{k|N}(\mathbf{x}_{k-1}) \le \lambda^{\mathsf{G}}_{k} V^{\theta}_{k-1|N}(\mathbf{x}_{k-1}) + K^{\mathsf{G}}_{k} \mathbb{1}_{C_{R_k}}(\mathbf{x}_{k-1})\,, \tag{40}$$

where $C_{R_k} = \{\mathbf{x} \in \mathbb{R}^d : V(\mathbf{x}) \le R_k\}$.

Proof. Since $0 < \tilde{g}_\ell \le 1$ for $0 \le \ell \le N-1$, we have

$$0 < \widetilde{\beta}^{\theta}_{k|N} \le 1\,, \qquad 0 \le k \le N\,.$$

Therefore,

$$\mathsf{Q}^{\theta,h}_{k} \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1}) \le 1\,.$$

Let $R > 0$, $\mathbf{x}_{k-1} \in C_R$, and $A \in \mathcal{B}(\mathbb{R}^d)$. By the definition of $\mathsf{G}^{\theta}_{k|N}$ and the minorization condition,

$$\begin{aligned}
\mathsf{G}^{\theta}_{k|N}(\mathbf{x}_{k-1}, A) &= \frac{\mathsf{Q}^{\theta,h}_{k}\left(\widetilde{\beta}^{\theta}_{k|N} \mathbb{1}_A\right)(\mathbf{x}_{k-1})}{\mathsf{Q}^{\theta,h}_{k} \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} \\
&\ge \mathsf{Q}^{\theta,h}_{k}\left(\widetilde{\beta}^{\theta}_{k|N} \mathbb{1}_A\right)(\mathbf{x}_{k-1}) \\
&\ge \varepsilon^{Q}_{k}(R) \nu^{Q}_{k,R}\left[\widetilde{\beta}^{\theta}_{k|N} \mathbb{1}_A\right]\,.
\end{aligned}$$

Moreover, using (38)

$$\nu^{Q}_{k,R}\left[\widetilde{\beta}^{\theta}_{k|N} \mathbb{1}_A\right] = \nu^{Q}_{k,R}\left[\widetilde{\beta}^{\theta}_{k|N}\right] \nu^{\mathsf{G}}_{k,R}(A) = c^{Q}_{k,R} \nu^{\mathsf{G}}_{k,R}(A)\,.$$

Thus

$$\mathsf{G}^{\theta}_{k|N}(\mathbf{x}_{k-1}, A) \ge \varepsilon^{Q}_{k}(R) c^{Q}_{k,R} \nu^{\mathsf{G}}_{k,R}(A)\,,$$

which proves (39).

For the drift condition, note that

$$\widetilde{\beta}^{\theta}_{k-1|N} = \tilde{g}_{k-1} \mathsf{Q}^{\theta,h}_{k} \widetilde{\beta}^{\theta}_{k|N} \le \mathsf{Q}^{\theta,h}_{k} \widetilde{\beta}^{\theta}_{k|N}\,. \tag{41}$$

Hence,

$$\frac{V(\mathbf{x}_{k-1})}{\mathsf{Q}^{\theta,h}_{k} \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} \le \frac{V(\mathbf{x}_{k-1})}{\widetilde{\beta}^{\theta}_{k-1|N}(\mathbf{x}_{k-1})} = V^{\theta}_{k-1|N}(\mathbf{x}_{k-1})\,.$$

Using the drift condition in Assumption C.4, we obtain

$$\begin{aligned}\mathsf{G}^{\theta}_{k|N} V^{\theta}_{k|N}(\mathbf{x}_{k-1}) &= \int \frac{V(\mathbf{x}_k)}{\widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_k)} \frac{\widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_k)\mathsf{Q}^{\theta,h}_k(\mathbf{x}_{k-1}, \mathrm{d}\mathbf{x}_k)}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} \\ &= \frac{\mathsf{Q}^{\theta,h}_k V(\mathbf{x}_{k-1})}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} \\ &\leq \lambda^Q_k V^{\theta}_{k-1|N}(\mathbf{x}_{k-1}) + \frac{K^Q_k}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} .\end{aligned}$$

It remains to control the offset term. Fix $R_k > K^Q_k/(1-\lambda^Q_k)$. If $\mathbf{x}_{k-1} \notin C_{R_k}$, then $V(\mathbf{x}_{k-1}) > R_k$, and

$$\frac{K^Q_k}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} = \frac{K^Q_k}{V(\mathbf{x}_{k-1})} \frac{V(\mathbf{x}_{k-1})}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} \leq \frac{K^Q_k}{R_k} V^{\theta}_{k-1|N}(\mathbf{x}_{k-1}) ,$$

where we used (41) in the last inequality. If $\mathbf{x}_{k-1} \in C_{R_k}$, then the minorization condition (37) gives

$$\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1}) \geq \varepsilon^Q_k(R_k) \nu^Q_{k,R_k}\left[\widetilde{\beta}^{\theta}_{k|N}\right] = \varepsilon^Q_k(R_k) c^Q_{k,R_k} .$$

Therefore,

$$\frac{K^Q_k}{\mathsf{Q}^{\theta,h}_k \widetilde{\beta}^{\theta}_{k|N}(\mathbf{x}_{k-1})} \leq \frac{K^Q_k}{\varepsilon^Q_k(R_k) c^Q_{k,R_k}} \mathbb{1}_{C_{R_k}}(\mathbf{x}_{k-1}) + \frac{K^Q_k}{R_k} V^{\theta}_{k-1|N}(\mathbf{x}_{k-1}) .$$

Combining this with the previous drift bound yields

$$\mathsf{G}^{\theta}_{k|N} V^{\theta}_{k|N}(\mathbf{x}_{k-1}) \leq \left(\lambda^Q_k + \frac{K^Q_k}{R_k}\right) V^{\theta}_{k-1|N}(\mathbf{x}_{k-1}) + \frac{K^Q_k}{\varepsilon^Q_k(R_k) c^Q_{k,R_k}} \mathbb{1}_{C_{R_k}}(\mathbf{x}_{k-1}) .$$

By definition,

$$\lambda^{\mathsf{G}}_k = \lambda^Q_k + \frac{K^Q_k}{R_k} , \qquad K^{\mathsf{G}}_k = \frac{K^Q_k}{\varepsilon^Q_k(R_k) c^Q_{k,R_k}} .$$

Since $R_k > K^Q_k/(1-\lambda^Q_k)$, we have $\lambda^{\mathsf{G}}_k \in (0,1)$. This proves (40). □

Corollary C.6 (Contraction in Lyapunov weigthed norms)**.** *Fix $N \geq 1$ and suppose that Assumption C.4 and Assumption 3.3 hold. Let $\lambda^{\mathsf{G}}_k \in (0,1)$ and $K^{\mathsf{G}}_k \in [0,\infty)$ be the constants of Lemma C.5. For every $1 \leq k \leq N$, choose*

$$R_k > \frac{2K^{\mathsf{G}}_k}{1-\lambda^{\mathsf{G}}_k} , \tag{42}$$

and define $\overline{\varepsilon}_k := \varepsilon^Q_k(R_k) c^Q_{k,R_k} \in (0,1]$.Choose $b_N \in (0,1]$ such that

$$2 b_N K^{\mathsf{G}}_k \leq \overline{\varepsilon}_k , \qquad 1 \leq k \leq N .$$

Such a choice exists since N is finite, $K^{\mathsf{G}}_k < \infty$, and $\overline{\varepsilon}_k > 0$. Define

$$\alpha_{k,N} := \max\left\{1 - \frac{\overline{\varepsilon}_k}{2} , \frac{1 + \frac{b_N}{2}\lambda^{\mathsf{G}}_k R_k + b_N K^{\mathsf{G}}_k}{1 + \frac{b_N}{2} R_k}\right\} \in (0,1) . \tag{43}$$

Then, for every $1 \leq k \leq N$ and probability measures μ_1, μ_2 with finite $V^{\theta}_{k-1|N}$-moments,

$$\left\| \mu_1 \mathsf{G}^{\theta}_{k|N} - \mu_2 \mathsf{G}^{\theta}_{k|N} \right\|_{b_N V^{\theta}_{k|N}} \leq \alpha_{k,N} \left\| \mu_1 - \mu_2 \right\|_{b_N V^{\theta}_{k-1|N}} . \tag{44}$$

Proof. The following result adapts the weighted-total-variation argument of Hairer & Mattingly (2008) to the time-inhomogeneous forward-smoothing kernels. Following the two-region argument of Hairer & Mattingly (2008), fix $1 \le k \le N$ and define the pairwise small set

$$\overline{C}_k := \left\{ (\mathbf{x}, \mathbf{x}') : V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}') \le R_k \right\} .$$

Let ψ be measurable and such that

$$|\psi| \le 1 + b_N V^\theta_{k|N} .$$

By Lemma C.5, and using $\mathbb{1}_{C_{R_k^{\mathrm{tr}}}} \le 1$,

$$\mathsf{G}^\theta_{k|N} V^\theta_{k|N}(\mathbf{x}) \le \lambda^{\mathsf{G}}_k V^\theta_{k-1|N}(\mathbf{x}) + K^{\mathsf{G}}_k .$$

Hence, for every $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$,

$$\left| \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}') \right| \le 2 + b_N \lambda^{\mathsf{G}}_k \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}') \right) + 2 b_N K^{\mathsf{G}}_k . \tag{45}$$

Outside $\overline{C}_k$. If $(\mathbf{x}, \mathbf{x}') \notin \overline{C}_k$, then

$$V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}') > R_k .$$

It follows that

$$\begin{aligned} \frac{\left| \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}') \right|}{2 + b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}') \right)} &\le \lambda^{\mathsf{G}}_k + \frac{2(1-\lambda^{\mathsf{G}}_k) + 2 b_N K^{\mathsf{G}}_k}{2 + b_N R_k} \\ &\le \frac{1 + \frac{b_N}{2}\lambda^{\mathsf{G}}_k R_k + b_N K^{\mathsf{G}}_k}{1 + \frac{b_N}{2} R_k} . \end{aligned} \tag{46}$$

By (42), $2K^{\mathsf{G}}_k < (1-\lambda^{\mathsf{G}}_k) R_k$, so the right-hand side of (46) is strictly smaller than one.

Inside $\overline{C}_k$. Since

$$V^\theta_{k-1|N}(\mathbf{x}) = \frac{V(\mathbf{x})}{\widetilde{\beta}^\theta_{k-1|N}(\mathbf{x})}$$

and $0 < \widetilde{\beta}^\theta_{k-1|N}(\mathbf{x}) \le 1$, if $(\mathbf{x}, \mathbf{x}') \in \overline{C}_k$, then

$$V(\mathbf{x}) \le V^\theta_{k-1|N}(\mathbf{x}) \le R_k , \qquad V(\mathbf{x}') \le V^\theta_{k-1|N}(\mathbf{x}') \le R_k .$$

Therefore, by the minorization part of Lemma C.5,

$$\mathsf{G}^\theta_{k|N}(\mathbf{x}, \cdot) \ge \overline{\varepsilon}_k \nu^{\mathsf{G}}_{k,R_k}(\cdot) , \qquad \mathsf{G}^\theta_{k|N}(\mathbf{x}', \cdot) \ge \overline{\varepsilon}_k \nu^{\mathsf{G}}_{k,R_k}(\cdot) .$$

Hence the measures

$$\mathsf{G}^\theta_{k|N}(\mathbf{x}, \cdot) - \overline{\varepsilon}_k \nu^{\mathsf{G}}_{k,R_k}(\cdot) , \qquad \mathsf{G}^\theta_{k|N}(\mathbf{x}', \cdot) - \overline{\varepsilon}_k \nu^{\mathsf{G}}_{k,R_k}(\cdot) ,$$

are nonnegative and both have total mass $1 - \overline{\varepsilon}_k$. Therefore, using $|\psi| \le 1 + b_N V^\theta_{k|N}$,

$$\begin{aligned} &\left| \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}') \right| \\ &= \left| \left(\mathsf{G}^\theta_{k|N}(\mathbf{x}, \cdot) - \overline{\varepsilon}_k \nu^{\mathsf{G}}_{k,R_k} \right)[\psi] - \left(\mathsf{G}^\theta_{k|N}(\mathbf{x}', \cdot) - \overline{\varepsilon}_k \nu^{\mathsf{G}}_{k,R_k} \right)[\psi] \right| \\ &\le 2(1-\overline{\varepsilon}_k) + b_N \mathsf{G}^\theta_{k|N} V^\theta_{k|N}(\mathbf{x}) + b_N \mathsf{G}^\theta_{k|N} V^\theta_{k|N}(\mathbf{x}') . \end{aligned}$$

Using Lemma C.5,

$$\left| \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}') \right| \le 2(1-\overline{\varepsilon}_k) + b_N \lambda^{\mathsf{G}}_k \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}') \right) + 2 b_N K^{\mathsf{G}}_k . \tag{47}$$

Dividing by $2 + b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}')\right)$ gives

$$\begin{aligned}\frac{\left|\mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}')\right|}{2 + b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}')\right)} &\leq \frac{2}{2 + b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}')\right)} \left(1 - \overline{\varepsilon}_k + b_N K^{\mathsf{G}}_k\right) \\ &\quad + \lambda^{\mathsf{G}}_k \frac{b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}')\right)}{2 + b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}')\right)} \\ &\leq \max\left\{1 - \overline{\varepsilon}_k + b_N K^{\mathsf{G}}_k \,,\, \lambda^{\mathsf{G}}_k\right\} .\end{aligned}$$

Since $2b_N K^{\mathsf{G}}_k \leq \overline{\varepsilon}_k$ by definition of b_N we have

$$1 - \overline{\varepsilon}_k + b_N K^{\mathsf{G}}_k \leq 1 - \frac{\overline{\varepsilon}_k}{2}.$$

Moreover,

$$\lambda^{\mathsf{G}}_k < \frac{1 + \frac{b_N}{2}\lambda^{\mathsf{G}}_k R_k + b_N K^{\mathsf{G}}_k}{1 + \frac{b_N}{2} R_k} .$$

Combining the two regions gives, for every $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and every measurable ψ satisfying $|\psi| \leq 1 + b_N V^\theta_{k|N}$,

$$\left|\mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}')\right| \leq \alpha_{k,N} \left[2 + b_N \left(V^\theta_{k-1|N}(\mathbf{x}) + V^\theta_{k-1|N}(\mathbf{x}')\right)\right] . \tag{48}$$

We now deduce the contraction between probability measures. Let $\eta := \mu_1 - \mu_2$ so that $\eta(\mathbb{R}^d) = 0$. Let

$$\eta = \eta^+ - \eta^-$$

be its Jordan decomposition. If $\eta = 0$, the result is immediate. Otherwise, since $\eta(\mathbb{R}^d) = 0$,

$$s := \eta^+(\mathbb{R}^d) = \eta^-(\mathbb{R}^d) > 0 \,,$$

and the measures

$$\pi^+ := \frac{\eta^+}{s} \,, \qquad \pi^- := \frac{\eta^-}{s}$$

belong to $\mathcal{P}(\mathbb{R}^d)$. Moreover,

$$\eta = s(\pi^+ - \pi^-) \,.$$

Hence, for every measurable $|\psi| \leq 1 + b_N V^\theta_{k|N}$,

$$\begin{aligned}\left|\left(\eta \mathsf{G}^\theta_{k|N}\right)[\psi]\right| &= s \left|\left(\pi^+ \mathsf{G}^\theta_{k|N}\right)[\psi] - \left(\pi^- \mathsf{G}^\theta_{k|N}\right)[\psi]\right| \\ &= s \left|\iint \left(\mathsf{G}^\theta_{k|N}\psi(\mathbf{x}) - \mathsf{G}^\theta_{k|N}\psi(\mathbf{x}')\right) \pi^+(\mathrm{d}\mathbf{x})\pi^-(\mathrm{d}\mathbf{x}')\right| \\ &\leq \alpha_{k,N} s \iint \left[2 + b_N V^\theta_{k-1|N}(\mathbf{x}) + b_N V^\theta_{k-1|N}(\mathbf{x}')\right] \pi^+(\mathrm{d}\mathbf{x})\pi^-(\mathrm{d}\mathbf{x}') \\ &= \alpha_{k,N} s \left[2 + b_N \pi^+\left[V^\theta_{k-1|N}\right] + b_N \pi^-\left[V^\theta_{k-1|N}\right]\right] \\ &= \alpha_{k,N} |\eta| \left[1 + b_N V^\theta_{k-1|N}\right] \\ &= \alpha_{k,N} \left\|\eta\right\|_{b_N V^\theta_{k-1|N}} .\end{aligned}$$

Taking the supremum over all measurable $|\psi| \leq 1 + b_N V^\theta_{k|N}$ yields

$$\left\|\eta \mathsf{G}^\theta_{k|N}\right\|_{b_N V^\theta_{k|N}} \leq \alpha_{k,N} \left\|\eta\right\|_{b_N V^\theta_{k-1|N}} . \tag{49}$$

Since $\eta = \mu_1 - \mu_2$, this is precisely (44). □

Corollary C.7 (Finite-horizon forward-smoothing stability)**.** *Under the assumptions of Corollary C.6, define*

$$\Gamma_{\ell,N} := \begin{cases} \dfrac{1}{b_N} \displaystyle\prod_{k=\ell+1}^{N} \alpha_{k,N}\,, & 0 \le \ell < N\,, \\ 1\,, & \ell = N\,. \end{cases} \tag{50}$$

Then, for every $0 \le \ell \le N$ and probability measures μ_1, μ_2 with finite $V^\theta_{\ell|N}$-moments,

$$\left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_V \le \Gamma_{\ell,N} \left\|\mu_1 - \mu_2\right\|_{V^\theta_{\ell|N}}\,. \tag{51}$$

In particular,

$$\left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_V \le \Gamma_{\ell,N} \left\{\mu_1[1 + V^\theta_{\ell|N}] + \mu_2[1 + V^\theta_{\ell|N}]\right\}\,.$$

Proof. The case $\ell = N$ is immediate. Let $0 \le \ell < N$, we first iterate the result of Corollary C.6

$$\left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_{b_N V^\theta_{N|N}} \le \left(\prod_{k=\ell+1}^{N} \alpha_{k,N}\right) \left\|\mu_1 - \mu_2\right\|_{b_N V^\theta_{\ell|N}}\,.$$

Since, by definition

$$V^\theta_{N|N} = V\,,$$

we obtain

$$\left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_{b_N V} \le \left(\prod_{k=\ell+1}^{N} \alpha_{k,N}\right) \left\|\mu_1 - \mu_2\right\|_{b_N V^\theta_{\ell|N}}\,. \tag{52}$$

For $0 \le \ell < N$, since $0 < b_N \le 1$,

$$1 + V \le \frac{1}{b_N}(1 + b_N V)\,, \qquad 1 + b_N V^\theta_{\ell|N} \le 1 + V^\theta_{\ell|N}\,.$$

Therefore, by (52),

$$\begin{aligned} \left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_V &\le \frac{1}{b_N} \left\|\mu_1 \mathsf{G}^\theta_{\ell:N} - \mu_2 \mathsf{G}^\theta_{\ell:N}\right\|_{b_N V} \\ &\le \frac{1}{b_N} \left(\prod_{k=\ell+1}^{N} \alpha_{k,N}\right) \left\|\mu_1 - \mu_2\right\|_{b_N V^\theta_{\ell|N}} \\ &\le \Gamma_{\ell,N} \left\|\mu_1 - \mu_2\right\|_{V^\theta_{\ell|N}}\,. \end{aligned}$$

The final bound follows from $|\mu_1 - \mu_2| \le \mu_1 + \mu_2$. □

D Monte Carlo Error

D.1 Detailed assumptions

To control the finite-particle approximation error in L_q norm. We need the following local moment condition ensuring that the selection and mutation empirical averages have finite L_q fluctuations.

Assumption D.1 (Local particle moments)**.** Let $q \ge 2$. Assume

$$\phi^\theta_0[(1 + V)^q] < \infty,$$

and, for every $1 \le k \le N$,

$$\sup_{M \ge 1} \mathbb{E}\left[\tau_{\hat{g}_{k-1}}(\phi^{\theta,M}_{k-1}) \left[\mathsf{Q}^{\theta,h}_k (1 + V)^q\right]\right] < \infty\,.$$

Under Assumption D.1, the local Monte Carlo fluctuations are of order $M^{-1/2}$; see Lemma D.3. Their propagation through the normalized Feynman–Kac flow additionally introduces random normalization factors. For $0 \le \ell \le N$, define the one-step predictive law

$$\eta_\ell^{M,\theta} := \begin{cases} \phi_0^\theta, & \ell = 0, \\ \mathbb{F}_\ell^\theta(\phi_{\ell-1}^{\theta,M}), & 1 \le \ell \le N\,. \end{cases}$$

Assumption D.2. Let $p \ge 1$. For every $0 \le \ell \le N$, assume that

$$\mathcal{A}_{\ell,N} := \sup_{M \ge 1} \left\| \frac{1}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}_{\ell|N}^\theta\right]} \left(1 + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^\theta + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^\theta]}\right) \right\|_{L_{2p}} < \infty\,.$$

For the L_p particle-error bound below, Assumption D.1 is applied with $q = 2p$. Under these conditions, the finite-particle approximation is controlled by Lemma D.4.

D.2 Proof

The proof follows the standard local-error propagation strategy for Feynman–Kac particle systems; see, for example, Del Moral (2004, Section 7.4.3) and Whiteley (2012).

For $0 \le \ell \le N$, define the future normalized approximate Feynman–Kac map

$$\mathbb{F}_{\ell:N}^\theta := \mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_{\ell+1}^\theta\,, \qquad \mathbb{F}_{N:N}^\theta := \mathrm{Id}\ .$$

Thus, for every probability measure ν,

$$\mathbb{F}_{\ell:N}^\theta(\nu) = \left(\mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_{\ell+1}^\theta\right)(\nu)\,.$$

Selected and propagated empirical measures. At time $k-1$, considering M particles, let

$$\mathcal{F}_{k-1}^\theta := \sigma\left(\mathbf{X}_\ell^{\theta,(i)},\, 1 \le i \le M,\, 0 \le \ell \le k-1\right)\,.$$

Given $\phi_{k-1}^{\theta,M}$, the *selection step* draws

$$\widetilde{\mathbf{X}}_{k-1}^{\theta,(1)}, \ldots, \widetilde{\mathbf{X}}_{k-1}^{\theta,(M)} \mid \mathcal{F}_{k-1}^\theta \overset{\text{i.i.d.}}{\sim} \tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\,.$$

We define

$$\widetilde{\phi}_{k-1}^{\theta,M} := \frac{1}{M}\sum_{i=1}^M \delta_{\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}}\,.$$

Let

$$\widetilde{\mathcal{F}}_{k-1}^\theta := \sigma\left(\mathcal{F}_{k-1}^\theta, \widetilde{\mathbf{X}}_{k-1}^{\theta,(1)}, \ldots, \widetilde{\mathbf{X}}_{k-1}^{\theta,(M)}\right)\,.$$

The *mutation step* draws, conditionally independently given $\widetilde{\mathcal{F}}_{k-1}^\theta$,

$$\mathbf{X}_k^{\theta,(i)} \mid \widetilde{\mathcal{F}}_{k-1}^\theta \sim \mathsf{Q}_k^{\theta,h}(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}, \mathrm{d}\mathbf{x})\,, \qquad i = 1, \ldots, M\,.$$

We define

$$\phi_k^{\theta,M} := \frac{1}{M}\sum_{i=1}^M \delta_{\mathbf{X}_k^{\theta,(i)}}\,.$$

Consequently, for every bounded measurable function ψ,

$$\mathbb{E}\left[\phi_k^{\theta,M}[\psi] \middle| \mathcal{F}_{k-1}^\theta\right] = \tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\mathsf{Q}_k^{\theta,h}[\psi] = \mathbb{F}_k^\theta(\phi_{k-1}^{\theta,M})[\psi].$$

Local selection and mutation errors. For every bounded measurable test function ψ, define

$$\phi_k^{\theta,M}[\psi] - \mathbb{F}_k^\theta(\phi_{k-1}^{\theta,M})[\psi] = M_k^M[\psi] + S_k^M[\psi] \ ,$$

where

$$M_k^M[\psi] := \phi_k^{\theta,M}[\psi] - \frac{1}{M}\sum_{i=1}^M \mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}) \, ,$$

is the *mutation error* and

$$S_k^M[\psi] := \frac{1}{M}\sum_{i=1}^M \mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}) - \left(\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\mathsf{Q}_k^{\theta,h}\right)[\psi] \, ,$$

is the *selection error*. Then

$$\mathbb{E}[M_k^M[\psi] \mid \widetilde{\mathcal{F}}_{k-1}^\theta] = 0 \, , \qquad \mathbb{E}[S_k^M[\psi] \mid \mathcal{F}_{k-1}^\theta] = 0 \, .$$

Local L_p control. We now quantify the size of the local empirical errors introduced at each selection and mutation step. Since the final estimates are stated in a V-weighted norm, we use the corresponding dual weighted supremum norm. For a measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$, define

$$\|\psi\|_{V,\infty} := \sup_{\mathbf{x}\in\mathbb{R}^d} \frac{|\psi(\mathbf{x})|}{1+V(\mathbf{x})} \, .$$

Lemma D.3. *Let $q \geq 2$ and suppose that Assumption D.1 holds. Then there exists a constant $B_q < \infty$, depending only on q, such that for every measurable $\psi : \mathbb{R}^d \to \mathbb{R}$ with $\|\psi\|_{V,\infty} < \infty$,*

$$\left\|(\phi_0^{\theta,M} - \phi_0^\theta)[\psi]\right\|_{L_q} \leq \frac{C_0^{(q)}}{\sqrt{M}}\|\psi\|_{V,\infty} \, , \tag{53}$$

and, for every $1 \leq k \leq N$,

$$\left\|\phi_k^{\theta,M}[\psi] - \left(\mathbb{F}_k^\theta(\phi_{k-1}^{\theta,M})\right)[\psi]\right\|_{L_q} \leq \frac{C_k^{S,(q)} + C_k^{M,(q)}}{\sqrt{M}}\|\psi\|_{V,\infty} \, , \tag{54}$$

where

$$C_0^{(q)} := 2B_q\phi_0^\theta[(1+V)^q]^{1/q} \ ,$$

$$C_k^{S,(q)} := 2B_q\left(\sup_{M\geq 1}\mathbb{E}\left[\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})[\mathsf{Q}_k^{\theta,h}(1+V)^q]\right]\right)^{1/q} \ ,$$

and

$$C_k^{M,(q)} := 2B_q\left(\sup_{M\geq 1}\mathbb{E}\left[\widetilde{\phi}_{k-1}^{\theta,M}\left[\mathsf{Q}_k^{\theta,h}(1+V)^q\right]\right]\right)^{1/q} \ .$$

Proof. The proof is an application of Lemma 7.7.3 (Chapter 7.3: Inequalities for independent Random Variables) in Del Moral (2004).

Initial empirical error. By construction, the initial particles $\mathbf{X}_0^{\theta,(1)}, \ldots, \mathbf{X}_0^{\theta,(M)}$ are i.i.d. with common law ϕ_0^θ. Define

$$U_i := \psi(\mathbf{X}_0^{\theta,(i)}) - \phi_0^\theta[\psi] \, , \qquad i = 1, \ldots, M \, .$$

Then the variables $U_1, \ldots, U_M$ are i.i.d. and centered, and

$$(\phi_0^{\theta,M} - \phi_0^\theta)[\psi] = \frac{1}{M}\sum_{i=1}^M U_i \, .$$

Therefore, by Marcinkiewicz–Zygmund inequality,

$$\left\|(\phi_0^{\theta,M} - \phi_0^\theta)[\psi]\right\|_{L_q} \leq \frac{B_q}{M}\left\|\left(\sum_{i=1}^M |U_i|^2\right)^{1/2}\right\|_{L_q} \, .$$

Since $q/2 \geq 1$, Jensen's inequality yields

$$\left(\sum_{i=1}^{M} |U_i|^2\right)^{q/2} \leq M^{\frac{q}{2}-1} \sum_{i=1}^{M} |U_i|^q \,.$$

Hence

$$\left\|(\phi_0^{\theta,M} - \phi_0^{\theta})[\psi]\right\|_{L_q} \leq \frac{B_q}{\sqrt{M}} \left(\frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[|U_i|^q]\right)^{1/q} .$$

Using $|a-b|^q \leq 2^{q-1}(|a|^q + |b|^q)$ and Jensen's inequality,

$$\mathbb{E}[|U_i|^q] \leq 2^{q-1}\left(\mathbb{E}[|\psi(\mathbf{X}_0^{\theta,(i)})|^q] + |\phi_0^{\theta}[\psi]\,|^q\right) \leq 2^q \phi_0^{\theta}[|\psi|^q] \,.$$

Since $|\psi(\mathbf{x})| \leq \|\psi\|_{V,\infty}(1+V(\mathbf{x}))$, we get

$$\phi_0^{\theta}[|\psi|^q] \leq \|\psi\|_{V,\infty}^q \, \phi_0^{\theta}\left[(1+V)^q\right] .$$

Combining the last displays proves (53).

Local selection error. By definition,

$$S_k^M[\psi] = \frac{1}{M}\sum_{i=1}^{M} \mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}) - \tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\mathsf{Q}_k^{\theta,h}\psi \,.$$

Conditionally on $\mathcal{F}_{k-1}^{\theta}$, the selected particles $\widetilde{\mathbf{X}}_{k-1}^{\theta,(1)}, \ldots, \widetilde{\mathbf{X}}_{k-1}^{\theta,(M)}$ are i.i.d. with common law $\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})$. Hence the variables

$$Y_i := \mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}) - \tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\mathsf{Q}_k^{\theta,h}\psi \,, \qquad i = 1, \ldots, M \,,$$

are conditionally i.i.d. and centered. Therefore, by Marcinkiewicz-Zygmund inequality, there exists B_q depending only on q such that,

$$\left\|S_k^M[\psi]\right\|_{L_q(.|\mathcal{F}_{k-1}^{\theta})} = \frac{1}{M}\left\|\sum_{i=1}^{M} Y_i\right\|_{L_q(.|\mathcal{F}_{k-1}^{\theta})} \leq \frac{B_q}{M}\left\|\left(\sum_{i=1}^{M}|Y_i|^2\right)^{1/2}\right\|_{L_q(.|\mathcal{F}_{k-1}^{\theta})}$$

where $L_q(.|\mathcal{F}_{k-1}^{\theta})$ means the L_q norms is taken conditional on $\mathcal{F}_{k-1}^{\theta}$. Since $q/2 \geq 1$, the map $x \mapsto x^{q/2}$ is convex. Hence, by Jensen,

$$\left(\sum_{i=1}^{M}|Y_i|^2\right)^{q/2} \leq M^{\frac{q}{2}-1}\sum_{i=1}^{M}|Y_i|^q \,,$$

Hence

$$\left\|S_k^M[\psi]\right\|_{L_q(.|\mathcal{F}_{k-1}^{\theta})} \leq \frac{B_q}{\sqrt{M}}\left(\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\left[|Y_i|^q\middle|\mathcal{F}_{k-1}^{\theta}\right]\right)^{1/q} .$$

Using $|a-b|^q \leq 2^{q-1}(|a|^q + |b|^q)$ and Jensen's inequality with $q \geq 2$

$$\begin{aligned}\mathbb{E}\left[|Y_i|^q\middle|\mathcal{F}_{k-1}^{\theta}\right] &\leq 2^{q-1}\left(\mathbb{E}\left[|\mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)})|^q\middle|\mathcal{F}_{k-1}^{\theta}\right] + \left|\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\mathsf{Q}_k^{\theta,h}\psi\right|^q\right) \\ &\leq 2^q \tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})\left[|\mathsf{Q}_k^{\theta,h}\psi|^q\right] .\end{aligned}$$

Since $|\psi(\mathbf{x})| \leq \|\psi\|_{V,\infty}(1+\mathrm{V}(\mathbf{x}))$ and using Jensen's inequality again

$$|\mathsf{Q}_k^{\theta,h}\psi(\mathbf{x})|^q \leq \mathsf{Q}_k^{\theta,h}|\psi|^q(\mathbf{x}) \leq \|\psi\|_{V,\infty}^q \mathsf{Q}_k^{\theta,h}(1+V)^q(\mathbf{x}) \,.$$

Therefore,

$$\left\|S_k^M[\psi]\right\|_{L_q(.|\mathcal{F}_{k-1}^{\theta})} \leq \frac{2B_q}{\sqrt{M}}\|\psi\|_{V,\infty}\left(\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})[\mathsf{Q}_k^{\theta,h}(1+V)^q]\right)^{1/q} .$$

Taking expectations and using the definition of $C_k^{S,(q)}$ yields

$$\left\|S_k^M[\psi]\right\|_{L_q} \leq \frac{C_k^{S,(q)}}{\sqrt{M}} \|\psi\|_{V,\infty} \,, \tag{55}$$

the claimed bound for the selection error.

Local mutation error. By definition,

$$M_k^M[\psi] = \phi_k^{\theta,M}[\psi] - \frac{1}{M}\sum_{i=1}^M \mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}) := \frac{1}{M}\sum_{i=1}^M Z_i \,,$$

where

$$Z_i := \psi(\mathbf{X}_k^{\theta,(i)}) - \mathsf{Q}_k^{\theta,h}\psi(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}) \,, \qquad i = 1, \ldots, M \,.$$

Conditionally on $\widetilde{\mathcal{F}}_{k-1}^{\theta}$, the propagated particles $\mathbf{X}_k^{\theta,(1)}, \ldots, \mathbf{X}_k^{\theta,(M)}$ are independent and satisfy

$$\mathbf{X}_k^{\theta,(i)} \mid \widetilde{\mathcal{F}}_{k-1}^{\theta} \sim \mathsf{Q}_k^{\theta,h}(\widetilde{\mathbf{X}}_{k-1}^{\theta,(i)}, \mathrm{d}\mathbf{x}) \,.$$

Thus the variables $Z_1, \ldots, Z_M$ are conditionally independent and centered. Applying again the conditional Marcinkiewicz–Zygmund inequality conditional on $\widetilde{\mathcal{F}}_{k-1}^{\theta}$ this time and then following the same steps as for the selection error yields

$$\left\|M_k^M[\psi]\right\|_{L_q(.|\widetilde{\mathcal{F}}_{k-1}^{\theta})} \leq \frac{2B_q}{\sqrt{M}} \|\psi\|_{V,\infty} \left(\widetilde{\phi}_{k-1}^{\theta,M}\left[\mathsf{Q}_k^{\theta,h}(1+V)^q\right]\right)^{1/q} .$$

Taking expectations and using the definition of $C_k^{M,(q)}$ yields

$$\left\|M_k^M[\psi]\right\|_{L_q} \leq \frac{C_k^{M,(q)}}{\sqrt{M}} \|\psi\|_{V,\infty} \,, \tag{56}$$

Finally, (54) follows from

$$\phi_k^{\theta,M}[\psi] - \mathbb{F}_k^{\theta}(\phi_{k-1}^{\theta,M})[\psi] = M_k^M[\psi] + S_k^M[\psi]$$

and Minkowski's inequality, together with the upper bound given (55) for the selection error and with upper bound (56) for the mutation error.

Integrability conditions. Finally, it follows from Assumption D.1 that the constants $C_0^{(q)}$ and $C_k^{S,(q)}$ are finite. For $C_k^{M,(q)}$, as

$$\sup_{M\geq 1} \mathbb{E}\left[\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})[\mathsf{Q}_k^{\theta,h}(1+V)^q]\right] < \infty \,,$$

by construction of the selected empirical measure,

$$\mathbb{E}\left[\widetilde{\phi}_{k-1}^{\theta,M}\left[\mathsf{Q}_k^{\theta,h}(1+V)^q\right]\middle|\mathcal{F}_{k-1}^{\theta}\right] = \tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})[\mathsf{Q}_k^{\theta,h}(1+V)^q] \,.$$

Hence,

$$\sup_{M\geq 1} \mathbb{E}\left[\widetilde{\phi}_{k-1}^{\theta,M}\left[\mathsf{Q}_k^{\theta,h}(1+V)^q\right]\right] = \sup_{M\geq 1} \mathbb{E}\left[\tau_{\hat{g}_{k-1}}(\phi_{k-1}^{\theta,M})[\mathsf{Q}_k^{\theta,h}(1+V)^q]\right] < \infty \,.$$

□

Lemma D.4 (Monte Carlo error under forward-smoothing forgetting)**.** *Let $p \geq 1$. Suppose that Assumptions 3.3 C.4, D.1 (with $q = 2p$) and D.2 hold. Then, for every measurable $\psi : \mathbb{R}^d \to \mathbb{R}$ such that $\|\psi\|_{V,\infty} < \infty$,*

$$\left\|\phi_N^{\theta,M}[\psi] - \phi_N^{\theta}[\psi]\right\|_{L_p} \leq \frac{\|\psi\|_{V,\infty}}{\sqrt{M}} \sum_{\ell=0}^N \mathcal{C}_\ell^{\mathrm{MC}} \mathcal{A}_{\ell,N} \Gamma_{\ell,N} \,,$$

where $\mathcal{C}_0^{\mathrm{MC}}$ and $\mathcal{C}_\ell^{\mathrm{MC}}$ are the constants of Lemma D.3 applied with exponent $2p$, namely

$$\mathcal{C}_0^{\mathrm{MC}} := C_0^{(2p)} \,, \qquad \mathcal{C}_\ell^{\mathrm{MC}} := C_\ell^{S,(2p)} + C_\ell^{M,(2p)} \,, \qquad 1 \leq \ell \leq N \,.$$

Proof. The proof follows the standard local-error propagation argument for Feynman–Kac particle systems; see, for instance, Del Moral (2004, Section 7.4.3) and Whiteley (2012). By a telescoping sum argument,

$$\phi_N^{\theta,M}[\psi] - \phi_N^{\theta}[\psi] = \sum_{\ell=0}^{N} \left\{ \mathbb{F}_{\ell:N}^{\theta}(\phi_\ell^{\theta,M})[\psi] - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi] \right\}, \tag{57}$$

where $\eta_\ell^{M,\theta}$ is defined in Assumption D.2. Fix $0 \le \ell \le N$. By Proposition C.1,

$$\mathbb{F}_{\ell:N}^{\theta}(\phi_\ell^{\theta,M})[\psi] = \frac{\phi_\ell^{\theta,M}\left[\widetilde{\beta}_{\ell|N}^{\theta}\mathsf{G}_{\ell:N}^{\theta}\psi\right]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]}.$$

Hence,

$$\mathbb{F}_{\ell:N}^{\theta}(\phi_\ell^{\theta,M})[\psi] - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi] = \frac{\phi_\ell^{\theta,M}\left[\widetilde{\beta}_{\ell|N}^{\theta}\left(\mathsf{G}_{\ell:N}^{\theta}\psi - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi]\right)\right]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]}. \tag{58}$$

Define the centered test function

$$\varphi_{\ell,N} := \frac{\widetilde{\beta}_{\ell|N}^{\theta}\left(\mathsf{G}_{\ell:N}^{\theta}\psi - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi]\right)}{\eta_\ell^{M,\theta}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]}.$$

By Proposition C.1,

$$\mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi] = \frac{\eta_\ell^{M,\theta}\left[\widetilde{\beta}_{\ell|N}^{\theta}\mathsf{G}_{\ell:N}^{\theta}\psi\right]}{\eta_\ell^{M,\theta}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]},$$

and therefore,

$$\eta_\ell^{M,\theta}[\varphi_{\ell,N}] = \frac{\eta_\ell^{M,\theta}\left[\widetilde{\beta}_{\ell|N}^{\theta}\mathsf{G}_{\ell:N}^{\theta}\psi\right] - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi]\eta_\ell^{M,\theta}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]}{\eta_\ell^{M,\theta}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]} = 0.$$

Hence combining the above with (58) yields

$$\mathbb{F}_{\ell:N}^{\theta}(\phi_\ell^{\theta,M})[\psi] - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi] = \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}_{\ell|N}^{\theta}\right]}\left(\phi_\ell^{\theta,M} - \eta_\ell^{M,\theta}\right)[\varphi_{\ell,N}]. \tag{59}$$

Applying Corollary C.7 to

$$\delta_{\mathbf{x}} - \frac{\widetilde{\beta}_{\ell|N}^{\theta}\eta_\ell^{M,\theta}}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]},$$

we obtain

$$\begin{aligned}\left|\mathsf{G}_{\ell:N}^{\theta}\psi(\mathbf{x}) - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})[\psi]\right| &\le \|\psi\|_{V,\infty}\left\|\delta_{\mathbf{x}}\mathsf{G}_{\ell:N}^{\theta} - \mathbb{F}_{\ell:N}^{\theta}(\eta_\ell^{M,\theta})\right\|_V \\ &\le \|\psi\|_{V,\infty}\Gamma_{\ell,N}\left\{1 + \frac{V(\mathbf{x})}{\widetilde{\beta}_{\ell|N}^{\theta}(\mathbf{x})} + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta} + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\right\}.\end{aligned}$$

Since $0 < \widetilde{\beta}_{\ell|N}^{\theta} \le 1$, the previous bound implies

$$\begin{aligned}\frac{|\varphi_{\ell,N}(\mathbf{x})|}{1+V(\mathbf{x})} &\le \|\psi\|_{V,\infty}\frac{\Gamma_{\ell,N}}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\frac{\widetilde{\beta}_{\ell|N}^{\theta}(\mathbf{x})}{1+V(\mathbf{x})}\left\{1 + \frac{V(\mathbf{x})}{\widetilde{\beta}_{\ell|N}^{\theta}(\mathbf{x})} + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta} + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\right\} \\ &= \|\psi\|_{V,\infty}\frac{\Gamma_{\ell,N}}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\left\{\frac{\widetilde{\beta}_{\ell|N}^{\theta}(\mathbf{x}) + V(\mathbf{x})}{1+V(\mathbf{x})} + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta} + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\frac{\widetilde{\beta}_{\ell|N}^{\theta}(\mathbf{x})}{1+V(\mathbf{x})}\right\} \\ &\le \|\psi\|_{V,\infty}\frac{\Gamma_{\ell,N}}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\left\{1 + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta} + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}_{\ell|N}^{\theta}]}\right\},\end{aligned}$$

Taking the supremum over $\mathbf{x} \in \mathbb{R}^d$, we obtain

$$\|\varphi_{\ell,N}\|_{V,\infty} \leq \|\psi\|_{V,\infty} \frac{\Gamma_{\ell,N}}{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]} \left\{1 + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N} + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]}\right\}. \tag{60}$$

Using (59) and Hölder's inequality we get

$$\begin{aligned}
&\left\|\mathbb{F}^\theta_{\ell:N}(\phi_\ell^{\theta,M})[\psi] - \mathbb{F}^\theta_{\ell:N}(\eta_\ell^{M,\theta})[\psi]\right\|_{L_p} \\
&\leq \left\| \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}^\theta_{\ell|N}\right]} \|\varphi_{\ell,N}\|_{V,\infty} \left| \left(\phi_\ell^{\theta,M} - \eta_\ell^{M,\theta}\right) \left[\frac{\varphi_{\ell,N}}{\|\varphi_{\ell,N}\|_{V,\infty}}\right] \right| \right\|_{L_p} \\
&\leq \left\| \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}^\theta_{\ell|N}\right]} \|\varphi_{\ell,N}\|_{V,\infty} \right\|_{L_{2p}} \left\| \left| \left(\phi_\ell^{\theta,M} - \eta_\ell^{M,\theta}\right) \left[\frac{\varphi_{\ell,N}}{\|\varphi_{\ell,N}\|_{V,\infty}}\right] \right| \right\|_{L_{2p}}
\end{aligned}$$

Using Lemma D.3,

$$\left\|\mathbb{F}^\theta_{\ell:N}(\phi_\ell^{\theta,M})[\psi] - \mathbb{F}^\theta_{\ell:N}(\eta_\ell^{M,\theta})[\psi]\right\|_{L_p} \leq \left\| \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}^\theta_{\ell|N}\right]} \|\varphi_{\ell,N}\|_{V,\infty} \right\|_{L_{2p}} \frac{\mathcal{C}_\ell^{\mathrm{MC}}}{\sqrt{M}},$$

Hence, it follows from (60) and Assumption D.2 that

$$\begin{aligned}
\left\| \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}^\theta_{\ell|N}\right]} \|\varphi_{\ell,N}\|_{V,\infty} \right\|_{L_{2p}} &\leq \Gamma_{\ell,N}\|\psi\|_{V,\infty} \left\| \frac{1}{\phi_\ell^{\theta,M}\left[\widetilde{\beta}^\theta_{\ell|N}\right]} \left(1 + \frac{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N} + V]}{\eta_\ell^{M,\theta}[\widetilde{\beta}^\theta_{\ell|N}]}\right) \right\|_{L_{2p}} \\
&\leq \Gamma_{\ell,N}\|\psi\|_{V,\infty}\mathcal{A}_{\ell,N}\,.
\end{aligned}$$

Therefore,

$$\left\|\mathbb{F}^\theta_{\ell:N}(\phi_\ell^{\theta,M})[\psi] - \mathbb{F}^\theta_{\ell:N}(\eta_\ell^{M,\theta})[\psi]\right\|_{L_p} \leq \frac{\mathcal{C}_\ell^{\mathrm{MC}}\mathcal{A}_{\ell,N}\Gamma_{\ell,N}}{\sqrt{M}}\|\psi\|_{V,\infty}\,.$$

Using (57) together with Minkowski's inequality yields

$$\left\|\phi_N^{\theta,M}[\psi] - \phi_N^\theta[\psi]\right\|_{L_p} \leq \frac{\|\psi\|_{V,\infty}}{\sqrt{M}} \sum_{\ell=0}^{N} \mathcal{C}_\ell^{\mathrm{MC}}\mathcal{A}_{\ell,N}\Gamma_{\ell,N}\,,$$

which finishes the proof. □

E Approximation bias

E.1 Local discrepancies and moment control

We first isolate the approximation error introduced at each SMC stage. Set

$$\mathcal{B}_0^{\mathrm{bias}} := \left\|\phi_0 - \phi_0^\theta\right\|_V,$$

and, for $1 \leq k \leq N$,

$$\begin{aligned}
\mathcal{B}_k^{\mathrm{bias}} &:= \left\|\mathbb{F}_k(\phi_{k-1}) - \mathbb{F}_k^\theta(\phi_{k-1})\right\|_V \\
&\leq \underbrace{\left\|\tau_{g_{k-1}}(\phi_{k-1})\left(\mathsf{Q}_k - \mathsf{Q}_k^{\theta,h}\right)\right\|_V}_{\mathcal{B}_k^{\mathrm{SGM}}} + \underbrace{\left\|\left(\tau_{\hat{g}_{k-1}}(\phi_{k-1}) - \tau_{g_{k-1}}(\phi_{k-1})\right)\mathsf{Q}_k^{\theta,h}\right\|_V}_{\mathcal{B}_k^{\mathrm{pot}}}.
\end{aligned} \tag{61}$$

Thus, $\mathcal{B}_k^{\mathrm{bias}}$ measures the discrepancy introduced locally at stage k when the ideal and implemented updates are initialized from the same ideal law ϕ_{k-1}. Its two components separate the SGM approximation error $\mathcal{B}_k^{\mathrm{SGM}}$, due to score approximation and numerical discretization, from the guidance-potential mismatch $\mathcal{B}_k^{\mathrm{pot}}$.

The following elementary moment bound ensures that the normalization factors appearing in the subsequent propagation argument are finite.

Lemma E.1 (Moment control)**.** *Suppose that Assumption C.4 and 3.3 hold. Assume moreover that*

$$\phi_0^\theta[V] < \infty\,, \qquad \mathcal{B}_0^{\mathrm{bias}} < \infty\,, \qquad \mathcal{B}_k^{\mathrm{bias}} < \infty\,, \quad 1 \le k \le N.$$

Then, for every $0 \le k \le N$*,*

$$\phi_k[V] < \infty\,,$$

and, for every $1 \le k \le N$*,*

$$\mathbb{F}_k^\theta(\phi_{k-1})[V] < \infty\,.$$

Proof. Since $\mathcal{B}_0^{\mathrm{bias}} < \infty$ and $\phi_0^\theta[V] < \infty$, we have $\phi_0[V] < \infty$. Assume that $\phi_{k-1}[V] < \infty$. By positivity and boundedness of $\hat{g}_{k-1}$,

$$\tau_{\hat{g}_{k-1}}(\phi_{k-1})[V] = \frac{\phi_{k-1}[\hat{g}_{k-1}V]}{\phi_{k-1}[\hat{g}_{k-1}]} < \infty\,.$$

Using the proposal drift condition,

$$\mathbb{F}_k^\theta(\phi_{k-1})[V] \le \lambda_k^Q \tau_{\hat{g}_{k-1}}(\phi_{k-1})[V] + K_k^Q < \infty\,.$$

Finally, since $\phi_k = \mathbb{F}_k(\phi_{k-1})$,

$$\phi_k[V] \le \mathbb{F}_k^\theta(\phi_{k-1})[V] + \mathcal{B}_k^{\mathrm{bias}} < \infty\,,$$

which closes the induction. □

E.2 Propagation of local approximation errors

We now propagate the local discrepancies introduced above through the remaining implemented Feynman–Kac dynamics. For $0 \le \ell \le N$, define

$$F_{\ell|N}^\theta := 1 + V_{\ell|N}^\theta.$$

Proposition E.2 (Approximation bias decomposition)**.** *For every* $N \ge 2$*,*

$$\begin{aligned}\phi_N - \phi_N^\theta = &\left(\mu_{0|N} - \mu_{0|N}^\theta\right) \mathsf{G}_{1|N}^\theta \cdots \mathsf{G}_{N|N}^\theta \\ &+ \sum_{\ell=1}^{N-1} \left(\mu_{\ell|N} - \mu_{\ell|N}^\theta\right) \mathsf{G}_{\ell+1|N}^\theta \cdots \mathsf{G}_{N|N}^\theta + \left(\mathbb{F}_N(\phi_{N-1}) - \mathbb{F}_N^\theta(\phi_{N-1})\right)\,,\end{aligned}$$

where for $1 \le \ell \le N-1$*,*

$$\mu_{\ell|N} := \frac{\widetilde{\beta}_{\ell|N}^\theta \mathbb{F}_\ell(\phi_{\ell-1})}{\mathbb{F}_\ell(\phi_{\ell-1})[\widetilde{\beta}_{\ell|N}^\theta]}\,, \qquad \mu_{\ell|N}^\theta := \frac{\widetilde{\beta}_{\ell|N}^\theta \mathbb{F}_\ell^\theta(\phi_{\ell-1})}{\mathbb{F}_\ell^\theta(\phi_{\ell-1})[\widetilde{\beta}_{\ell|N}^\theta]}\,,$$

and

$$\mu_{0|N} := \frac{\widetilde{\beta}_{0|N}^\theta \phi_0}{\phi_0\left[\widetilde{\beta}_{0|N}^\theta\right]}\,, \qquad \mu_{0|N}^\theta := \frac{\widetilde{\beta}_{0|N}^\theta \phi_0^\theta}{\phi_0^\theta\left[\widetilde{\beta}_{0|N}^\theta\right]}\,.$$

Equivalently,

$$\phi_N - \phi_N^\theta = \sum_{\ell=0}^{N} \left(\mu_{\ell|N} - \mu_{\ell|N}^\theta\right) \mathsf{G}_{\ell:N}^\theta\,, \tag{62}$$

where $\mathsf{G}_{N:N}^\theta = \mathrm{Id}$*, and*

$$\mu_{N|N} := \mathbb{F}_N(\phi_{N-1})\,, \qquad \mu_{N|N}^\theta := \mathbb{F}_N^\theta(\phi_{N-1})\,.$$

Proof. We write

$$\phi_N = (\mathbb{F}_N \circ \cdots \circ \mathbb{F}_1)(\phi_0), \qquad \phi_N^\theta = \left(\mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_1^\theta\right)(\phi_0^\theta).$$

A standard telescoping argument yields for any $N \geq 2$,

$$\begin{aligned}\phi_N - \phi_N^\theta = &\left(\mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_1^\theta\right)(\phi_0) - \left(\mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_1^\theta\right)(\phi_0^\theta) \\ &+ \sum_{\ell=1}^{N-1} \left\{ \left(\mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_{\ell+1}^\theta\right)(\phi_\ell) - \left(\mathbb{F}_N^\theta \circ \cdots \circ \mathbb{F}_{\ell+1}^\theta\right)\left(\mathbb{F}_\ell^\theta(\phi_{\ell-1})\right) \right\} \\ &+ \mathbb{F}_N(\phi_{N-1}) - \mathbb{F}_N^\theta(\phi_{N-1}).\end{aligned}$$

Then, applying proposition C.1 finishes the proof. □

Lemma E.3 (Kernel bias propagation under forward-smoothing forgetting)**.** *Suppose that C.4 3.3 hold, that $\phi_0^\theta[V] < \infty$, and that $\mathcal{B}_0^{\mathrm{bias}} < \infty$ and $\mathcal{B}_k^{\mathrm{SGM}}, \mathcal{B}_k^{\mathrm{pot}} < \infty$ for $1 \leq k \leq N$. For $1 \leq \ell \leq N-1$, define*

$$\Lambda_{\ell,N} := \frac{1 + \mu_{\ell|N}^\theta\left[F_{\ell|N}^\theta\right]}{\mathbb{F}_\ell(\phi_{\ell-1})\left[\widetilde{\beta}_{\ell|N}^\theta\right]},$$

and

$$\Lambda_{0,N} := \frac{1 + \mu_{0|N}^\theta\left[F_{0|N}^\theta\right]}{\phi_0\left[\widetilde{\beta}_{0|N}^\theta\right]}, \qquad \Lambda_{N,N} := 1.$$

Then, for every $N \geq 2$,

$$\begin{aligned}\left\|\phi_N - \phi_N^\theta\right\|_V &\leq \Gamma_{0,N}\Lambda_{0,N}\left\|\phi_0 - \phi_0^\theta\right\|_V + \sum_{\ell=1}^{N} \Gamma_{\ell,N}\Lambda_{\ell,N}\left\|\mathbb{F}_\ell(\phi_{\ell-1}) - \mathbb{F}_\ell^\theta(\phi_{\ell-1})\right\|_V \\ &\leq \Gamma_{0,N}\Lambda_{0,N}\mathcal{B}_0^{\mathrm{bias}} + \sum_{\ell=1}^{N} \Gamma_{\ell,N}\Lambda_{\ell,N}\left(\mathcal{B}_\ell^{\mathrm{SGM}} + \mathcal{B}_\ell^{\mathrm{pot}}\right), \qquad (63)\end{aligned}$$

with for $0 \leq \ell \leq N$, $\Gamma_{\ell,N}$ defined as in Corollary C.7 and $\Lambda_{\ell,N} < \infty$.

Proof. By Proposition E.2,

$$\phi_N - \phi_N^\theta = \sum_{\ell=0}^{N} \left(\mu_{\ell|N} - \mu_{\ell|N}^\theta\right) \mathsf{G}_{\ell:N}^\theta.$$

Using the triangle inequality and Corollary C.7, we get

$$\left\|\phi_N - \phi_N^\theta\right\|_V \leq \sum_{\ell=0}^{N} \Gamma_{\ell,N}\left\|\mu_{\ell|N} - \mu_{\ell|N}^\theta\right\|_{V_{\ell|N}^\theta}.$$

Recall that for $\mathbf{x} \in \mathbb{R}^d$,

$$F_{\ell|N}^\theta(\mathbf{x}) = 1 + V_{\ell|N}^\theta(\mathbf{x}) = 1 + \frac{V(\mathbf{x})}{\widetilde{\beta}_{\ell|N}^\theta(\mathbf{x})}, \qquad 0 \leq \ell \leq N.$$

For $0 \leq \ell \leq N-1$, we use the following estimate. Let μ, ν be probability measures such that $\mu[\widetilde{\beta}_{\ell|N}^\theta] > 0$ and $\nu[\widetilde{\beta}_{\ell|N}^\theta] > 0$. Then

$$\left\|\frac{\widetilde{\beta}_{\ell|N}^\theta \mu}{\mu[\widetilde{\beta}_{\ell|N}^\theta]} - \frac{\widetilde{\beta}_{\ell|N}^\theta \nu}{\nu[\widetilde{\beta}_{\ell|N}^\theta]}\right\|_{V_{\ell|N}^\theta} \leq \frac{1 + \left(\frac{\widetilde{\beta}_{\ell|N}^\theta \nu}{\nu[\widetilde{\beta}_{\ell|N}^\theta]}\right)[F_{\ell|N}^\theta]}{\mu[\widetilde{\beta}_{\ell|N}^\theta]}\|\mu - \nu\|_V. \qquad (64)$$

Indeed, for every measurable ψ such that $|\psi| \le F^\theta_{\ell|N}$,

$$\left|\frac{\mu[\widetilde{\beta}^\theta_{\ell|N}\psi]}{\mu[\widetilde{\beta}^\theta_{\ell|N}]} - \frac{\nu[\widetilde{\beta}^\theta_{\ell|N}\psi]}{\nu[\widetilde{\beta}^\theta_{\ell|N}]}\right| \le \frac{|\mu-\nu|[\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N}]}{\mu[\widetilde{\beta}^\theta_{\ell|N}]} + \frac{\nu[\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N}]}{\nu[\widetilde{\beta}^\theta_{\ell|N}]}\frac{|\mu-\nu|[\widetilde{\beta}^\theta_{\ell|N}]}{\mu[\widetilde{\beta}^\theta_{\ell|N}]}$$
$$\le \frac{1+\left(\frac{\widetilde{\beta}^\theta_{\ell|N}\nu}{\nu[\widetilde{\beta}^\theta_{\ell|N}]}\right)\left[F^\theta_{\ell|N}\right]}{\mu[\widetilde{\beta}^\theta_{\ell|N}]}|\mu-\nu|[\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N}]\,,$$

where we used in the last inequality that $F^\theta_{\ell|N} \ge \mathbb{1}$, and therefore $|\mu-\nu|[\widetilde{\beta}^\theta_{\ell|N}] \le |\mu-\nu|[\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N}]$. Also, since

$$\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N} = \widetilde{\beta}^\theta_{\ell|N} + V \le 1 + V\,,$$

we obtain (64). Applying (64) with

$$\mu = \phi_0\,, \qquad \nu = \phi^\theta_0\,,$$

gives

$$\left\|\mu_{0|N} - \mu^\theta_{0|N}\right\|_{V^\theta_{0|N}} \le \Lambda_{0,N}\left\|\phi_0 - \phi^\theta_0\right\|_V\,,$$

The initial term satisfies

$$\left\|\phi_0 - \phi^\theta_0\right\|_V \le \mathcal{B}^{\mathrm{bias}}_0\,.$$

Similarly, for $1 \le \ell \le N-1$, applying (64) with

$$\mu = \mathbb{F}_\ell(\phi_{\ell-1})\,, \qquad \nu = \mathbb{F}^\theta_\ell(\phi_{\ell-1})\,,$$

gives

$$\left\|\mu_{\ell|N} - \mu^\theta_{\ell|N}\right\|_{V^\theta_{\ell|N}} \le \Lambda_{\ell,N}\left\|\mathbb{F}_\ell(\phi_{\ell-1}) - \mathbb{F}^\theta_\ell(\phi_{\ell-1})\right\|_V\,.$$

For every $1 \le k \le N$, by definition of the local discrepancy,

$$\left\|\mathbb{F}_k(\phi_{k-1}) - \mathbb{F}^\theta_k(\phi_{k-1})\right\|_V = \mathcal{B}^{\mathrm{bias}}_k\,.$$

Moreover,

$$\mathcal{B}^{\mathrm{bias}}_k \le \mathcal{B}^{\mathrm{SGM}}_k + \mathcal{B}^{\mathrm{pot}}_k\,.$$

Finally, for $\ell = N$,

$$\widetilde{\beta}^\theta_{N|N} = \mathbb{1}\,, \qquad \mathsf{G}^\theta_{N:N} = \mathrm{Id}\,, \qquad \Lambda_{N,N} := 1\,,$$

and therefore

$$\left\|\mu_{N|N} - \mu^\theta_{N|N}\right\|_V = \left\|\mathbb{F}_N(\phi_{N-1}) - \mathbb{F}^\theta_N(\phi_{N-1})\right\|_V\,.$$

It remains to prove the finiteness of $\Lambda_{\ell,N}$. Since $0 < \widetilde{\beta}^\theta_{\ell|N} \le 1$, for any probability measure μ

$$0 < \mu[\widetilde{\beta}^\theta_{\ell|N}] \le 1\,,$$

and

$$\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N} = \widetilde{\beta}^\theta_{\ell|N} + V \le 1 + V\,.$$

Hence, for $\ell = 0$, using Assumption 3.3,

$$\mu^\theta_{0|N}[F^\theta_{0|N}] = \frac{\phi^\theta_0\left[\widetilde{\beta}^\theta_{0|N}F^\theta_{0|N}\right]}{\phi^\theta_0\left[\widetilde{\beta}^\theta_{0|N}\right]} \le \frac{1+\phi^\theta_0[V]}{\phi^\theta_0\left[\widetilde{\beta}^\theta_{0|N}\right]} < \infty\,.$$

For $1 \le \ell \le N-1$, the same argument holds with Lemma E.1,

$$\mu^\theta_{\ell|N}[F^\theta_{\ell|N}] = \frac{\mathbb{F}^\theta_\ell(\phi_{\ell-1})[\widetilde{\beta}^\theta_{\ell|N}F^\theta_{\ell|N}]}{\mathbb{F}^\theta_\ell(\phi_{\ell-1})[\widetilde{\beta}^\theta_{\ell|N}]} \le \frac{1+\mathbb{F}^\theta_\ell(\phi_{\ell-1})[V]}{\mathbb{F}^\theta_\ell(\phi_{\ell-1})[\widetilde{\beta}^\theta_{\ell|N}]} < \infty\,.$$

Finally, $\Lambda_{N,N} = 1$, which concludes the proof. □

E.3 LOCAL ERROR CONTROLS

We now derive explicit bounds for the two approximation terms appearing in (61). Applying such estimates to Lemma E.3 concludes the proof for the approximation bias.

Lemma E.4 (Local SGM approximation error)**.** *Fix $1 \le k \le N$. Suppose Assumptions 3.1, 3.3, and 3.5 hold. Define*

$$\varepsilon_k^{\mathrm{score}}(\theta,h)^2 := \sum_{n=(k-1)J}^{kJ-1} \Delta_n \, \mathbb{E}_k\left[\left\|\nabla \log p_{T-s_n}(\overleftarrow{\mathbf{X}}_{s_n}) - \mathrm{s}_\theta\left(\overleftarrow{\mathbf{X}}_{s_n}, T-s_n\right)\right\|^2\right],$$

where $\mathbb{E}_k$ denotes expectation under the exact reverse dynamics (2) *initialized from $\tau_{g_{k-1}}(\phi_{k-1})$ at time t_{k-1}. Then there exist $C_k^{\mathrm{discr}}, C_k^{\mathrm{net}} < \infty$, independent of the numerical refinement, such that*

$$\mathcal{B}_k^{\mathrm{SGM}} := \left\|\phi_k - \tau_{g_{k-1}}(\phi_{k-1})\mathsf{Q}_k^{\theta,h}\right\|_V \le C_k^{\mathrm{discr}}\sqrt{h} + C_k^{\mathrm{net}}\varepsilon_k^{\mathrm{score}}(\theta,h)\,.$$

Proof. To reuse the local Girsanov–Hellinger argument of Strasman et al. (2026, Section H.2), we need to verify its moment and score-regularity requirements under this common distribution $\tau_{g_{k-1}}(\phi_{k-1})$ as

$$\mathcal{B}_k^{\mathrm{SGM}} = \left\|\tau_{g_{k-1}}(\phi_{k-1})\mathsf{Q}_k - \tau_{g_{k-1}}(\phi_{k-1})\mathsf{Q}_k^{\theta,h}\right\|_V .$$

In the SMC setting, these estimates are therefore required along a guided exact flow rather than under the unconditional diffusion marginals. We first transfer the corresponding moment and score-regularity estimates from the unconditional marginals p_{T-u} of (2) to this SMC exact flow.

Control of the V-weighted moments. Since $\phi_0 = p_T$ and the ideal potentials are bounded by Assumption 3.3, induction along the ideal Feynman–Kac flow gives

$$\tau_{g_{k-1}}(\phi_{k-1}) \le \prod_{j=0}^{k-1} \frac{\|g_j\|_\infty}{\phi_j[g_j]} p_{T-t_{k-1}}\,,$$

Propagating both sides with the same exact reverse dynamics therefore yields, for $u \in [t_{k-1}, t_k]$,

$$\mathcal{L}_k^{\mathrm{ex}}(\overleftarrow{\mathbf{X}}_u) = \tau_{g_{k-1}}(\phi_{k-1})\mathsf{Q}_{t_{k-1},u} \le D_k\, p_{T-t_{k-1}}\mathsf{Q}_{t_{k-1},u} = D_k\, p_{T-u}\,,$$

where by assumption

$$D_k := \prod_{j=0}^{k-1} \frac{\|g_j\|_\infty}{\phi_j[g_j]} < \infty\,.$$

Hence any integrability estimate available under the unconditional marginal p_{T-u} transfers to the exact reverse dynamics initialized from $\tau_{g_{k-1}}(\phi_{k-1})$. In particular, π_{data} has finite fourth order moment and so

$$\sup_{u\in[t_{k-1},t_k]} \mathbb{E}_k^{\mathrm{ex}}\left[(1+V(\overleftarrow{\mathbf{X}}_u))^2\right] < \infty\,.$$

The linear-growth condition in Assumption 3.1 implies

$$\sup_{u\in[t_{k-1},t_k]} \mathbb{E}\left[(1+V(\bar{\mathbf{X}}_u^\theta))^2\right] < \infty,$$

uniformly over the numerical refinement, where $(\bar{\mathbf{X}}_u^\theta)_{u\in[t_{k-1},t_k]}$ denotes the continuous-time Euler interpolation initialized from $\tau_{g_{k-1}}(\phi_{k-1})$.

Hessian integrability condition. The freezing argument of Strasman et al. (2026, Section H.2) further requires the Jacobian of the true score to be square-integrable along the exact backward flow. By the same domination,

$$\begin{aligned}
&\int_{t_{k-1}}^{t_k} \bar\beta_u\, \mathbb{E}_k^{\mathrm{ex}}\left[\|\nabla^2 \log p_{T-u}(\overleftarrow{\mathbf{X}}_u)\|_{\mathrm{F}}^2\right] \mathrm{d}u \\
&\qquad \le D_k \int_{t_{k-1}}^{t_k} \bar\beta_u \int_{\mathbb{R}^d} \|\nabla^2 \log p_{T-u}(\mathbf{x})\|_{\mathrm{F}}^2 p_{T-u}(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}u \\
&\qquad = D_k \int_{T-t_k}^{T-t_{k-1}} \beta(t) \int_{\mathbb{R}^d} \|\nabla^2 \log p_t(\mathbf{x})\|_{\mathrm{F}}^2 p_t(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}t < \infty. \qquad (65)
\end{aligned}$$

The finiteness of (65) is a standard consequence of the Fisher-information evolution identity along the VE/VP forward flow (Bakry et al., 2014).

Local error estimate We can therefore repeat, on each numerical subinterval, the Girsanov–freezing argument of Strasman et al. (2026, Section H.2).

$$\mathrm{KL}\Big(\phi_k \,\Big\|\, \tau_{g_{k-1}}(\phi_{k-1})\mathsf{Q}_k^{\theta,h}\Big) \lesssim h + \varepsilon_k^{\mathrm{score}}(\theta,h)^2\,,$$

Finally, the weighted Hellinger argument of Strasman et al. (2026, Section H.2), together with the uniform fourth-moment bounds above, gives

$$\mathcal{B}_k^{\mathrm{SGM}} \le C_k^{\mathrm{discr}}\sqrt{h} + C_k^{\mathrm{net}}\varepsilon_k^{\mathrm{score}}(\theta,h)\,,$$

with $C_k^{\mathrm{discr}}, C_k^{\mathrm{net}} < \infty$ independent of the numerical refinement. This proves the claim. □

Proposition E.5 (Potential mismatch)**.** *Suppose that Assumptions 3.3 and 3.5 hold, and that the implemented proposals satisfy the drift bound in C.4. Consider the ideal flow initialized from $\phi_0 = p_T$. For $1 \le k \le N$, define*

$$\big(\varepsilon_k^{\mathrm{pot}}\big)^2 := \phi_{k-1}\left[\left(\frac{\hat g_{k-1}}{\phi_{k-1}[\hat g_{k-1}]} - \frac{g_{k-1}}{\phi_{k-1}[g_{k-1}]}\right)^2\right]. \tag{66}$$

Then,

$$\begin{aligned}
\mathcal{B}_k^{\mathrm{pot}} &:= \Big\|\big(\tau_{\hat g_{k-1}}(\phi_{k-1}) - \tau_{g_{k-1}}(\phi_{k-1})\big)\,\mathsf{Q}_k^{\theta,h}\Big\|_V \\
&\le \phi_{k-1}\left[\left|\frac{\hat g_{k-1}}{\phi_{k-1}[\hat g_{k-1}]} - \frac{g_{k-1}}{\phi_{k-1}[g_{k-1}]}\right|(1 + K_k^Q + \lambda_k^Q V)\right] &(67)\\
&\le \varepsilon_k^{\mathrm{pot}}\left\{\phi_{k-1}\Big[(1 + K_k^Q + \lambda_k^Q V)^2\Big]\right\}^{1/2}, &(68)
\end{aligned}$$

where in the main text we used

$$C_k^{\mathrm{pot}} := \left\{\phi_{k-1}\Big[(1 + K_k^Q + \lambda_k^Q V)^2\Big]\right\}^{1/2}.$$

Proof. We first establish finiteness of the quantities in the statement. Strict positivity and boundedness of the potentials imply

$$0 < \phi_j[g_j] < \infty\,, \qquad 0 < \phi_j[\hat g_j] < \infty\,.$$

Thus the two normalized potentials in (66) are bounded, and $\varepsilon_k^{\mathrm{pot}} < \infty$. Since $\phi_0 = p_T$, induction using bounded ideal selection and the identity $p_{T-t_j}\mathsf{Q}_{j+1} = p_{T-t_{j+1}}$ gives

$$\phi_j \le \left(\prod_{i=0}^{j-1}\frac{\|g_i\|_\infty}{\phi_i[g_i]}\right) p_{T-t_j}\,, \qquad 0 \le j \le N\,.$$

The Gaussian representation of the forward diffusion together with the fact that π_{data} has finite fourth moment imply $\int_{\mathbb{R}^d}(1+V(\mathbf{x}))^2\,p_t(\mathbf{x})\,\mathrm{d}\mathbf{x} < \infty$ for every $t \in [0,T]$. Consequently, $\phi_j\big[(1+V)^2\big] < \infty$ and right-hand side of (68) is finite.

By definition of the weighted total-variation norm and using $|\mu Q| \le |\mu|Q$ for Markov kernels,

$$\begin{aligned}
\mathcal{B}_k^{\mathrm{pot}} &= \int_{\mathbb{R}^d}(1+V(\mathbf{u}))\left|\big(\tau_{\hat g_{k-1}}(\phi_{k-1}) - \tau_{g_{k-1}}(\phi_{k-1})\big)\,\mathsf{Q}_k^{\theta,h}\right|(\mathrm{d}\mathbf{u}) \\
&\le \int_{\mathbb{R}^d}\left|\frac{\hat g_{k-1}(\mathbf{x})}{\phi_{k-1}[\hat g_{k-1}]} - \frac{g_{k-1}(\mathbf{x})}{\phi_{k-1}[g_{k-1}]}\right|\mathsf{Q}_k^{\theta,h}(1+V)(\mathbf{x})\,\phi_{k-1}(\mathrm{d}\mathbf{x}) \\
&= \phi_{k-1}\left[\left|\frac{\hat g_{k-1}}{\phi_{k-1}[\hat g_{k-1}]} - \frac{g_{k-1}}{\phi_{k-1}[g_{k-1}]}\right|\mathsf{Q}_k^{\theta,h}(1+V)\right].
\end{aligned}$$

The proposal drift bound gives

$$\mathsf{Q}_k^{\theta,h}(1+V) \le 1 + K_k^Q + \lambda_k^Q V\,,$$

which proves (67). Cauchy–Schwarz under ϕ_{k-1} then yields (68). □

Remark E.6 (Interpretation). The quantity $\varepsilon_k^{\mathrm{pot}}$ measures the mean-square discrepancy between the normalized potentials under the common ideal input law ϕ_{k-1}. It is invariant under separate positive rescaling of either potential and vanishes when they are proportional ϕ_{k-1}-almost surely. The bound only penalizes potential mismatch in regions that are relevant under the current ideal filtering distribution, with errors further weighted by the Lyapunov cost induced by the next mutation step. The data assumption is used here only through its fourth-moment condition.

F Numerical illustration

F.1 Gaussian-mixture target and closed form diffusion model

We consider a synthetic benchmark for which all ideal quantities are available in closed form. This allows us to measure separately the finite-particle error and the deterministic proposal bias appearing in Theorem 3.6.

Data distribution. Let π_{data} be a Gaussian mixture on $\mathbb{R}^d$,

$$\pi_{\mathrm{data}}(\mathrm{d}\mathbf{x}) = \sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}; \mu_j, \Sigma_j)\,\mathrm{d}\mathbf{x}\,, \qquad w_j > 0\,, \qquad \sum_{j=1}^{K} w_j = 1\,. \tag{69}$$

Forward diffusion. We use the variance-preserving Ornstein–Uhlenbeck forward diffusion

$$\mathrm{d}\overrightarrow{\mathbf{X}}_t = -\beta \overrightarrow{\mathbf{X}}_t\,\mathrm{d}t + \sqrt{2\beta}\mathrm{d}B_t\,, \qquad \overrightarrow{\mathbf{X}}_0 \sim \pi_{\mathrm{data}}\,,$$

with constant $\beta > 0$. This corresponds to the setting $\alpha = 1$ and $\beta_t = \beta$ in (1). For $0 \le s \le t \le T$, write

$$m_{s,t} := \mathrm{e}^{-\beta(t-s)}\,, \qquad \sigma_{s,t}^2 := 1 - m_{s,t}^2\,.$$

Hence

$$\overrightarrow{\mathbf{X}}_t = m_{s,t}\overrightarrow{\mathbf{X}}_s + \sigma_{s,t} Z\,, \qquad Z \sim \mathcal{N}(0, \mathbf{I}_d)\,, \qquad Z \perp \overrightarrow{\mathbf{X}}_s\,. \tag{70}$$

In particular, under mixture component j, that is conditionally on $J = j$,

$$\overrightarrow{\mathbf{X}}_t \mid J = j \sim \mathcal{N}(\mu_{j,t}, \Sigma_{j,t})\,, \tag{71}$$

where

$$\mu_{j,t} := m_{0,t}\mu_j\,, \qquad \Sigma_{j,t} := m_{0,t}^2\Sigma_j + \sigma_{0,t}^2\mathbf{I}_d\,.$$

Therefore the forward marginal is again a Gaussian mixture:

$$p_t(\mathbf{x}) = \sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}; \mu_{j,t}, \Sigma_{j,t})\,.$$

Exact score. The exact score of the noised GMM is available analytically as

$$\nabla \log p_t(\mathbf{x}) = \sum_{j=1}^{K} r_j(t,\mathbf{x})\left[-\Sigma_{j,t}^{-1}(\mathbf{x} - \mu_{j,t})\right]\,,$$

with

$$r_j(t,\mathbf{x}) := \frac{w_j \mathcal{N}(\mathbf{x}; \mu_{j,t}, \Sigma_{j,t})}{\sum_{\ell=1}^{K} w_\ell \mathcal{N}(\mathbf{x}; \mu_{\ell,t}, \Sigma_{\ell,t})}\,.$$

Exact reverse transition kernels. Let $0 = t_0 < t_1 < \cdots < t_N = T$ be the reverse-time grid. The reverse process $\overleftarrow{\mathbf{X}}_{t_k}$ has the same distribution as $\overrightarrow{\mathbf{X}}_{T-t_k}$ and therefore,

$$\mathsf{Q}_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) = \mathcal{L}\left(\overleftarrow{\mathbf{X}}_{t_{k+1}} \in \mathrm{d}\mathbf{z} \middle| \overleftarrow{\mathbf{X}}_{t_k} = \mathbf{x}\right) = \mathcal{L}\left(\overrightarrow{\mathbf{X}}_{T-t_{k+1}} \in \mathrm{d}\mathbf{z} \middle| \overrightarrow{\mathbf{X}}_{T-t_k} = \mathbf{x}\right)\,.$$

Set $u_k := T - t_k$, $u_{k+1} := T - t_{k+1}$, so that $u_{k+1} < u_k$ and

$$\mathsf{Q}_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) = \mathcal{L}\left(\overrightarrow{\mathbf{X}}_{u_{k+1}} \in \mathrm{d}\mathbf{z} \middle| \overrightarrow{\mathbf{X}}_{u_k} = \mathbf{x}\right)\,.$$

The pair $\left(\overrightarrow{\mathbf{X}}_{u_{k+1}}, \overrightarrow{\mathbf{X}}_{u_k}\right) | J = j$ is jointly Gaussian with mean

$$\begin{pmatrix} \mu_{j,u_{k+1}} \\ \mu_{j,u_k} \end{pmatrix},$$

and covariance matrix using (70)

$$\begin{pmatrix} \Sigma_{j,u_{k+1}} & m_{u_{k+1},u_k}\Sigma_{j,u_{k+1}} \\ m_{u_{k+1},u_k}\Sigma_{j,u_{k+1}} & \Sigma_{j,u_k} \end{pmatrix}.$$

By the Gaussian conditioning formula, for every $\mathbf{x} \in \mathbb{R}^d$,

$$\overrightarrow{\mathbf{X}}_{u_{k+1}} \mid \left\{\overrightarrow{\mathbf{X}}_{u_k} = \mathbf{x},\ J = j\right\} \sim \mathcal{N}\left(\mu_{j,u_{k+1}|u_k}(\mathbf{x}), \Sigma_{j,u_{k+1}|u_k}\right),$$

where

$$\begin{aligned} \mu_{j,u_{k+1}|u_k}(\mathbf{x}) &:= \mu_{j,u_{k+1}} + m_{u_{k+1},u_k}\Sigma_{j,u_{k+1}}\Sigma_{j,u_k}^{-1}(\mathbf{x} - \mu_{j,u_k}), \\ \Sigma_{j,u_{k+1}|u_k} &:= \Sigma_{j,u_{k+1}} - m_{u_{k+1},u_k}^2\Sigma_{j,u_{k+1}}\Sigma_{j,u_k}^{-1}\Sigma_{j,u_{k+1}}. \end{aligned}$$

By Bayes' formula, the probability that $J = j$ given $\overrightarrow{\mathbf{X}}_{u_k} = \mathbf{x}$ is

$$r_j(u_k, \mathbf{x}) = \frac{w_j\mathcal{N}(\mathbf{x}; \mu_{j,u_k}, \Sigma_{j,u_k})}{\sum_{\ell=1}^K w_\ell\mathcal{N}(\mathbf{x}; \mu_{\ell,u_k}, \Sigma_{\ell,u_k})}.$$

As a consequence

$$\mathsf{Q}_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) = \sum_{j=1}^K r_j(u_k, \mathbf{x})\mathcal{N}\left(\mathbf{z}; \mu_{j,u_{k+1}|u_k}(\mathbf{x}), \Sigma_{j,u_{k+1}|u_k}\right)\mathrm{d}\mathbf{z}. \tag{72}$$

Sampling with these exact kernels introduces no time-discretization error. In the unconditional case, initializing from p_T and applying $\mathsf{Q}_1 \cdots \mathsf{Q}_N$ recovers π_{data} up to Monte Carlo error, consistently with the ideal diffusion construction in Section 2.

Approximate proposals: separating discretization and score errors. The closed-form kernel Q_{k+1} in (72) will be used as the ideal reverse kernel. In order to disentangle time-discretization error from score approximation error, we introduce two time-changed Euler–Maruyama proposal kernels. First, define the oracle Euler proposal $\overline{\mathsf{Q}}_{k+1}$, obtained by applying one Euler step to the reverse SDE using the exact GMM score:

$$\overline{\mathsf{Q}}_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) = \mathcal{N}\left(\mathbf{z}; \mathbf{x} + \Delta_k\left[\alpha\mathbf{x} + 2\nabla\log p_{u_k}(\mathbf{x})\right], 2\Delta_k\mathbf{I}_d\right)\mathrm{d}\mathbf{z}, \qquad \Delta_k := \beta(t_{k+1} - t_k). \tag{73}$$

This kernel is biased only because it discretizes the reverse dynamics. Second, define the perturbed-score Euler proposal $\overline{\mathsf{Q}}_{k+1}^\theta$ by replacing the exact score with

$$\mathrm{s}_\theta(\mathbf{x}, u_k) = \nabla\log p_{u_k}(\mathbf{x}) + \varepsilon_{\mathrm{net}}\lambda(\mathbf{x}),$$

where $\lambda : \mathbb{R}^d \to \mathbb{R}^d$ is a fixed bounded vector field, while $\varepsilon_{\mathrm{net}}$ controls the bias magnitude. This perturbation is used as a controlled proxy for a learned-score error. The corresponding proposal is

$$\overline{\mathsf{Q}}_{k+1}^\theta(\mathbf{x}, \mathrm{d}\mathbf{z}) = \mathcal{N}\left(\mathbf{z}; \mathbf{x} + \Delta_k\left[\alpha\mathbf{x} + 2\,\mathrm{s}_\theta(\mathbf{x}, u_k)\right], 2\Delta_k\mathbf{I}_d\right)\mathrm{d}\mathbf{z}. \tag{74}$$

The difference between Q_{k+1} and $\overline{\mathsf{Q}}_{k+1}$ isolates the *time-discretization bias*, whereas the difference between $\overline{\mathsf{Q}}_{k+1}$ and $\overline{\mathsf{Q}}_{k+1}^\theta$ isolates *the score approximation bias*. In the experiments, the approximate mutation kernel used by SMC is $\overline{\mathsf{Q}}_{k+1}^\theta$.

In the reported experiments in Section F.3 we use bounded $\tanh$ perturbations. In the two-dimensional experiment, the globally biased Euler curve uses

$$\lambda_{\mathrm{bias}}(\mathbf{x}) = (0, \tanh(x_2)),$$

whereas the one-step forgetting diagnostic uses

$$\lambda_{\mathrm{loc}}(\mathbf{x}) = (\tanh(x_1), \tanh(x_2)).$$

In the 50-dimensional experiment in Section F.4, the same perturbations are applied only in the visible coordinates:

$$\lambda_{\mathrm{bias}}(\mathbf{x}) = (0, \tanh(x_2), 0, \ldots, 0), \qquad \lambda_{\mathrm{loc}}(\mathbf{x}) = (\tanh(x_1), \tanh(x_2), 0, \ldots, 0).$$

F.2 CONDITIONAL SAMPLING MODEL: NOISY INPAINTING.

Observation model. Let $d_{\mathrm{y}} \geq 1$ be the observation dimension. We consider a linear Gaussian observation of the clean variable,

$$Y = H\overrightarrow{\mathbf{X}}_0 + \varepsilon\,, \qquad H \in \mathbb{R}^{d_{\mathrm{y}} \times d}\,, \qquad \varepsilon \sim \mathcal{N}(0, R)\,, \qquad \varepsilon \perp \overrightarrow{\mathbf{X}}_0\,, \tag{75}$$

where R is symmetric positive definite with compatible dimensions. We condition on the observed value $Y = y \in \mathbb{R}^{d_{\mathrm{y}}}$. The target distribution is

$$\phi_N(\mathrm{d}\mathbf{x}) = \mathcal{L}(\overrightarrow{\mathbf{X}}_0 \in \mathrm{d}\mathbf{x} \mid Y = y)\,.$$

Since the prior is a Gaussian mixture and the likelihood is linear Gaussian, this target is again a Gaussian mixture and is available in closed form.

Lemma F.1 (Closed-form posterior for a GMM under linear Gaussian observation)**.** *Let π_{data} be defined as in* (69) *and Y as in* (75)*. Then, for every observation $y \in \mathbb{R}^{d_{\mathrm{y}}}$, the conditional law $\mathcal{L}(\overrightarrow{\mathbf{X}}_0 \in \mathrm{d}\mathbf{x} \mid Y = y)$ is the Gaussian mixture*

$$\phi_N(\mathrm{d}\mathbf{x}) = \sum_{j=1}^{K} \widetilde{w}_j \mathcal{N}\left(\mathbf{x}; \widetilde{\mu}_j, \widetilde{\Sigma}_j\right) \mathrm{d}\mathbf{x}\,, \tag{76}$$

where $\widetilde{\mu}_j$, $\widetilde{\Sigma}_j$ and $\widetilde{w}_j$ are defined in the proof.

Proof. Conditionally on $J = j$, the observation model $Y = H\overrightarrow{\mathbf{X}}_0 + \varepsilon$ is linear. Hence the joint density of $(\overrightarrow{\mathbf{X}}_0, Y, J)$ satisfies

$$p(\mathbf{x}, y, J = j) = w_j \mathcal{N}(\mathbf{x}; \mu_j, \Sigma_j)\, \mathcal{N}(y; H\mathbf{x}, R)\,. \tag{77}$$

For each j, completing the square in $\mathbf{x}$ yields

$$\mathcal{N}(\mathbf{x}; \mu_j, \Sigma_j)\, \mathcal{N}(y; H\mathbf{x}, R) \propto_{\mathbf{x}} \mathcal{N}\left(\mathbf{x}; \widetilde{\mu}_j, \widetilde{\Sigma}_j\right)\,, \tag{78}$$

with

$$\widetilde{\mu}_j = \widetilde{\Sigma}_j\left(\Sigma_j^{-1}\mu_j + H^\top R^{-1} y\right)\,, \qquad \widetilde{\Sigma}_j = \left(\Sigma_j^{-1} + H^\top R^{-1} H\right)^{-1}\,.$$

So we get

$$\mathcal{L}(\overrightarrow{\mathbf{X}}_0 \in \mathrm{d}\mathbf{x} \mid Y = y, J = j) = \mathcal{N}\left(\mathbf{x}; \widetilde{\mu}_j, \widetilde{\Sigma}_j\right) \mathrm{d}\mathbf{x}\,.$$

Moreover, since

$$Y \mid J = j \sim \mathcal{N}(H\mu_j, S_j)\,, \qquad S_j := H\Sigma_j H^\top + R\,, \tag{79}$$

we obtain

$$\mathcal{N}(\mathbf{x}; \mu_j, \Sigma_j)\, \mathcal{N}(y; H\mathbf{x}, R) = \mathcal{N}(y; H\mu_j, S_j)\, \mathcal{N}\left(\mathbf{x}; \widetilde{\mu}_j, \widetilde{\Sigma}_j\right)\,. \tag{80}$$

Integrating (77) with respect to $\mathbf{x}$, we get

$$p(y, J = j) = w_j \mathcal{N}(y; H\mu_j, S_j)\,.$$

Consequently, $p(y) = \sum_{\ell=1}^{K} w_\ell \mathcal{N}(y; H\mu_\ell, S_\ell)$. By Bayes' formula,

$$\widetilde{w}_j := \mathbb{P}(J = j \mid Y = y) = \frac{w_j \mathcal{N}(y; H\mu_j, S_j)}{\sum_{\ell=1}^{K} w_\ell \mathcal{N}(y; H\mu_\ell, S_\ell)}\,.$$

Hence, we obtain

$$\mathcal{L}(\overrightarrow{\mathbf{X}}_0 \in \mathrm{d}\mathbf{x} \mid Y = y) = \sum_{j=1}^{K} \widetilde{w}_j \mathcal{N}\left(\mathbf{x}; \widetilde{\mu}_j, \widetilde{\Sigma}_j\right) \mathrm{d}\mathbf{x}\,.$$

□

Bridge likelihoods. For $0 \le k \le N$, define the likelihood of the observation given the current reverse-time state by

$$h_k(\mathbf{x}) := p(y \mid \overleftarrow{\mathbf{X}}_{t_k} = \mathbf{x}) = p(y \mid \overrightarrow{\mathbf{X}}_{T-t_k} = \mathbf{x}). \tag{81}$$

Set $u_k := T - t_k$, conditionally on $J = j$, the pair $\left(Y, \overrightarrow{\mathbf{X}}_{u_k}\right)$ is jointly Gaussian, using (71) and (79), we have

$$Y \mid J = j \sim \mathcal{N}\left(H\mu_j, S_j\right), \qquad S_j := H\Sigma_j H^\top + R,$$

and

$$\overrightarrow{\mathbf{X}}_{u_k} \mid J = j \sim \mathcal{N}\left(\mu_{j,u_k}, \Sigma_{j,u_k}\right), \qquad \mu_{j,u_k} = m_{0,u_k}\mu_j, \qquad \Sigma_{j,u_k} = m_{0,u_k}^2\Sigma_j + \sigma_{0,u_k}^2\mathbf{I}_d.$$

Hence,

$$\mathrm{Cov}\left(Y, \overrightarrow{\mathbf{X}}_{u_k} \middle| J = j\right) = m_{0,u_k} H\Sigma_j.$$

Therefore, by the Gaussian conditioning formula,

$$Y \mid \left\{\overrightarrow{\mathbf{X}}_{u_k} = \mathbf{x},\ J = j\right\} \sim \mathcal{N}\left(\mu_{j,k}^y(\mathbf{x}), S_{j,k}^y\right),$$

with

$$\mu_{j,k}^y(\mathbf{x}) := H\mu_j + m_{0,u_k} H\Sigma_j\Sigma_{j,u_k}^{-1}(\mathbf{x} - \mu_{j,u_k}),$$
$$S_{j,k}^y := S_j - m_{0,u_k}^2 H\Sigma_j\Sigma_{j,u_k}^{-1}\Sigma_j H^\top.$$

The posterior probability of component j given the current state $\overrightarrow{\mathbf{X}}_{u_k} = \mathbf{x}$ is

$$r_j(u_k, \mathbf{x}) = \frac{w_j\mathcal{N}\left(\mathbf{x}; \mu_{j,u_k}, \Sigma_{j,u_k}\right)}{\sum_{\ell=1}^K w_\ell\mathcal{N}\left(\mathbf{x}; \mu_{\ell,u_k}, \Sigma_{\ell,u_k}\right)}.$$

Hence,

$$h_k(\mathbf{x}) = \sum_{j=1}^K r_j(u_k, \mathbf{x})\mathcal{N}\left(y; \mu_{j,k}^y(\mathbf{x}), S_{j,k}^y\right). \tag{82}$$

At the terminal reverse time $k = N$, we have $u_N = 0$. Hence

$$\mu_{j,u_N} = \mu_j, \qquad \Sigma_{j,u_N} = \Sigma_j,$$

and the conditional observation law reduces to

$$Y \mid \{\overrightarrow{\mathbf{X}}_0 = \mathbf{x}, J = j\} \sim \mathcal{N}\left(H\mathbf{x}, R\right).$$

Bridge weights and numerical implementation. The abstract Feynman–Kac model in Appendix B.3 is written in selection-before-propagation form: at time k, particles are first selected according to a state potential $g_k(\mathbf{x})$, and are then propagated with the mutation kernel Q_{k+1}. In this controlled numerical conditional-sampling experiment, the observation is more naturally introduced through a bridge weight depending on two consecutive states. We introduce,

$$g_k^{\mathrm{br}}(\mathbf{x}, \mathbf{z}) := \frac{h_{k+1}(\mathbf{z})}{h_k(\mathbf{x})}, \qquad k = 0, \dots, N-1, \tag{83}$$

The reason for using the ratio (83) is the telescoping identity

$$h_0(\overleftarrow{\mathbf{X}}_{t_0}) \prod_{k=0}^{N-1} g_k^{\mathrm{br}}(\overleftarrow{\mathbf{X}}_{t_k}, \overleftarrow{\mathbf{X}}_{t_{k+1}}) = h_N(\overleftarrow{\mathbf{X}}_{t_N}).$$

Hence, for every bounded measurable test function ψ,

$$\frac{\mathbb{E}\left[\psi(\overleftarrow{\mathbf{X}}_{t_N})h_0(\overleftarrow{\mathbf{X}}_{t_0})\prod_{k=0}^{N-1} g_k^{\mathrm{br}}(\overleftarrow{\mathbf{X}}_{t_k}, \overleftarrow{\mathbf{X}}_{t_{k+1}})\right]}{\mathbb{E}\left[h_0(\overleftarrow{\mathbf{X}}_{t_0})\prod_{k=0}^{N-1} g_k^{\mathrm{br}}(\overleftarrow{\mathbf{X}}_{t_k}, \overleftarrow{\mathbf{X}}_{t_{k+1}})\right]} = \frac{\mathbb{E}\left[\psi(\overleftarrow{\mathbf{X}}_{t_N})h_N(\overleftarrow{\mathbf{X}}_{t_N})\right]}{\mathbb{E}\left[h_N(\overleftarrow{\mathbf{X}}_{t_N})\right]}.$$

The right-hand side is precisely

$$\mathbb{E}[\psi(\overleftarrow{\mathbf{X}}_{t_N}) \mid Y = y].$$

Strictly speaking, (83) depends on both endpoints of the transition, whereas the abstract presentation uses state potentials $g_k(\mathbf{x})$. This is only a notational issue: the bridge formulation can be written as a standard Feynman–Kac model on the augmented state $(\overleftarrow{\mathbf{X}}_{t_k}, \overleftarrow{\mathbf{X}}_{t_{k+1}})$ as we will show in the next paragraph.

Relation with the selection-before-propagation form. For some generic proposal kernel Q_{k+1}, define

$$g_k^\star(\mathbf{x}) := \int Q_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) g_k^{\mathrm{br}}(\mathbf{x}, \mathbf{z}) = \frac{Q_{k+1} h_{k+1}(\mathbf{x})}{h_k(\mathbf{x})} ,$$

and, whenever $g_k^\star(\mathbf{x}) > 0$,

$$Q_{k+1}^\star(\mathbf{x}, \mathrm{d}\mathbf{z}) := \frac{Q_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) g_k^{\mathrm{br}}(\mathbf{x}, \mathbf{z})}{g_k^\star(\mathbf{x})} .$$

Then $Q_{k+1}^\star$ is a Markov kernel and

$$Q_{k+1}(\mathbf{x}, \mathrm{d}\mathbf{z}) g_k^{\mathrm{br}}(\mathbf{x}, \mathbf{z}) = g^\star(\mathbf{x}) Q_{k+1}^\star(\mathbf{x}, \mathrm{d}\mathbf{z}) .$$

Hence the transition-potential update defined by (83) can be written in the same selection-before-propagation form as Appendix B.3: first select with the state potential $\bar{g}_k^Q(\mathbf{x})$, then propagate with the twisted Markov kernel $\bar{Q}_{k+1}$.

F.3 Two-dimensional Gaussian mixture

The prior target is the Gaussian mixture

$$p_0(\mathbf{x}) = \sum_{j=1}^{4} w_j \mathcal{N}(\mathbf{x}; \mu_j, \Sigma_j) , \qquad \mathbf{x} \in \mathbb{R}^2,$$

with weights $(w_1, w_2, w_3, w_4) = (0.05, 0.15, 0.30, 0.50)$, means

$$\mu_1 = (-5, -4)^\top , \qquad \mu_2 = (-4, 4.5)^\top , \qquad \mu_3 = (4, -3.5)^\top , \qquad \mu_4 = (5, 4)^\top ,$$

and covariance matrices

$$\Sigma_1 = \begin{pmatrix} 0.20 & 0.12 \\ 0.12 & 0.50 \end{pmatrix} , \qquad \Sigma_2 = \begin{pmatrix} 0.80 & -0.35 \\ -0.35 & 0.30 \end{pmatrix} ,$$
$$\Sigma_3 = \begin{pmatrix} 0.35 & 0.20 \\ 0.20 & 0.90 \end{pmatrix} , \qquad \Sigma_4 = \begin{pmatrix} 0.50 & -0.25 \\ -0.25 & 0.40 \end{pmatrix} .$$

The components are deliberately separated and anisotropic, so that conditioning on a one-dimensional observation produces a posterior with two visible modes. The observation model is scalar and acts only on the first coordinate:

$$Y = H \overrightarrow{\mathbf{X}}_0 + \varepsilon , \qquad H = (1, 0) , \qquad \varepsilon \sim \mathcal{N}(0, R) , \qquad R = 0.16 .$$

In the reported experiment we condition on the realization $y = 3.8$. This observation is chosen so that the two left components receive negligible posterior mass, while the two right components remain visible with unequal weights.

We test different SMC sampling methods to approximate the posterior distribution. To make the comparison quantitative, we choose a visible region, the red rectangle, around one posterior mode. For each method, we estimate the posterior probability of this region by the weighted fraction of final particles falling inside it, and compare this estimate with the exact posterior probability, which is accurately estimated by standard numerical routines for the Gaussian CDF, as

$$\phi_N[\psi_R] = \sum_{j=1}^{4} \widetilde{w}_j \mathbb{P}(Z_j \in \mathcal{R}) , \qquad Z_j \sim \mathcal{N}\left(\widetilde{\mu}_j, \widetilde{\Sigma}_j\right) . \tag{84}$$

The selected rectangle $\mathcal{R}$ is the red rectangle in Figure 2. The left panel shows samples from the original GMM and the right panel shows samples from the exact conditional posterior. Points are colored according to whether $\psi_R(\mathbf{x}) = 1$. Moreover, we show a representative conditional particle cloud for the three mutation mechanisms used in the experiment. The first panel in Figure 3 shows independent samples from the exact posterior GMM. The remaining panels show the final SMC particles obtained when the mutation kernels are, respectively, the exact reverse kernels $\{\mathsf{Q}_{k+1}\}_{k=0}^{N-1}$ in (72), the oracle Euler proposals $\{\overline{\mathsf{Q}}_{k+1}\}_{k=0}^{N-1}$ in (73), and the perturbed-score Euler proposals $\{\overline{\mathsf{Q}}_{k+1}^{\theta}\}_{k=0}^{N-1}$ in (74). This plot is meant as a qualitative diagnostic: it shows whether the particle system places mass in the correct posterior regions and around the selected event $\{\psi_R = 1\}$.

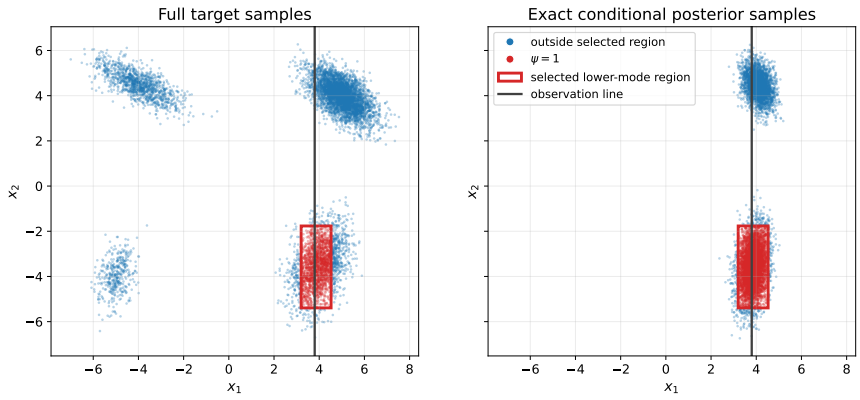


Figure 2: Two-dimensional Gaussian-mixture benchmark. The red rectangle defines the event $\{\psi_{\mathcal{R}} = 1\}$ used to evaluate the samplers. Left: samples from the original Gaussian mixture. Right: samples from the exact conditional posterior. Blue dots correspond to points outside the selected region, while red dots correspond to points inside it. The black vertical line indicates the observed value $y = 3.8$.

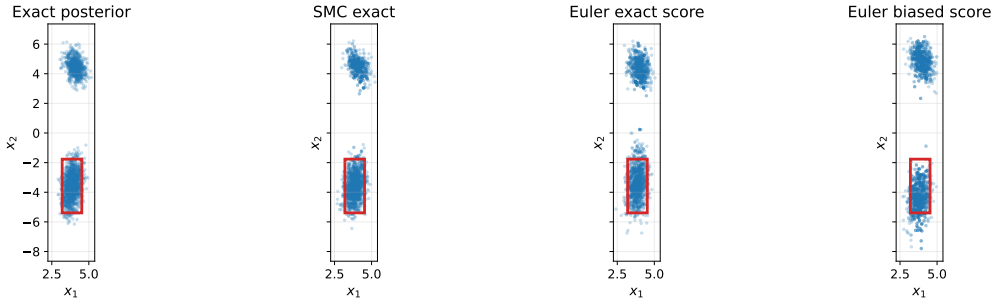


Figure 3: Conditional samples in the two-dimensional benchmark. The first panel shows samples from the exact posterior GMM. The other panels show final SMC particle clouds obtained with the exact reverse kernels Q_{k+1}, the oracle Euler kernels $\overline{\mathsf{Q}}_{k+1}$, and the perturbed-score Euler kernels $\overline{\mathsf{Q}}^{\theta}_{k+1}$. The red rectangle is the modal region defining ψ_R.

Illustration of Theorem 3.6 We next quantify the two effects predicted by Theorem 3.6: *the finite-particle Monte Carlo error* and *the deterministic bias* induced by approximate mutation kernels.

For each particle count M, we run independent SMC repetitions and report the empirical L_2 error with respect to the exact posterior value $\phi_N[\psi_{\mathcal{R}}]$ given in (84). The three curves in Figure 4 (Left) and in Figure 5 (same plot in log scale) correspond to SMC with the exact reverse kernels Q_{k+1}, the Euler kernels with true score function $\overline{\mathsf{Q}}_{k+1}$, and the Euler kernels with perturbed-score $\overline{\mathsf{Q}}^{\theta}_{k+1}$. The exact reverse kernels introduce no time-discretization error and therefore exhibit the expected Monte Carlo decay. The Euler kernels with true score function isolate the effect of time discretization, while the perturbed-score Euler kernels exhibit an additional non-vanishing bias plateau.

We then isolate the forgetting of a local kernel perturbation, which includes both discretization and score-approximation effects. In this experiment all mutation kernels are exact, except at one selected reverse step where Q_{k+1} is replaced by a perturbed-score Euler kernel $\overline{\mathsf{Q}}^{\theta}_{k+1}$. We vary the location of this single perturbation and measure the resulting absolute bias in the final estimate of $\phi_N[\psi_{\mathcal{R}}]$. The horizontal axis in Figure 4 (Right) is the number of remaining reverse steps after the perturbation. Moving to the right therefore corresponds to placing the local perturbation earlier in the reverse samlping process, closer to the initial distribution p_T and farther from the data distribution p_0, leaving more subsequent kernels through which its effect can be forgotten.

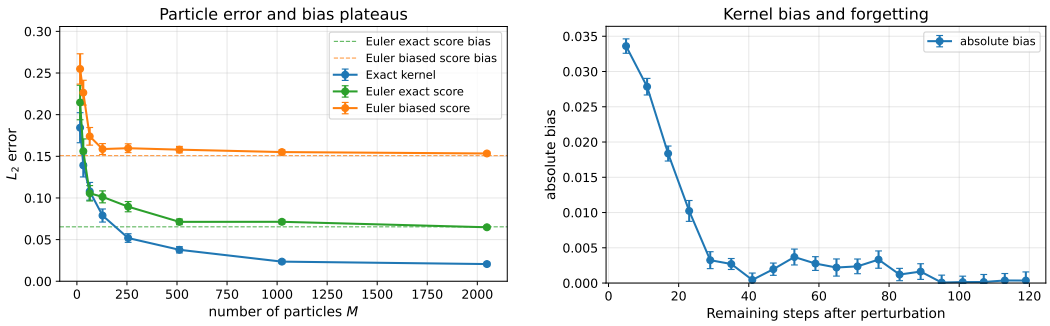


Figure 4: Illustration of the two terms in Theorem 3.6. Left: empirical L_2 error as a function of the number of particles M, showing Monte Carlo decay and deterministic bias plateaus for approximate mutation kernels. Right: propagated effect of a single local kernel perturbation as a function of the number of remaining bridge steps, illustrating forward-smoothing forgetting. On the right panel, the left side corresponds to perturbations close to the data distribution p_0, while the right side corresponds to perturbations closer to the initial noised distribution p_T. Error bars show empirical standard errors across independent repetitions.

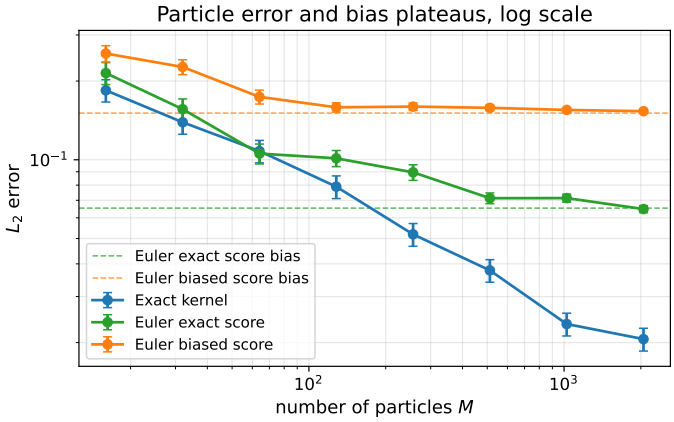


Figure 5: Log-scale version of the particle error and kernel bias plot for the two-dimensional GMM. The empirical L_2 error is plotted as a function of the number of particles M, using the same repetitions and methods as in Figure 4. The log scale makes the initial Monte Carlo decay and the subsequent bias plateaus easier to distinguish.

As an additional diagnostic, Figure 6 varies the number of reverse steps N while keeping the particle count fixed. This plot checks that the discrepancy of the oracle Euler kernel is indeed tied to the time discretization of the reverse dynamics. Increasing N refines the reverse-time grid, and therefore reduces the discretization component of the error, up to the remaining finite-particle variability and score perturbation bias.

F.4 A 25-component Gaussian mixture in dimension 50

We next consider a higher-dimensional benchmark. The data distribution is a Gaussian mixture on $\mathbb{R}^{50}$ with $K = 25$ components,

$$p_0(\mathbf{x}) = \sum_{j=1}^{25} w_j \mathcal{N}(\mathbf{x}; \mu_j, \Sigma_j) , \qquad \mathbf{x} \in \mathbb{R}^{50} .$$

The component means are arranged on a 5×5 grid in the first two coordinates and are zero in the remaining coordinates. More precisely, if $(a_j, b_j)_{j=1}^{25}$ denotes the grid $\{-10, -5, 0, 5, 10\}^2$, then

$$\mu_j = (a_j, b_j, 0, \ldots, 0)^\top \in \mathbb{R}^{50}.$$

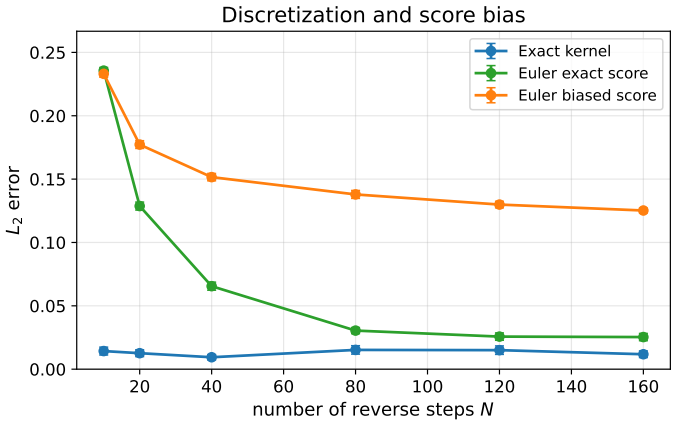


Figure 6: Discretization diagnostic for the two-dimensional GMM. The empirical L_2 error is plotted as a function of the number of reverse steps N, with the particle count kept fixed. This separates the effect of refining the Euler time discretization from the particle-count experiment in Figure 4.

The weights are generated once from independent chi-square random variables with three degrees of freedom and then normalized. The covariance matrices are full and anisotropic: for each component,

$$\Sigma_j = U_j \,\mathrm{diag}(1, 2^{-1}, \ldots, 50^{-1}) U_j^\top ,$$

where U_j is an orthogonal matrix obtained from a random Gaussian matrix. All random quantities defining the mixture are fixed using the same seed in all runs.

As in the two-dimensional experiment, we use a scalar observation acting only on the first coordinate,

$$Y = H\overrightarrow{\mathbf{X}}_0 + \varepsilon\,, \qquad H = (1, 0, \ldots, 0)\,, \qquad \varepsilon \sim \mathcal{N}(0, R)\,, \qquad R = 1\,.$$

In the reported experiment we condition on the fixed realization $y = 5.0$. The observation is not resampled across repetitions: all methods are compared on the same conditional target $\mathcal{L}(\overrightarrow{\mathbf{X}}_0 \mid Y = y)$.

Although the sampler evolves in the full ambient space $\mathbb{R}^{50}$, the test function is defined through the first two coordinates. Let $\Pi_{1:2}\mathbf{x} = (x_1, x_2)$ and let $\mathcal{R} \subset \mathbb{R}^2$ be an axis-aligned rectangle centered at the dominant posterior mode in the projected (x_1, x_2) plane. We define, $\psi_{\mathcal{R}}(\mathbf{x}) := 1_{\mathcal{R}}(\Pi_{1:2}\mathbf{x})$. Thus $\psi_{\mathcal{R}}$ measures the posterior probability assigned to one projected modal region, while still testing an SMC sampler operating in dimension 50.

The exact posterior value is again computed deterministically from the closed-form posterior GMM. If $\widetilde{\mu}_{j,1:2}$ and $\widetilde{\Sigma}_{j,1:2}$ denote the first two coordinates and the corresponding 2×2 marginal covariance of posterior component j, then

$$\phi_N[\psi_{\mathcal{R}}] = \sum_{j=1}^{25} \widetilde{w}_j \mathbb{P}(Z_j \in \mathcal{R})\,, \qquad Z_j \sim \mathcal{N}\left(\widetilde{\mu}_{j,1:2}, \widetilde{\Sigma}_{j,1:2}\right)\,.$$

Figure 7 displays the projected geometry of the test function. The figure is only a two-dimensional visualization of a 50-dimensional experiment: the SMC sampler, the weights, and the proposal kernels all act in the full ambient space.

We also show representative projected conditional particle clouds in Figure 8. The first panel shows independentssamples from the exact posterior GMM, while the remaining panels show the final SMC particles obtained with the exact reverse kernels Q_{k+1}, the oracle Euler kernels $\overline{\mathsf{Q}}_{k+1}$, and the perturbed-score Euler kernels $\overline{\mathsf{Q}}^{\theta}_{k+1}$. The projection makes it possible to inspect whether the particle system places mass in the correct visible posterior regions.

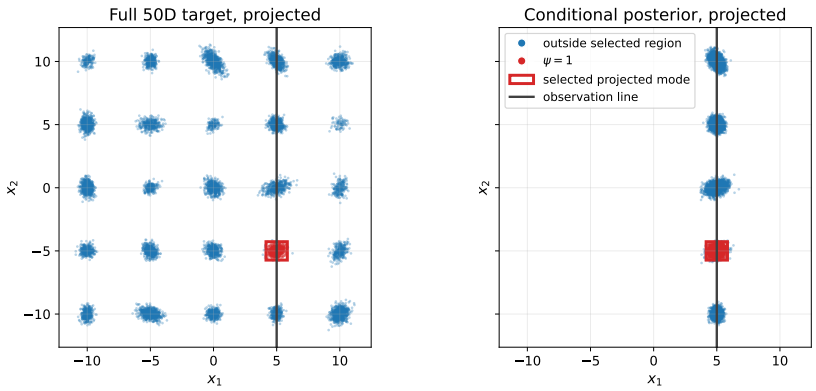


Figure 7: Projected geometry of the 50-dimensional test function. The sampler evolves in $\mathbb{R}^{50}$, but the plotted coordinates are (x_1, x_2). The red rectangle defines the projected event $\{\psi_{\mathcal{R}} = 1\}$ around the selected posterior mode. Left: samples from the original GMM projected onto the first two coordinates. Right: samples from the exact conditional posterior projected onto the same plane. The black vertical line indicates the observed value $y = 5$

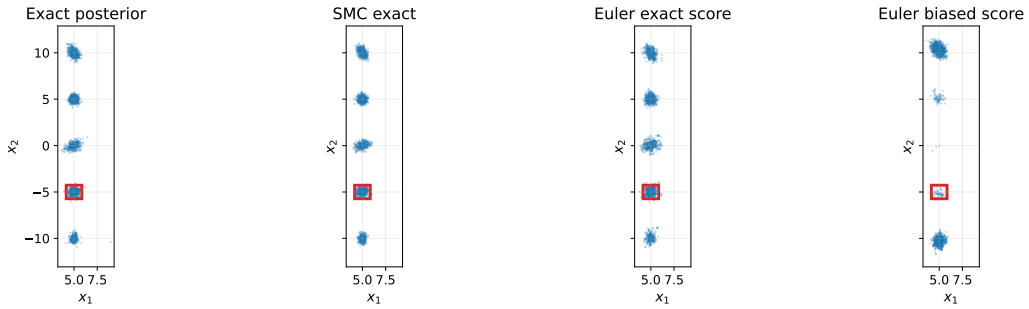


Figure 8: Conditional samples in the 50-dimensional benchmark, projected onto the first two coordinates. The first panel shows samples from the exact posterior GMM. The other panels show final SMC particle clouds obtained with the exact reverse kernels Q_{k+1}, the oracle Euler kernels $\overline{\mathsf{Q}}_{k+1}$, and the perturbed-score Euler kernels $\overline{\mathsf{Q}}^{\theta}_{k+1}$. The red rectangle is the projected modal region defining $\psi_{\mathcal{R}}$.

Illustration of Theorem 3.6 We repeat the same particle-error experiment as in the two-dimensional benchmark. For each particle count M, we run independent SMC repetitions and report the empirical L_2 error with respect to the exact posterior value $\phi_N[\psi_{\mathcal{R}}]$. The three curves in Figure 9 correspond to the exact reverse kernels, the oracle Euler kernels, and the perturbed-score Euler kernels. This experiment checks whether the same Monte Carlo decay and deterministic bias plateau remain visible in a substantially higher-dimensional ambient space.

We finally report the local-perturbation diagnostic in the 50-dimensional setting. As before, all mutation kernels are exact except at one selected reverse step, where the exact kernel is replaced by a perturbed-score Euler kernel. This diagnostic is more demanding than in the two-dimensional example: the sampler evolves in the full 50-dimensional space, while the test function is a discontinuous indicator depending only on a projected modal region. As a result, the attenuation pattern is less smooth, but the experiment still illustrates how the effect of a local kernel perturbation depends on the number of subsequent remaining steps.

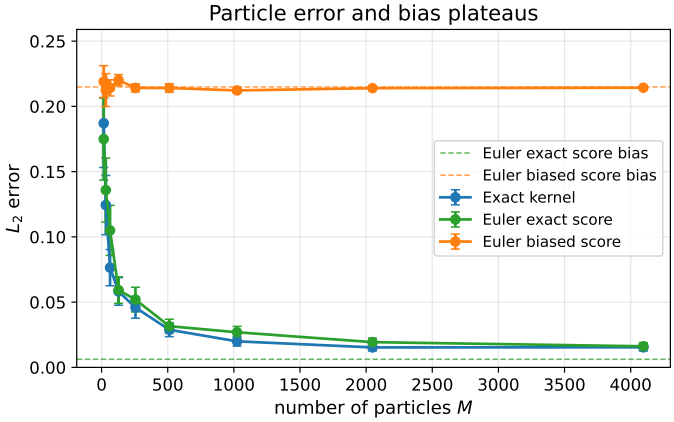


Figure 9: Particle error and kernel bias for the 50-dimensional benchmark. The empirical L_2 error is plotted as a function of the number of particles M, using 30 independent SMC repetitions for each value of M and each method. The horizontal dashed lines show large-particle estimates of the Euler bias plateaus, computed with $M_{\mathrm{ref}} = 40\,000$ particles and 4 independent reference runs. Error bars show empirical standard errors across repetitions.

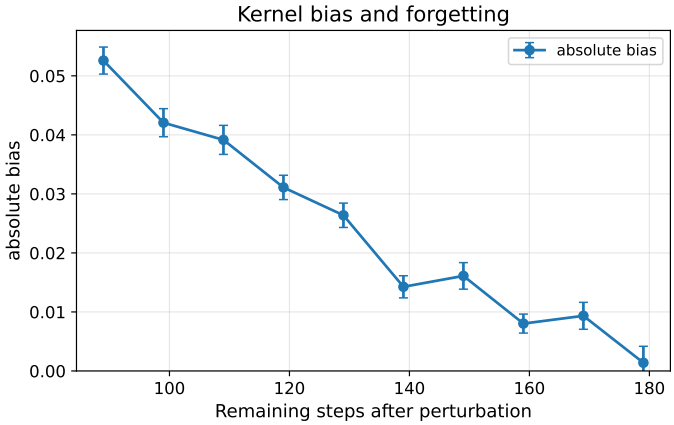


Figure 10: Kernel bias and forgetting for the 50-dimensional benchmark. A single reverse transition is replaced by the perturbed-score Euler kernel $\overline{\mathrm{Q}}^{\theta}_{k+1}$. Error bars show empirical standard errors across independent repetitions.