ITERATING CONSISTENCY MODELS: STABILITY, ERROR BOUNDS AND NOISE SCHEDULES

Anonymous authors
Paper under double-blind review

ABSTRACT

Consistency models (CMs) have become a leading approach for generating high-quality samples in few steps. However, adding steps can improve or degrade sample quality in ways that are highly sensitive to the schedule and that existing theory does not fully explain. To provide accuracy guarantees and guide CM sampler design, we analyze multistep CM sampling as a composition of noising and approximate denoising operators. Under explicit, verifiable stability assumptions, we derive a non-asymptotic error bound that separates contraction of the initialization error from accumulation of approximation error. The bound assigns distinct roles to the schedule: large early noise levels drive contraction, while small late noise levels control the residual bias. As a corollary, we obtain explicit constants for strongly log-concave and semi-log-concave targets. We further establish a complementary guarantee whose assumptions, one-step accuracy and stability, can be estimated for a given trained model. Experiments show that the contraction and approximation profiles entering our bounds can be reliably measured and closely match the predicted functional forms. Together, these results provide a meaningful convergence theory for multi-step CMs and a practical route to sampler design.

1 INTRODUCTION

Diffusion models (DMs) (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Song et al., 2021; Ho et al., 2020) have revolutionized generative machine learning, achieving remarkable results in image (Ramesh et al., 2022; Rombach et al., 2022; Zhang et al., 2023), video (Ho et al., 2022; Blattmann et al., 2023; Wan et al., 2025), 3D (Liu et al., 2023b; Poole et al., 2022; Wang et al., 2023), and audio generation (Evans et al., 2024; Liu et al., 2023a). Early DMs required hundreds to thousands of neural function evaluations (NFEs) per sample. However, distillation substantially reduces this cost, such as direct (Luhman & Luhman, 2021; Zheng et al., 2022), adversarial (Wang et al., 2022; Sauer et al., 2024), or variational score distillation (Wang et al., 2023; Yin et al., 2024b;a; Luo et al., 2023b; Salimans et al., 2024). Within this context, consistency models (CMs) (Song et al., 2023; Song & Dhariwal, 2023; Kim et al., 2023) have emerged as particularly effective few-step DMs. A CM learns a *consistency function* that maps noisy points along a probability-flow ODE trajectory directly to clean data. It can be distilled from a pre-trained DM (consistency distillation) or trained from scratch (consistency training).

Training CMs is nonetheless delicate. Discrete-time training introduces a discretization error that degrades sample quality, whereas naive continuous-time training is notoriously unstable (Song et al., 2023; Song & Dhariwal, 2023; Geng et al., 2024). sCM (Lu & Song, 2025) addressed this with TrigFlow, a formulation that greatly improves stability. A closely related line of work (Geng et al., 2025; Sabour et al., 2025; Zhou et al., 2025) learns *flow maps* (Boffi et al., 2025a), which transport samples between any two points of the trajectory and include CMs as an important special case. Methods such as IMM (Zhou et al., 2025), MeanFlow (Geng et al., 2025), and SCoT (Wu et al., 2025) make training more stable still by adding constraints to the learning objective.

In principle, CMs can generate a sample with a single network evaluation, but practical samplers often apply the model several times to improve sample quality. Each iteration re-noises the current sample and maps it back to the target with the learned consistency function. Multistep CMs, however, are brittle: additional steps do not always help, sample quality is highly sensitive to the noise schedule,

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.

and stopping early sometimes yields better samples than continuing. Existing theory for multistep samplers is limited and offers little guidance to practitioners designing or deploying CMs.

Intuitively, multistep CM sampling trades off two competing effects. On the one hand, each additional step can reduce the discrepancy between the generated and target distributions. On the other hand, each step introduces a new error from model bias. The key question is thus whether the contraction gained at each step outweighs the error it introduces. This paper makes that balance explicit.

To this end, we study multistep samplers that alternately apply stochastic noising and deterministic denoising operators to probability distributions. Our main contribution is a non-asymptotic convergence theory for multistep CMs that explains how the schedule governs sample quality and provides a principled basis for sampler design. Specifically:

- **Error bound and the role of the schedule.** We derive a non-asymptotic error bound that separates contraction of the initialization error from accumulated approximation error. The schedule plays two roles: large early steps drive contraction, while small late steps control the residual bias. For strongly log-concave and semi-log-concave targets, we obtain explicit constants.
- **Guarantees for trained CMs.** We provide a complementary guarantee whose assumptions, one-step accuracy and stability of the learned map, can be estimated for any trained model.
- **Empirical validation.** Experiments show that the contraction and approximation profiles in our bounds can be measured reliably and closely match the functional forms predicted by our theory.

Relation to existing work. Convergence theory for CM sampling is recent. Lyu et al. (2024) analyze one-step and multistep sampling under Ornstein–Uhlenbeck noising. Assuming that the learned map is Lipschitz and that the score and consistency training errors are small, they bound the Wasserstein error of the sampler and show that repeating a suitable noise level reduces this error geometrically, down to a plateau set by the training errors. Chen et al. (2025) derive multistep guarantees for bounded targets via a Kullback–Leibler recursion, and other work addresses statistical accuracy and consistency training (Dou et al., 2024; Li et al., 2025).

We build on the error recursion of Lyu et al. (2024) but establish its ingredients under different assumptions. First, as they suggest, we place the stability condition required for contraction on the exact denoiser rather than on the learned network, and show that it holds under natural conditions on the target, such as semi-log-concavity. Second, we allow general Gaussian noising processes and arbitrary schedules. Third, we adopt a more realistic error model in which the score error may diverge as the noise level tends to zero (De Bortoli, 2022, Assumption A3), where noising no longer smooths the target, rather than remaining uniformly small. We make the correspondence between the two sets of results precise after Proposition 3.1, in Section 3.3, and in Appendix C.

2 Background

Notation. We work on $\mathbb{R}^d$ with the Euclidean norm $\|\cdot\|$, and denote by $\mathcal{P}_2(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$ with finite second moment. For a measurable map F and $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, the pushforward $F_\#\rho$ is the law of $F(X)$ with $X \sim \rho$. We identify a measure with its Lebesgue density whenever the latter exists. For $r > 0$, we define the Gaussian smoothing $\rho_r := \kappa_r * \rho$, where $\kappa_r(x) := (2\pi r^2)^{-d/2} \exp\big(-\|x\|^2/(2r^2)\big)$; equivalently, ρ_r is the law of $X + rZ$ with $X \sim \rho$ and $Z \sim \mathcal{N}(0, \mathbb{I})$ independent. The 2-Wasserstein distance between $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is $W_2^2(\rho_1, \rho_2) := \inf_{\Gamma \in \Pi(\rho_1, \rho_2)} \int \|x - y\|^2 \, \Gamma(\mathrm{d}x, \mathrm{d}y)$, where $\Pi(\rho_1, \rho_2)$ is the set of couplings of ρ_1 and ρ_2. A dot denotes differentiation in time, e.g. $\dot{x}_t = \mathrm{d}x_t/\,\mathrm{d}t$.

Diffusion models. Diffusion models (DMs) draw approximate samples from a target distribution $\pi \in \mathcal{P}_2(\mathbb{R}^d)$ by reversing an $\mathbb{R}^d$-valued noising process $X_t = \alpha_t X_0 + \sigma_t Z$, where $X_0 \sim \pi$ and $Z \sim \mathcal{N}(0, \mathbb{I})$ are independent. The schedules α and σ are non-negative and differentiable, with $\alpha_0 = 1$, $\sigma_0 = 0$ and $\alpha_t > 0$ on the sampling interval. The noise-to-signal ratio $\gamma_t := \sigma_t/\alpha_t$ is non-decreasing, with $\gamma_t > 0$ for $t > 0$. For $t > 0$, the law of X_t has density $\varpi_t(x) = \alpha_t^{-d} \pi_{\gamma_t}(x/\alpha_t)$. The reversal is typically implemented via the probability-flow ODE (Song et al., 2021)

$$\dot{x}_t = a_t x_t - \tfrac{1}{2} b_t^2 \nabla \log \varpi_t(x_t), \qquad a_t = \dot{\alpha}_t/\alpha_t, \quad b_t^2 = 2\alpha_t \sigma_t \dot{\gamma}_t, \tag{1}$$

whose solutions preserve the marginals $(\varpi_t)_{t>0}$, when the flow is solved backwards from X_t. Let $\Phi_{t\to 0}$ denote the resulting map from time t to time 0, note that $(\Phi_{t\to 0})_{\#}\varpi_t = \pi$. Although easy to train, early DMs approximate the score $\nabla \log \varpi_t$ rather than the solution of (1), and must therefore integrate (1) numerically to 0, at a cost of hundreds of neural function evaluations (NFEs) per sample. This drawback has been addressed through so-called distilled DMs, notably consistency models.

Consistency models. Consistency models (CMs) approximate the flow map $\Phi_{t\to 0}$, predicting the endpoint at time 0 of a state at time t in a single NFE. They are trained either by distilling a pre-trained DM or directly from data via consistency training (Song et al., 2023; Song & Dhariwal, 2023; Boffi et al., 2025b), using two properties of the flow map: self-consistency, $\Phi_{t\to 0}(x_t) = \Phi_{s\to 0}(x_s)$ for states on the same trajectory of (1), and the boundary condition $\Phi_{0\to 0}(x) = x$. The choice of α_t and σ_t also affects training stability. Two common choices are the variance-exploding (VE) schedule, $\alpha_t = 1$ and $\sigma_t = t$ for $t \in [0, \infty)$ (Song et al., 2023; Song & Dhariwal, 2023), and the TrigFlow schedule, $\alpha_t = \cos t$ and $\sigma_t = \sin t$ for $t \in [0, \pi/2)$ (Geng et al., 2024; Lu & Song, 2025).

Since the learned CM map $\widetilde{\Phi}^{\theta}_{t\to 0}$, a neural network with weights θ, only approximates $\Phi_{t\to 0}$, a single CM step can yield insufficiently accurate samples. Multistep sampling alternates forward noising with reverse transport. Song et al. (2023) and Heek et al. (2024) show that these iterations sometimes improve sample quality. Namely, starting from an initialisation $X_0^{(0)}$, iteration $k \geq 1$ computes

$$X_{t_k}^{(k)} = \alpha_{t_k} X_0^{(k-1)} + \sigma_{t_k} Z^{(k)}, \qquad X_0^{(k)} = \widetilde{\Phi}^{\theta}_{t_k\to 0}\big(X_{t_k}^{(k)}\big), \tag{2}$$

where $X_0^{(0)}, Z^{(1)}, \ldots, Z^{(n)}$ are independent, with $Z^{(k)} \sim \mathcal{N}(0, \mathrm{Id})$, and $t_1 > \cdots > t_n > 0$. In practice, 2–8 iterations often suffice for excellent samples, yet more iterations can improve or degrade quality in ways that depend sensitively on the schedule and that existing theory does not fully explain.

We herein close this gap with non-asymptotic bounds on the distance between the law of $X_0^{(k)}$ and π, explicit in the initialization, schedule and time steps. The bounds separate contraction of the initialization error, driven by large early noise levels, from accumulation of approximation error, controlled by small late noise levels. Crucially, their assumptions can be estimated directly from a pre-trained model, so the bounds can be evaluated in practice rather than serving only as qualitative guidance. For strongly log-concave and semi-log-concave targets, all constants are explicit.

3 Main results

3.1 Convergence of multistep consistency models

Let $\pi \in \mathcal{P}_2(\mathbb{R}^d)$ be the target. We describe the sampler (2) through its action on probability distributions. For $t > 0$, the noising operator $N_t : \mathcal{P}_2(\mathbb{R}^d) \to \mathcal{P}_2(\mathbb{R}^d)$ maps the law of X to the law of $\alpha_t X + \sigma_t Z$, with $Z \sim \mathcal{N}(0, \mathrm{Id})$ independent of X, so that $N_t(\pi) = \varpi_t$; the exact denoising operator is the pushforward by the flow map of (1), $D_{\pi,t}(\rho) = (\Phi_{t\to 0})_{\#}\rho$; and the learned denoising operator is $\widetilde{D}^{\theta}_{\pi,t}(\rho) = (\widetilde{\Phi}^{\theta}_{t\to 0})_{\#}\rho$. The exact and learned sampling steps are $\mathcal{L}_t = D_{\pi,t} \circ N_t$ and $\widetilde{\mathcal{L}}^{\theta}_t = \widetilde{D}^{\theta}_{\pi,t} \circ N_t$, so that the law ρ_k of $X_0^{(k)}$ in (2) satisfies $\rho_k = \widetilde{\mathcal{L}}^{\theta}_{t_k}(\rho_{k-1})$, $k \geq 1$, where ρ_0 is the law of $X_0^{(0)}$. The results of this section use only the properties of these operators stated below, and therefore hold for any maps $D_{\pi,t}, \widetilde{D}^{\theta}_{\pi,t} : \mathcal{P}_2(\mathbb{R}^d) \to \mathcal{P}_2(\mathbb{R}^d)$ with $\widetilde{\mathcal{L}}^{\theta}_t$ preserving $\mathcal{P}_2(\mathbb{R}^d)$, which is the case when $\widetilde{\Phi}^{\theta}_{t\to 0}$ has at most linear growth. Fix finite nonnegative profiles R_t, L_t, E_t for noising stability, denoising stability and approximation error, respectively.

A1. *For every $t > 0$, $\mathcal{L}_t(\pi) = \pi$.*

A2. *For all $t > 0$ and $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$, $W_2\big(N_t(\rho_1),\, N_t(\rho_2)\big) \;\leq\; R_t\, W_2(\rho_1, \rho_2)$.*

A3. *For all $t > 0$ and $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$, $W_2\big(D_{\pi,t}(\rho_1),\, D_{\pi,t}(\rho_2)\big) \;\leq\; L_t\, W_2(\rho_1, \rho_2)$.*

A4. *For every $t > 0$ and $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, $W_2\big(\mathcal{L}_t(\rho),\, \widetilde{\mathcal{L}}^{\theta}_t(\rho)\big) \leq E_t$.*

Proposition 3.1 (Finite-horizon sampling error)**.** *Let A1–A4 hold and set $q_t := R_t L_t$. Fix $T > 0$ and an integer $n \geq 1$. Let $t_1, \ldots, t_n \in (0, T]$ be an arbitrary schedule and $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ an initialization. Define $\rho_k := \widetilde{\mathcal{L}}^{\theta}_{t_k}(\rho_{k-1})$ for $k = 1, \ldots, n$ and $Q_{k,j} := \prod_{i=k}^{j} q_{t_i}$ for $1 \leq k \leq j \leq n$, with $Q_{j+1,j} = 1$.*

Then the error after n steps satisfies

$$W_2(\rho_n, \pi) \le Q_{1,n} W_2(\rho_0, \pi) + \sum_{k=1}^{n} Q_{k+1,n} E_{t_k} . \tag{3}$$

If, additionally, the schedule satisfies the balance condition $q_{t_k} E_{t_{k-1}} \le \xi_n E_{t_k}$ for $k = 2, \dots, n$ and some $\xi_n \in (0, 1)$, then

$$\sum_{k=1}^{n} Q_{k+1,n} E_{t_k} \le E_{t_n} \frac{1 - \xi_n^n}{1 - \xi_n} \le \frac{E_{t_n}}{1 - \xi_n}. \tag{4}$$

The schedule may repeat noise levels and need not be decreasing. The balance condition also covers zero error bounds; for $n = 1$ it is empty and the sum equals E_{t_1}.

The proof is in Appendix B.

The proposition uses a uniform error bound for the learned step and stability of the exact step; it does not require the learned step to be contractive. The result alone implies neither an optimal finite number of steps nor deterioration when more steps are used.

Repeating a fixed noise level. For a fixed $t > 0$ with $0 \le q_t < 1$, let $\rho_n = (\widetilde{\mathcal{L}}_t^\theta)^n(\rho_0)$. The geometric sum in (3) gives $W_2(\rho_n, \pi) \le q_t^n W_2(\rho_0, \pi) + E_t(1 - q_t^n)/(1 - q_t)$ for $n \ge 1$. Iteration attenuates the initial-distance term, while the accumulated-error bound approaches $E_t/(1 - q_t)$ (for their factors, this is the constant-level bound of Lyu et al. (2024, Corollary 3.11), minimized over the level). Repeating a noisy step can therefore lower the certificate for a large initial error, while reducing the final bound may require a level with a smaller ratio $E_t/(1 - q_t)$.

Adding one step. Given a certificate $\bar{d} \ge W_2(\rho, \pi)$, a step at level t yields $q_t \bar{d} + E_t$, which improves on $\bar{d}$ exactly when $E_t < (1 - q_t)\bar{d}$, i.e., when $\bar{d}$ exceeds the plateau $E_t/(1 - q_t)$; no improvement is possible for $q_t \ge 1$. This is the case distinction in the proof of Lyu et al. (2024, Corollary 3.11), here valid at every level with $q_t < 1$ and for both placements of the hypotheses (Proposition 3.9). An extra step pays off only if its attenuation outweighs its error budget, and lowering the noise level reduces E_t but weakens attenuation (Section 3.2). These statements compare certificates, not actual errors.

Chen et al. (2025, Theorem 2 and Section 3.3) use a related decomposition: appending a lower-noise step increases their accumulated-error term but reduces the final denoising-error term, a trade-off they exploit to choose two-step OU and geometric VE schedules. Their recursion, however, propagates the KL divergence between noisy marginals with coefficient one, while our factors q_t quantify when later steps attenuate earlier errors. Remark 3.8 and Appendix C compare the assumptions and bounds.

3.2 Noise-dependent stability and approximation error

We specialize A1–A4 to the operators of Section 3.1 when the learned map is an approximate flow of the probability-flow ODE. Replacing $\nabla \log \varpi_t(\cdot)$ in (1) by an approximation $s_\theta(\cdot, t)$ gives the approximate-score ODE $\dot{x}_t = a_t x_t - \frac{1}{2} b_t^2 s_\theta(x_t, t)$. In this section the learned map $\widetilde{\Phi}_{t\to 0}^\theta$ is its flow map from time t to 0, so that $\widetilde{D}_{\pi,t}^\theta(\rho) = (\widetilde{\Phi}_{t\to 0}^\theta)_\# \rho$ is the denoiser induced by the exact flow of the approximate score. The score is not learned directly. If $\widetilde{\Phi}_{t\to 0}^\theta$ is a diffeomorphism for every t, depends differentiably on t and is the identity at $t = 0$, the curves $t \mapsto (\widetilde{\Phi}_{t\to 0}^\theta)^{-1}(x_0)$ have a velocity field V_t, and $s_\theta(x, t) := 2(a_t x - V_t(x))/b_t^2$ is the score implied by the learned map wherever $b_t > 0$.

With these definitions, A1 holds by construction: $N_t(\pi) = \varpi_t$ and, since the probability-flow ODE (1) transports ϖ_t to $\varpi_0 = \pi$ (Song et al., 2021), $\mathcal{L}_t(\pi) = (\Phi_{t\to 0})_\# \varpi_t = \pi$ for all $t > 0$.

We now verify A2–A4 for these operators, starting with the stability of N_t.

Lemma 3.2. *Let α_t be the signal coefficient of the noising process of Section 2, so that N_t maps the law of X to that of $\alpha_t X + \sigma_t Z$. Then N_t is α_t-Lipschitz for W_2, and α_t is its exact Lipschitz constant; that is, A2 holds with $R_t = \alpha_t$.*

The proof is in Appendix B; we now turn to A3 and A4.

Theorem 3.3. *Let $\pi \in \mathcal{P}_2(\mathbb{R}^d)$ and $f_r = -\log \pi_r$, where $\pi_r = \kappa_r * \pi$. Assume α and $\gamma = \sigma/\alpha$ are absolutely continuous on every sampling interval $[0,t]$, with $\alpha_0 = 1$, $\gamma_0 = 0$, $\alpha_t, \gamma_t > 0$ for $t > 0$, and γ nondecreasing. Suppose (1) and the approximate-score ODE, with $s_\theta(x,t) = \alpha_t^{-1}\widetilde{s}_\theta(x/\alpha_t, \gamma_t)$, admit Borel backward flow maps whose trajectories, for every terminal state, are locally absolutely continuous on $(0,t]$ and extend continuously to $[0,t]$. Assume $\nabla^2 f_r \succeq \mu_r \,\mathrm{Id}$ for all $r > 0$, and define*

$$e_r := \sup_{x \in \mathbb{R}^d} \|\widetilde{s}_\theta(x,r) - \nabla \log \pi_r(x)\| < \infty. \tag{5}$$

Suppose μ_r, e_r are measurable and $r\mu_r^-, re_r \in L^1_{\mathrm{loc}}([0,\infty))$, where $\mu_r^- := \max\{-\mu_r, 0\}$. Then the pushforward operators $D_{\pi,t}$ preserve $\mathcal{P}_2(\mathbb{R}^d)$, and A3 and A4 hold with

$$L_t = \frac{1}{\alpha_t} \exp\left(-\int_0^{\gamma_t} u\,\mu_u \,\mathrm{d}u\right), \quad E_t = \int_0^{\gamma_t} r\, e_r \exp\left(-\int_0^r u\mu_u \,\mathrm{d}u\right) \mathrm{d}r\,. \tag{6}$$

The integrals from zero are improper integrals: since only the negative part of $u\mu_u$ is assumed integrable near zero, $\int_0^r u\mu_u \,\mathrm{d}u$ may equal $+\infty$, in which case the exponential is set to zero.

The proof is in Appendix B.

Strong log-concavity gives an explicit family of stability profiles. We say that π is m-strongly log-concave, for $m \geq 0$, if it has a positive Lebesgue density and $f = -\log \pi$ is m-strongly convex.

Corollary 3.4 (Strongly log-concave case)**.** *Assume the conditions of Theorem 3.3 hold and that there exist $M_0, M_1 \geq 0$ such that the score approximation error (5) satisfies $e_r \leq M_0 + M_1/r$, for all $r > 0$. Assume further that π is m-strongly log-concave with $m > 0$ and that $f = -\log \pi$ is C^2. Then, A3 and A4 hold with $L_t = \alpha_t^{-1}(1 + m\gamma_t^2)^{-1/2}$ and $E_t \leq \frac{M_0}{m}\left(\sqrt{1+m\gamma_t^2} - 1\right) + \frac{M_1}{\sqrt{m}} \operatorname{arcsinh}(\sqrt{m}\gamma_t)$.*

The proof is in Appendix B.

Remark 3.5 (Variance-exploding & TrigFlow CMs)**.** *Under the assumptions of Corollary 3.4, define $B_m(r) := \frac{M_0}{m}\left(\sqrt{1+mr^2} - 1\right) + \frac{M_1}{\sqrt{m}} \operatorname{arcsinh}(\sqrt{m}\, r)$. Lemma 3.2 and the corollary give $R_t = 1$, $L_t = (1+mt^2)^{-1/2}$ and $E_t = B_m(t)$ for VE ($\alpha_t = 1$, $\sigma_t = t$), and $R_t = \cos t$, $L_t = (\cos^2 t + m \sin^2 t)^{-1/2}$ and $E_t = B_m(\tan t)$ for TrigFlow ($\alpha_t = \cos t$, $\sigma_t = \sin t$, $0 < t < \pi/2$); in both cases $\mathcal{L}_t$ is q_t-contractive in W_2 with $q_t = (1 + m\gamma_t^2)^{-1/2} < 1$ for all admissible $t > 0$.*

Beyond strong log-concavity. We now apply Theorem 3.3 to nonconvex potentials, extending the strongly log-concave case of Corollary 3.4 at the cost of weaker contraction. For $\ell \geq 0$ define

$$\psi_\ell(u) := 2\sqrt{\ell}\, u \tanh(\sqrt{\ell}\, u/2), \qquad u \geq 0, \tag{7}$$

and consider, for $m > 0$, the relaxed curvature condition

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq m\|x-y\|^2 - \psi_\ell(\|x-y\|), \qquad x, y \in \mathbb{R}^d. \tag{8}$$

For $\ell = 0$ this is m-strong convexity. For $\ell > 0$, the defect ψ_ℓ is quadratic near zero and grows linearly at infinity, so the quadratic term set by m dominates at large separations while ℓ controls the curvature defect near zero. With m as in (8) and score-error constants $M_0, M_1 \geq 0$, let B_m be as in Remark 3.5 and set $A_{m,\ell}(r) := \exp\{\ell r^2/[2(1+mr^2)]\}$ for $r \geq 0$.

Theorem 3.6 (Stability)**.** *Assume the parametrization and backward-flow hypotheses of Theorem 3.3. Let $\pi \in \mathcal{P}_2(\mathbb{R}^d)$ have a positive density with $f = -\log \pi \in C^1$ satisfying (8). Assume the score error e_r in (5) is measurable and satisfies $e_r \leq M_0 + M_1/r$ for every $r > 0$. Then A2–A4 hold with*

$$R_t = \alpha_t, \qquad L_t = \frac{A_{m,\ell}(\gamma_t)}{\alpha_t\sqrt{1+m\gamma_t^2}}, \qquad E_t = A_{m,\ell}(\gamma_t) B_m(\gamma_t). \tag{9}$$

Consequently, Proposition 3.1 applies with $q_t = R_t L_t$. The exact sampling step $\mathcal{L}_t$ is q_t-Lipschitz in W_2, with $q_t < 1$ if and only if

$$\frac{\ell\gamma_t^2}{1+m\gamma_t^2} < \log(1+m\gamma_t^2). \tag{10}$$

The proof is in Appendix B. When $\ell \le m$, (10) holds at every noise ratio, since $\log(1+z) > z/(1+z)$ for $z > 0$; when $\ell > m$ it fails for small γ_t, as $q_t = 1 + (\ell - m)\gamma_t^2/2 + O(\gamma_t^4)$, but holds for large γ_t, since $q_t \to 0$: larger noise restores certified contraction at the cost of a larger error budget in (9).

A concrete sufficient condition for (8) is that $f \in C^2$ has $\nabla^2 f \succeq m\,\mathrm{Id}$ outside a ball of radius $\varrho > 0$ and $\nabla^2 f \succeq -\beta\,\mathrm{Id}$ inside, with $\beta \ge 0$. One may then take $\ell = 2\max\{m + \beta, (m+\beta)^2\varrho^2\}$; the segment argument justifying this choice is given in Appendix A.

Remark 3.7 (Comparison with Lyu et al. (2024)). *Their analysis already establishes a multistep error recursion and geometric reduction toward an error plateau under a Lipschitz-type condition on the trained map. We instead derive the propagation factor from the target's curvature, following the direction proposed in their Section 4. Under strong log-concavity, Corollary 3.4 gives $q_t = (1+m\gamma_t^2)^{-1/2} < 1$ at every $\gamma_t > 0$, and Theorem 3.6 identifies the levels that certify contraction under a curvature defect. Combined with E_t, these factors turn the balance condition of Proposition 3.1 into a quantitative criterion for reducing the noise while controlling accumulated error.*

The two analyses rest on different assumptions: our exact-step recursion uses the uniform approximation bound A4, whereas their estimates use L^2 training losses and include numerical discretization. Theorem 3.11 connects our flow estimates to consistency losses, with network Lipschitz constants still weighting the cellwise flow errors, and Section 3.3 recovers their affine stability recursion as the learned-map alternative, with constants detailed in Appendix C.

Remark 3.8 (Comparison with Chen et al. (2025)). *Their Theorem 2 assumes bounded target samples and learned outputs, and L^2 consistency along exact probability-flow trajectories, but neither Lipschitz denoisers nor target curvature. It therefore covers maps lacking the stability required by Proposition 3.1. Their Appendix D truncates suitable unbounded targets.*

The difference is how errors propagate. Their Eq. (6) accumulates errors in a KL bound between noisy marginals and converts it to W_2 using boundedness, taking a fourth root. Our direct W_2 recursion requires exact-denoiser stability and uniform approximation (A3 and A4), but no bounded support. These controls retain the attenuation by later steps: for two levels, the first-step error contributes $q_{t_2}E_{t_1}$ and is attenuated when $q_{t_2} < 1$.

The gain can be substantial when both sets of controls are available. For the bounded VE example with $\pi = \mathrm{Unif}[-1,1]$ in Appendix C, the same learned maps, initialization and two sampling levels give our certificate 0.1746, versus 0.7659 from their bound, even after the refinements detailed there. Thus stability can give a sharper guarantee; this comparison of upper bounds for the same sampler does not order actual sampling errors or establish a general dominance.

3.3 Multistep bounds estimated from a trained model

An error estimate for a trained model, such as Theorem 3.11 below, typically controls one learned sampling step started from π. During iteration, however, the input law generally differs from π. To handle these inputs, Proposition 3.1 assumes that the learned and exact sampling steps produce distributions at W_2-distance at most E_t, whatever their common input law (A4). The estimate at π does not guarantee this. We instead control how the learned denoiser responds to a change in its input law. Fix finite nonnegative profiles E_t^π and $\widetilde{L}_t$ for target error and stability, and set $\widetilde{q}_t := R_t\widetilde{L}_t$.

A5. *For any $t > 0$, $W_2\big(\pi, \widetilde{\mathcal{L}}_t^\theta(\pi)\big) \le E_t^\pi$.*

A6. *For all $t > 0$ and $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$, $W_2\big(\widetilde{D}_{\pi,t}^\theta(\rho_1),\ \widetilde{D}_{\pi,t}^\theta(\rho_2)\big) \le \widetilde{L}_t\, W_2(\rho_1, \rho_2)$.*

Compared with Proposition 3.1, the tradeoff is that $R_t\widetilde{L}_t$ replaces R_tL_t. Contraction of the exact step does not control either learned quantity.

Proposition 3.9 (Learned-map recursion). *Let A2, A5, and A6 hold. Then, for any $\rho \in \mathcal{P}_2(\mathbb{R}^d)$,*

$$W_2\big(\widetilde{\mathcal{L}}_t^\theta(\rho), \pi\big) \le \widetilde{q}_t\, W_2(\rho, \pi) + E_t^\pi\,, \quad \textit{for all} \quad t > 0 \tag{11}$$

Hence, the results of Proposition 3.1, its balance condition and geometric bound (4), and the results of Section 3.1 on repeated levels and added steps hold verbatim with (q_t, E_t) replaced by $(\widetilde{q}_t, E_t^\pi)$.

The proof is in Appendix B.

Remark 3.10 (Affine stability). *A6 may be relaxed to* $W_2(\widetilde{D}^\theta_{\pi,t}(\rho_1), \widetilde{D}^\theta_{\pi,t}(\rho_2)) \le \widetilde{L}_t W_2(\rho_1,\rho_2)+\beta_t$, *with an additive defect* $\beta_t \ge 0$. *This covers maps that are Lipschitz only up to an offset: the pointwise bound of Lyu et al. (2024, Assumption 3.5) gives it by coupling and Minkowski's inequality (Appendix C), and outputs in a ball of radius R, as in Chen et al. (2025, Theorem 2), give* $\widetilde{L}_t = 0$ *and* $\beta_t = 2R$. *In this relaxation,* $\widetilde{L}_t$ *is only the multiplicative constant of the inequality, not a Lipschitz constant of the learned map: bounded outputs, for instance, are not* 0*-Lipschitz. The proof of Proposition 3.9 then gives (11) with* $E^\pi_t + \beta_t$ *in place of* E^π_t, *and the recursion, the balance condition and the plateau follow as before. What is lost is the fixed point of Remark B.3: for* $\beta_t > 0$ *the learned step need not be a contraction, and its iterates need not converge.*

When both sets of assumptions hold, the certificates combine level by level (Corollary B.2). For $\widetilde{q}_t < 1$, the repeated-level bound plateaus at $E^\pi_t/(1-\widetilde{q}_t)$, and $\widetilde{\mathcal{L}}^\theta_t$ has a unique fixed point within this distance of π (Remark B.3).

From the approximate-score flow to a trained consistency model. We transfer the bounds of Section 3.2 to a trained CM, whose learned map $\widetilde{\Phi}^\theta_{t\to 0}$ need not be the flow of an ODE. Let $\widetilde{\Phi}_{t\to s}$ and $\Phi_{t\to s}$ denote the approximate-score and exact flow maps from t to s, respectively. Here the approximate-score ODE uses, in place of s_θ, the score approximation of the solver used in training, such as a pretrained score model in consistency distillation, let e_r in (5) denote its error. On a grid $0 = \tau_0 < \cdots < \tau_N = T$, consistency training compares $\widetilde{\Phi}^\theta_{\tau_{i+1}\to 0}(x)$ with $\widetilde{\Phi}^\theta_{\tau_i\to 0}(\widetilde{\Phi}_{\tau_{i+1}\to\tau_i}(x))$ for $x \sim \varpi_{\tau_{i+1}}$ (Song et al., 2023; Lyu et al., 2024); here we idealize the numerical solver by the approximate-score flow. Define the cell-wise error

$$E_{s,t} := \alpha_s \int_{\gamma_s}^{\gamma_t} r e_r \exp\Big(-\int_{\gamma_s}^{r} u\mu_u \,\mathrm{d}u\Big)\,\mathrm{d}r, \qquad 0 \le s \le t. \tag{12}$$

For $s = 0$ it is the profile (6) of this score approximation and bounds $\|\widetilde{\Phi}_{t\to s}(x) - \Phi_{t\to s}(x)\|$ for every x (Lemma A.1): it controls the solver's flow, not the network. For each training cell, define the root-mean-square consistency error

$$\varepsilon_i := \Big(\mathbb{E}_{X\sim\varpi_{\tau_{i+1}}} \big\|\widetilde{\Phi}^\theta_{\tau_{i+1}\to 0}(X) - \widetilde{\Phi}^\theta_{\tau_i\to 0}(\widetilde{\Phi}_{\tau_{i+1}\to\tau_i}(X))\big\|^2\Big)^{1/2}, \qquad 0 \le i < N.$$

Theorem 3.11 (Error of a trained CM). *Assume the hypotheses of Theorem 3.3 for the score approximation of the solver, with* $(\Phi_{t\to s})_\# \varpi_t = \varpi_s$ *for* $0 \le s \le t$. *Suppose* $\widetilde{\Phi}^\theta_{0\to 0}$ *is the identity, so that* $\widetilde{L}_{\tau_0} = 1$, *and* $\widetilde{\Phi}^\theta_{\tau_i\to 0}$ *is* $\widetilde{L}_{\tau_i}$*-Lipschitz with* $\widetilde{L}_{\tau_i} < \infty$ *for* $1 \le i \le N$. *Then, for* $1 \le k \le N$,

$$W_2\big(\pi, \widetilde{\mathcal{L}}^\theta_{\tau_k}(\pi)\big) \le \sum_{i=0}^{k-1} \varepsilon_i + \sum_{i=0}^{k-1} \widetilde{L}_{\tau_i} E_{\tau_i,\tau_{i+1}}. \tag{13}$$

Thus, on the grid, A5 holds with $E^\pi_{\tau_k}$ the right-hand side of (13), and A6 holds with the Lipschitz constants $\widetilde{L}_{\tau_k}$ of the network, without requiring uniformly bounded consistency residuals. A5 can also be obtained without the theorem: A1 and A4 give $E^\pi_t := E_t$, since then $W_2(\pi, \widetilde{\mathcal{L}}^\theta_t(\pi)) = W_2(\mathcal{L}_t(\pi), \widetilde{\mathcal{L}}^\theta_t(\pi))$, and the target error can be estimated from samples (Section 4). The full statement, including the uniform bound with sup-norm consistency errors, and its proof are in Theorem B.4.

In both (13) and its uniform version (28), the factor $\widetilde{L}_{\tau_i}$ replaces the contraction of the exact flow built into $E_{0,t}$: if $\widetilde{L}_{\tau_i} = L_{\tau_i}$ the second sum equals E_{0,τ_k} exactly. Under the assumptions of Corollary 3.4, $E_{s,t} \le \alpha_s\sqrt{1+m\gamma_s^2}\,\big(B_m(\gamma_t) - B_m(\gamma_s)\big)$. The uniform form (28), needed for the exact-map route of Proposition 3.1, requires finite sup-norm consistency errors. The target bound (13) instead gives A5 directly. Using a numerical solver in place of $\widetilde{\Phi}_{\tau_{i+1}\to\tau_i}$ adds $\widetilde{L}_{\tau_i}$ times its local error to the consistency budget: this error must be measured in $L^2(\varpi_{\tau_{i+1}})$ for (13) and in the sup norm for (28).

3.4 A Gaussian example, and tightness of certificates

For a Gaussian target and an affine learned map, the CM iterates and their W_2 errors are explicit. This case shows that the certificate is attained for errors in the mean, overstates errors in the variance,

and ignores cancellations between variance errors of opposite signs. Derivations and further regimes are given in Appendix E. Let $\pi = \mathcal{N}(0, m^{-1}\,\mathrm{Id}_d)$. The profiles of Remark 3.5 are then attained: $\Phi_{t\to 0}(x) = L_t x$ and $q_t = (1 + m\gamma_t^2)^{-1/2}$. Let the learned map be $\widetilde{\Phi}^{\theta}_{t\to 0}(x) = \widetilde{L}_t x + \omega_t v$, where $\widetilde{L}_t$ is L_t with a precision $\widetilde{m} > 0$ in place of m, v is a unit vector and $\omega_t \geq 0$. These parameters play the role of θ: $\widetilde{m}$ sets the variance the model believes in, and ω_t shifts its mean. The score it implies, $\widetilde{s}_\theta(x, r) = -\widetilde{\mu}_r x + \eta_r v$ with $\widetilde{\mu}_r = \widetilde{m}/(1 + \widetilde{m}r^2)$ and an offset η_r determined by ω_t, is also affine (Section E.1): the example specifies the map, and the score is derived from it. Gaussian laws remain Gaussian: from $\rho_0 = \mathcal{N}(\zeta_0 v, s_0^2\,\mathrm{Id})$ and with $\widetilde{q}_t = \alpha_t \widetilde{L}_t$, the iterates are $\rho_k = \mathcal{N}(\zeta_k v, s_k^2\,\mathrm{Id})$, where $\zeta_k = \widetilde{q}_{t_k}\zeta_{k-1} + \omega_{t_k}$, $s_k = \widetilde{L}_{t_k}\left(\alpha_{t_k}^2 s_{k-1}^2 + \sigma_{t_k}^2\right)^{1/2}$ and $W_2(\rho_k, \pi)^2 = \zeta_k^2 + d\,(s_k - m^{-1/2})^2$.

Errors in the mean attain the bounds exactly: if $\widetilde{m} = m$, the learned map is the exact one plus the shift $\omega_t v$, so $E_t = \omega_t$ is the best constant in A4; if also $s_0 = m^{-1/2}$, the variance stays exact and $W_2(\rho_k, \pi) = \zeta_k$ follows $\zeta_k = q_{t_k}\zeta_{k-1} + E_{t_k}$: (3) holds with equality for every schedule. The trade-off of Section 3.1 then governs the actual error. Along a decreasing schedule it first falls and then rises, since $q_{t_k} \to 1$ faster than $E_{t_k} \to 0$; at a repeated level it approaches the plateau $E_t/(1-q_t)$ from either side, so one denoiser improves a poor initialization and degrades a good one.

Errors in the variance are overstated. The constant $R_t = \alpha_t$ is attained by two laws that differ by a shift, but adding independent noise also compresses differences of standard deviation: for $\nu_s = \mathcal{N}(0, s^2\,\mathrm{Id}_d)$, $W_2(N_t(\nu_s), N_t(\nu_{m^{-1/2}})) = \alpha_t W_2(\nu_s, \nu_{m^{-1/2}})/\kappa_t(s)$ with $\kappa_t \geq 1$, and $\kappa_t = q_t^{-1}$ at $s = m^{-1/2}$. The certificate misses this factor, and the loss compounds over steps: from a wrong variance and with an exact CM, it decays like q_t^n at a repeated level while the error eventually decays like q_t^{2n} (Fig. 1). An error in the mean, being certified exactly, hides this loss.

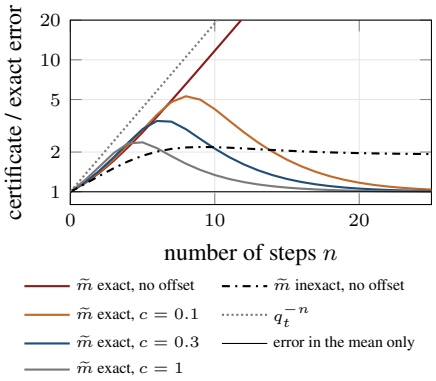


Figure 1: Certificate over exact error at the repeated VE level $t = 1$, from $\mathcal{N}(0, 2.5^2\,\mathrm{Id})$, with $m = 0.8$ and, when inexact, $\widetilde{m} = 0.55$; offsets are $\omega_t = c\,B_{\widetilde{m}}(\gamma_t)$ with $M_0 = 0.03$ and $M_1 = 0.05$.

When a trained model does not reproduce the variance of π exactly, i.e., $\widetilde{m} \neq m$, the learned and exact maps differ by $(\widetilde{L}_t - L_t)x$, which is unbounded in x, so no uniform E_t satisfies A4. Proposition 3.9 still applies, with $\widetilde{L}_t$ and the target error E_t^π of (48), both attained. As before, the certificate overstates the inherited variance error by κ_t, which an error in the mean partly hides; at a repeated level its plateau $E_t^\pi/(1-\widetilde{q}_t)$ exceeds the limiting error by a factor that decreases to one as ω_t grows (Fig. 1 shows the case without offset). Finally, the certificate cannot exploit cancellation: when the initialization and the trained model err in the variance in opposite directions, their errors partly cancel in the law but add up in the certificate (cf. Section E.3).

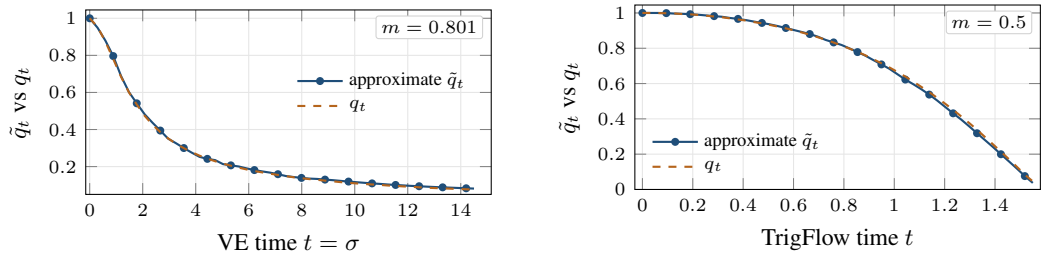
4 Numerical validation on Stable Diffusion (SD) 1.5

We now show empirically that the contraction and approximation profiles entering our bounds can be measured reliably on trained CMs and closely match the functional forms predicted by the theory, indicating that our assumptions hold for real CMs and that we can quantify the constants involved.

We conduct this validation on a latent consistency model (LCM) distilled from Stable Diffusion (SD) 1.5 (Rombach et al., 2022) by discrete-time consistency distillation, using a variance-exploding (VE) parametrization.[1] To cover the continuous-time setting, we also report results for an LCM trained on CelebA (Liu et al., 2015) with the TrigFlow parametrization (Lu & Song, 2025) in the latent space of a KL-regularized VAE (Rombach et al., 2022). For each model, we approximate the target error E_t^π and the contraction profile $\tilde{q}_t = R_t \tilde{L}_t$ of the learned map from A2, A5, and A6 by using optimal transport W_2 estimators (see Appendix D), and compare them with the predictions of Corollary 3.4 (see Remark 3.5) for strongly log-concave targets. We focus on LCMs for three

[1] See Appendix D for details and implementation link.

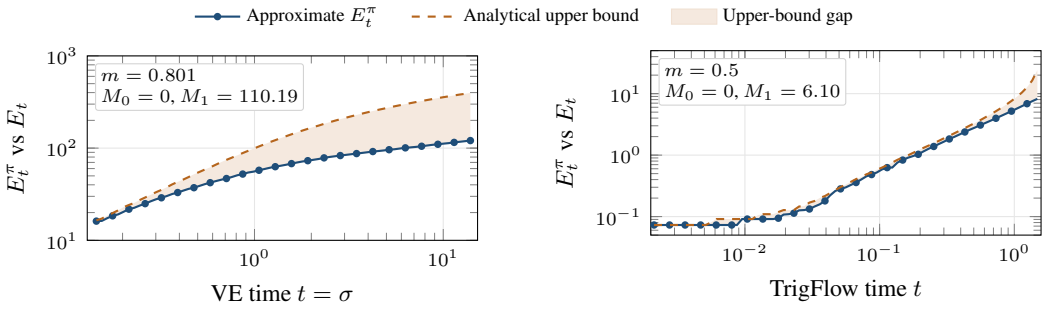
reasons: LCMs underlie state-of-the-art few-step image generators (Luo et al., 2023a); the reduced dimensionality of the latent space makes empirical estimation of W_2 distances and Lipschitz constants more reliable; and the KL-regularized VAE of latent models (Rombach et al., 2022) pushes the latent distribution π towards a Gaussian, so the functional forms derived for the strongly log-concave setting of Corollary 3.4 may suffice.



(a) SD 1.5 LCM in VE coordinates, $m = 0.801$.

(b) TrigFlow LCM on CelebA, $m = 0.5$.

Figure 2: Approximate $\tilde{q}_t$ and the predicted profile q_t of Corollary 3.4. The analytical curve is used as a fitted functional form; it is not a certificate of the learned map's global contraction factor.



(a) VE SD 1.5 LCM on AFHQ

(b) TrigFlow LCM on CelebA.

Figure 3: Approximate target-only one-step error E_t^π for the TrigFlow and VE LCMs, compared with the analytical VE shape of Corollary 3.4 used as an upper envelope.

Figure 2 shows the approximate contraction profile $\tilde{q}_t$ for both LCMs, alongside those predicted by Corollary 3.4 for a strongly log-concave target. They are contractive ($\tilde{q}_t < 1$) and closely follow the predicted forms, with fitted strong log-concavity constants $m \approx 0.8$ for the SD 1.5 LCM and $m \approx 0.5$ for the CelebA LCM. These agree with independent estimates from the covariance of the latent target, which give $m > 0.6$ and $m > 0.3$, respectively. This agreement supports the conclusion that our assumptions hold for these LCMs and that $\tilde{L}_t \approx L_t$. Moreover, Figure 3 shows the approximate error profile E_t^π for both LCMs, alongside the prediction E_t of Corollary 3.4. Since E_t bounds the approximation error uniformly over inputs, whereas E_t^π measures it only at the target, $E_t \geq E_t^\pi$ by definition, so the predicted E_t should exceed the approximate E_t^π. Lastly, Appendix F combines the estimated SD 1.5 profiles into the plug-in one-step W_2 bound and compares it with the empirical W_2 error on held-out data. The obtained bound lies above the error and tracks its variation with t.

5 Conclusion

This paper derived a finite-horizon W_2 recursion for multistep CM sampling that separates contraction of the initialization error from accumulation of approximation error. Placing stability on the exact denoiser makes the contraction factor q_t a property of the target, explicit under strong log-concavity and, above a noise threshold, under semi-log-concavity. Under a balance condition, the recursion yields a sharp geometric bound, and a complementary guarantee extends it to trained CMs under estimable assumptions. The experiments corroborate this analysis, as the estimated profiles closely follow the predicted functional forms. Future work includes extending the analysis to zero-shot samplers with inference-time guidance or rewards, and optimizing schedules directly from the bound.

6 AI USE DISCLOSURE

In this work, we used generative AI tools for help in the writing of proofs. We have not used generative AI tools to generate synthetic data sets, to help develop theoretical models or conceptual frameworks, to formulate mathematical claims, to provide critical ingredients for proving mathematical claims, to propose or refine hypotheses, to design or provide feedback on research methodology or experiments, to implement methods, or to interpret results. Moreover, the use of generative AI tools to assist with translation, clean and reformat dataset, or to support qualitative and thematic data analysis, are not applicable to this work. Additionally, we used generative AI tools to polish writing and draft sections of the paper, notably in the appendix. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023.

Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Transactions on Machine Learning Research*, 2025a. ISSN 2835-8856. arXiv:2406.07507.

Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025b. arXiv:2505.18825.

Soravit Changpinyo, Piyush Sharma, Nan Ding, and Radu Soricut. Conceptual 12m: Pushing web-scale image-text pre-training to recognize long-tail visual concepts, 2021. URL https://arxiv.org/abs/2102.08981.

Yiding Chen, Yiyi Zhang, Owen Oertell, and Wen Sun. Convergence of consistency model with multistep sampling under general data assumptions. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 9764–9782. PMLR, 2025. URL https://proceedings.mlr.press/v267/chen25ck.html.

Giovanni Conforti. Weak semiconvexity estimates for schrödinger potentials and logarithmic sobolev inequality for schrödinger bridges. *Probability Theory and Related Fields*, 189(3):1045–1071, 2024.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. arXiv:2208.05314.

Zehao Dou, Minshuo Chen, Mengdi Wang, and Zhuoran Yang. Theory of consistency diffusion models: Distribution estimation meets fast sampling. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pp. 11592–11612. PMLR, 2024.

Zach Evans, CJ Carr, Josiah Taylor, Scott H Hawley, and Jordi Pons. Fast timing-conditioned latent audio diffusion. In *Forty-first International Conference on Machine Learning*, 2024.

Zhengyang Geng, Ashwini Pokle, William Luo, Justin Lin, and J Zico Kolter. Consistency models made easy. *arXiv preprint arXiv:2406.14548*, 2024.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. *arXiv preprint arXiv:2505.13447*, 2025.

Jonathan Heek, Emiel Hoogeboom, and Tim Salimans. Multistep consistency models, 2024. URL https://arxiv.org/abs/2403.06807.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. URL https://arxiv.org/abs/2006.11239.

Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. Imagen video: High definition video generation with diffusion models, 2022. URL https://arxiv.org/abs/2210.02303.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ode trajectory of diffusion. *arXiv preprint arXiv:2310.02279*, 2023.

Gen Li, Zhihan Huang, and Yuting Wei. Towards a mathematical theory for consistency training in diffusion models. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1621–1629. PMLR, 2025. arXiv:2402.07802.

Haohe Liu, Zehua Chen, Yi Yuan, Xinhao Mei, Xubo Liu, Danilo Mandic, Wenwu Wang, and Mark D Plumbley. Audioldm: Text-to-audio generation with latent diffusion models. *arXiv preprint arXiv:2301.12503*, 2023a.

Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9298–9309, 2023b.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models, 2025. URL https://arxiv.org/abs/2410.11081.

Eric Luhman and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *ArXiv*, abs/2101.02388, 2021. URL https://api.semanticscholar.org/CorpusID:230799531.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference, 2023a. URL https://arxiv.org/abs/2310.04378.

Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhihua Zhang. Diff-instruct: A universal approach for transferring knowledge from pre-trained diffusion models. *ArXiv*, abs/2305.18455, 2023b. URL https://api.semanticscholar.org/CorpusID:258967550.

Junlong Lyu, Zhitang Chen, and Shoubo Feng. Sampling is as easy as keeping the consistency: convergence guarantee for consistency models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 33664–33685. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/lyu24b.html.

Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

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.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your flow: Scaling continuous-time flow map distillation. *arXiv preprint arXiv:2506.14603*, 2025.

Tim Salimans, Thomas Mensink, Jonathan Heek, and Emiel Hoogeboom. Multistep distillation of diffusion models via moment matching. *ArXiv*, abs/2406.04103, 2024. URL https://api.semanticscholar.org/CorpusID:270285800.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pp. 87–103. Springer, 2024.

Adrien Saumard and Jon A Wellner. Log-concavity and strong log-concavity: a review. *Statistics surveys*, 8:45, 2014.

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

Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models, 2023. URL https://arxiv.org/abs/2310.14189.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021. URL https://arxiv.org/abs/2011.13456.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/song23a.html.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, Jianyuan Zeng, Jiayu Wang, Jingfeng Zhang, Jingren Zhou, Jinkai Wang, Jixuan Chen, Kai Zhu, Kang Zhao, Keyu Yan, Lianghua Huang, Mengyang Feng, Ningyi Zhang, Pandeng Li, Pingyu Wu, Ruihang Chu, Ruili Feng, Shiwei Zhang, Siyang Sun, Tao Fang, Tianxing Wang, Tianyi Gui, Tingyu Weng, Tong Shen, Wei Lin, Wei Wang, Wei Wang, Wenmeng Zhou, Wente Wang, Wenting Shen, Wenyuan Yu, Xianzhong Shi, Xiaoming Huang, Xin Xu, Yan Kou, Yangyu Lv, Yifei Li, Yijing Liu, Yiming Wang, Yingya Zhang, Yitong Huang, Yong Li, You Wu, Yu Liu, Yulin Pan, Yun Zheng, Yuntao Hong, Yupeng Shi, Yutong Feng, Zeyinzi Jiang, Zhen Han, Zhi-Fan Wu, and Ziyu Liu. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

Zhendong Wang, Huangjie Zheng, Pengcheng He, Weizhu Chen, and Mingyuan Zhou. Diffusion-gan: Training gans with diffusion. *ArXiv*, abs/2206.02262, 2022. URL https://api.semanticscholar.org/CorpusID:249395201.

Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. *Advances in neural information processing systems*, 36:8406–8441, 2023.

Zhangkai Wu, Xuhui Fan, Hongyu Wu, and Longbing Cao. Scot: Unifying consistency models and rectified flows via straight-consistent trajectories. *arXiv preprint arXiv:2502.16972*, 2025.

Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Fredo Durand, and Bill Freeman. Improved distribution matching distillation for fast image synthesis. *Advances in neural information processing systems*, 37:47455–47487, 2024a.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6613–6623, 2024b.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3836–3847, 2023.

Hongkai Zheng, Weili Nie, Arash Vahdat, Kamyar Azizzadenesheli, and Anima Anandkumar. Fast sampling of diffusion models via operator learning. In *International Conference on Machine Learning*, 2022. URL https://api.semanticscholar.org/CorpusID:254017822.

Linqi Zhou, Stefano Ermon, and Jiaming Song. Inductive moment matching. *arXiv preprint arXiv:2503.07565*, 2025.

A Curvature under Gaussian smoothing and stability of the probability flow

The approximate-score ODE introduced in Section 3.2 is

$$\dot{x}_t = a_t x_t - \tfrac{1}{2} b_t^2 s_\theta(x_t, t). \tag{14}$$

Lemma A.1. *Under the assumptions of Theorem 3.3, let x_u, y_u solve (1) and (14), respectively. For $0 \le s \le t$,*

$$\|x_s - y_s\| \le L_{s,t}\|x_t - y_t\| + E_{s,t},$$

where

$$L_{s,t} = \frac{\alpha_s}{\alpha_t} \exp\left(- \int_{\gamma_s}^{\gamma_t} r\mu_r \, \mathrm{d}r \right),$$

$$E_{s,t} = \alpha_s \int_{\gamma_s}^{\gamma_t} r e_r \exp\left(- \int_{\gamma_s}^{r} u\mu_u \, \mathrm{d}u \right) \mathrm{d}r.$$

Both constants are finite, with the improper-integral convention of Theorem 3.3 at $s = 0$.

Proof. Fix $0 < s < t$. On this interval set $z_u = x_u/\alpha_u$, $w_u = y_u/\alpha_u$ and $c_u = \gamma_u \dot{\gamma}_u \ge 0$. The identity $\nabla \log \varpi_u(x) = -\alpha_u^{-1} \nabla f_{\gamma_u}(x/\alpha_u)$ and $b_u^2 = 2\alpha_u^2 c_u$ give, almost everywhere,

$$\dot{z}_u = c_u \nabla f_{\gamma_u}(z_u), \qquad \dot{w}_u = -c_u \widetilde{s}_\theta(w_u, \gamma_u).$$

By the curvature bound and (5), the absolutely continuous function $h_u = \|z_u - w_u\|$ satisfies

$$\dot{h}_u \ge c_u \mu_{\gamma_u} h_u - c_u e_{\gamma_u} \quad \text{almost everywhere.}$$

For $h_u > 0$, this follows by taking the scalar product of $\dot{z}_u - \dot{w}_u$ with $(z_u - w_u)/h_u$; on $\{h_u = 0\}$, $\dot{h}_u = 0$ almost everywhere, so the inequality still holds. Gaussian smoothing makes $\nabla^2 f_r(0)$ continuous for $r > 0$; hence $\mu_r \le \lambda_{\min}(\nabla^2 f_r(0))$ bounds μ_r^+ on compact positive-noise intervals. Together with the integrability assumptions and substitution for an absolutely continuous nondecreasing γ, this ensures that $c_u \mu_{\gamma_u}$ and $c_u e_{\gamma_u}$ are integrable on $[s, t]$. A backward integrating factor therefore gives

$$h_s \le e^{-\int_s^t c_u \mu_{\gamma_u} \, \mathrm{d}u} h_t + \int_s^t c_v e_{\gamma_v} e^{-\int_s^v c_u \mu_{\gamma_u} \, \mathrm{d}u} \, \mathrm{d}v.$$

Substituting $r = \gamma_v$ and multiplying by α_s gives the stated constants. This substitution needs no inverse of γ and allows constant positive-noise intervals.

To pass to $s = 0$, let $s \downarrow 0$. For $0 < a \le r \le \gamma_t$,

$$\exp\left(- \int_a^r u\mu_u \, \mathrm{d}u \right) \le C_t := \exp\left(\int_0^{\gamma_t} u\mu_u^- \, \mathrm{d}u \right) < \infty.$$

Since $\int_0^{\gamma_t} r e_r \, \mathrm{d}r < \infty$, dominated convergence yields the stated $E_{0,t}$, even if $\int_0^r u\mu_u \, \mathrm{d}u = +\infty$. The same improper-integral limit gives $L_{0,t}$; continuity of the trajectories and $\alpha_0 = 1$ give the left-hand side. These bounds also prove finiteness. The case $s = t$ is immediate. □

Lemma A.2 (Strongly log-concave case, pathwise form)**.** *Under the parametrization and flow-existence assumptions of Theorem 3.3, let $f = -\log \pi$ be C^2 and m-strongly convex, $m > 0$. Let x_u, y_u be the exact and approximate trajectories of Lemma A.1. Assume that there exist $M_0, M_1 \ge 0$ such that*

$$e_r = \sup_{x \in \mathbb{R}^d} \|\widetilde{s}_\theta(x, r) - \nabla \log \pi_r(x)\| \le M_0 + M_1/r$$

for all $r > 0$, with e_r measurable. Then

$$\|x_0 - y_0\| \le L_t \|x_t - y_t\| + E_t$$

where

$$L_t = \alpha_t^{-1}\big(1 + m\gamma_t^2\big)^{-1/2}$$

and

$$E_t = \frac{M_0}{m}\left(\sqrt{1 + m\gamma_t^2} - 1\right) + \frac{M_1}{\sqrt{m}}\,\operatorname{arcsinh}\big(\sqrt{m}\gamma_t\big)\ .$$

Proof. By Lemma A.3, f_r is μ_r-strongly convex for $\mu_r = \frac{m}{mr^2+1}$, hence by Lemma A.1, we have the result once we substitute μ_r, $\alpha_0 = 1, \sigma_0 = 0$ and integrate. □

Lemma A.3 (Curvature under Gaussian smoothing, (Saumard & Wellner, 2014, Theorem 3.7-(b)))**.** *Suppose that π is m-strongly log-concave, for $m \ge 0$. Then for any $r > 0$, π_r is μ_r-strongly log-concave with $\mu_r = m/(1 + mr^2)$. Since $\pi_r = \kappa_r * \pi$ is positive and C^∞, this means that $f_r = -\log \pi_r \in C^\infty(\mathbb{R}^d)$ satisfies $\nabla^2 f_r \succeq \mu_r \operatorname{Id}$ on $\mathbb{R}^d$.*

We use ψ_ℓ and the relaxed curvature condition from (7)–(8), allowing $m = 0$ here. For $m, \ell \ge 0$, a positive density π with $f = -\log \pi \in C^1$ satisfying (8) is called (m, ℓ)-*strongly log-concave outside a ball* below. This terminology refers to the condition on pairs of points; a Hessian bound outside a spatial ball is a sufficient condition, as shown below. The following lemmas establish the curvature bounds used in Theorem 3.6.

Lemma A.4 (Elementary bounds on ψ_ℓ)**.** *Let ψ_ℓ be as in (7).*

- *(a) For all $\ell \ge 0$ and $u \ge 0$, $\psi_\ell(u) \le \min\{\ell u^2,\ 2\ell^{1/2}u\}$.*
- *(b) Let $\beta', B' \ge 0$ and let $\Delta : [0,\infty) \to \mathbb{R}$ satisfy $\Delta(u) \le \min\{\beta' u^2,\ 2B'u\}$ for all $u \ge 0$. Then $\Delta \le \psi_\ell$ on $[0,\infty)$ for $\ell = 2\max\{\beta',\ B'^2\}$.*

Proof. (a) is $\tanh z \le \min\{z, 1\}$ with $z = u\ell^{1/2}/2$ (and is trivial when $\ell = 0$).

(b) The function $\tanh$ is concave on $[0,\infty)$ with $\tanh 0 = 0$, so $\tanh z \ge z \tanh 1$ for $z \in [0,1]$ and $\tanh z \ge \tanh 1$ for $z \ge 1$, i.e. $\tanh z \ge \tanh(1)\min\{z, 1\}$. With $z = u\ell^{1/2}/2$ this gives the companion lower bound to (a),

$$\tanh(1)\ \min\{\ell u^2,\ 2\ell^{1/2}u\}\ \le\ \psi_\ell(u), \qquad u \ge 0, \tag{15}$$

and, $\min$ being nondecreasing in each argument, $\Delta(u) \le \min\{\beta' u^2, 2B'u\} \le \psi_\ell(u)$ for all $u \ge 0$ as soon as $\tanh(1)\ell \ge \beta'$ and $\tanh(1)\ell^{1/2} \ge B'$. Both hold for $\ell = 2\max\{\beta', B'^2\}$, because $2\tanh 1 > 1$ and $\sqrt{2}\,\tanh 1 > 1$. □

A sufficient condition for the relaxed curvature bound. Suppose $f \in C^2$, with $\nabla^2 f \succeq m \operatorname{Id}$ outside a ball of radius $\varrho > 0$ and $\nabla^2 f \succeq -\beta \operatorname{Id}$ inside, where $m > 0$ and $\beta \ge 0$. For $u = \|x - y\| > 0$, the part of the segment $[y, x]$ inside the ball has length at most $\min(u, 2\varrho)$. Integrating the Hessian along this segment therefore gives

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge mu^2 - (m + \beta)\min\{u^2, 2\varrho u\}.$$

The case $x = y$ is immediate. Applying Lemma A.4(b) with $\beta' = m + \beta$ and $B' = (m + \beta)\varrho$ proves (8) with $\ell = 2\max\{m + \beta, (m + \beta)^2\varrho^2\}$.

Lemma A.5 (Weak curvature under Gaussian smoothing)**.** *Suppose π is (m, ℓ)-strongly log-concave outside a ball as in (8), for some $m, \ell \ge 0$, let $f = -\log \pi$ and $c_r = 1 + mr^2$ for $r \ge 0$. Then,*

- *(a) With the convention $\pi_0 = \pi$, for any $r \ge 0$, π_r is (m_r, ℓ_r)-strongly log-concave outside a ball, with*

$$m_r = \frac{m}{c_r} = \frac{m}{1 + mr^2}\,, \qquad \ell_r = \frac{\ell}{c_r^2} = \frac{\ell}{(1 + mr^2)^2}\,; \tag{16}$$

that is, $f_r = -\log \pi_r$ is C^1 and satisfies (8) with (m_r, ℓ_r) in place of (m, ℓ).

(b) For any $r > 0$, f_r is C^∞ and $\nabla^2 f_r(x) \succeq (m_r - \ell_r)\,\mathrm{Id}$ for any $x \in \mathbb{R}^d$.

Proof. (a) The claim is the hypothesis when $r = 0$, so let $r > 0$ and define $h(x) = f(x) - \frac{m}{2}\|x\|^2$, which is continuously differentiable. For any $x, y \in \mathbb{R}^d$, (8) yields

$$\langle \nabla h(x) - \nabla h(y), x - y\rangle = \langle \nabla f(x) - \nabla f(y), x - y\rangle - m\|x - y\|^2 \geq -\psi_\ell(\|x - y\|)\,. \tag{17}$$

Step 1: e^{-h} grows at most exponentially. Fix a unit vector $\theta \in \mathbb{R}^d$ and $t > 0$. Applying (17) with $x = t\theta$ and $y = 0$, dividing by t, and using $\psi_\ell(t) \leq 2\ell^{1/2}t$ (Lemma A.4(a)), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} h(t\theta) = \langle \nabla h(t\theta), \theta\rangle \geq \langle \nabla h(0), \theta\rangle - \frac{\psi_\ell(t)}{t} \geq \langle \nabla h(0), \theta\rangle - 2\ell^{1/2}\,.$$

Integrating over $t \in [0, \|x\|]$ along the ray $\theta = x/\|x\|$ gives, for all $x \in \mathbb{R}^d$,

$$h(x) \geq h(0) + \langle \nabla h(0), x\rangle - 2\ell^{1/2}\|x\|\,. \tag{18}$$

Consequently $e^{-h(y)} \leq e^{-h(0)}e^{A\|y\|}$ with $A = \|\nabla h(0)\| + 2\ell^{1/2}$. Hence, for every $s > 0$, the function

$$h_s := -\log\big(\kappa_s * e^{-h}\big) \tag{19}$$

is well defined and finite: the Gaussian convolution of a positive function with at most exponential growth is finite, positive, and C^∞, since differentiation under the integral sign is justified by the domination $|\partial_x^k \kappa_s(x - y)|e^{A\|y\|} \leq C_{k,s}(1 + \|x - y\|)^k \kappa_{2s}(x - y)e^{A\|y\|}$, which is integrable in y locally uniformly in x. We also set $h_0 = h$.

Step 2: propagation of the weak-convexity bound. We claim that, for every $s \geq 0$ and $x, y \in \mathbb{R}^d$,

$$\langle \nabla h_s(x) - \nabla h_s(y), x - y\rangle \geq -\psi_\ell(\|x - y\|)\,. \tag{20}$$

For $s = 0$ this is (17), so let $s > 0$. Assume first $\ell > 0$. Let $(P_t)_{t\geq 0}$ be the heat semigroup of a standard d-dimensional Brownian motion, $P_t g = \kappa_{\sqrt{t}} * g$, so that $h_s = -\log P_{s^2} e^{-h}$. By (17), h belongs to the class $\mathcal{F}_\ell$ of (Conforti, 2024, Theorem 2.1), namely the set of continuously differentiable $g : \mathbb{R}^d \to \mathbb{R}$ such that $\langle \nabla g(x) - \nabla g(y), x - y\rangle \geq -\psi_\ell(\|x - y\|)$ for all $x, y \in \mathbb{R}^d$. That theorem states that $\mathcal{F}_\ell$ is invariant under the flow $g \mapsto -\log P_t e^{-g}$ for every $t \in [0, \infty)$; applied with $t = s^2$, it gives $h_s \in \mathcal{F}_\ell$, which is (20). Assume now $\ell = 0$. Since $\psi_0 = 0 \geq -\psi_L$, h belongs to $\mathcal{F}_L$ for every $L > 0$, and the previous case yields $\langle \nabla h_s(x) - \nabla h_s(y), x - y\rangle \geq -\psi_L(\|x - y\|) \geq -L\|x - y\|^2$ for every $L > 0$, the last step by Lemma A.4(a). Letting $L \downarrow 0$ gives (20) with $\psi_0 = 0$.

Step 3: rescaling. Since $f(y) = h(y) + m\|y\|^2/2$, setting $\tau_r := r/c_r^{1/2}$, so that $\tau_r^2 = r^2/c_r$, and completing the square, we have for all $x, y \in \mathbb{R}^d$

$$\frac{\|x - y\|^2}{2r^2} + \frac{m}{2}\|y\|^2 = \frac{\|y - x/c_r\|^2}{2\tau_r^2} + \frac{m}{2c_r}\|x\|^2\,. \tag{21}$$

Consequently, using $\pi = e^{-f}$ and $(2\pi\tau_r^2)^{d/2}/(2\pi r^2)^{d/2} = c_r^{-d/2}$,

$$\begin{aligned}
\pi_r(x) &= (2\pi r^2)^{-d/2} \int_{\mathbb{R}^d} e^{-\|x-y\|^2/(2r^2) - m\|y\|^2/2 - h(y)}\,\mathrm{d}y \\
&= c_r^{-d/2} e^{-m\|x\|^2/(2c_r)} \big(\kappa_{\tau_r} * e^{-h}\big)\,(x/c_r) \\
&= c_r^{-d/2} e^{-m\|x\|^2/(2c_r)} e^{-h_{\tau_r}(x/c_r)}\,.
\end{aligned}$$

Taking logarithms,

$$f_r(x) = \frac{d}{2}\log c_r + \frac{m}{2c_r}\|x\|^2 + h_{\tau_r}(x/c_r)\,, \tag{22}$$

so that, by Step 1, f_r is C^∞ and

$$\nabla f_r(x) = \frac{m}{c_r}\,x + \frac{1}{c_r}\,\nabla h_{\tau_r}(x/c_r)\,.$$

Therefore, for any $x, y \in \mathbb{R}^d$, applying (20) with $s = \tau_r$ at the points x/c_r and y/c_r,

$$\begin{aligned}
\langle \nabla f_r(x) - \nabla f_r(y), x - y\rangle &= \frac{m}{c_r}\|x - y\|^2 + \langle \nabla h_{\tau_r}(x/c_r) - \nabla h_{\tau_r}(y/c_r), x/c_r - y/c_r\rangle \\
&\geq \frac{m}{c_r}\|x - y\|^2 - \psi_\ell(\|x - y\|/c_r) = m_r\|x - y\|^2 - \psi_{\ell_r}(\|x - y\|)\,,
\end{aligned}$$

where the last equality follows from the identity $\psi_\ell(u/c) = \psi_{\ell/c^2}(u)$, valid for all $u \geq 0$, $c > 0$ and $\ell \geq 0$, which is immediate from the definition of ψ_ℓ since $(\ell/c^2)^{1/2} = \ell^{1/2}/c$. This proves (a).

(b) Let $r > 0$. By (22) and Step 1, f_r is C^∞. Fix $x \in \mathbb{R}^d$, a unit vector $v \in \mathbb{R}^d$ and $\varepsilon > 0$. Applying (a) at the points $x + \varepsilon v$ and x, and dividing by ε^2, gives

$$\frac{1}{\varepsilon}\,\langle \nabla f_r(x + \varepsilon v) - \nabla f_r(x), v\rangle \geq m_r - \frac{\psi_{\ell_r}(\varepsilon)}{\varepsilon^2}\,.$$

By Lemma A.4(a), $\psi_{\ell_r}(\varepsilon)/\varepsilon^2 \leq \ell_r$ for every $\varepsilon > 0$ (including when $\ell_r = 0$). Letting $\varepsilon \downarrow 0$, the left-hand side converges to $v^\top \nabla^2 f_r(x) v$ because f_r is C^2, and we obtain $v^\top \nabla^2 f_r(x) v \geq m_r - \ell_r$. Since v is an arbitrary unit vector, the claimed Hessian bound follows. □

B Proofs of the main results

Proof of Proposition 3.1. For $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, the triangle inequality, $\mathcal{L}_t(\pi) = \pi$ (A1) and A2–A4 give $W_2(\widetilde{\mathcal{L}}_t^\theta(\rho), \pi) \leq W_2(\widetilde{\mathcal{L}}_t^\theta(\rho), \mathcal{L}_t(\rho)) + W_2(\mathcal{L}_t(\rho), \mathcal{L}_t(\pi)) \leq E_t + L_t R_t\, W_2(\rho, \pi)$. With $d_k := W_2(\rho_k, \pi)$ this yields $d_k \leq q_{t_k} d_{k-1} + E_{t_k}$, and unrolling proves (3). Under the balance condition, $a_k := Q_{k+1,n} E_{t_k}$ satisfies $a_{k-1} \leq \xi_n a_k$ for $k = 2, \dots, n$ and $a_n = E_{t_n}$, so $a_k \leq \xi_n^{n-k} E_{t_n}$, and summing the geometric series gives (4). □

Remarks on the proof of Proposition 3.1. The one-step estimate $W_2(\widetilde{\mathcal{L}}_t^\theta(\rho), \pi) \leq E_t + q_t W_2(\rho, \pi)$ is applied to $\rho = \rho_{k-1}$, a law produced by approximate steps; this is where the uniformity in the input law of A4 is used. No division by a stability factor occurs, so zero factors are allowed, and neither $q_{t_k} < 1$ nor an ordering of the levels is needed for (3). The balance condition

$$q_{t_k} E_{t_{k-1}} \leq \xi_n E_{t_k}, \qquad k = 2, \dots, n, \tag{23}$$

enters only through $a_{k-1} = Q_{k+1,n} q_{t_k} E_{t_{k-1}} \leq \xi_n Q_{k+1,n} E_{t_k} = \xi_n a_k$, obtained by multiplying (23) by $Q_{k+1,n} \geq 0$; it covers zero error bounds, and if $E_{t_n} = 0$ all a_k vanish, even when some stability factors are zero. For $n = 1$ the induction is empty and (4) reads $E_{t_1} \leq E_{t_1}$.

Proof of Proposition 3.9. Inserting $\widetilde{\mathcal{L}}_t^\theta(\pi)$ and using A2, A5, and A6, $W_2(\widetilde{\mathcal{L}}_t^\theta(\rho), \pi) \leq W_2\big(\widetilde{\mathcal{L}}_t^\theta(\rho), \widetilde{\mathcal{L}}_t^\theta(\pi)\big) + W_2\big(\widetilde{\mathcal{L}}_t^\theta(\pi), \pi\big) \leq \widetilde{L}_t\, W_2\big(N_t(\rho), N_t(\pi)\big) + E_t^\pi \leq \widetilde{q}_t\, W_2(\rho, \pi) + E_t^\pi$. All iterates have finite second moments (Section 3.1), so this bound applies at every step. The remaining claims follow as in the proof of Proposition 3.1. □

Lemma B.1 (Fixed-point tracking and comparison of error bounds)**.** *Let A2, A5, and A6 hold. Fix $T > 0$, an integer $n \geq 1$ and levels $t_0, \dots, t_n \in (0, T]$ such that $\widetilde{q}_{t_k} := R_{t_k} \widetilde{L}_{t_k} \in [0, 1)$ and $E_{t_k}^\pi < \infty$ for $k = 0, \dots, n$. Suppose finite constants $H_k \geq 0$ satisfy*

$$W_2\big(\widetilde{\mathcal{L}}_{t_k}^\theta(\rho), \widetilde{\mathcal{L}}_{t_{k-1}}^\theta(\rho)\big) \leq H_k, \qquad \rho \in \mathcal{P}_2(\mathbb{R}^d), \quad k = 1, \dots, n.$$

Set $F_k := E_{t_k}^\pi/(1 - \widetilde{q}_{t_k})$ and $\widetilde{Q}_{k,j} := \prod_{i=k}^{j} \widetilde{q}_{t_i}$, with empty products equal to one. For $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$, define $\rho_k := \widetilde{\mathcal{L}}_{t_k}^\theta(\rho_{k-1})$ for $k = 1, \dots, n$. Then

$$W_2(\rho_n, \pi) \leq B_{\mathrm{fp}} := \widetilde{Q}_{1,n}\big(W_2(\rho_0, \pi) + F_0\big) + F_n + \sum_{k=1}^{n} \frac{\widetilde{Q}_{k,n}}{1 - \widetilde{q}_{t_k}} H_k. \tag{24}$$

The recursive bound $B_{\mathrm{rec}} := \widetilde{Q}_{1,n} W_2(\rho_0, \pi) + \sum_{k=1}^{n} \widetilde{Q}_{k+1,n} E_{t_k}^\pi$ satisfies the identity

$$B_{\mathrm{fp}} - B_{\mathrm{rec}} = \widetilde{Q}_{1,n}\Big(F_0 + F_1 + \frac{H_1}{1 - \widetilde{q}_{t_1}}\Big) + \sum_{k=2}^{n} \widetilde{Q}_{k,n}\Big(F_k - F_{k-1} + \frac{H_k}{1 - \widetilde{q}_{t_k}}\Big). \tag{25}$$

In particular, if

$$\widetilde{q}_{t_k} \geq \widetilde{q}_{t_{k-1}} \quad \text{and} \quad E_{t_{k-1}}^\pi \leq E_{t_k}^\pi + H_k, \qquad k = 2, \dots, n,$$

then $B_{\mathrm{fp}} \geq B_{\mathrm{rec}} + \widetilde{Q}_{1,n}(F_0 + F_1)$. The error condition holds when $E_{t_k}^\pi = W_2(\pi, \widetilde{\mathcal{L}}_{t_k}^\theta(\pi))$ at every visited level; for arbitrary upper bounds it must be checked separately. At a repeated level $t_0 = \cdots = t_n = t$, choosing $H_k = 0$ gives $B_{\mathrm{fp}} - B_{\mathrm{rec}} = 2\widetilde{q}_t^{\,n} E_t^\pi/(1 - \widetilde{q}_t)$. When $n = 1$, the sum over $k \geq 2$ and the comparison conditions are empty.

Proof. By A2 and A6, each $\widetilde{\mathcal{L}}^{\theta}_{t_k}$ is a contraction of the complete metric space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$. Let ν_k be its unique fixed point. Applying (11) to ν_k and rearranging gives

$$W_2(\pi, \nu_k) \le F_k, \qquad k = 0, \dots, n. \tag{26}$$

Inserting $\widetilde{\mathcal{L}}^{\theta}_{t_k}(\nu_{k-1})$ between the fixed points yields

$$W_2(\nu_{k-1}, \nu_k) \le H_k + \widetilde{q}_{t_k} W_2(\nu_{k-1}, \nu_k),$$

hence $W_2(\nu_{k-1}, \nu_k) \le H_k/(1-\widetilde{q}_{t_k})$. Contraction at level t_k and the triangle inequality therefore give

$$W_2(\rho_k, \nu_k) \le \widetilde{q}_{t_k} W_2(\rho_{k-1}, \nu_{k-1}) + \frac{\widetilde{q}_{t_k}}{1-\widetilde{q}_{t_k}} H_k.$$

Unrolling it with $W_2(\rho_0, \nu_0) \le W_2(\rho_0, \pi) + F_0$ and $W_2(\rho_n, \pi) \le W_2(\rho_n, \nu_n) + F_n$ gives (24).

Since $E^{\pi}_{t_k} = (1-\widetilde{q}_{t_k})F_k$, summation by parts gives

$$\sum_{k=1}^{n} \widetilde{Q}_{k+1,n} E^{\pi}_{t_k} = F_n - \widetilde{Q}_{1,n} F_1 - \sum_{k=2}^{n} \widetilde{Q}_{k,n}(F_k - F_{k-1}).$$

Subtracting B_{rec} from B_{fp} proves (25). Under the two comparison conditions, for each $k = 2, \dots, n$,

$$F_k - F_{k-1} + \frac{H_k}{1-\widetilde{q}_{t_k}} = \frac{E^{\pi}_{t_k} + H_k}{1-\widetilde{q}_{t_k}} - \frac{E^{\pi}_{t_{k-1}}}{1-\widetilde{q}_{t_{k-1}}} \ge E^{\pi}_{t_{k-1}} \Big(\frac{1}{1-\widetilde{q}_{t_k}} - \frac{1}{1-\widetilde{q}_{t_{k-1}}} \Big) \ge 0.$$

All weights and H_1 are nonnegative, proving the comparison. For the exact defects $e_k := W_2(\pi, \widetilde{\mathcal{L}}^{\theta}_{t_k}(\pi))$, the reverse triangle inequality gives $|e_k - e_{k-1}| \le W_2(\widetilde{\mathcal{L}}^{\theta}_{t_k}(\pi), \widetilde{\mathcal{L}}^{\theta}_{t_{k-1}}(\pi)) \le H_k$, which verifies the error condition when $E^{\pi}_{t_k} = e_k$. At a repeated level, $H_k = 0$ is admissible and F_k is constant; (25) then reduces to the stated formula. □

A uniform Hölder bound on the sampling operators supplies $H_k = H|t_k - t_{k-1}|^h$, but only the consecutive-level bounds above are needed. A decreasing schedule with $\widetilde{q}_t$ nonincreasing in t satisfies the monotonicity condition; the comparison still requires compatible error bounds.

Corollary B.2 (Combining the two routes)**.** *Let A1–A6 hold. For any schedule $t_1, \dots, t_n$ and any $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$, the certificates $\bar{d}_0 := W_2(\rho_0, \pi)$ and*

$$\bar{d}_k := \min\big\{ q_{t_k} \bar{d}_{k-1} + E_{t_k},\ \widetilde{q}_{t_k} \bar{d}_{k-1} + E^{\pi}_{t_k} \big\}, \qquad k = 1, \dots, n,$$

satisfy $W_2(\rho_k, \pi) \le \bar{d}_k$.

Proof. Both one-step estimates hold at every step (proofs of Proposition 3.1 and Proposition 3.9) and are nondecreasing in the previous certificate; induct on k. □

Remark B.3 (Fixed points of the learned step)**.** *Under A2, A5, and A6, if $\widetilde{q}_t < 1$, $\widetilde{\mathcal{L}}^{\theta}_t$ is a contraction of $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ with a unique fixed point $\widetilde{\pi}^{(t)}$, and (11) at $\rho = \widetilde{\pi}^{(t)}$ gives $W_2(\pi, \widetilde{\pi}^{(t)}) \le E^{\pi}_t/(1-\widetilde{q}_t)$: the plateau of the repeated-level bound is the certified distance from π to the law the repeated sampler converges to, and the Gaussian example of Section 3.4 attains it. The fixed-point argument of Lemma B.1 additionally requires bounds on the change between consecutive sampling operators, uniformly over input laws. Its bound is no smaller than the unrolled recursion under the lemma's monotonicity and error-compatibility conditions; there is no such ordering for arbitrary error bounds.*

Proof of Lemma 3.2. Let (X, Y) be an optimal coupling of (ρ, ν) and $Z \sim \mathcal{N}(0, \mathrm{Id})$ be independent of it. Then $\alpha_t X + \sigma_t Z$ and $\alpha_t Y + \sigma_t Z$ couple $N_t(\rho)$ and $N_t(\nu)$, so $W_2(N_t(\rho), N_t(\nu)) \le \alpha_t \big(\mathbb{E}\|X - Y\|^2\big)^{1/2} = \alpha_t W_2(\rho, \nu)$, with equality for $\rho = \delta_x$ and $\nu = \delta_y$ with $x \ne y$, whose images are Gaussian with covariance $\sigma_t^2\,\mathrm{Id}$ and means $\alpha_t x$, $\alpha_t y$. □

Proof sketch for Theorem 3.3. Dividing trajectories by α_u removes the linear drift. For exact and approximate trajectories x_u, y_u, the rescaled distance $h_u := \|x_u - y_u\|/\alpha_u$ satisfies $\dot{h}_u \geq \gamma_u \dot{\gamma}_u(\mu_{\gamma_u} h_u - e_{\gamma_u})$ almost everywhere at positive times. Thus curvature controls backward stability, while score errors act as a forcing term. A backward integrating factor and the substitution $r = \gamma_u$ yield (6). Comparing two exact trajectories removes the forcing and gives L_t; using the same terminal state for exact and approximate trajectories gives E_t. These estimates imply linear growth, preserving $\mathcal{P}_2(\mathbb{R}^d)$, and transfer to W_2 by transporting input couplings. The endpoint limit is justified by the integrability assumptions in Lemma A.1.

Proof of Theorem 3.3. Apply Lemma A.1 at $s = 0$, first to two exact trajectories ($e_r = 0$), then to an exact and an approximate trajectory with the same terminal state. Since $\alpha_0 = 1$ and $\gamma_0 = 0$, this gives

$$\|\Phi_{t\to 0}(x) - \Phi_{t\to 0}(y)\| \leq L_t \|x - y\|,$$
$$\|\widetilde{\Phi}^{\theta}_{t\to 0}(x) - \Phi_{t\to 0}(x)\| \leq E_t.$$

Both constants are finite by the lemma. These inequalities imply linear growth of both Borel maps, since $\|\widetilde{\Phi}^{\theta}_{t\to 0}(x)\| \leq \|\Phi_{t\to 0}(0)\| + L_t\|x\| + E_t$; their pushforwards therefore preserve $\mathcal{P}_2(\mathbb{R}^d)$. Pushing an optimal coupling of ρ_1, ρ_2 through the exact map gives $W_2(D_{\pi,t}(\rho_1), D_{\pi,t}(\rho_2)) \leq L_t W_2(\rho_1, \rho_2)$. Finally, coupling the exact and approximate outputs through the same $X_t \sim N_t(\rho)$ gives $W_2(\mathcal{L}_t(\rho), \widetilde{\mathcal{L}}^{\theta}_t(\rho)) \leq E_t$, uniformly over $\rho \in \mathcal{P}_2(\mathbb{R}^d)$. □

Proof of Corollary 3.4. By Lemma A.3, f_r is C^∞ and μ_r-strongly convex with $\mu_r = m/(1 + mr^2)$; substituting μ_r and $e_r \leq M_0 + M_1/r$ into (6) and integrating gives the result. □

Proof of Theorem 3.6. Lemma A.5 gives $\nabla^2 f_r \succeq \mu_r \,\mathrm{Id}$ with $\mu_r = m/(1 + mr^2) - \ell/(1 + mr^2)^2$. This profile is bounded and $re_r \leq M_0 r + M_1$, so the integrability requirements of Theorem 3.3 hold. Direct integration gives $\exp\big(-\int_0^r u\mu_u \,\mathrm{d}u\big) = A_{m,\ell}(r)/\sqrt{1 + mr^2}$. Substitution into (6) yields L_t. Since $A_{m,\ell}$ is nondecreasing, the error integral there is at most $A_{m,\ell}(\gamma_t)B_m(\gamma_t)$. Lemma 3.2 gives R_t; rearranging $R_t L_t < 1$ gives (10). □

Theorem B.4 (Error of a trained consistency model, full statement)**.** *Let $\widetilde{\Phi}_{t\to s}$ and $\Phi_{t\to s}$, $0 \leq s \leq t$, be the backward flow maps from t to s of the approximate-score ODE (14), with s_θ replaced by the score approximation used by the training solver, and of the probability-flow ODE (1). Assume the hypotheses of Theorem 3.3 for this score approximation, with e_r in (5) denoting its error, and that the exact flow transports the noised marginals, $(\Phi_{t\to s})_{\#}\varpi_t = \varpi_s$ for $0 \leq s \leq t$. Fix a grid $0 = \tau_0 < \cdots < \tau_N = T$ and let $E_{s,t}$ be defined by (12) with this e_r. Let the learned maps $\widetilde{\Phi}^{\theta}_{\tau_i\to 0} : \mathbb{R}^d \to \mathbb{R}^d$, $0 \leq i \leq N$, be Borel, with $\widetilde{\Phi}^{\theta}_{0\to 0}$ the identity, so that $\widetilde{L}_{\tau_0} = 1$, and $\widetilde{\Phi}^{\theta}_{\tau_i\to 0}$ Lipschitz with finite constant $\widetilde{L}_{\tau_i}$ for $1 \leq i \leq N$. For $0 \leq i < N$, define*

$$\varepsilon_i := \Big(\mathbb{E}_{X\sim\varpi_{\tau_{i+1}}} \big\|\widetilde{\Phi}^{\theta}_{\tau_{i+1}\to 0}(X) - \widetilde{\Phi}^{\theta}_{\tau_i\to 0}(\widetilde{\Phi}_{\tau_{i+1}\to\tau_i}(X))\big\|^2\Big)^{1/2},$$
$$\varepsilon_i^{\infty} := \sup_{x\in\mathbb{R}^d} \big\|\widetilde{\Phi}^{\theta}_{\tau_{i+1}\to 0}(x) - \widetilde{\Phi}^{\theta}_{\tau_i\to 0}(\widetilde{\Phi}_{\tau_{i+1}\to\tau_i}(x))\big\|.$$

The operators $\widetilde{D}^{\theta}_{\pi,\tau_k}$ and $\widetilde{\mathcal{L}}^{\theta}_{\tau_k}$ of Section 3.1 preserve $\mathcal{P}_2(\mathbb{R}^d)$, all ε_i are finite, and, for $k = 1, \ldots, N$,

$$W_2\big(\pi, \widetilde{\mathcal{L}}^{\theta}_{\tau_k}(\pi)\big) \leq \sum_{i=0}^{k-1} \varepsilon_i + \sum_{i=0}^{k-1} \widetilde{L}_{\tau_i} E_{\tau_i,\tau_{i+1}}, \tag{27}$$

If additionally $\varepsilon_i^{\infty} < \infty$ for $0 \leq i < k$, then

$$\sup_{\rho\in\mathcal{P}_2(\mathbb{R}^d)} W_2\big(\mathcal{L}_{\tau_k}(\rho), \widetilde{\mathcal{L}}^{\theta}_{\tau_k}(\rho)\big) \leq \sum_{i=0}^{k-1} \varepsilon_i^{\infty} + \sum_{i=0}^{k-1} \widetilde{L}_{\tau_i} E_{\tau_i,\tau_{i+1}}. \tag{28}$$

Thus, for schedules in the grid, A5 holds with the right-hand side of (27), and A6 holds with the constants $\widetilde{L}_{\tau_k}$. If $\varepsilon_i^{\infty} < \infty$ for every $0 \leq i < N$, then A4 also holds, with the right-hand side of (28).

Proof sketch for Theorem B.4. The proof telescopes $\widetilde{\Phi}^{\theta}_{\tau_k\to 0}(x) - \Phi_{\tau_k\to 0}(x)$ along the exact trajectory through the grid: each cell contributes its consistency error and the discrepancy between the two flows over the cell, carried to time zero by $\widetilde{\Phi}^{\theta}_{\tau_i\to 0}$.

Proof of Theorem B.4. By Lemma A.1, each exact cell flow is Lipschitz and the approximate-score cell flow is at a uniformly bounded distance from it. Both have at most linear growth. Since the network is Lipschitz and $\varpi_{\tau_i} \in \mathcal{P}_2(\mathbb{R}^d)$, all ε_i are finite; the network pushforwards also preserve $\mathcal{P}_2(\mathbb{R}^d)$. Fix k and $x \in \mathbb{R}^d$, and set $x_{\tau_i} := \Phi_{\tau_k\to\tau_i}(x)$ for $0 \le i \le k$, so that $x_{\tau_k} = x$ and $x_{\tau_0} = \Phi_{\tau_k\to 0}(x)$. Since $\widetilde{\Phi}^{\theta}_{0\to 0}$ is the identity,

$$\widetilde{\Phi}^{\theta}_{\tau_k\to 0}(x) - \Phi_{\tau_k\to 0}(x) = \sum_{i=0}^{k-1} \big[\widetilde{\Phi}^{\theta}_{\tau_{i+1}\to 0}(x_{\tau_{i+1}}) - \widetilde{\Phi}^{\theta}_{\tau_i\to 0}(x_{\tau_i})\big].$$

Write $\hat{x}_{\tau_i} := \widetilde{\Phi}_{\tau_{i+1}\to\tau_i}(x_{\tau_{i+1}})$ and split each summand as $[\widetilde{\Phi}^{\theta}_{\tau_{i+1}\to 0}(x_{\tau_{i+1}}) - \widetilde{\Phi}^{\theta}_{\tau_i\to 0}(\hat{x}_{\tau_i})] + [\widetilde{\Phi}^{\theta}_{\tau_i\to 0}(\hat{x}_{\tau_i}) - \widetilde{\Phi}^{\theta}_{\tau_i\to 0}(x_{\tau_i})]$. The restriction of the exact trajectory $u \mapsto \Phi_{\tau_k\to u}(x)$ to $[\tau_i, \tau_{i+1}]$ and the trajectory $u \mapsto \widetilde{\Phi}_{\tau_{i+1}\to u}(x_{\tau_{i+1}})$ solve (1) and (14) with the same terminal state $x_{\tau_{i+1}}$, so Lemma A.1, in its $s = 0$ form for $i = 0$, gives $\|\hat{x}_{\tau_i} - x_{\tau_i}\| \le E_{\tau_i,\tau_{i+1}}$, and the Lipschitz property of $\widetilde{\Phi}^{\theta}_{\tau_i\to 0}$, the identity for $i = 0$, bounds the second bracket by $\widetilde{L}_{\tau_i} E_{\tau_i,\tau_{i+1}}$. If $\varepsilon_i^{\infty} < \infty$ for all $i < k$, the first bracket is bounded by ε_i^{∞} pointwise, which proves $\sup_x \|\widetilde{\Phi}^{\theta}_{\tau_k\to 0}(x) - \Phi_{\tau_k\to 0}(x)\| \le \sum_i \varepsilon_i^{\infty} + \sum_i \widetilde{L}_{\tau_i} E_{\tau_i,\tau_{i+1}}$. For $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $Y \sim N_{\tau_k}(\rho)$, the pair $(\Phi_{\tau_k\to 0}(Y), \widetilde{\Phi}^{\theta}_{\tau_k\to 0}(Y))$ couples $\mathcal{L}_{\tau_k}(\rho)$ and $\widetilde{\mathcal{L}}^{\theta}_{\tau_k}(\rho)$, which gives (28). For (27), take $X \sim \varpi_{\tau_k}$, so that $x_{\tau_{i+1}} = \Phi_{\tau_k\to\tau_{i+1}}(X) \sim \varpi_{\tau_{i+1}}$ by the transport hypothesis and the first bracket of cell i has L^2 norm ε_i. Minkowski's inequality in L^2, the pathwise bound on the second brackets, $(\Phi_{\tau_k\to 0})_{\#}\varpi_{\tau_k} = \pi$ and the coupling $(\Phi_{\tau_k\to 0}(X), \widetilde{\Phi}^{\theta}_{\tau_k\to 0}(X))$ of π and $\widetilde{\mathcal{L}}^{\theta}_{\tau_k}(\pi)$ give (27). □

Early stopping. For a grid $0 < \delta = \tau_0 < \cdots < \tau_N = T$ with learned maps $\widetilde{\Phi}^{\theta}_{\tau_i\to\delta}$ that return to time δ and $\widetilde{\Phi}^{\theta}_{\delta\to\delta}$ the identity, the target is ϖ_δ and noising must start from time δ. Define, for $t \ge \delta$,

$$N_t^{(\delta)}(\rho) := \mathrm{Law}\left(\frac{\alpha_t}{\alpha_\delta} X + \alpha_t \sqrt{\gamma_t^2 - \gamma_\delta^2}\, Z\right), \qquad X \sim \rho, \quad Z \sim \mathcal{N}(0, \mathrm{Id}), \quad X \perp Z.$$

At each grid level, use $\mathcal{L}_t^{(\delta)} := (\Phi_{t\to\delta})_{\#} \circ N_t^{(\delta)}$ and $\widetilde{\mathcal{L}}_t^{\theta,(\delta)} := (\widetilde{\Phi}^{\theta}_{t\to\delta})_{\#} \circ N_t^{(\delta)}$. Under the remaining hypotheses of Theorem B.4, (27) and (28) hold with $(\pi, \mathcal{L}_t, \widetilde{\mathcal{L}}_t^{\theta})$ replaced by $(\varpi_\delta, \mathcal{L}_t^{(\delta)}, \widetilde{\mathcal{L}}_t^{\theta,(\delta)})$, using this grid to define the residuals, the same cell errors $E_{s,t}$ for $s \ge \delta$, and $\widetilde{L}_{\tau_0} = 1$. Indeed, $N_t^{(\delta)}(\varpi_\delta) = \varpi_t$, and the proof telescopes from $\tau_0 = \delta$. The noising factor becomes $R_t^{(\delta)} = \alpha_t/\alpha_\delta$.

C Comparison with other multistep bounds

The multistep bound of Lyu et al. (2024) Their sampler is (2) for the OU process restarted from p_δ, with $\alpha_t = e^{-(t-\delta)}$, $\sigma_t^2 = 1 - e^{-2(t-\delta)}$ and a learned map returning to $\delta > 0$. Thus $\pi = p_\delta$ and $R_t = e^{-(t-\delta)}$. Their initial law $\mathcal{N}(0, \mathrm{Id})$ is OU-invariant, so their first step is $\widetilde{\mathcal{L}}^{\theta}_T(\mathcal{N}(0, \mathrm{Id}))$.

Their Assumption 3.5 is an affine stability bound on the training grid: for a universal constant C_{Lip}, one may write $\|f_\theta(x,t) - f_\theta(y,t)\| \le \Lambda\|x - y\| + \beta_t$, where $\Lambda := C_{\mathrm{Lip}} L_f$ and $\beta_t := C_{\mathrm{Lip}} t(L_f \varepsilon_{\mathrm{sc}} + \varepsilon_{\mathrm{cm}})$. Coupling the inputs and applying Minkowski's inequality gives the affine stability of Remark 3.10 with $\widetilde{L}_t = \Lambda$ and this β_t. Using the approximation bound stated in their Theorem 3.7, a coupling through the exact denoiser gives A5 with $E_t^{\pi} := \mathrm{err}(t)$ at grid levels, where

$$\mathrm{err}(t) := C\big[t^2(\varepsilon_{\mathrm{cm}} + L_f \varepsilon_{\mathrm{sc}}) + t(\varepsilon_{\mathrm{cm}} + L_f \varepsilon_{\mathrm{sc}} + L_f L_s^{3/2} d^{1/2} h) + t^{1/2} L_f L_s d^{1/2} h\big],$$

where C is universal, $\varepsilon_{\mathrm{sc}}, \varepsilon_{\mathrm{cm}}$ are their L^2 score and consistency errors, h the training-grid step and L_s the Lipschitz constant of the noised scores. By Remark 3.10, Proposition 3.9 then applies with factor ΛR_t and error budget $\bar{E}_t := \mathrm{err}(t) + \beta_t$, recovering the recursion underlying their Theorem 3.8 and Corollary 3.10. For two levels $t_1 \ge t_2 > \delta$, this yields

$$W_2(\rho_2, p_\delta) \le \Lambda^2 e^{-(t_1+t_2-2\delta)} W_2(\rho_0, p_\delta) + \Lambda e^{-(t_2-\delta)} \bar{E}_{t_1} + \bar{E}_{t_2}.$$

The distance to p_{data} adds a term of order $(\sqrt{d} \vee \mathfrak{m})\sqrt{\delta}$, with $\mathfrak{m}^2$ its second moment (their Eq. (35)).

Proposition 3.1 instead gives $W_2(\rho_2, \pi) \le q_{t_2} q_{t_1} W_2(\rho_0, \pi) + q_{t_2} E_{t_1} + E_{t_2}$, with factors certified from the exact denoiser. For an m-strongly log-concave p_δ, since $\gamma_t^2 = e^{2(t-\delta)} - 1$, Corollary 3.4 gives $q_t = (1 + m(e^{2(t-\delta)} - 1))^{-1/2} < 1$ for every $t > \delta$ and $E_t \le B_m(\sqrt{e^{2(t-\delta)} - 1})$. The affine recursion has a coefficient below one when $t > \delta + \log \Lambda$. These thresholds use different assumptions: ours requires curvature control and a uniform approximation bound; theirs requires the affine stability condition and target-marginal error bounds. Moreover, $\bar{E}_t$ has polynomial growth in t, whereas the displayed bound on E_t grows like e^t when $M_0 > 0$, so the comparison does not order the error budgets. Their Theorem C.1 constructs the affine stability bound under bounded support and uniform training errors; those extra hypotheses are not required by their main results.

The multistep bound of Chen et al. (2025) Use VE noising $X_r = X_0 + rZ$, with $Z \sim \mathcal{N}(0, \mathrm{Id})$ independent of X_0, and sampling levels $r_1 > r_2 > 0$ on a uniform training grid of spacing Δr. Write $c := \epsilon_{\mathrm{CM}}/\Delta r$, where ϵ_{CM} is their L^2 consistency error, and $D_0 := W_2(\delta_0, \pi)$. For target samples and learned outputs bounded in norm by R, their Theorem 2 gives the two-step certificate

$$B^{(2)}_{\mathrm{Chen}} := \min\left\{2R,\ 2R\left(\frac{D_0^2}{4r_1^2} + \frac{c^2 r_1^2}{4r_2^2}\right)^{1/4} + c r_2\right\}. \tag{29}$$

Here we strengthen their Eq. (6) by retaining the actual second moment D_0^2 in their proof, Eq. (15), and by applying the support bound $2R$. The initial noisy law is $\mathcal{N}(0, r_1^2 \,\mathrm{Id})$ in both analyses: in our notation it is $N_{r_1}(\delta_0)$. Under A1–A4, Proposition 3.1 instead gives

$$B^{(2)}_{\mathrm{ours}} := q_{r_2} q_{r_1} D_0 + q_{r_2} E_{r_1} + E_{r_2}. \tag{30}$$

Their consistency error alone does not imply the uniform bounds E_r. The following example verifies these additional bounds and shows a quantitative gain from contraction.

Take $\pi = \mathrm{Unif}[-1, 1]$. Let F_r be the CDF of $U + rZ$ for $U \sim \pi$ independent of Z, and let Φ_{N} be the standard normal CDF. The one-dimensional probability flow preserves quantiles, so its exact endpoint map is $g_r(x) = 2F_r(x) - 1$. Differentiating gives $g_r'(x) = \Phi_{\mathrm{N}}((x+1)/r) - \Phi_{\mathrm{N}}((x-1)/r)$, maximized at $x = 0$. Thus, with $a := \sqrt{2/\pi}$,

$$q_r = \mathrm{Lip}(g_r) = 2\Phi_{\mathrm{N}}(1/r) - 1 \le a/r, \qquad D_0 = 1/\sqrt{3}.$$

For $\kappa > 0$, use the learned maps $\widetilde{g}_r = (1 - \kappa r) g_r$ at $0 < r \le T < 1/\kappa$, with $\widetilde{g}_0(x) = x$. Their positive-time outputs lie in $[-1, 1]$.[2] For the exact probability-flow trajectory (Y_r) started from U, $g_r(Y_r) = U$. Each grid residual therefore has L^2 norm $\kappa \Delta r/\sqrt{3}$, so $c = \kappa/\sqrt{3}$. The pointwise approximation error is at most κr, while the learned map sends the noised target to $\mathrm{Law}((1 - \kappa r)U)$. Hence admissible profiles for our two recursions are

$$E_r = \sqrt{3}\, cr, \qquad E_r^\pi = cr, \qquad \widetilde{q}_r = (1 - \kappa r) q_r.$$

This uses the abstract propositions; no full-support curvature assumption is needed. At $\kappa = 0.01$, $T = r_1 = 2$, $r_2 = 1$ and $\Delta r = 1$, (30) is 0.1746, whereas (29) is 0.7659. The learned-map certificate of Proposition 3.9 is 0.1600.

The gain also concerns the error rate. For an integer $j \ge 3$, take $c = j^{-4}$, a unit training grid up to $T = j^2$, and the two sampling levels $r_1 = j^2$, $r_2 = j$. These choices satisfy $\kappa T < 1$. Using $q_r \le a/r$ in (30) gives

$$B^{(2)}_{\mathrm{ours}} \le K c^{3/4}, \qquad K := a^2/\sqrt{3} + \sqrt{3}(1 + a) < 3\sqrt{2}.$$

In contrast, for any decreasing schedule with $N \ge 2$ steps on this same grid, the sum inside the fourth root in their Eq. (6) contains a term $c^2 (r_{i-1}/r_i)^2/4 \ge c^2/4$. Their certificate is therefore at least $\sqrt{2}\, c^{1/2}$, even with the second-moment refinement and the support cap in (29). Since $K/j < \sqrt{2}$, our two-step certificate is strictly smaller than every such multistep certificate. This example shows that additional stability and uniform accuracy can yield a sharper guarantee. These comparisons concern upper bounds, not an ordering of actual sampling errors.

[2] We use boundedness at positive sampling times, as required by the proof of their Theorem 2. The printed output bound also includes time zero, incompatible with the global identity condition in their Assumption 1.

D EXPERIMENTAL ESTIMATORS AND IMPLEMENTATION DETAILS

This appendix details the diagnostics used in Section 4. They are finite-sample estimators of the quantities in Proposition 3.9. To distill SD 1.5, we run the `diffusers` consistency distillation script (https://github.com/huggingface/diffusers/tree/main/examples/consistency_distillation) for 1000 epochs on Conceptual 12M (Changpinyo et al., 2021). We evaluate all quantities at 50 VE noise levels $\{\sigma_j\}_{j=1}^{50}$ from the schedule Karras et al. (2022).

D.1 LATENT COORDINATES AND EXPERIMENTAL SETUP

For a clean latent $z \sim \pi$ and $\epsilon \sim \mathcal{N}(0, \mathrm{Id})$, write $r = \gamma_t = \sigma_t/\alpha_t$ and use the normalized noisy coordinate

$$u = z + r\epsilon, \qquad F_t(u) := \widetilde{\Phi}^{\theta}_{t\to 0}(\alpha_t u). \tag{31}$$

Thus $F_t(u)$ is the clean-latent prediction of the learned denoiser. In the VE experiment $\alpha_t = 1$ and $r = t = \sigma$; for TrigFlow $\alpha_t = \cos t$, $\sigma_t = \sin t$, and $r = \tan t$.

For SD 1.5, AFHQ-cat images are resized to 512×512 and encoded with the SD 1.5 VAE, producing scaled latents of size $4 \times 64 \times 64$ ($d = 16384$ after flattening). The current pool contains 2048 target latents and uses the fixed prompt `a high quality photo of a cat`, without classifier-free guidance. The VE grid contains 50 positive levels, from approximately 0.13445 to 14.61465. All reported distances are in unsquared latent Euclidean units.

For the CelebA TrigFlow experiment, images are resized and center-cropped with the transform used to train the local VAE, and the VAE posterior mean is used directly as the latent representation. We use at most $50{,}000$ CelebA images and a fixed $90/10$ train/validation split, giving $45{,}000$ calibration latents and $5{,}000$ held-out latents in the run used for the plots. The latent dimension is $d = 32$. The profiles shown in Figs. 2b and 3 are calibrated on the training latent pool.

The CelebA LCM uses the trigonometric noising map

$$x_t = \cos(t)x + \sigma_{\mathrm{data}} \sin(t)z, \qquad \sigma_{\mathrm{data}} = 1, \tag{32}$$

with $z \sim \mathcal{N}(0, I)$. We evaluate 100 TrigFlow times obtained by first constructing a Karras grid with $\sigma \in [0.002, 80]$ and $\rho = 7$, and then mapping it to TrigFlow time through $t = \arctan(\sigma/\sigma_{\mathrm{data}})$. Hence the grid spans approximately $t \in [0.002, 1.558]$.

D.2 PAIRWISE STABILITY ESTIMATOR

The population quantity entering Proposition 3.9 is $\widetilde{q}_t = R_t \widetilde{L}_t$, where $\widetilde{L}_t$ is a Lipschitz constant of the learned denoising operator. At each level, we draw noisy particles $u_i = z_i + r\epsilon_i$ and compute $v_i = F_t(u_i)$. For random index pairs (a, b) we form the secant ratios

$$s_{ab}(t) = \frac{\|v_a - v_b\|_2}{\|u_a - u_b\|_2 + 10^{-12}}. \tag{33}$$

Taking $\widetilde{L}_t$ to be the optimal Lipschitz constant gives $\mathrm{Lip}(F_t) = \alpha_t \widetilde{L}_t$; for the Gaussian noising maps considered here $R_t = \alpha_t$, so the same ratio directly estimates the learned-step factor $\widetilde{q}_t$ when the normalized coordinate u is used.

Let $Q_p(t)$ denote the empirical p-quantile of the ratios. We then estimate

$$\widehat{q}_t = 2Q_{0.995}(t) - Q_{0.99}(t), \tag{34}$$

with 512 noisy particles and 8000 sampled pairs per level for SD 1.5, and with 2048 noisy particles and 12000 sampled pairs per level for CelebA. This high-quantile extrapolation is designed to be less sensitive than the sample maximum while probing the upper tail of the secants.

D.3 TARGET-ONLY ERROR ESTIMATORS

The quantity required by A5 is the target defect

$$e_t^{\pi} := W_2\Big(\pi, \widetilde{\mathcal{L}}_t^{\theta}(\pi)\Big). \tag{35}$$

We estimate this quantity with the same exact finite-sample OT estimator used for $\widehat{D}_0$. For repeat r, draw target particles $X_{r,i} \sim \pi$ and noises $Z_{r,i} \sim \mathcal{N}(0, I)$, and evaluate each level directly from the same clean targets,

$$Y_{r,i}(t) = \widetilde{\Phi}^{\theta}_{t\to 0}(\alpha_t X_{r,i} + \sigma_t Z_{r,i}). \tag{36}$$

For two equal-size clouds $A = \{a_i\}_{i=1}^N$ and $B = \{b_i\}_{i=1}^N$, we use the exact empirical transport distance

$$\widehat{W}_2(A, B) = \left[\frac{1}{N} \min_{\tau \in S_N} \sum_{i=1}^{N} \|a_i - b_{\tau(i)}\|_2^2\right]^{1/2}. \tag{37}$$

The target-only error estimate for repeat r is therefore

$$\widehat{E}^{\pi}_{t,r} := \widehat{W}_2\big(\{X_{r,i}\}_{i=1}^N, \{Y_{r,i}(t)\}_{i=1}^N\big), \tag{38}$$

and the reported profile is the average across repeats,

$$\widehat{E}^{\pi}_t := \frac{1}{R} \sum_{r=1}^{R} \widehat{E}^{\pi}_{t,r}. \tag{39}$$

The resulting $\widehat{E}^{\pi}_t$ is a finite-sample plug-in estimate of the distributional target defect e^{π}_t, rather than a Monte Carlo estimate of a fixed-coupling upper cost.

Within each repeat, the clean particles and noise realizations are held fixed across noise levels, while evaluations at different levels are not chained. For the CelebA TrigFlow experiment we use the same empirical-OT routine, cloud size, and repeat count as for $\widehat{D}_0$; in the run used for the plots this corresponds to $N = 2048$ and $R = 3$. The SD 1.5 experiment likewise uses the same empirical-OT estimator and OT budget as its corresponding $\widehat{D}_0$ computation.

D.4 Relation between E_t and E^{π}_t

The distinction used in Section 4 is important. The error E_t in A4 must satisfy

$$W_2\Big(\mathcal{L}_t(\rho), \widetilde{\mathcal{L}}^{\theta}_t(\rho)\Big) \le E_t \qquad \text{for every } \rho \in \mathcal{P}_2(\mathbb{R}^d), \tag{40}$$

and the analytical expression $B_m(\gamma_t)$ in Corollary 3.4 and Remark 3.5 is derived for this uniform quantity from a pointwise score-approximation bound. By contrast, E^{π}_t in A5 only controls the learned step at the target. Since the exact step preserves π, any admissible uniform E_t also gives an admissible target-only bound $E^{\pi}_t \le E_t$, but the converse need not hold.

For the two parametrizations we compare the measurements with the common analytical expression

$$B_m(r) = \frac{M_0}{m}\big(\sqrt{1 + mr^2} - 1\big) + \frac{M_1}{\sqrt{m}} \operatorname{arcsinh}(\sqrt{m}\, r), \qquad r = \gamma_t. \tag{41}$$

For VE, $r = t = \sigma$. We keep $m = 0.801$ from the stability fit and choose $M_0 = 0$, $M_1 = 110.19$, giving the upper envelope shown in Fig. 3. For the CelebA TrigFlow experiment, $r = \tan t$; keeping the independently fitted stability value $m = 0.5$ and using $M_0 = 0$, $M_1 = 6.10$ gives the analytical overlay shown in the TrigFlow panel. These M_0, M_1 values are empirical envelope fits, whereas the functional dependence on γ_t is the analytical shape predicted by Remark 3.5.

D.5 One-step empirical OT comparison

For the one-step experiment, the initialization is $\rho_0 = \mathcal{N}(0, I_d)$. For each repeat and each VE level t, we sample $G^0_i \sim \rho_0$ and generate

$$G_i(t) = \widetilde{\Phi}^{\theta}_{t\to 0}(G^0_i + tZ_i). \tag{42}$$

Each time is evaluated independently from the same initialization law; there is no preceding step at a larger noise level. The current sweep uses $N = 128$ particles and 3 repeats.

The empirical output-to-target distance is computed with the same exact linear-assignment estimator (37). The same estimator also provides $\widehat{D}_0$ between a Gaussian reference cloud and a target cloud. We then evaluate the plug-in version of (11),

$$\widehat{B}_1(t) = \widehat{q}_t \widehat{D}_0 + \widehat{E}_t^{\pi}. \tag{43}$$

The target-error profile $\widehat{E}_t^{\pi}$ is calibrated separately using the same empirical-OT construction in (39). Consequently, Fig. 6 compares the empirical finite-cloud OT of the one-step samples with a recursive estimate whose two distance terms, $\widehat{D}_0$ and $\widehat{E}_t^{\pi}$, are computed with the same OT estimator, while $\widehat{q}_t$ is calibrated independently from the secant ratios.

E The Gaussian example: derivations and further regimes

This appendix derives the statements of Section 3.4, in its notation, and treats further regimes.

E.1 Model

Let $\pi = \mathcal{N}(0, m^{-1}\,\mathrm{Id}_d)$. Then $\pi_r = \mathcal{N}(0, (m^{-1} + r^2)\,\mathrm{Id})$ and $\nabla^2 f_r = \mu_r\,\mathrm{Id}$ with $\mu_r = m/(1 + mr^2)$, so Lemma A.3 is attained, and the probability-flow ODE (1) integrates to $\Phi_{t\to 0}(x) = L_t x$ with

$$R_t = \alpha_t, \qquad L_t = \frac{1}{\alpha_t\sqrt{1 + m\gamma_t^2}}, \qquad q_t = \frac{1}{\sqrt{1 + m\gamma_t^2}}, \tag{44}$$

the profiles of Remark 3.5. A1–A3 hold with these constants, which are exact operator norms (R_t is attained on Dirac masses).

The learned map may err in its slope, which encodes the variance it ascribes to the target, and by an offset along a unit vector v, which displaces the mean:

$$\widetilde{\Phi}^{\theta}_{t\to 0}(x) = \widetilde{L}_t\, x + \omega_t\, v, \qquad \widetilde{L}_t = \frac{1}{\alpha_t\sqrt{1 + \widetilde{m}\gamma_t^2}}, \tag{45}$$

with a learned precision $\widetilde{m} > 0$ and offsets $\omega_t \geq 0$, which play the role of θ. Without offset, $\omega \equiv 0$, this is the exact map of $\mathcal{N}(0, \widetilde{m}^{-1}\,\mathrm{Id})$, a model of the right mean and the wrong variance.

Two affine objects appear: the learned map (45), which the example specifies, and the score (46) it implies; in each, the slope encodes the variance and the offset the mean. To connect the map with Theorem 3.3, take offsets of the form $\omega_t = \int_0^{\gamma_t} r\,\eta_r(1 + \widetilde{m}r^2)^{-1/2}\,\mathrm{d}r$ with $\eta_r \geq 0$. In the rescaled coordinates of Lemma A.1 the approximate-score ODE is then linear, with integrating factor $(1 + \widetilde{m}\gamma_u^2)^{-1/2}$, and (45) is its flow for the implied affine score

$$\widetilde{s}_\theta(x, r) = -\widetilde{\mu}_r\, x + \eta_r\, v, \qquad \widetilde{\mu}_r = \frac{\widetilde{m}}{1 + \widetilde{m}r^2}. \tag{46}$$

Each $\widetilde{s}_\theta(\cdot, r)$ is the score of a Gaussian, of variance $\widetilde{m}^{-1} + r^2$ and mean $\eta_r(\widetilde{m}^{-1} + r^2)v$; with an offset, the mean depends on r, which Theorem 3.3 allows since it treats the levels separately. The score error (5) is $e_r = \eta_r$ if $\widetilde{m} = m$, and $e_r = +\infty$ otherwise, the slope mismatch $(\mu_r - \widetilde{\mu}_r)x$ being unbounded: the offset is the score error only when the slope is right. For $\widetilde{m} = m$, ω_t is therefore the error profile E_t of (6). The figures use $\eta_r = M_0 + M_1/r$, for which $\omega_t = B_{\widetilde{m}}(\gamma_t)$ with B_m as in Remark 3.5; for $\widetilde{m} = m$ this offset is the score-error envelope of Corollary 3.4, attained with equality. The learned step is then the exact one followed by a shift, which moves every law by the same vector; hence E_t, $E_t^{\pi} = E_t$ and $\widetilde{L}_t = L_t$ are the best constants for both Proposition 3.1 and Proposition 3.9, whose certificates coincide.

The mean/variance recursion of Section 3.4 follows, since noising maps $\mathcal{N}(\zeta v, s^2\,\mathrm{Id})$ to $\mathcal{N}(\alpha_t\zeta v, (\alpha_t^2 s^2 + \sigma_t^2)\,\mathrm{Id})$, the learned map is affine, and two Gaussians with proportional covariances are compared coordinatewise. For $\widetilde{m} = m$, $s = m^{-1/2}$ is preserved.

E.2 How the two errors propagate

By the mean/variance recursion of Section 3.4, an error in the mean contracts at exactly the certified $\widetilde{q}_t$. An error in the variance contracts more strongly: for $\nu_s = \mathcal{N}(0, s^2\,\mathrm{Id}_d)$, rationalizing the difference

of standard deviations after noising gives

$$\frac{W_2\big(N_t(\nu_s), N_t(\nu_{m^{-1/2}})\big)}{W_2(\nu_s, \nu_{m^{-1/2}})} = \frac{\alpha_t}{\kappa_t(s)}, \qquad \kappa_t(s) = \frac{(\alpha_t^2 s^2 + \sigma_t^2)^{1/2} + (\alpha_t^2/m + \sigma_t^2)^{1/2}}{\alpha_t\,(s + m^{-1/2})}. \tag{47}$$

Since $(\alpha_t^2 x^2 + \sigma_t^2)^{1/2} \geq \alpha_t x$ and its derivative lies in $[0, \alpha_t]$, $\kappa_t \geq 1$ is nonincreasing; it equals q_t^{-1} at $s = m^{-1/2}$, tends to $q_t^{-1} + \sqrt{m}\,\gamma_t$ as $s \to 0$ and to 1 as $s \to \infty$. The constant $R_t = \alpha_t$ is exact on Dirac masses, which differ by a shift, but misses the factor κ_t on the laws the sampler visits.

Let $S_t(x) = \widetilde{L}_t(\alpha_t^2 x^2 + \sigma_t^2)^{1/2}$ be the standard deviation after one learned step from standard deviation x. After a step from $\rho = \mathcal{N}(\zeta v, s^2\,\mathrm{Id})$, $\zeta \geq 0$, the error in the standard deviation is the sum of a part inherited from ρ, $\Delta_t^{\mathrm{init}} = S_t(s) - S_t(m^{-1/2})$, and a part created by the step, $\Delta_t^{\mathrm{step}} = S_t(m^{-1/2}) - m^{-1/2}$; each is positive if it makes the law too wide and negative if it makes it too narrow. Since $\widetilde{\mathcal{L}}_t^\theta(\pi) = \mathcal{N}\big(\omega_t v, S_t(m^{-1/2})^2\,\mathrm{Id}\,\big)$, the target error of A5 is attained at

$$E_t^\pi = \|\big(\omega_t,\ \sqrt{d}\,\Delta_t^{\mathrm{step}}\big)\| = \Big(\omega_t^2 + \frac{d}{m}\Big[\Big(\frac{1 + m\gamma_t^2}{1 + \widetilde{m}\gamma_t^2}\Big)^{1/2} - 1\Big]^2\Big)^{1/2}. \tag{48}$$

Its two entries are the errors in the mean and in the standard deviation of one learned step from π; as $\gamma_t \to \infty$ the second tends to that of the learned target $\mathcal{N}(0, \widetilde{m}^{-1}\,\mathrm{Id})$. Since $|\Delta_t^{\mathrm{init}}| = \widetilde{q}_t|s - m^{-1/2}|/\kappa_t(s)$ by (47), the exact error and the certificate after one step are

$$\begin{aligned} W_2\big(\widetilde{\mathcal{L}}_t^\theta(\rho), \pi\big) &= \|\big(\widetilde{q}_t\zeta + \omega_t,\ \sqrt{d}\,(\Delta_t^{\mathrm{init}} + \Delta_t^{\mathrm{step}})\big)\|, \\ \widetilde{q}_t\, W_2(\rho, \pi) + E_t^\pi &= \|\big(\widetilde{q}_t\zeta,\ \sqrt{d}\,\kappa_t(s)|\Delta_t^{\mathrm{init}}|\big)\| + \|\big(\omega_t,\ \sqrt{d}\,|\Delta_t^{\mathrm{step}}|\big)\|. \end{aligned} \tag{49}$$

The mean entries agree, so an error in the mean is certified exactly, while the inherited variance error is overstated by κ_t, a bounded factor per step that compounds over steps. Adding two norms instead of combining their entries costs at most a further factor $\sqrt{2}$, and adding the sizes of Δ_t^{init} and Δ_t^{step} instead of their signed values ignores cancellation (Section E.3).

At a repeated level the iterates converge to $\rho_\star = \mathcal{N}(\zeta_\star v, s_\star^2\,\mathrm{Id})$, with $\zeta_\star = \omega_t/(1 - \widetilde{q}_t)$ and $s_\star = \widetilde{L}_t\sigma_t(1 - \widetilde{q}_t^2)^{-1/2}$. Writing $\Delta_t = \sqrt{d}\,|\Delta_t^{\mathrm{step}}|$ and $\delta_t = \sqrt{d}\,|s_\star - m^{-1/2}|$, the plateau of Remark B.3 exceeds the exact limit by

$$\frac{E_t^\pi/(1 - \widetilde{q}_t)}{W_2(\rho_\star, \pi)} = \sqrt{\frac{\omega_t^2 + \Delta_t^2}{\omega_t^2 + (1 - \widetilde{q}_t)^2\delta_t^2}}, \tag{50}$$

which is at least one, by (11) at $\rho_\star$, and decreases in ω_t.

E.3 Regimes

Table 1 lists what the certificate overstates per regime; the extremes isolate a single error source.

Regime	Initialization	Learned step	Certificate overstates	Outcome
Mean	mean	mean	nothing	exact
Initial variance	variance	none	inherited variance, by κ_t	ratio $\sim q_t^{-n}$
Learned variance	variance or none	variance	inherited variance, by κ_t	no uniform E_t
Both	both kinds	both kinds	the same, diluted by the mean	loose, then tight

Table 1: Regimes of Section E.3. The second and third columns give the kind of error introduced by the initialization and by the learned step, *none* meaning exact. The certificate also adds the sizes of variance errors of opposite directions, which may cancel (Section E.2).

Error in the mean. With $\widetilde{m} = m$ and $s_0 = m^{-1/2}$ the two lines of (49) agree, and so do the unrolled (11) and (3). Figure 4 shows the statements of Section 3.4 in both parametrizations; the plateau of Remark B.3 is attained. The geometric form (4) also needs the balance condition (23), which for $t_k = t_1\lambda^{k-1}$ holds only for slowly decaying schedules, a range that narrows as n grows.

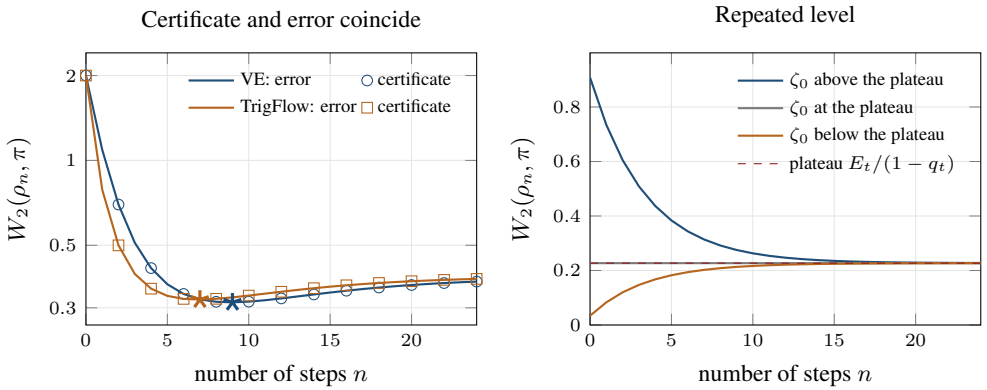


Figure 4: Error in the mean only: $m = 0.8$, the curvature fitted in Section 4, $\widetilde{m} = m$, $\eta_r = M_0 + M_1/r$ with $M_0 = 0.03$, $M_1 = 0.05$, and $s_0 = m^{-1/2}$. Left: from $\zeta_0 = 2$, along $t_k = 2(0.85)^{k-1}$ in VE coordinates and $t_k = 1.3(0.85)^{k-1}$ in TrigFlow coordinates, the certificate (3) (markers) and the exact error (lines) coincide; stars mark the optimal step count. Right: at the repeated VE level $t = 1$, the error moves to the plateau $E_t/(1 - q_t)$ from above, at, or below it.

Variance error from the initialization. With an exact CM ($\widetilde{m} = m$, $\eta \equiv 0$, hence $E_t = E_t^\pi = 0$) started from $\mathcal{N}(0, s_0^2 \,\mathrm{Id})$, only the factor κ_t acts: the certificate is the pure contraction $Q_{1,n} W_2(\rho_0, \pi)$, while the error contracts at $q_{t_k}/\kappa_{t_k}(s_{k-1})$.

Variance error from the learned step. With $\widetilde{m} \neq m$ and $\eta \equiv 0$, the model has learned $\mathcal{N}(0, \widetilde{m}^{-1} \,\mathrm{Id})$: the right mean and the wrong variance. This is the extreme of a learned step that errs only in the variance, and it holds exactly when π is symmetric and the model respects the symmetry, since an odd affine score has no offset. The score is nonetheless wrong, $e_r = +\infty$, and no finite E_t satisfies A4, since $W_2\big(\mathcal{L}_t(\nu_s), \widetilde{\mathcal{L}}_t^\theta(\nu_s)\big) = \sqrt{d}\,|L_t - \widetilde{L}_t|(\alpha_t^2 s^2 + \sigma_t^2)^{1/2} \to \infty$ as $s \to \infty$. A5 and A6 still hold, with $\widetilde{L}_t$ and $E_t^\pi = \Delta_t$, so Proposition 3.9 applies where Proposition 3.1 does not: a CM that learned a slightly wrong variance, a first-order description of the models of Section 4, admits no uniform error bound, however accurate at the target. Its certificate overstates the inherited variance error by κ_t and keeps a gap at the plateau, where (50) equals $\Delta_t/[(1 - \widetilde{q}_t)\delta_t]$ (Fig. 1).

Both errors. The mean error contracts at $\widetilde{q}_{t_k}$ and the inherited variance error at $\widetilde{q}_{t_k}/\kappa_{t_k}$, so the variance component dominates early and the mean component late: the certificate is loose, then tight, and the crossover comes earlier as ω_t grows and later as the initial variance error grows (Fig. 1). At the plateau the mean error dilutes the gap, (50) decreasing from its value at $\omega_t = 0$ to one. A tight certificate thus indicates that the dominant error is in the mean, not that the denoiser is accurate.

The same balance governs the choice of the next level. Figure 5 (left) takes one step at a large fixed level t_1 and then sweeps the level t_2 of a second step, for an initialization and a learned step that are both too wide, so that no cancellation occurs. Without an error in the mean, a larger t_2 contracts the remaining variance error more strongly, and the exact error and the certificate both favour the largest level. An error in the mean makes each step add a bias ω_{t_2} that grows with t_2, so the exact error has an interior optimum. The certificate, which overstates the inherited variance error by κ_{t_2}, prefers more noise than the optimum, and for a large mean error it still takes a step when none helps. Its choice nonetheless costs only a few percent of the error.

Cancellation. When the inherited and created variance errors have opposite directions, the exact error in (49) adds them with their signs while the certificate adds their sizes, and the certificate can exceed the error by an arbitrary factor: if $\widetilde{m} > m$ and $\eta \equiv 0$, the step makes laws too narrow, and the too wide law with $S_t(s) = m^{-1/2}$ is mapped exactly onto π while its certificate is positive. This needs a variance error entering the step in the direction opposite to the one the step creates. A model whose variance error keeps one direction across levels produces such an error only through the initialization, which a large first level largely erases. The right panel of Fig. 5 repeats the comparison with a learned variance that is too small, against the same too wide initialization. After the first

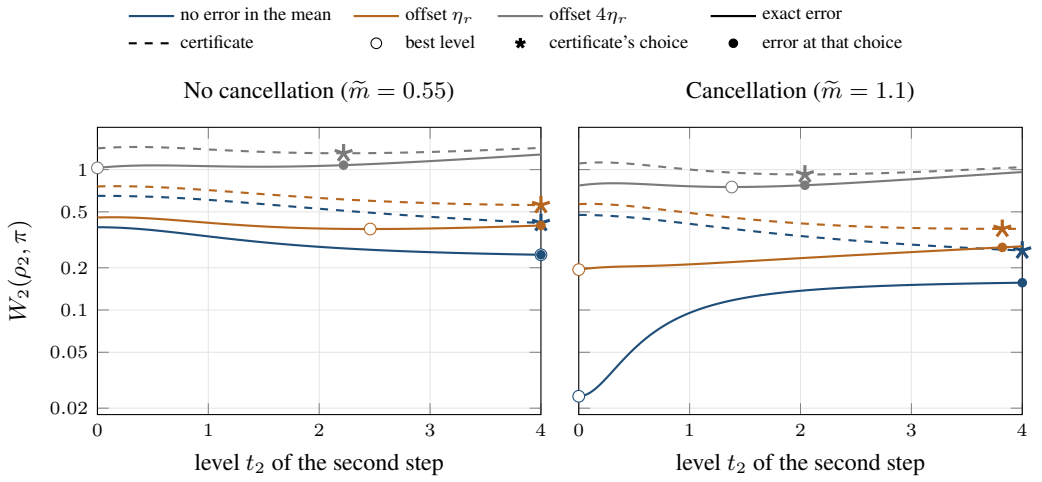


Figure 5: Choosing the level of a second step, in VE coordinates with $m = 0.8$ and $d = 1$: a first step at $t_1 = 4$ from $\mathcal{N}(0, 2.5^2\,\mathrm{Id})$, then a second at level t_2, for no error in the mean and for the offsets η_r and $4\eta_r$, with η_r as in Fig. 4, on a logarithmic scale. Exact error solid, certificate dashed; circles mark the best level ($t_2 = 0$: no second step), stars the certificate's choice and dots the exact error at that choice. Left: the learned variance is too large ($\widetilde{m} = 0.55$), like the initialization, so no cancellation occurs. Right: it is too small ($\widetilde{m} = 1.1$), so the learned step makes the law too narrow while the initialization is too wide, and the two variance errors cancel.

step the two variance errors nearly cancel, so the law is closer to π than its certificate says, and the certificate recommends a second step that makes the error worse. The error in the mean, certified exactly, limits the damage: with no error in the mean the law after one step is almost exactly π and the recommended step multiplies the error several times; with the offset η_r it still increases the error; with $4\eta_r$ a second step helps, and following the certificate costs little.

Scope. For $\pi = \mathcal{N}(0, \Sigma)$ the profiles hold with $m = \lambda_{\max}(\Sigma)^{-1}$ and the recursions extend coordinatewise; the certificate is then driven by the flattest direction while the error averages over all of them, a further, dimension-dependent slack. The model does not show that these mechanisms are at work in Section 4, which would require measuring the error across step counts for a trained model.

F Validating the one-step bound on SD 1.5

We combine the estimated profiles $\widehat{q}_t$ and $\widehat{E}_t^{\pi}$ (Appendix D) of the SD 1.5 LCM into the one-step W_2 bound that underpins our recursion, and compare this bound with the empirical W_2 error measured by an optimal transport estimator. More precisely, starting from $\rho_0 = \mathcal{N}(0, I)$, we apply a single noising–denoising step at each VE level t separately, and compute the empirical W_2 distance between the resulting latent samples and latent samples from a target dataset (different from the one used to calibrate $\widehat{q}_t$ and $\widehat{E}_t^{\pi}$), using the exact optimal assignment between the two point clouds. We compare this distance with the plug-in one-step bound

$$\widehat{B}_1(t) := \widehat{q}_t\, \widehat{D}_0 + \widehat{E}_t^{\pi}, \qquad \widehat{D}_0 \approx W_2(\rho_0, \pi). \tag{51}$$

As shown in Fig. 6, $\widehat{B}_1(t)$ lies above the observed W_2 error over the full range of VE times and follows its variation with t. Since $\widehat{B}_1(t)$ is assembled from independently estimated profiles rather than fitted to the measured error, its proximity to that error indicates that the bound is reasonably tight. Interestingly, the gap is smallest near the optimal value of t, which is the regime that matters for schedule design.

Unlike in the Gaussian case, we do not report multistep bounds for the considered CMs. Already for two steps, the bound drops to the level of the statistical bias of the empirical W_2 estimator, so it

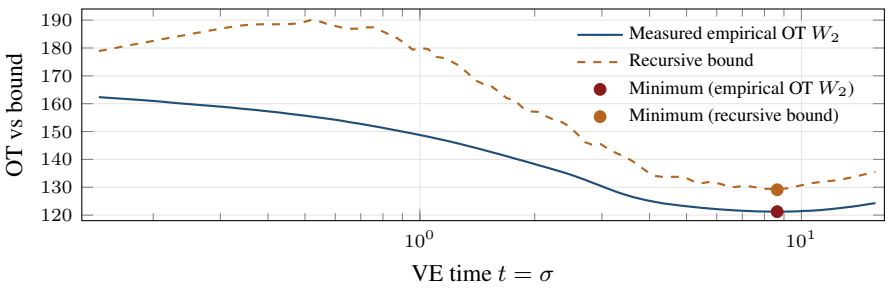


Figure 6: One-step SD 1.5 experiment. The measured empirical W2 error is compared with the plug-in recursive estimate $\widehat{q}_t \widehat{D}_0 + \widehat{E}_t^{\pi}$.

cannot be meaningfully compared with the measured error. Resolving this would require vastly more samples, since empirical W_2 estimates converge at rate that is notoriously slow in high dimensions.