NON-ASYMPTOTIC ERROR BOUNDS FOR PROBABILITY FLOW ODES UNDER WEAK LOG-CONCAVITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Score-based generative modeling, implemented through probability flow ODEs, has shown impressive results in numerous practical settings. However, most convergence guarantees rely on restrictive regularity assumptions on the target distribution—such as strong log-concavity or bounded support—or focus on the specific case of constant drift and diffusion coefficients. This work establishes non-asymptotic convergence bounds in 2-Wasserstein distance for probability flow ODEs when the target is only weakly log-concave, allowing multimodal distributions such as Gaussian mixtures and time-dependent noise schedules. Our bounds account for initialization, discretization, and score approximation errors. We show that, for fixed regularity parameters of the target distribution, weakening strong to weak log-concavity leaves the asymptotic dependence on the time horizon, step size, and score matching error unchanged. Yet, the relaxation incurs a potentially large non-asymptotic price: the bounds contain an exponential dependence on the deviation from log-concavity. We complement the upper bound with a lower bound construction, showing that this exponential dependence is unavoidable under weak log-concavity alone.

1 INTRODUCTION

Diffusion models are a powerful class of generative models designed to sample from complex data distributions. They operate by reversing a forward stochastic process that progressively transforms data into noise. The generative process is typically modeled using a reverse-time stochastic differential equation (SDE) or an equivalent deterministic probability flow ordinary differential equation (ODE) that preserves the same marginal distributions (Song & Ermon, 2019; Song et al., 2021; Ho et al., 2020). The key idea is to use a learned score function—an estimate of the gradient (with respect to the data) of the log-density—to guide the reverse dynamics. Samples are then generated by integrating this reverse process from pure noise back to the data manifold.

The key issue in diffusion models is under what assumptions and in which settings these reverse processes provably approximate the target distribution. While a growing body of literature addresses this issue, often distinguishing between stochastic and deterministic samplers, most analyses impose restrictive assumptions on the unknown target distribution—such as log-concavity or bounded support (Block et al., 2020; De Bortoli, 2022; Lee et al., 2023; Gao et al., 2025)—or focus on constant-coefficient forward process rather than the time-varying noise schedules used in practice (Gentiloni-Silveri & Ocello, 2025; Bruno & Sabanis, 2025). A natural and intriguing question is: to what extent can these assumptions be relaxed and what is the resulting cost in the convergence guarantees?

In this paper, we provide an answer to this question for probability flow ODEs. We replace the standard strong log-concavity assumption by weak log-concavity, thereby covering a substantially broader class of target distributions, including multimodal Gaussian mixtures, while explicitly covering time-dependent drift and diffusion coefficients. Crucially, we characterize the price of this relaxation: weak log-concavity preserves the asymptotic dependence of the sampling error on the time horizon, discretization step size, and score matching error, but introduces an additional dependence on the degree of non-log-concavity, which we show must, in general, be exponential.

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.

Contributions We study the distance between the approximated and the true sample distribution for a general class of probability flow ODEs, while relaxing the standard strong log-concavity assumption. Our main contributions are:

1. We establish non-asymptotic 2-Wasserstein convergence guarantees for probability flow ODEs under weak log-concavity, accounting simultaneously for initialization, discretization, and score matching errors and explicitely covering time-dependent variance-preserving and -exploding noise schedules (Theorem 5).
2. We show that, for fixed weak log-concavity parameters, relaxing the strong log-concavity assumption leaves the asymptotic dependence on time horizon, discretization step size, and score matching error unchanged (Proposition 6).
3. While the asymptotic rates remain unchanged, the weaker assumption leads to an exponential dependence on the strength of deviation from log-concavity in the non-asymptotic error bounds. We show that this dependence is intrinsic and unavoidable under the weak log-concavity assumption alone (Proposition 7).

A key ingredient in our analysis is a regime shifting property of the forward process, whereby weak log-concavity evolves into strong log-concavity after a finite time. We establish this phenomenon for a broad class of time-dependent drift and diffusion functions and explicitly quantify the transition time (Proposition 3).

1.1 Related Work

Existing studies of the convergence of trained score-based generative models (SGMs) invoke a variety of different distances. Total Variation (TV) distance and Kullback–Leibler (KL) divergence are the most commonly used in theoretical analyses. For instance, theoretical guarantees for diffusion models in terms of TV or KL have been studied in Lee et al. (2022); Wibisono & Yang (2022); Chen et al. (2023c;a;b;d); Gentiloni Silveri et al. (2024); Li et al. (2024); Conforti et al. (2025a). However, these metrics often fail to capture perceptual similarity in applications such as image generation. In contrast, the 2-Wasserstein distance is often the preferred evaluation metric, as it better reflects the underlying geometry of the data distribution. One of the most popular performance metrics for the quality of generated samples in image applications, the Fréchet inception distance (FID), measures the Wasserstein distance between the distributions of generated images and the distribution of real images (Heusel et al., 2017). Importantly, convergence in TV or KL does not generally imply convergence in Wasserstein distance unless strong conditions are satisfied (Gibbs & Su, 2002).

A smaller number of works go further to analyze convergence in Wasserstein distances, though these typically require additional assumptions like compact support or uniform moment bounds, see e.g. Block et al. (2020); De Bortoli (2022); Lee et al. (2023); Gao et al. (2025) for SDE-based samplers. For example, Beyler & Bach (2025) establish 2-Wasserstein convergence guarantees for diffusion-based generative models, treating both stochastic and deterministic sampling via early-stopping analysis. Their results are limited to bounded perturbations of Gaussian targets and diffusion models with constant drift and diffusion coefficients. On the other hand, the error bounds for reverse SDEs and reverse ODEs established in Gao & Zhu (2025) and Gao et al. (2025), respectively, incorporate time-dependent noise schedules. However, they assume the data distribution to be strongly log-concave, which does not hold for many distributions of practical interest, including Gaussian mixture models with well-separated modes.

Recently, there has been growing interest in relaxing the common assumption of strong log-concavity in the analysis of SGMs. For example, Gentiloni-Silveri & Ocello (2025) derived 2-Wasserstein convergence guarantees for SDE-based samplers assuming the target distribution to be Lipschitz smooth and weakly log-concave, a milder assumption than strong log-concavity, allowing for well-behaved multi-modal distributions. Bruno & Sabanis (2025) established error bounds in 2-Wasserstein distance under a similar semiconvexity assumption, allowing for data distributions with potentially discontinuous gradients, whereas Brigati & Pedrotti (2024) proposed a different weakening of the log-concavity assumption, in the form of a Lipschitz perturbation of a log-concave distribution. A key feature of these analyses is the emergence of a *regime shifting* behavior (also referred to as *creation of log-concavity* or *eventual log-concavity*), whereby the smoothing effect of the flow renders the distribution log-concave after some time. However, all of these works con-

sider only constant drift and diffusion functions. A central part of our theoretical analysis establishes quantative control over this phenomenon for more general noise schedules, going beyond the constant-coefficient case treated in prior work.

Despite the growing body of literature, most existing convergence results—whether for stochastic or deterministic samplers— typically suffer from at least one of the following limitations: they use less suitable distance measures (in particular TV or KL), are restricted to constant drift and diffusion functions, or rely on strong structural assumptions, such as log-concavity or bounded support of the data distribution. It therefore remains unclear how the convergence rates of deterministic samplers change once these restrictions are lifted and multimodal target distributions as well as general noise schedules are allowed.

Paper outline Section 2 introduces SGMs, highlighting the approximations that are necessary to enable sampling from the probability flow ODE. In Section 3, we investigate the weak log-concavity assumption and establish a regime shifting property, which is crucial for the proof of our error bound. Section 4 presents our main result, a non-asymptotic error bound for the 2-Wasserstein distance of the true and approximated sample distribution under weak log-concavity. We provide a result for a specific time-dependent noise schedule which is commonly used in practice, yielding a directly interpretable bound, and a general result that applies to any choice of the drift and diffusion function. Moreover, we compare our results to those under the stricter assumption of strong log-concavity and quantify the price to pay for our weaker assumption. Finally, in Section 5, we summarize our results and provide an outlook into related future research directions. Additional technical results and detailed proofs are provided in the Appendix.

Notation For $a, b \in \mathbb{R}$, we write $a \wedge b$ as a shorthand for $\min\{a, b\}$ and $a \vee b$ for $\max\{a, b\}$. Given a random variable $X \in \mathbb{R}^d$, we denote its law by $\mathcal{L}(X)$ and its L_2-norm as $\|X\|_{L_2} := \sqrt{\mathbb{E}(\|X\|^2)}$, where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. For any two probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the space of measures on $\mathbb{R}^d$ with finite second moment, the 2-Wasserstein distance is defined as

$$\mathcal{W}_2(\mu, \nu) := \left(\inf_{X \sim \mu, Y \sim \nu} \mathbb{E}\|X - Y\|^2 \right)^{\frac{1}{2}},$$

where the infimum is taken over all possible couplings of μ and ν.

2 Preliminaries on score-based generative models

This section introduces SGMs and their ODE-based implementation of the sampling process (the probability flow ODE), which provides the framework for our analysis. Denote with $p_0 \in \mathcal{P}(\mathbb{R}^d)$ an unknown probability distribution on $\mathbb{R}^d$. Our goal is to generate new samples from p_0 given a data set of independent and identically distributed observations. SGMs use a two-stage procedure to achieve this. First, noisy samples are progressively generated by means of a diffusion-type stochastic process. Then, in order to reverse this process, a model is trained to approximate the score, enabling the generation of new samples.

Noisy samples are generated from the forward process $\{X_t\}_{t\in[0,T]}$, solution to the SDE

$$\mathrm{d}X_t = -f(t)X_t \,\mathrm{d}t \; + \; g(t)\,\mathrm{d}B_t, \qquad X_0 \sim p_0, \tag{1}$$

where $f, g : [0, T] \to \mathbb{R}_{\geq 0}$ are continuous and non-negative, $g(t)$ is positive and non-decreasing for all $t > 0$, and B_t is a standard d-dimensional Brownian motion. Through this process, the unknown data distribution p_0 progressively evolves over time into the family $\{p_t\}_{t\geq 0}$, where p_t denotes the marginal law of the process X_t. The solution to (1) is given by (see e.g. Karatzas & Shreve, 2012, Chapter 5.6)

$$X_t = e^{-\int_0^t f(s)\,\mathrm{d}s} X_0 + \int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s. \tag{2}$$

Note that the stochastic integral in (2) has a Gaussian distribution independent of p_0, namely

$$\int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s \sim \mathcal{N}\left(0, \int_0^t e^{-2\int_s^t f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s \cdot I_d\right) =: \hat{p}_t.$$

Common instances used in score-based generative modeling are *variance-exploding* (VE) and *variance-preserving* (VP) SDEs (Song et al., 2021). In a VE-SDE, we have

$$f(t) \equiv 0 \quad \text{and} \quad g(t) = \sqrt{\frac{\mathrm{d}[\sigma^2(t)]}{\mathrm{d}t}}, \tag{3}$$

whereas in a VP-SDE, it holds that

$$f(t) = \frac{1}{2}\beta(t) \quad \text{and} \quad g(t) = \sqrt{\beta(t)} \tag{4}$$

for some non-negative non-decreasing functions $\sigma(t)$ and $\beta(t)$. The name "variance-preserving" can be justified by noting that noise is added in the forward process in a way that exactly offsets the drift's tendency to contract the variance. Namely, the integral of $f(t)$ diverges while the variance of $\hat{p}_t$ tends to 1 as $T \to \infty$. Therefore, in the VP-case, X_t has stationary distribution $\mathcal{N}(0, I_d)$.

Next, score matching is performed, i.e. the unknown true score function $\nabla_x \log p_t$ is estimated by training a model in some family $\{s_\theta(t,x)\}_{\theta\in\Theta}$, typically a deep neural network. This is achieved by minimizing a denoising score matching objective of the form (Song & Ermon, 2019)

$$L(\theta) = \int_0^T \mathbb{E}\Big[\big\| s_\theta(X_t, t) \; - \; \nabla_x \log p_t(X_t)\big\|_2^2\Big] \, \mathrm{d}t \, . \tag{5}$$

Practical implementations of (5) typically introduce a time-dependent weighting function and rewrite the objective in terms of conditional expectations to make the optimization viable. These modifications do not affect our analysis; the only requirement is that a sufficiently accurate model is available (see Assumption 3).

The key idea behind SGMs is that the dynamics of the reverse process are explicitly characterized, allowing for new sample generation. In this work, we focus on the ODE formulation of this time reversal. According to Song et al. (2021), the time-reversed process $\{\tilde{X}_t\}_{t\in[0,T]}$ satisfying $\mathcal{L}(\tilde{X}_t) = \mathcal{L}(X_{T-t})$ is defined as the solution to the so-called *probability flow ODE*

$$\frac{\mathrm{d}\tilde{X}_t}{\mathrm{d}t} = f(T-t)\tilde{X}_t + \frac{1}{2}\, g^2(T-t)\, \nabla \log p_{T-t}(\tilde{X}_t), \qquad \tilde{X}_0 \sim p_T. \tag{6}$$

Three approximations are needed in order to use this ODE to create new samples in practice. First, note that the distribution p_T of the final state X_T is unknown. We therefore approximate it with a tractable law from which samples can be generated efficiently. Following Gao & Zhu (2025), we replace p_T with $\hat{p}_T$ and consider the probability flow

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = f(T-t)Y_t + \frac{1}{2}\, g^2(T-t)\, \nabla \log p_{T-t}(Y_t), \qquad Y_0 \sim \hat{p}_T. \tag{7}$$

The only difference between $\tilde{X}_t$ and Y_t lies in their initial distribution. In the VP case, one might also start the reverse process from the invariant distribution $\mathcal{N}(0, I_d)$.

Second, we employ a numerical discretization method to approximate the solution of ODE (7), as it is not generally available in closed form. Similarly to Gao & Zhu (2025), we consider an exponential integrator discretization for this purpose. This method has been shown to be faster than other options such as Euler method or RK45, as it is more stable with respect to taking larger step sizes (Zhang & Chen, 2023). Specifically, the interval $[0, T]$ is split into discrete time steps $t_k = kh$ for $k \in \{0, 1, \ldots, K\}$ and step size $h > 0$. Without loss of generality, we assume that $T = Kh$ for some positive integer K. On each interval $t_{k-1} \le t \le t_k$, ODE (7) is then approximated by

$$\frac{\mathrm{d}\widehat{Y}_t}{\mathrm{d}t} = f(T-t)\widehat{Y}_t + \frac{1}{2} g^2(T-t) \nabla \log p_{T-t_{k-1}}\Big(\widehat{Y}_{t_{k-1}}\Big). \tag{8}$$

Since the non-linear term is not dependent on t anymore, this ODE can be explicitly solved on each interval. As in (7), the initial distribution is given by $\widehat{Y}_0 \sim \hat{p}_T$.

Finally, since the score function $\nabla \log p_t$ is unknown in practice, we approximate it by the score model $s_\theta(x,t)$. This leads to an approximation of (8) given by

$$\frac{\mathrm{d}\widehat{Z}_t}{\mathrm{d}t} = f(T-t)\widehat{Z}_t + \frac{1}{2} g^2(T-t) s_\theta\Big(\widehat{Z}_{t_{k-1}}, T - t_{k-1}\Big) \tag{9}$$

	$E_0(f, g, T)$	$E_1(f, g, K, h)$	$E_2(f, g, K, h, \mathcal{E})$
Error source	Initialization	Discretization	Score matching
Vanishes with	$T \to \infty$	$h \to 0$	$\mathcal{E} \to 0$
VP-SDE*	$\mathcal{O}\Big(e^{-T^{\rho+1}}\sqrt{d}\Big)$	$\mathcal{O}\Big(e^{T^{2\rho+1}h}T^{2\rho+1}h\Big(\sqrt{d}+T\Big)\Big)$	$\mathcal{O}\Big(e^{T^{2\rho+1}h}T^{\rho+1}\mathcal{E}\Big)$
Error $\leq \varepsilon$ if*	$T \geq \mathcal{O}\Bigg(\Big(\log\Big(\frac{\sqrt{d}}{\varepsilon}\Big)\Big)^{\frac{1}{\rho+1}}\Bigg)$	$h \leq \mathcal{O}\Big(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\Big)$	$\mathcal{E} \leq \mathcal{O}\Big(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\Big)$

Table 1: In our main result (Theorem 5), we show that the error $\mathcal{W}_2(\mathcal{L}(\widehat{Z}_{t_K}), p_0)$ can be bounded by the sum of three error components E_0, E_1, and E_2. The table provides a summary of the main properties of these terms and their specific heuristics in the case of a variance-preserving SDE with noise schedule $\beta(t) = (b + at)^\rho$ and $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$, indicated by an asterisk* (see Theorem 4).

with $\widehat{Z}_0 \sim \hat{p}_T$.

This means that, effectively—after replacing the initial distribution, discretizing, and learning the score—one is able to sample from the law $\mathcal{L}(\widehat{Z}_{t_K})$, which serves as a viable approximation of the unknown data distribution p_0. Our objective is then to quantify the accuracy of the method by providing bounds on the 2-Wasserstein distance $\mathcal{W}_2(\mathcal{L}(\widehat{Z}_{t_K}), p_0)$ between the generated samples and the target distribution. A first overview of the components of our error bound is given in Table 1.

3 Weak concavity

Our main result establishes an error bound for the probability flow ODE, relying on a weaker assumption than strong log-concavity of the target density p_0. In particular, we use the notion of weak concavity which is defined as follows.

Definition 1 (Weak convexity). *The weak convexity profile of a function $g \in C^1(\mathbb{R}^d)$ is defined as*

$$\kappa_g(r) = \inf_{x,y\in\mathbb{R}:\|x-y\|=r} \left\{ \frac{\langle \nabla g(x) - \nabla g(y), x - y\rangle}{\|x-y\|^2} \right\}, \quad r > 0.$$

We say that g is (α, M)-weakly convex if

$$\kappa_g(r) \geq \alpha - r^{-1} f_M(r) \quad \textit{for all } r > 0$$

for some constants $\alpha, M > 0$ and $f_M(r) = 2\sqrt{M}\tanh\Big(\sqrt{M}r/2\Big)$. Moreover, we say that g is (α, M)-weakly concave if $-g$ is (α, M)-weakly convex.

The weak convexity assumption means that the function is approximately convex at "large scales" (large r), while allowing small non-convex fluctuations at short distances (small r). More specifically, a function is weakly convex if and only if it is semiconvex on a ball and strongly convex outside the ball (Bruno & Sabanis, 2025, Proposition 13). Importantly, (α, M)-weak concavity implies $(\alpha - M)$-strong concavity if $\alpha - M > 0$, as laid out in Lemma 8, meaning that it is in fact a more general assumption. A relevant example for a family of distributions that are weakly but not strongly log-concave are Gaussian mixture models with well-separated modes (Gentiloni-Silveri & Ocello, 2025, Proposition 4.1). A specific example of such a mixture model including graphs of the log-density and score function is given in Example 1 in Appendix B. Note that, due to their strong log-concavity at large scales, weakly log-concave distributions necessarily need to have sub-gaussian tails. On the other hand, not every distribution with sub-gaussian tails is weakly log-concave, as the constructed Example 2 in Appendix B with a score exhibiting an excessively steep increase at one point shows. As noted in Remark 13 in Appendix B, other forms of f_M are possible, keeping all results unaffected.

In the following, we investigate the log-concavity of p_t given that p_0 is weakly log-concave. In other words, we establish results on how the weak log-concavity propagate through time following

the forward SDE (1) with general noise schedules f and g. A similar result for the propagation in time of Lipschitz smoothness is given in Proposition 16 in Appendix B. Our main result heavily relies on these findings.

Proposition 2 (Propagation in time of weak log-concavity). *If p_0 is (α_0, M_0)-weakly log-concave, then p_t is $(\alpha(t), M(t))$-weakly log-concave with $\alpha(t)$ and $M(t)$ explicitly defined in Appendix A.*

In fact, a stronger statement holds: Not only does the forward flow preserve weak log-concavity, it actively induces strong log-concavity, so that p_t becomes strongly log-concave after a finite amount of time, even when p_0 is only weakly log-concave. We call this the *regime shift property*, and it plays a central role in establishing convergence guarantees for the probability flow. The forthcoming Proposition 3 formalizes this property and explicitly quantifies the time at which the change takes place. For $\alpha, M \in \mathbb{R}$, let

$$\tau(\alpha, M) := \begin{cases} 0, & \alpha - M > 0 \\ \inf\left\{ t > 0 : \int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s > \dfrac{M-\alpha}{\alpha^2} \right\}, & \alpha - M \leq 0 \end{cases} \tag{10}$$

Since g is assumed to be non-decreasing, we have $\tau(\alpha, M) < \infty$.

Proposition 3 (Regime shifting). *For $0 < t < T$, it holds that*

$$\begin{cases} p_t \text{ is weakly log-concave}, & t \in (0, \tau(\alpha_0, M_0) \wedge T] \\ p_t \text{ is strongly log-concave}, & t \in (\tau(\alpha_0, M_0) \wedge T, T)\,. \end{cases}$$

The weak log-concavity constant $\alpha(t) - M(t)$ being negative for $t = 0$ and becoming positive for $t = \tau(\alpha_0, M_0)$ rises the question whether this transition progresses monotonously. This is, in fact, not necessarily the case, as Figure 3 in Appendix B illustrates.

We emphasize that, while similar results have been used in prior work for the specific case $f(t) = 1$ and $g(t) = \sqrt{2}$ (Gentiloni-Silveri & Ocello, 2025; Bruno & Sabanis, 2025), to the best of our knowledge this is the first result to hold for more general drift and diffusion functions, explicitely quantifying the transition time for time-dependent noise schedules used in practice. As shown in Example 3 in Appendix B, our result recovers the regime shifting property of the Ornstein-Uhlenbeck process, confirming that it is a proper generalization to time-dependent noise schedules.

4 Main result

This section presents our main result, a non-asymptotic error bound for the approximated probability flow (9) with general drift and diffusion functions. In order for our error bound to hold, we impose the following assumptions.

Assumption 1 (Regularity of the target). *The density of the data distribution p_0 is twice differentiable and positive everywhere. Moreover, $\log p_0$ is (α_0, M_0)-weakly concave in the sense of Definition 1 as well as L_0-Lipschitz continuous, meaning that for all $x, y \in \mathbb{R}^d$, it holds that*

$$\|\nabla \log p_0(x) - \nabla \log p_0(y)\| \leq L_0 \|x - y\|.$$

The first part of Assumption 1 has been employed in previous works such as Gentiloni-Silveri & Ocello (2025). Notably, it is a relaxed version of strong log-concavity which is the prevailing assumption in related works, e.g. Bruno et al. (2025); Li et al. (2022); Gao & Zhu (2025); Gao et al. (2025). The second part, i.e. the Lipschitz continuity of the score function, is a standard regularity condition that ensures the gradient of the log-density varies smoothly and is also considered in a large number of previous works, for example, Chen et al. (2023a); Gao & Zhu (2025); Taheri & Lederer (2025); Gao et al. (2025). In particular, Gentiloni-Silveri & Ocello (2025, Proposition 4.1) show that Gaussian mixtures—a widely used and expressive model class—satisfy both the weak log-concavity and log-Lipschitz conditions.

Assumption 2 (Lipschitz continuity in time). *There exists some $L_1 > 0$ such that for all $x \in \mathbb{R}^d$*

$$\sup_{\substack{k \in \{1,\dots,K\} \\ t_{k-1} \leq t \leq t_k}} \left\| \nabla \log p_{T-t}(x) - \nabla \log p_{T-t_{k-1}}(x) \right\| \leq L_1 h (1 + \|x\|).$$

Assumption 2 imposes a Lipschitz condition on the score function with respect to time, ensuring that the scores vary smoothly over time. This assumption is mainly employed to bound the discretization error (see proof of Proposition 21) and has been invoked widely (Gao & Zhu, 2025; Gao et al., 2025). A straightforward motivation is the idealized setting $X_0 \sim \mathcal{N}(0, \sigma^2 I_d)$, in which case its validity has been shown in Gao et al. (2025, p. 8-9).

Assumption 3 (Score-matching error)**.** *There exists some $\mathcal{E} > 0$ such that*

$$\sup_{k \in \{1,\dots,K\}} \left\| \nabla \log p_{T-t_{k-1}}\left(\widehat{Z}_{t_{k-1}}\right) - s_\theta\left(\widehat{Z}_{t_{k-1}}, T - t_{k-1}\right) \right\|_{L_2} \leq \mathcal{E}.$$

Assumption 3 ensures the accuracy of the learned score function. Just as in similar papers on the topic (Gao & Zhu, 2025; Gao et al., 2025; Gentiloni-Silveri & Ocello, 2025), it allows us to separate the convergence properties of the sampling algorithm from the challenges of score estimation. Our work focuses on the algorithmic aspects under idealized score estimates; the statistical error due to learning the score from data is the subject of another rich line of research (Zhang et al., 2024; Wibisono et al., 2024; Dou et al., 2024).

4.1 Error bound for variance-preserving SDEs

Since our main result, a general error bound accounting for all possible drift and diffusion functions f and g, is rather complex, we first consider VP SDEs. In particular, this includes the VP-SDE (4) with $\beta(t) = (b + at)^\rho$. For $\rho = 1$, this corresponds to the continuous time limit of the arithmetic noise schedule in DDPMs as described in Song et al. (2021, Appendix C).

Theorem 4 (Error bound for VP-SDE)**.** *For a VP-SDE with $\beta(t) = \Omega(1)$ and $\beta(T)h = o(1)$, it holds that*

$$\begin{aligned}
&\mathcal{W}_2(\mathcal{L}(\widehat{Z}_T), p_0) \\
&\quad \leq \mathcal{O}\Big(\underbrace{e^{-\frac{1}{2}\int_0^T \beta(t)\,\mathrm{d}t} \|X_0\|_{L_2}}_{\text{Initialization error}} + \underbrace{e^{T\beta^2(T)h} T\beta^2(T) h \Big(\|X_0\|_{L_2} + \sqrt{d} + T \Big)}_{\text{Discretization error}} + \underbrace{e^{T\beta^2(T)h} T\beta(T)\mathcal{E}}_{\text{Propag. score-match. error}} \Big).
\end{aligned}$$

The proof of this result is provided in Appendix C. The theorem implies that, in order to achieve a given accuracy level ε, meaning that $\mathcal{W}_2(\mathcal{L}(\widehat{Z}_T), p_0) \leq \varepsilon$, we need (1) the time scale T to be large enough for the initialization error to be small, (2) the step size h to be small enough for the discretization error to be small, and (3) the score-matching error $\mathcal{E}$ to be small enough for the propagated score-matching error to be small. The concrete heuristics for how the hyperparameters need to be chosen in case $\beta(t) = (b + at)^\rho$ and $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$, as it is the case for strongly log-concave distributions, are given in Table 1. Notably, these complexities coincide with those in Gao & Zhu (2025, Table 1) after translating the lower bound for T to a bound for $K = T/h$. This is remarkable as our results do not assume strong log-concavity of the data distribution and thus account for more general settings. In fact, this finding is not specific to the specific VP-SDE under consideration but applies to all other VP and also VE SDEs, as we will show in Section 4.3. Note that the additional assumptions on β in Theorem 4 are merely imposed to simplify the bounds and, in particular, implied by the conditions listed in Table 1.

Comparing the rates of our bound for ODE models as given in Table 1 with the analogous results for SDE-based models given in Gao et al. (2025, Table 2), yields an interesting result. In order to achieve a small sampling error that tends to zero, our bounds imply that one requires the step size to scale polynomially in T, while the bounds for the SDE-based sampling require the step size to scale exponentially in T. Corresponding derivations are provided in Appendix C. This comparison of the available bounds suggests that probability flow models may be more efficient than their SDE counterparts, as they can achieve the same accuracy under much less restrictive step-size requirements.

4.2 Error bound for general noise schedules

Now, we state the error bound for general drift and diffusion functions f and g.

Theorem 5 (Error bound for general probability flow ODEs)**.** *Under Assumptions 1, 2, and 3, it holds that*

$$\mathcal{W}_2\Big(\mathcal{L}\Big(\widehat{Z}_T\Big), p_0\Big) \leq \underbrace{E_0(f,g,T)}_{\textit{Initialization error}} + \underbrace{E_1(f,g,K,h)}_{\textit{Discretization error}} + \underbrace{E_2(f,g,K,h,\mathcal{E})}_{\textit{Propagated score-matching error}},$$

where

$$E_0(f,g,T) \coloneqq C(\alpha_0, M_0) e^{-\frac{1}{2}\int_0^T g^2(t)|\alpha(t)-M(t)|\,\mathrm{d}t} \|X_0\|_{L_2}, \tag{11}$$

$$E_1(f,g,K,h) \coloneqq \sum_{k=1}^{K} \left(\prod_{j=k+1}^{K} \gamma_{j,h} \right) e^{\int_{t_k}^{T} f(T-t)\,\mathrm{d}t} \cdot \left(\frac{1}{2} L_1 h (1 + \theta(T) + \omega(T)) \phi_{k,h} + \frac{1}{2}\sqrt{h}\nu_{k,h}\sqrt{\psi_{k,h}} \right), \tag{12}$$

$$E_2(f,g,K,h,\mathcal{E}) \coloneqq \sum_{k=1}^{K} \left(\prod_{j=k+1}^{K} \gamma_{j,h} \right) e^{\int_{t_k}^{T} f(T-t)\,\mathrm{d}t} \left(\frac{1}{2}\mathcal{E}\phi_{k,h} \right), \tag{13}$$

the functions $\alpha(t)$*,* $M(t)$*, and* $\tau(\alpha, M)$ *are defined in* (15)*,* (16)*, and* (10)*,*

$$C(\alpha_0, M_0) \coloneqq \exp\left(\frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \xi(\tau(\alpha_0, M_0)) \int_0^{\tau(\alpha_0, M_0)} g^2(t)\,\mathrm{d}t \right), \tag{14}$$

and all other quantities are defined in Appendix A.

Note that the error terms E_0, E_1, and E_2 also depend on the weak concavity and Lipschitz constants α_0, M_0, L_0 and, L_1 from Assumptions 1 and 2. However, since these are determined by the data distribution p_0 and thus cannot be controlled by the user, we do not explicitly include them in the arguments.

Although the error bound in Theorem 5 looks rather complex, we can identify its key properties as follows. According to (11), E_0 depends on the drift f, the diffusion coefficient g, and the time horizon T. It decreases exponentially with T and increases with factors related to the target distribution, namely α_0, M_0, and $\|X_0\|_{L_2}$. Thus, in practice, for sufficiently large T, the error E_0 can be neglected. As stated in (12), E_1 depends on f, g, K, and also on the step size h. At its core lies a product over $\gamma_{j,h}$. Depending on the regime shift, each $\gamma_{j,h}$ takes values either less than or greater than one (see Proposition 18 in Appendix D). A sufficiently small step size h is necessary to control that product when the factors exceed one. In particular, E_1 vanishes as h goes to zero, which matches with intuition as it corresponds to the discretization error. Note that it increases with the Lipschitz constant of the target L_1, $\|X_0\|_{L_2}$, and the dimensionality of the data d (we refer to Taheri & Lederer (2025), who employ regularization techniques to reduce d to a much smaller sparsity level for diffusion models). Finally, the propagated score-matching error E_2, defined in (13), depends on f, g, K, h, and additionally on the score-matching error $\mathcal{E}$. It also involves the product over $\gamma_{j,h}$, as in E_1. As $\mathcal{E} \to 0$, this error vanishes. Thus, to prevent this source of error from blowing up, the score-matching error $\mathcal{E}$ must be sufficiently small. Note that, as opposed to T and h, the admissible score-matching error $\mathcal{E}$ cannot be directly chosen, but rather determines how to pick $s_\theta(x,t)$. When using a neural network, for example, $\mathcal{E}$ might affect its architecture, the number of epochs used for training, and the necessary number of training samples. For a closer understanding of how large the time horizon T and how small the score-matching error $\mathcal{E}$ and step size h need to be, see the discussion following Theorem 4 for the VP-SDE case, and Table 2 in Appendix D for specific VE and VP SDEs.

The proof of Theorem 5 is based on splitting the overall error into the initialization error and the combined discretization and propagated score-matching error. While the former one can be bounded directly, the upper bound for the latter one is based on a recursive error propagation property and the choice of a sepecific coupling. A detailed proof is given in Appendix E.

4.3 The price of weak log-concavity

It is instructive to compare our result to the strongly log-concave case analyzed in Gao & Zhu (2025). In fact, our Theorem 5 matches their Theorem 3.4 when p_0 is strongly log-concave, i.e. $M_0 = 0$. A

detailed comparison of the two error bounds is given in Appendix D. Differences are mostly due to the altered concavity and Lipschitz parameters of $\log p_t$. Analyzing the effects of these differences on the asymptotic behavior of the error bound in case p_0 is weakly log-concave leads to the following result. Its proof is also given in Appendix D.

Proposition 6 (Weak log-concavity preserves the T, h and $\mathcal{E}$ asymptotics)**.** *For any choice of f and g according to a VP-SDE* (3) *or VE-SDE* (4)*, the following holds. Even if p_0 is only weakly log-concave, the asymptotics of the error bound in Theorem 5 with respect to T, h, and $\mathcal{E}$ are the same as for the bound given in Gao & Zhu (2025, Theorem 3.4), which relies on the stricter assumption of strong log-concavity.*

This is a striking result: the error $\mathcal{W}_2(\mathcal{L}(\hat{Z}_T), p_0)$ scales in T, h, and $\mathcal{E}$ exactly as under the more restrictive strong log-concavity assumption. This means, in particular, that the heuristics for choosing these hyperparameters remain exactly the same. Consequently, we do not have to derive the heuristics from scratch but can reuse the results from Gao & Zhu (2025, Section 3.3). A more detailed discussion on how the results can be transferred, including a table for the heuristics under general noise schedules, is provided in Appendix D.

While the asymptotics remain the same as compared to the strongly log-concave case, the weaker assumption does have an adverse effect on the non-asymptotic error bounds. In particular, the initialization error bound of Theorem 5 involves a coefficient that scales exponentially in the strength of deviation from log-concavity given by M_0/α_0. As the following result shows, an exponential dependence on this quantity is unavoidable.

Proposition 7 (Exponential dependence is unavoidable)**.** *Let f and g be of VP-SDE type, and let $\mathcal{P}_{\alpha_0, M_0}$ denote the class of d-dimensional (α_0, M_0)-weakly log-concave densities. Then, there exists a universal constant $c > 0$ such that, for $M_0 \geq \alpha_0$ (excluding the strongly log-concave setting),*

$$\liminf_{T\to\infty} \sup_{p_0 \in \mathcal{P}_{\alpha_0, M_0}} \frac{\mathcal{W}_2(\mathcal{L}(Y_T), p_0)}{e^{-\frac{1}{2}\int_0^T \beta(t)\,\mathrm{d}t} \|X_0\|_{L_2}} \geq \frac{c}{(\alpha_0 M_0)^{\frac{1}{4}}} \exp\left(\frac{M_0}{4\alpha_0}\right).$$

Thus, the exponential dependence on the weak log-concavity parameters arising in our upper bounds reflects a genuine instability of the reverse probability flow for multimodal targets and cannot, under the weak log-concavity assumption alone, be replaced by a dependence that is merely polynomial in α_0 and M_0. We note that the present lower bound establishes the necessity of an exponential dependence itself, not the optimality of the precise exponential constant in our upper bound.

5 Conclusion

This paper extends convergence theory for score-based generative models to more realistic data distributions and practical ODE solvers, providing concrete guarantees for the efficiency and correctness of the sampling algorithm in practical applications such as image generation. In particular, our results extend existing 2-Wasserstein convergence bounds for probability flow ODEs with general time-dependent drift and diffusion coefficients to a significantly broader class of distributions (including multi-modal mixtures), by relaxing the strong log-concavity assumption on the data distribution. Remarkably, the asymptotic rates in the time horizon, step size, and acceptable score-matching error remain the same as in the strongly log-concave case. At the same time, the weaker assumption necessarily introduces an additional multiplicative factor in the error bound that scales exponentially with the degree of deviation from log-concavity. These findings suggest that the strong performance of diffusion models in practice does not hinge on log-concavity of the data distribution, since comparable gurantees hold much more broadly, at an explicit and quantifiable cost.

In future work, it would be interesting to see if the assumptions can be even further relaxed and how this would influence the error bound. Moreover, it may be possible to extend the results to the more general case of vector-valued drift functions f and matrix-valued diffusion functions g. Another promising line of research concerns reducing the (potentially very large) dimensionality d to the intrinsic one of a lower-dimensional manifold on which the data lie. It remains to be seen whether the error bounds presented here can be adapted to this setting.

AI USE STATEMENT

In this work, we used generative AI tools for providing critical ingredients for proving Proposition 7. We have not used generative AI tools for helping develop theoretical models or conceptual frameworks, formulating mathematical claims, assisting in the writing of proofs, propose or refine hypotheses, designing or providing feedback on research methodology or experiments, assisting with translation, or interpreting results, and generating synthetical data sets, implementing methods, cleaning and reformating datasets, supporting qualitative and thematic data analysis are not applicable to this work. We have reviewed all AI-assisted work. In particular, we checked the LLM-generated proof step by step, refined and extended it, and then rewrote it from scratch to arrive at a more readable and rigorous form. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Eliot Beyler and Francis Bach. Convergence of deterministic and stochastic diffusion-model samplers: A simple analysis in wasserstein distance. *arXiv preprint arXiv:2508.03210*, 2025.

A. Block, Y. Mroueh, and A. Rakhlin. Generative modeling with denoising auto-encoders and langevin sampling. *arXiv preprint arxiv:2002.00107*, 2020.

Giovanni Brigati and Francesco Pedrotti. Heat flow, log-concavity, and lipschitz transport maps. *arXiv preprint arXiv:2404.15205*, 2024.

Stefan Bruno and Sotirios Sabanis. Wasserstein convergence of score-based generative models under semiconvexity and discontinuous gradients. *Transactions on Machine Learning Research*, 2025.

Stefano Bruno, Ying Zhang, Dong-Young Lim, Ömer Deniz Akyildiz, and Sotirios Sabanis. On diffusion-based generative models and their error bounds: The log-concave case with full convergence estimates. *Transactions on Machine Learning Research*, 2025.

H. Chen, H. Lee, and J. Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *ICML*, pp. 4735–4763, 2023a.

S. Chen, S. Chewi, H. Lee, Y. Li, J. Lu, and A. Salim. The probability flow ode is provably fast. In *NeurIPS*, volume 36, pp. 68552–68575, 2023b.

S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, and A. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *ICLR*, 2023c.

S. Chen, G. Daras, and A. Dimakis. Restoration-degradation beyond linear diffusions: A non-asymptotic analysis for ddim-type samplers. In *ICML*, pp. 4462–4484. PMLR, 2023d.

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.

Giovanni Conforti, Alain Durmus, and Marta Gentiloni-Silveri. Kl convergence guarantees for score diffusion models under minimal data assumptions. *SIAM Journal on Mathematics of Data Science*, 7(1):86–109, 2025a.

Giovanni Conforti, Daniel Lacker, and Soumik Pal. Projected langevin dynamics and a gradient flow for entropic optimal transport. *Journal of the European Mathematical Society*, 2025b.

V. De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022.

Z. Dou, S. Kotekal, Z. Xu, and H. Zhou. From optimal score matching to optimal sampling. *arXiv preprint arXiv:2409.07032*, 2024.

Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow odes of diffusion models in wasserstein distances. In *AISTATS*, 2025.

Xuefeng Gao, Hoang M Nguyen, and Lingjiong Zhu. Wasserstein convergence guarantees for a general class of score-based generative models. *Journal of machine learning research*, 26(43): 1–54, 2025.

Marta Gentiloni-Silveri and Antonio Ocello. Beyond log-concavity and score regularity: Improved convergence bounds for score-based generative models in w2-distance. In *ICML*, pp. 55631–55656. PMLR, 2025.

Marta Gentiloni Silveri, Alain Durmus, and Giovanni Conforti. Theoretical guarantees in kl for diffusion flow matching. In *NeurIPS*, volume 37, pp. 138432–138473, 2024.

Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, volume 33, pp. 6840–6851, 2020.

Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

H. Lee, J. Lu, and Y. Tan. Convergence for score-based generative modeling with polynomial complexity. In *NeurIPS*, volume 35, pp. 22870–22882, 2022.

H. Lee, J. Lu, and Y. Tan. Convergence of score-based generative modeling for general data distributions. *ALT*, pp. 946–985, 2023.

G. Li, Y. Wei, Y. Chen, and Y. Chi. Towards non-asymptotic convergence for diffusion-based generative models. In *ICLR*, 2024.

Ruilin Li, Hongyuan Zha, and Molei Tao. Sqrt (d) dimension dependence of langevin monte carlo. In *ICLR*, 2022.

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

Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *NeurIPS*, volume 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. In *ICLR*, 2021.

M. Taheri and J. Lederer. Regularization can make diffusion models more efficient. *arXiv preprint arXiv:2502.09151*, 2025.

A. Wibisono and K. Yang. Convergence in KL divergence of the inexact langevin algorithm with application to score-based generative models. *arXiv preprint arxiv:2211.01512*, 2022.

A. Wibisono, Y. Wu, and K. Yang. Optimal score estimation via empirical bayes smoothing. In *COLT*, 2024.

K. Zhang, C. Yin, F. Liang, and J. Liu. Minimax optimality of score-based diffusion models: Beyond the density lower bound assumptions. In *ICML*, 2024.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *ICLR*, 2023.

APPENDIX

The Appendix is structured as follows:

- Appendix A contains the definitions of the key quantities in our main result, namely the error bound for general noise schedules given in Theorem 5.
- Appendix B provides the proofs of Proposition 2, 3, and 16 dealing with the propagation in time of Assumption 1. We start by establishing general results on weak concavity that are used in these proofs and also include bounds for the weak concavity constant $K(t) = \alpha(t) - M(t)$ and the Lipschitz constant $L(t)$. Moreover, we provide an example of a (constructed) distribution that is sub-gaussian but not weakly log-concave.
- Appendix C treats the specific case of VP-SDEs with $\beta(t) = (b + at)^\rho$ and provides the derivation of the corresponding error bound given in Theorem 4.
- Appendix D deals with the interpretation of our main result (Theorem 5). We establish a regime shift result for the contraction rate $\gamma_{k,h}$, derive a bound for $\theta(T)$ that is used in the arguments of Section 4.3, and compare our error bound to the one in Gao & Zhu (2025), which imposes a strong log-concavity assumption. Finally, we provide the proofs of Proposition 6 and 7.
- Appendix E provides the proofs of Proposition 20 and 21, which establish bounds for the different error sources and constitute the key ingredients for the proof of our main result (Theorem 5).

A KEY QUANTITIES

In the following, we provide the definitions of the key quantities in Theorem 5. For convenience, we also include variables that have already been defined in the main text. In particular, we have

$$\tau(\alpha, M) = \begin{cases} 0, & \alpha - M > 0 \\ \inf\left\{ t > 0 : \int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s > \dfrac{M-\alpha}{\alpha^2} \right\}, & \alpha - M \le 0 \end{cases},$$

$$C(\alpha_0, M_0) = \exp\left(\frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \xi(\tau(\alpha_0, M_0)) \int_0^{\tau(\alpha_0, M_0)} g^2(t)\,\mathrm{d}t \right).$$

The weak concavity and Lipschitz parameters of $\log p_t$ are given by

$$\alpha(t) \coloneqq \frac{1}{\frac{1}{\alpha_0} e^{-2\int_0^t f(s)ds} + \int_0^t e^{-2\int_s^t f(v)dv} g^2(s)ds}, \tag{15}$$

$$M(t) \coloneqq \frac{M_0 e^{2\int_0^t f(s)ds}}{\left(1 + \alpha_0 \int_0^t e^{2\int_0^s f(v)dv} g^2(s)ds\right)^2}, \tag{16}$$

$$L(t) \coloneqq \max\left\{ \min\left\{ \left(\int_0^t e^{-2\int_s^t f(v)dv} g^2(s)ds \right)^{-1}, e^{2\int_0^t f(s)ds} L_0 \right\}, \right.$$
$$\left. - \Big(\alpha(t) - M(t)\Big) \right\}, . \tag{17}$$

Further, we define

$$\xi(T) \coloneqq \sup_{0\le t\le T} \min\left\{ e^{2\int_0^t f(s)ds}, \frac{e^{2\int_0^t f(s)ds}}{(\int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s)^2} \right\}, \tag{18}$$

$$\gamma_{k,h} \coloneqq 1 - \int_{t_{k-1}}^{t_k} \delta_k(T-t)\,\mathrm{d}t + \frac{1}{2}L_1 h \int_{t_{k-1}}^{t_k} g^2(T-t)\,\mathrm{d}t, \tag{19}$$

$$\begin{aligned}\delta_k(T-t) \coloneqq{}& \frac{1}{2}e^{-\int_{t_{k-1}}^t f(T-s)\,\mathrm{d}s} g^2(T-t)\big(\alpha(T-t) - M(T-t)\big) \\ &- \frac{1}{8}hg^4(T-t)L^2(T-t),\end{aligned} \tag{20}$$

$$\theta(T) \coloneqq \sup_{0\le t\le T} e^{-\frac{1}{2}\int_0^t g^2(T-s)(\alpha(T-s)-M(T-s))-2f(T-s)\,\mathrm{d}s} e^{-\int_0^T f(s)\,\mathrm{d}s} \|X_0\|_{L_2}, \tag{21}$$

$$\omega(T) \coloneqq \sup_{0\le t\le T} \left(e^{-2\int_0^t f(s)\,\mathrm{d}s} \|X_0\|_{L_2}^2 + d\int_0^t e^{-2\int_s^t f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s \right)^{\frac{1}{2}}, \tag{22}$$

$$\begin{aligned}\nu_{k,h} \coloneqq{}& (\theta(T)+\omega(T)) \int_{t_{k-1}}^{t_k} \left[f(T-s) + \frac{1}{2}g^2(T-s)L(T-s) \right] \mathrm{d}s, \\ &+ (L_1(T+h) + \|\nabla \log p_0(\mathbf{0})\|) \int_{t_{k-1}}^{t_k} \frac{1}{2} g^2(T-s)\,\mathrm{d}s,\end{aligned} \tag{23}$$

$$\phi_{k,h} \coloneqq \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t, \tag{24}$$

$$\psi_{k,h} \coloneqq \int_{t_{k-1}}^{t_k} \left(e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)L(T-t) \right)^2 \mathrm{d}t. \tag{25}$$

B Propagation in time of Assumption 1

We start this section with general properties of weak concavity that will be used in the proof for its propagation in time. The following result relates the weak convexity profile $\kappa_g(r)$ introduced in Definition 1 to the classical definition of strong convexity. In particular, it says that (α, M)-weak concavity implies $(\alpha - M)$-strong concavity if $\alpha - M > 0$.

Lemma 8. *Let $g \in C^1(\mathbb{R}^d)$ and $k : [0,\infty) \to \mathbb{R}$. The following two statements are equivalent:*

(i) $\kappa_g(r) \ge k(r)$ *for all* $r > 0$,

(ii) $\langle \nabla g(x) - \nabla g(y), x - y\rangle \ge k(\|x-y\|)\|x-y\|^2$ *for all* $x, y \in \mathbb{R}^d$.

In particular, if g is (α, M)-weakly concave, then

$$\begin{aligned}\langle \nabla g(x) - \nabla g(y), x - y\rangle &\le -\alpha\|x-y\|^2 + \|x-y\| f_M(\|x-y\|) \\ &\le -(\alpha - M)\|x-y\|^2.\end{aligned}$$

Proof of Lemma 8. We can rewrite $\kappa_g(r) \ge k(r)$ as

$$\inf_{x,y\in\mathbb{R}:\|x-y\|=r} \{\langle \nabla g(x) - \nabla g(y), x - y\rangle\} \ge k(r)r^2, \quad r > 0.$$

Since the infimum over a set is bounded below by a constant if and only if each element of the set is greater than or equal to this constant, and the inequality holds for all possible values of r, the above display is equivalent to

$$\langle \nabla g(x) - \nabla g(y), x - y\rangle \ge k(\|x-y\|)\|x-y\|^2$$

for all $x, y \in \mathbb{R}^d$.

The second part of the statement follows from the fact that $\tanh(t) \le t$ for any $t > 0$ and hence $f_M(\|x-y\|) = 2\sqrt{M}\tanh\left(\frac{1}{2}\sqrt{M}\|x-y\|\right) \le M\|x-y\|$. □

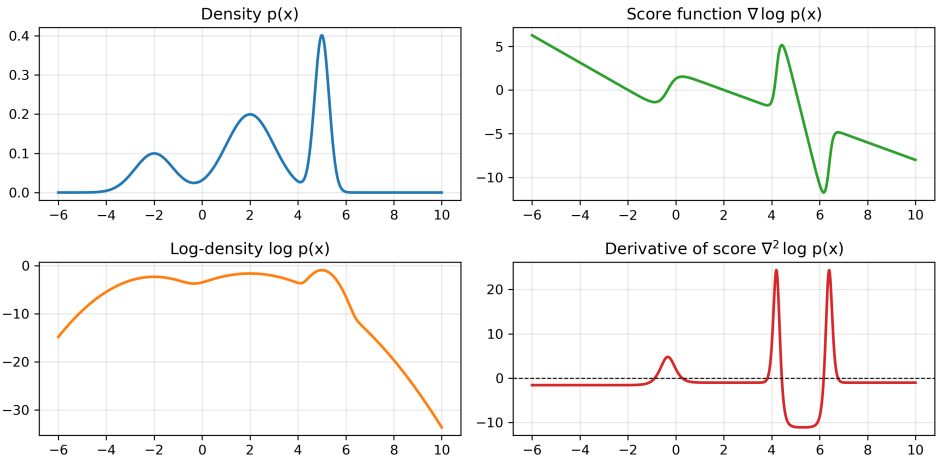

Figure 1: Plots corresponding to a Gaussian mixture model. See Example 1 for more details.

The next result establishes an equivalence between convexity of a function and boundedness of its Hessian.

Lemma 9. *Let $g \in C^2(\mathbb{R}^d)$ and $\beta \in \mathbb{R}$. The following two statements are equivalent:*

(i) $\langle \nabla g(x) - \nabla g(y), x - y \rangle \geq \beta \|x - y\|^2$ *for all* $x, y \in \mathbb{R}^d$,

(ii) $\nabla^2 g(x) \succeq \beta I_d$ *for all* $x \in \mathbb{R}^d$.

Proof of Lemma 9. First, assume that (i) holds. Then, for any $v \in \mathbb{R}^d$, we have

$$v^T \nabla^2 g(x) v = \lim_{t \to 0} \frac{\langle \nabla g(x + tv) - \nabla g(x), v \rangle}{t} \geq \lim_{t \to 0} \frac{\beta \|tv\|^2}{t^2} = \beta \|v\|^2 = v^t (\beta I_d) v.$$

On the other hand, assume that (ii) holds, and define

$$h(t) = \langle \nabla g(y + t(x - y)), x - y \rangle ,$$

so that

$$h'(t) = (x - y)^T \nabla^2 g(y + t(x - y))(x - y).$$

By the mean value theorem, it follows that

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle = h(1) - h(0) = h'(\tau)$$

for some $\tau \in [0, 1]$, and hence

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle = (x - y)^T \nabla^2 g(y + \tau(x - y))(x - y) \geq \beta \|x - y\|^2. \qquad \square$$

An example for weakly log-concave distribution are Gaussian mixture models.

Example 1. Let $p(x)$ denote the density function of a one-dimensional Gaussian mixture model with three components given by

$$0.2 \cdot \mathcal{N}(-2, 0.8^2) + 0.5 \cdot \mathcal{N}(2, 1^2) + 0.3 \cdot \mathcal{N}(5, 0.3^2).$$

As proved in Gentiloni-Silveri & Ocello (2025, Proposition 4.1), this is an example of a weakly log-concave distribution. An illustration of the density, log-density, score and derivative of the score is given in Figure 1. It clearly shows that the log-density is strongly concave at "large scales" with some local fluctuations. Accordingly, the Hessian $\nabla^2 \log p(x)$ is negative for large enough values of $|x|$ and globally bounded from above.

Next, we provide an example of a probability density function that has sub-gaussian tails but does not satisfy the weak log-concavity assumption. Note that it is a very constructed example explicitly meant to reveal the nature of our assumption.

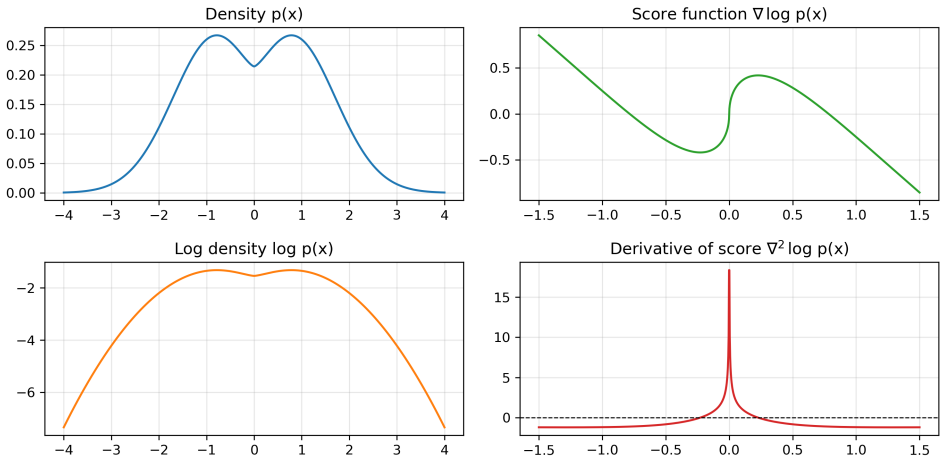

Figure 2: Plots corresponding to a constructed probability density function that is sub-gaussian but not weakly log-concave. See Example 2 for more details.

Example 2. Consider the probability density function

$$p(x) = \frac{1}{Z} e^{-x^2} \left(1 + |x|^{\frac{3}{2}}\right), \quad x \in \mathbb{R},$$

where the normalization constant $Z = \int_{-\infty}^{\infty} e^{-x^2}(1 + |x|^{\frac{3}{2}})\,\mathrm{d}x < \infty$ guarantees its total mass of one. Since, for any $x \in \mathbb{R}$,

$$p(x) \leq \frac{1}{Z} e^{-x^2} \left(1 + e^{\frac{1}{2}x^2}\right) \leq \frac{2}{Z} e^{-\frac{1}{2}x^2},$$

the corresponding distribution is sub-gaussian. However, as

$$\log p(x) = -\log(Z) - x^2 + \log\left(1 + |x|^{\frac{3}{2}}\right)$$

and thus

$$\nabla \log p(x) = \begin{cases} -2x + \frac{\frac{3}{2}\operatorname{sign}(x)\sqrt{|x|}}{1+|x|^{3/2}}, & x \neq 0 \\ 0, & x = 0 \end{cases},$$

the score function is infinitely steep at $x = 0$. Hence, the Hessian $\nabla^2 \log p(x)$ is unbounded, implying that the distribution cannot be weakly log-concave (cf. Lemma 8 and 9). An illustration of the involved functions is given in Figure 2.

In the following lemma, we list several properties of the convexity profile $\kappa_g(r)$ introduced in Definition 1. Since the proofs are rather trivial, we do not explicitly state them here.

Lemma 10. *Let $r > 0, \gamma \in \mathbb{R}, c \in \mathbb{R}^d$, and $g, g_1, g_2 \in C^1(\mathbb{R}^d)$. It holds*

(i) $\kappa_{g_1+g_2}(r) \geq \kappa_{g_1}(r) + \kappa_{g_2}(r)$,

(ii) $\kappa_{\gamma g}(r) = \gamma \kappa_g(r)$ *for* $\gamma > 0$,

(iii) $\kappa_{g+c}(r) = \kappa_g(r)$,

(iv) $\kappa_{g(\gamma x)}(r) = \gamma^2 \kappa_g(|\gamma| r)$,

(v) $\kappa_{\gamma\|\cdot\|^2}(r) = 2\gamma$.

As we will see in the proof of Proposition 2, the density p_t can be written as a convolution of p_0 with a Gaussian distribution. We are interested in how the weak log-concavity of p_0 is carried over to p_t by this transformation. The following theorem provides an important result in this context.

Theorem 11. *(Conforti, 2024, Theorem 2.1) Fix $M > 0$ and define*

$$\mathcal{F}_M = \{g \in C^1(\mathbb{R}^d) : \kappa_g(r) \geq -r^{-1} f_M(r)\}.$$

Then for all $0 < v < \infty$, it holds that

$$-\log g \in \mathcal{F}_M \Rightarrow -\log(S_v g) \in \mathcal{F}_M,$$

where $(S_v)_{v \geq 0}$ denotes the semigroup generated by a standard Brownian motion on $\mathbb{R}^d$, defined as

$$S_v g(x) = \int (2\pi v)^{-d/2} \exp\left(-\frac{\|x-y\|^2}{2v}\right) g(y) dy.$$

The connection between $\mathcal{F}_M$ and weak convexity is revealed in the following lemma.

Lemma 12. *If h is (α, M)-weakly convex, then $h - \frac{1}{2}\alpha\|\cdot\|^2 \in \mathcal{F}_M$.*

Proof of Lemma 12. By Lemma 10(i) and (v) together with the weak convexity of h, we have

$$\kappa_{h-\frac{1}{2}\alpha\|\cdot\|^2}(r) \geq \kappa_h(r) + \kappa_{-\frac{1}{2}\alpha\|\cdot\|^2} \geq \alpha - r^{-1} f_M(r) + 2\left(-\frac{1}{2}\alpha\right) = -r^{-1} f_M(r). \qquad \square$$

Remark 13 (General $f_M(r)$)**.** *As stated by Conforti et al. (2025b, Theorem 5.4), a general class for f_M is possible, provided that $f_M \in \widehat{\mathcal{G}} := \{g \in \mathcal{G} \text{ such that } g' \geq 0,\ 2g'' + gg' \leq 0\}$, where*

$$\mathcal{G} := \left\{ g \in C^2\big((0,\infty), \mathbb{R}_+\big) : r \mapsto r^{1/2} g(r^{1/2}) \text{ is non-decreasing and concave, } \lim_{r \downarrow 0} r g(r) = 0 \right\}.$$

We also need that there exists an $M > 0$ such that $rg(r) \leq Mr^2$ in order for the second part of Lemma 8 to hold. Naively speaking, the set $\widehat{\mathcal{G}}$ consists of smooth, non-negative, non-decreasing functions $g(r)$ defined on $(0, \infty)$ that grow in a controlled way and do not bend upward too rapidly. The transformation $r \mapsto r^{1/2} g(r^{1/2})$ must be non-decreasing and concave, ensuring mild growth behavior. The condition $2g'' + gg' \leq 0$ further constrains how sharply the function is allowed to curve upward.

B.1 Propagation in time of weak log-concavity

Now, we are ready to present the proof of Proposition 2, establishing the weak log-concavity of p_t given that p_0 is (α_0, M_0)-weakly log-concave.

Proof of Proposition 2. Observe that $p_t(x) = \int p_{t|0}(x|y) p_0(y) dy$, where $p_{t|0}(\cdot|y)$ denotes the conditional density of X_t given $X_0 = y$. From equation (2), it follows that

$$p_{t|0}(x|y) = (2\pi c_1(t))^{-d/2} \exp\left(-\frac{\left\|x - c_0^{-1}(t) y\right\|^2}{2c_1(t)}\right)$$

with

$$c_0(t) = e^{\int_0^t f(s) ds}, \quad c_1(t) = \int_0^t e^{-2\int_s^t f(v) dv} g^2(s) ds, \tag{26}$$

which yields

$$p_t(x) = \int (2\pi c_1(t))^{-d/2} \exp\left(-\frac{\left\|x - c_0^{-1}(t) y\right\|^2}{2c_1(t)}\right) p_0(y) dy. \tag{27}$$

We can write the argument of the exponential function within p_t as

$$-\frac{\left\|x - c_0^{-1}(t) y\right\|^2}{2c_1(t)} = -\frac{\|x\|^2 - 2c_0^{-1}(t)\langle x, y\rangle + c_0^{-2}(t)\|y\|^2 + \alpha_0 c_1(t)\|y\|^2}{2c_1(t)} + \frac{1}{2}\alpha_0 \|y\|^2.$$

Defining $c_\alpha(t) = c_0^{-2}(t) + \alpha c_1(t)$ and completing the square further yields

$$\begin{aligned}
-\frac{\left\|x - c_0^{-1}(t)y\right\|^2}{2c_1(t)} &= -\frac{\|x\|^2 - 2c_0^{-1}(t)\langle x, y\rangle + c_{\alpha_0}(t)\|y\|^2}{2c_1(t)} + \frac{1}{2}\alpha_0\|y\|^2 \\
&= -\frac{c_{\alpha_0}^{-1}(t)\|x\|^2 - 2(c_{\alpha_0}(t)c_0(t))^{-1}\langle x, y\rangle + \|y\|^2}{2c_{\alpha_0}^{-1}(t)c_1(t)} + \frac{1}{2}\alpha_0\|y\|^2 \\
&= -\frac{\left\|(c_{\alpha_0}(t)c_0(t))^{-1}x - y\right\|^2}{2c_{\alpha_0}^{-1}(t)c_1(t)} + \frac{(c_{\alpha_0}(t)c_0(t))^{-2} - c_{\alpha_0}^{-1}(t)}{2c_{\alpha_0}^{-1}(t)c_1(t)}\|x\|^2 + \frac{1}{2}\alpha_0\|y\|^2 \\
&= -\frac{\|c_4(\alpha_0, t)x - y\|^2}{2c_3(\alpha_0, t)} - \frac{1}{2}c_2(\alpha_0, t)\|x\|^2 + \frac{1}{2}\alpha_0\|y\|^2,
\end{aligned}$$

where

$$\begin{aligned}
c_2(\alpha, t) &\coloneqq -\frac{(c_\alpha(t)c_0(t))^{-2} - c_\alpha^{-1}(t)}{c_\alpha^{-1}(t)c_1(t)}, \\
c_3(\alpha, t) &\coloneqq c_\alpha^{-1}(t)c_1(t), \\
c_4(\alpha, t) &\coloneqq (c_\alpha(t)c_0(t))^{-1}.
\end{aligned}$$

Altogether, we get

$$\begin{aligned}
p_t(x) &= \int (2\pi c_1(t))^{-d/2} \exp\left(-\frac{\|c_4(\alpha_0, t)x - y\|^2}{2c_3(\alpha_0, t)} - \frac{1}{2}c_2(\alpha_0, t)\|x\|^2 + \frac{1}{2}\alpha_0\|y\|^2\right) p_0(y)dy \\
&= \exp\left(-\frac{1}{2}c_2(\alpha_0, t)\|x\|^2\right)\left(\frac{c_1(t)}{c_3(t)}\right)^{-d/2} \\
&\quad \cdot \int (2\pi c_3(t))^{-d/2} \exp\left(-\frac{\|c_4(\alpha_0, t)x - y\|^2}{2c_3(\alpha_0, t)} + \frac{1}{2}\alpha_0\|y\|^2\right) p_0(y)dy \\
&= \exp\left(-\frac{1}{2}c_2(\alpha_0, t)\|x\|^2\right)\left(\frac{c_1(t)}{c_3(t)}\right)^{-d/2} S_{c_3(\alpha_0,t)}\left(\exp\left(\frac{1}{2}\alpha_0\|\cdot\|^2\right)p_0\right)(c_4(\alpha_0, t)x),
\end{aligned}$$

or equivalently

$$-\log p_t(x) = \frac{1}{2}c_2(\alpha_0, t)\|x\|^2 + \frac{d}{2}\log\left(\frac{c_1(t)}{c_3(t)}\right) - \log S_{c_3(\alpha_0,t)}\left(\exp\left(\frac{1}{2}\alpha_0\|\cdot\|^2\right)p_0\right)(c_4(\alpha_0, t)x).$$

By Lemma 10(i) and (iv), this implies that

$$\kappa_{-\log p_t}(r) \geq \kappa_{\frac{1}{2}c_2(\alpha_0,\tilde{t})\|\cdot\|^2}(r) + c_4^2(\alpha_0, t)\kappa_{-\log S_{c_3(\alpha_0,t)}\left(\exp\left(\frac{1}{2}\alpha_0\|\cdot\|^2\right)p_0\right)}\big(c_4(\alpha_0, t)r\big).$$

Since p_0 is assumed to be (α_0, M_0)-weakly log-concave, it follows by Lemma 12 that

$$-\log\left(\exp\left(\frac{1}{2}\alpha_0\|\cdot\|^2\right)p_0\right) = -\log p_0 - \frac{1}{2}\alpha_0\|\cdot\|^2 \in \mathcal{F}_{M_0}$$

and thus, by Theorem 11, that

$$-\log S_{c_3(\alpha_0,t)}\left(\exp\left(\frac{1}{2}\alpha_0\|\cdot\|^2\right)p_0\right) \in \mathcal{F}_{M_0}.$$

This result together with Lemma 10(v) further yields

$$\begin{aligned}
\kappa_{-\log p_t}(r) &\geq c_2(\alpha_0, t) - c_4^2(\alpha_0, t)\big(c_4(\alpha_0, t)r\big)^{-1} f_{M_0}\big(c_4(\alpha_0, t)r\big) \\
&= c_2(\alpha_0, t) - r^{-1} f_{M_0 c_4^2(\alpha_0,t)}(r),
\end{aligned}$$

where in the last equality we used the fact that by definition $cf_M(cr) = f_{c^2 M}(r)$ for any $c, M, r > 0$.

The following simple but tedious calculations finally show that $\alpha(t) = c_2(\alpha_0, t)$ and $M(t) = M_0 c_4^2(\alpha_0, t)$, completing the proof. In particular, we have

$$c_2(\alpha, t) = -\frac{(c_\alpha(t)c_0(t))^{-2} - c_\alpha^{-1}(t)}{c_\alpha^{-1}(t)c_1(t)} = \frac{1 - c_\alpha^{-1}(t)c_0^{-2}(t)}{c_1(t)}$$

$$= \frac{1}{c_1(t)}\left(1 - \frac{1}{\big(c_0^{-2}(t) + \alpha c_1(t)\big)c_0^2(t)}\right)$$
$$= \frac{1}{c_1(t)}\left(1 - \frac{1}{1 + \alpha c_0^2(t)c_1(t)}\right)$$
$$= \frac{\alpha c_0^2(t)}{1 + \alpha c_0^2(t)c_1(t)} = \frac{1}{\alpha^{-1}c_0^{-2}(t) + c_1(t)},$$

and

$$c_4(\alpha, t) = \frac{1}{c_\alpha(t)c_0(t)} = \frac{1}{\big(c_0^{-2}(t) + \alpha c_1(t)\big)c_0(t)} = \frac{c_0(t)}{1 + \alpha c_0^2(t)c_1(t)}. \qquad \square$$

Note that p_t being $(\alpha(t), M(t))$-weakly log-concave implies that

$$\langle \nabla \log p_t(x) - \nabla \log p_t(y), x - y\rangle \leq -(\alpha(t) - M(t))\|x - y\|^2.$$

Consequently, it is a generalization of the result in Gao et al. (2025, Equation (5.4)) since $\alpha(t) = a(t)$ and $M(t) = 0$ if and only if $M_0 = 0$.

Remark 14. *It can be easily checked that $c_0(t)$, $c_\alpha(t)$, $c_2(\alpha, t)$, and $c_4(\alpha, t)$ are positive for any $\alpha > 0$ and $t \geq 0$. Moreover, $c_1(t)$ and $c_3(\alpha, t)$ are strictly positive for any $\alpha > 0$ and $t > 0$ and zero for $t = 0$.*

Next, we prove the regime shifting result, Proposition 3, dealing with the switch of p_t from being weakly to becoming strongly log-concave around $t = \tau(\alpha_0, M_0)$.

Proof of Proposition 3. If $\alpha_0 - M_0 > 0$, the result trivially holds with $\tau(\alpha_0, M_0) = 0$, due to the log-concavity preservation result in Gao et al. (2025, Proposition 7). So we only need to consider the case that $\alpha_0 - M_0 \leq 0$. By Lemma 8, p_t is $(\alpha(t) - M(t))$-strongly log-concave if $\alpha(t) - M(t) > 0$ which holds if and only if

$$\frac{\alpha_0 c_0^2(t)}{1 + \alpha_0 c_0^2(t)c_1(t)} - \frac{M_0 c_0^2(t)}{\big(1 + \alpha_0 c_0^2(t)c_1(t)\big)^2} > 0$$
$$\Leftrightarrow \quad \alpha_0 + \alpha_0^2 c_0^2(t)c_1(t) > M_0$$
$$\Leftrightarrow \quad c_0^2(t)c_1(t) > \frac{M_0 - \alpha_0}{\alpha_0^2}. \tag{28}$$

By recalling the definition (26) of c_0 and c_1, we have that

$$c_1(t) = \int_0^t e^{-2\int_s^t f(v)dv} g^2(s)\,\mathrm{d}s = \frac{1}{c_0^2(t)} \int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s. \tag{29}$$

Hence, condition (28) can be rewritten as

$$\int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s > \frac{M_0 - \alpha_0}{\alpha_0^2}. \qquad \square$$

The following lemma provides a lower bound for the weak concavity constant $K(t) = \alpha(t) - M(t)$. It is used at several occasions within the paper: when comparing our error bound to the strongly log-concave case in Section 4.3, to establish the more accessible error bound for the VP-SDE in Theorem 4, and in the proof of Proposition 20 bounding the initialization error.

Lemma 15. *Let $K(t) = \alpha(t) - M(t)$, $t \geq 0$. Then the following holds:*

$$K(t) \geq -|\alpha_0 - M_0| \min\left\{e^{2\int_0^t f(s)ds}, \frac{e^{2\int_0^t f(s)ds}}{(\alpha_0 \int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s)^2}\right\}, \qquad t \geq 0. \tag{30}$$

In particular, for any finite time $T > 0$, it holds that

$$\inf_{0 \leq t \leq T} K(t) \geq -\frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1}\xi(T), \tag{31}$$

where $\xi(T)$ is defined in (18)*.*

For example, in the VP case, for small t, (30) would read $K(t) \geq -|\alpha_0 - M_0| e^{\mathcal{B}(t)}$. This is very tight around $t = 0$, where the bound is close to the exact value $K(0) = \alpha_0 - M_0$. For large t, (30) would read

$$K(t) \geq -\frac{|\alpha_0 - M_0|}{\alpha_0^2} \frac{e^{\mathcal{B}(t)}}{(e^{\mathcal{B}(t)} - 1)^2},$$

which is close to zero for large t. This is enough for our purpose, as, intuitively, our results only require a control of $K(t) = \alpha(t) - M(t)$ when it is negative, that is when p_t deviates from strong log-concavity. But, thanks to the regime shifting result in Proposition 3, we know this can only happen up to a finite time $\tau(\alpha_0, M_0)$. See also Example 3 below for more details on the VP case.

Proof of Lemma 15. If $\alpha_0 - M_0 > 0$, it holds that $K(t) > 0$ as a consequence of log-concavity preservation (Gao et al., 2025, Proposition 7) and then (30) is trivially satisfied. So we only need to consider the case that $\alpha_0 - M_0 < 0$. For any $t \geq 0$, by means of simple algebra, we can write

$$\begin{aligned} K(t) &= \frac{\alpha_0 c_0^2(t)}{1 + \alpha_0 c_0^2(t) c_1(t)} - \frac{M_0 c_0^2(t)}{\big(1 + \alpha_0\, c_0^2(t) c_1(t)\big)^2} \\ &\geq \frac{\alpha_0 c_0^2(t)}{(1 + \alpha_0 c_0^2(t) c_1(t))^2} - \frac{M_0 c_0^2(t)}{\big(1 + \alpha_0\, c_0^2(t) c_1(t)\big)^2} \\ &= -\frac{|\alpha_0 - M_0| c_0^2(t)}{(1 + \alpha_0 c_0^2(t) c_1(t))^2} \\ &\geq -|\alpha_0 - M_0| c_0^2(t). \end{aligned} \tag{32}$$

Alternatively, starting from (32), we have

$$K(t) \geq -\frac{|\alpha_0 - M_0|}{\alpha_0^2} \frac{c_0^2(t)}{(c_0^2(t) c_1(t))^2}.$$

Finally, by combining the inequalities above we conclude:

$$K(t) \geq -|\alpha_0 - M_0| \min\left\{ c_0^2(t), \frac{1}{(\alpha_0 c_0(t) c_1(t))^2} \right\} \tag{33}$$

$$\geq -\frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \min\left\{ c_0^2(t), \frac{1}{(c_0(t) c_1(t))^2} \right\}. \tag{34}$$

Equation (33) can be rewritten as (30) by recalling the definitions of c_0 and c_1 given in (26). By taking infima over $t \geq 0$ in (34), we get (31). □

Example 3. We derive explicit expressions for the regime-shift time and the weak-concavity constant in the VP case, i.e. for $f(t) = \beta(t)/2$ and $g(t) = \sqrt{\beta(t)}$.

Let $\mathcal{B}(t) = \int_0^t \beta(s)\,\mathrm{d}s$. Then, from the definition (10) of $\tau(\alpha_0, M_0)$, we get

$$\int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s = e^{\mathcal{B}(t)} - 1 > \frac{M_0 - \alpha_0}{\alpha_0^2},$$

and consequently

$$\tau(\alpha_0, M_0) = \mathcal{B}^{-1}\left(\log\left(\frac{\alpha_0^2 + M_0 - \alpha_0}{\alpha_0^2} \right) \right), \tag{35}$$

where the inverse function $\mathcal{B}^{-1}(\cdot)$ is well-defined as $\mathcal{B}(\cdot)$ is continuous and strictly increasing. In particular, for the Ornstein-Uhlenbeck process, i.e. $f(t) = 1$ and $g(t) = \sqrt{2}$, we have

$$\tau(\alpha_0, M_0) = \log \sqrt{\frac{\alpha_0^2 + M_0 - \alpha_0}{\alpha_0^2}},$$

which matches Gentiloni-Silveri & Ocello (2025, equation (26)).

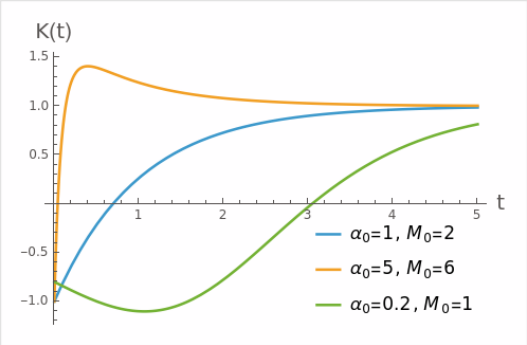

Figure 3: Plot of $K(t) = \alpha(t) - M(t)$, $t \geq 0$ for different values of α_0, M_0, in the OU case.

Next, we turn to the weak concavity constant $K(t)$. By recalling the definition (26) of $c_0(t)$, $c_1(t)$, and by relation (29), we have

$$c_0^2(t) = e^{\mathcal{B}(t)}, \qquad c_0^2(t)c_1(t) = e^{\mathcal{B}(t)} - 1.$$

Hence, from the definitions (15), (16) of $\alpha(t)$, $M(t)$ we get

$$K(t) = \alpha(t) - M(t) = \frac{\alpha_0 e^{\mathcal{B}(t)}}{1 + \alpha_0(e^{\mathcal{B}(t)} - 1)} - \frac{M_0 e^{\mathcal{B}(t)}}{\left(1 + \alpha_0(e^{\mathcal{B}(t)} - 1)\right)^2}, \quad t \geq 0, \tag{36}$$

for positive α_0, M_0. We remark that, as $t \to \infty$, if $\mathcal{B}(t) \to \infty$, one has $K(t) \to 1$, in agreement with the limiting standard Gaussian behavior of the forward diffusion process. If, in addition, $\alpha_0 = 1$, then $K(t)$ is guaranteed to be strictly increasing, since $\mathcal{B}(t)$ is strictly increasing. See Figure 3 for a graphical representation of possible behaviors.

B.2 Propagation in time of Lipschitz continuity

Assuming weak log-concavity of p_0 also guarantees Lipschitz continuity of the score function $\nabla \log(p_0)$ to propagate through the forward SDE (1) as the following result shows.

Proposition 16 (Propagation of Lipschitz continuity in time)**.** *If p_0 is (α_0, M_0)-weakly log-concave and $\nabla \log p_0$ is L_0-Lipschitz continuous, i.e.*

$$\|\nabla \log p_0(x) - \nabla \log p_0(y)\| \leq L_0 \|x - y\|,$$

then $\nabla \log p_t$ is $L(t)$-Lipschitz continous, i.e.

$$\|\nabla \log p_t(x) - \nabla \log p_t(y)\| \leq L(t)\|x - y\|$$

with

$$L(t) = \max\left\{\min\left\{\left(\int_0^t e^{-2\int_s^t f(v)dv} g^2(s)ds\right)^{-1}, e^{2\int_0^t f(s)ds} L_0\right\}, -\Big(\alpha(t) - M(t)\Big)\right\},$$

as defined in (17)*.*

This is a proper generalization of a corresponding result for strongly log-concave distributions p_0 given in Gao et al. (2025, Lemma 9), as, in that case, $\alpha(t) - M(t) > 0$ for all $t \in [0, T]$ and the maximum in (17) is achieved at the first term matching the definition of $L(t)$ in Gao et al. (2025).

Proof of Proposition 16. We use similar arguments as in the proof of (Gao et al., 2025, Lemma 9). With a change of variable, we can rewrite (27) as

$$p_t(x) = (c_0(t))^d \int (2\pi c_1(t))^{-d/2} \exp\left(-\frac{\|x - y\|^2}{2c_1(t)}\right) p_0(c_0(t)y)dy$$

with $c_0(t)$ and $c_1(t)$ defined in (26). Letting

$$q_0^t(x) \coloneqq p_0(c_0(t)x), \quad q_1^t(x) \coloneqq (2\pi c_1(t))^{-d/2} \exp\left(-\frac{\|x\|^2}{2c_1(t)}\right),$$

and $q_0^t * q_1^t$ denote their convolution, this implies that

$$\nabla^2 \log p_t(x) = \nabla^2 \log \big(q_0^t * q_1^t\big)(x).$$

We further define $\varphi_k^t = -\log q_k^t$ for $k \in \{0, 1\}$. An intermediate result of Saumard & Wellner (2014, Proposition 7.1), that does not make use of the strong log-concavity assumption, yields

$$\begin{aligned}\nabla^2(-\log p_t)(z) &= -\mathrm{Var}\big(\nabla\varphi_0^t(X)|X+Y=z\big) + \mathbb{E}\big[\nabla^2\varphi_0^t(X)|X+Y=z\big] \\ &= -\mathrm{Var}\big(\nabla\varphi_1^t(Y)|X+Y=z\big) + \mathbb{E}\big[\nabla^2\varphi_1^t(Y)|X+Y=z\big].\end{aligned}$$

Let $v \in \mathbb{R}^d$. By Cauchy-Schwartz inequality and the L_0-Lipschitz continuity of $\nabla \log p_0$, we have

$$\begin{aligned}v^t\nabla^2\varphi_0^t(x)v &= \lim_{t\to 0} \frac{\langle \nabla\varphi_0^t(x+tv) - \nabla\varphi_0^t(x), v\rangle}{t} \le \lim_{t\to 0} \frac{\|\nabla\varphi_0^t(x+tv) - \nabla\varphi_0^t(x)\| \cdot \|v\|}{t} \\ &= \lim_{t\to 0} \frac{|c_0(t)| \cdot \|\nabla \log p_0^t(c_0(t)(x+tv)) - \nabla \log p_0^t(c_0(t)x)\| \cdot \|v\|}{t} \\ &\le \lim_{t\to 0} \frac{|c_0(t)| \cdot L_0\|c_0(t)tv\| \cdot \|v\|}{t} = c_0^2(t)L_0\|v\|^2 = v^T c_0^2(t)L_0 I_d v.\end{aligned}$$

Hence, for all $x \in \mathbb{R}^d$,

$$\nabla^2\varphi_0^t(x) \preceq c_0^2(t)L_0 I_d.$$

Moreover, recall that

$$\varphi_1^t(x) = -\log q_1^t(x) = \frac{d}{2}\log(2\pi c_1(t)) + \frac{\|x\|^2}{2c_1(t)}$$

and thus

$$\nabla^2\varphi_1^t(x) = \frac{1}{c_1(t)} I_d$$

for all $x \in \mathbb{R}^d$. Since covariance matrices are always positive semi-definite, this finally leads to

$$\nabla^2(-\log p_t)(z) \preceq \min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\} I_d.$$

Note that from $A \preceq B$, we cannot directly follow that $\|A\| \le \|B\|$. However, if $C \preceq A \preceq B$, then we have $\|A\| \le \max\{\|B\|, \|C\|\}$. In particular, $0 \preceq A \preceq B$ implies $\|A\| \le \|B\|$. This result can be easily proven using the fact that the (spectral) norm of a symmetric matrix is given by its largest absolute eigenvalue.

From Proposition 2 together with Lemma 9, we get

$$\nabla^2(-\log p_t)(z) \succeq (\alpha(t) - M(t)) I_d.$$

In case $\alpha(t) - M(t) > 0$, this yields

$$\begin{aligned}\big\|\nabla^2 \log p_t(x)\big\| &\le \min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\} \\ &= \max\left\{\min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\}, -(\alpha(t) - M(t))\right\} = L(t).\end{aligned}$$

If, on the other hand, $\alpha(t) - M(t) < 0$, it follows that

$$\begin{aligned}\big\|\nabla^2 \log p_t(x)\big\| &\le \max\left\{\min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\}, |\alpha(t) - M(t)|\right\} \\ &= \max\left\{\min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\}, -(\alpha(t) - M(t))\right\} = L(t).\end{aligned}$$

□

The following lemma provides an upper bound for the Lipschitz constant $L(t)$. It is used when comparing our error bound to the strongly log-concave case in Section 4.3 and to establish the more accessible error bound for the VP-SDE in the proof of Theorem 4.

Lemma 17 (Upper bound for $L(t)$)**.** *It holds that*

$$\sup_{0\le t\le T} L(t) \le \max\left\{L_0 \vee 1, \frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1}\right\}\eta(T),$$

where

$$\eta(T) \coloneqq \sup_{0\le t\le T} \min\left\{e^{2\int_0^t f(s)ds}, \frac{1}{\int_0^t e^{-2\int_s^t f(v)dv} g^2(s)ds}\right\}. \tag{37}$$

Moreover, for $\xi(T)$ defined in Lemma 15, we have $\xi(T) \le \eta(T)$.

Proof of Lemma 17. Using the definition of $c_0(t)$ and $c_1(t)$ in (26), Proposition 16 states that

$$L(t) = \max\left\{\min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\}, -(\alpha(t) - M(t))\right\}.$$

By Lemma 15, it holds that

$$\sup_{0\le t\le T} -(\alpha(t) - M(t)) \le \frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1}\xi(T)$$

with (see (34))

$$\xi(T) = \sup_{0\le t\le T} \min\left\{c_0^2(t), \frac{1}{(c_0(t)c_1(t))^2}\right\}.$$

Furthermore, we have

$$\begin{aligned}\sup_{0\le t\le T} \min\left\{\frac{1}{c_1(t)}, c_0^2(t)L_0\right\} &\le (L_0 \vee 1) \sup_{0\le t\le T} \min\left\{\frac{1}{c_1(t)}, c_0^2(t)\right\} \\ &= (L_0 \vee 1)\eta(T).\end{aligned}$$

The result follows if we can show that $\xi(T) \le \eta(T)$. For that, consider $t \ge 0$ for which $c_1^{-1}(t) \le c_0^2(t)$. For those values of t, it also holds that

$$\frac{1}{(c_0(t)c_1(t))^2} = \frac{1}{c_0^2(t)}\left(\frac{1}{c_1(t)}\right)^2 \le \frac{1}{c_0^2(t)}\left(c_0^2(t)\right)^2 = c_0^2(t)$$

and

$$\frac{1}{(c_0(t)c_1(t))^2} = \frac{1}{c_0^2(t)}\frac{1}{c_1(t)}\frac{1}{c_1(t)} \le \frac{1}{c_0^2(t)}\frac{1}{c_1(t)}c_0^2(t) = \frac{1}{c_1(t)}.$$

It follows that

$$\min\left\{c_0^2(t), \frac{1}{(c_0(t)c_1(t))^2}\right\} \le \min\left\{\frac{1}{c_1(t)}, c_0^2(t)\right\}$$

and consequently $\xi(T) \le \eta(T)$. □

C Error bound for VP-SDEs

In this section, we derive the explicit error bound given in Theorem 4 for VP-SDEs with $\beta(T)h = o(1)$ and $\beta(t) = \Omega(1)$. With a slight abuse of $\mathcal{O}$-notation, we neglect constants in the exponents.

Proof of Theorem 4. From (36), using the substitution $B(t) = 1 + \alpha_0(e^{\mathcal{B}(t)} - 1)$ with $\mathcal{B}(t) = \int_0^s \beta(s)\,\mathrm{d}s$, we can write

$$\begin{aligned}
&\int_{t_k}^{T} \beta(T-t)(\alpha(T-t) - M(T-t))\,\mathrm{d}t \\
&\quad= \int_0^{T-t_k} \frac{\alpha_0\beta(t)e^{\mathcal{B}(t)}}{1+\alpha_0\big(e^{\mathcal{B}(t)}-1\big)} - \frac{M_0\beta(t)e^{\mathcal{B}(t)}}{\big(1+\alpha_0\big(e^{\mathcal{B}(t)}-1\big)\big)^2}\,\mathrm{d}t \\
&\quad= \int_0^{T-t_k} \frac{B'(t)}{B(t)} - \frac{M_0}{\alpha_0}\frac{B'(t)}{B^2(t)}\,\mathrm{d}t \\
&\quad= \left[\log(B(T-t_k)) - \log(B(0)) - \frac{M_0}{\alpha_0}\left(\frac{1}{B(0)} - \frac{1}{B(T-t_k)}\right)\right] \\
&\quad= \log\Big(1+\alpha_0\Big(e^{\mathcal{B}(T-t_k)}-1\Big)\Big) - \frac{M_0}{\alpha_0}\left(1 - \frac{1}{1+\alpha_0\big(e^{\mathcal{B}(T-t_k)}-1\big)}\right) \\
&\quad\geq \log\Big(1+\alpha_0\Big(e^{\mathcal{B}(T-t_k)}-1\Big)\Big) - \frac{M_0}{\alpha_0}
\end{aligned} \tag{38}$$

For the initialization error (11), this yields

$$\begin{aligned}
E_0(f,g,T) &= C(\alpha_0,M_0)e^{-\frac{1}{2}\int_0^T \beta(t)(\alpha(t)-M(t))\,\mathrm{d}t}\|X_0\|_{L_2} \\
&\leq C(\alpha_0,M_0)\Big(1+\alpha_0\Big(e^{\mathcal{B}(T)}-1\Big)\Big)^{-\frac{1}{2}} e^{\frac{M_0}{2\alpha_0}}\|X_0\|_{L_2} \\
&= \mathcal{O}\Big(e^{-\frac{1}{2}\mathcal{B}(T)}\|X_0\|_{L_2}\Big)
\end{aligned}$$

Next, we turn to the discretization error (12). According to the definitions (22) and (37), we have

$$\omega(T) = \sup_{0\leq t\leq T}\Big(e^{-\mathcal{B}(t)}\|X_0\|_{L^2}^2 + d(1-e^{-\mathcal{B}(t)})\Big)^{\frac{1}{2}} \leq \|X_0\|_{L^2} + \sqrt{d} \tag{39}$$

and

$$\eta(T) = \sup_{0\leq t\leq T}\min\left\{e^{\mathcal{B}(t)}, \frac{1}{1-e^{-\mathcal{B}(t)}}\right\} = \begin{cases} e^{\mathcal{B}(T)} & T < \mathcal{B}^{-1}(\log 2) \\ 2 & T \geq \mathcal{B}^{-1}(\log 2) \end{cases} \leq 2.$$

By Lemma 15 and 17, it follows that

$$\begin{aligned}
\sup_{0\leq t\leq T} L(t) &\leq \max\left\{(L_0\vee 1)\eta(T), \frac{|\alpha_0 - M_0|}{\alpha_0^2\wedge 1}\xi(T)\right\} \leq 2\mathfrak{a}_0, \\
\sup_{0\leq t\leq T} -(\alpha(t) - M(t)) &\leq \frac{|\alpha_0 - M_0|}{\alpha_0^2\wedge 1}\xi(T) \leq 2\mathfrak{a}_0,
\end{aligned} \tag{40}$$

where we define

$$\mathfrak{a}_0 \coloneqq \max\left\{(L_0\vee 1), \frac{|\alpha_0 - M_0|}{\alpha_0^2\wedge 1}\right\}.$$

Recall that β is positive and non-decreasing. The upper bound for $L(t)$ in (40) together with the definition of $\nu_{k,h}$ in (23) as well as Lemma 19 further yields

$$\begin{aligned}
\nu_{k,h} &= (\theta(T)+\omega(T))\frac{1}{2}\int_{t_{k-1}}^{t_k}\beta(T-s)(1+L(T-s))\,\mathrm{d}s,\\
&\quad+\frac{1}{2}(L_1(T+h)+\|\nabla\log p_0(\mathbf{0})\|)\int_{t_{k-1}}^{t_k}\beta(T-s)\,\mathrm{d}s,\\
&\le h\beta(T-t_{k-1})\cdot\Big(\big(\theta(T)+\omega(T)\big)\Big(1+\sup_{t_{k-1}\le t\le t_k}L(T-t)\Big)\\
&\qquad+L_1(T+h)+\|\nabla\log p_0(\mathbf{0})\|\Big)\\
&\le h\beta(T-t_{k-1})\cdot\Big(\big(\sqrt{C(\alpha_0,M_0)}\|X_0\|_{L_2}+\|X_0\|_{L^2}+\sqrt{d}\big)\big(1+2\mathfrak{a}_0\big)\\
&\qquad+L_1(T+h)+\|\nabla\log p_0(\mathbf{0})\|\Big)\\
&= \mathcal{O}\Big(h\beta(T)\big(\|X_0\|_{L_2}+\sqrt{d}+T\big)\Big)
\end{aligned}\tag{41}$$

and

$$\begin{aligned}
\phi_{k,h} &= \int_{t_{k-1}}^{t_k} e^{\frac{1}{2}\int_t^{t_k}\beta(T-s)\,\mathrm{d}s}\beta(T-t)\,\mathrm{d}t = 2\Big(e^{\frac{1}{2}\int_{t_{k-1}}^{t_k}\beta(T-s)\,\mathrm{d}s}-1\Big)\\
&\le 2(e^{h\beta(T)}-1),
\end{aligned}\tag{42}$$

$$\begin{aligned}
\psi_{k,h} &= \int_{t_{k-1}}^{t_k}\Big(e^{\frac{1}{2}\int_t^{t_k}\beta(T-s)\,\mathrm{d}s}\beta(T-t)L(T-t)\Big)^2\,\mathrm{d}t\\
&\le \beta(T-t_{k-1})\sup_{t_{k-1}\le t\le t_k}L^2(T-t)\int_{t_{k-1}}^{t_k}e^{\int_t^{t_k}\beta(T-s)\,\mathrm{d}s}\beta(T-t)\,\mathrm{d}t\\
&\le 4\mathfrak{a}_0^2\beta(T)(e^{h\beta(T)}-1).
\end{aligned}\tag{43}$$

Moreover, since $1-x\le e^{-x}$ for all x, we have

$$\begin{aligned}
\prod_{j=k+1}^{K}\gamma_{j,h} &= \prod_{j=k+1}^{K}\left(1-\int_{t_{j-1}}^{t_j}\delta_j(T-t)\,\mathrm{d}t+\frac{1}{2}L_1h\int_{t_{j-1}}^{t_j}\beta(T-t)\,\mathrm{d}t\right)\\
&\le \exp\left(-\sum_{j=k+1}^{K}\int_{t_{j-1}}^{t_j}\delta_j(T-t)\,\mathrm{d}t\right)\exp\left(\frac{1}{2}L_1h\int_{t_k}^{T}\beta(T-t)\,\mathrm{d}t\right)
\end{aligned}\tag{44}$$

Further, we can compute

$$\begin{aligned}
\sum_{j=k+1}^{K}\int_{t_{j-1}}^{t_j}\delta_j(T-t)\,\mathrm{d}t &= \sum_{j=k+1}^{K}\int_{t_{j-1}}^{t_j}\Big[\frac{1}{2}e^{-\frac{1}{2}\int_{t_{j-1}}^{t}\beta(T-s)\,\mathrm{d}s}\beta(T-t)\big(\alpha(T-t)-M(T-t)\big)\\
&\qquad-\frac{1}{8}h\beta^2(T-t)L^2(T-t)\Big]\,\mathrm{d}t\\
&\ge \frac{1}{2}e^{-\frac{h}{2}\beta(T)}\int_{t_k}^{T}\beta(T-t)(\alpha(T-t)-M(T-t))\,\mathrm{d}t\\
&\quad-\frac{1}{2}\mathfrak{a}_0^2h\beta(T-t_k)\int_{t_k}^{T}\beta(T-t)\,\mathrm{d}t\\
&\ge \frac{1}{2}e^{-h\beta(T)}\left[\log\Big(1+\alpha_0\Big(e^{\mathcal{B}(T-t_k)}-1\Big)\Big)-\frac{M_0}{\alpha_0}\right]\\
&\quad-\frac{1}{2}\mathfrak{a}_0^2h\beta(T)\mathcal{B}(T).
\end{aligned}\tag{45}$$

Combining (38), (44), (45) and using the upper bound for $L(t)$ given in (40), we get, by neglecting constants,

$$\begin{aligned}\prod_{j=k+1}^{K}\gamma_{j,h} &\leq \exp\left(-\frac{1}{2}e^{-h\beta(T)}\left[\log\left(1+\alpha_0\Big(e^{\mathcal{B}(T-t_k)}-1\Big)\right)-\frac{M_0}{\alpha_0}\right]\right)\\ &\quad\cdot\exp\left(\frac{1}{2}\mathfrak{a}_0^2h\beta(T)\mathcal{B}(T)+\frac{1}{2}hL_1\int_{t_k}^{T}\beta(T-t)\,\mathrm{d}t\right)\\ &\leq \Big(1+\alpha_0\Big(e^{\mathcal{B}(T-t_k)}-1\Big)\Big)^{-\frac{1}{2}e^{-h\mathcal{B}(T)}}e^{\frac{M_0}{\alpha_0}e^{-h\mathcal{B}(T)}}\\ &\quad\cdot\exp\left(\frac{1}{2}\mathfrak{a}_0^2h\beta(T)\mathcal{B}(T)+\frac{1}{2}hL_1\mathcal{B}(T)\right)\\ &= \mathcal{O}\Big(\exp\Big(-\mathcal{B}(T-t_k)e^{-h\beta(T)}+e^{-h\beta(T)}+h\beta(T)\mathcal{B}(T)\Big)\Big).\end{aligned}$$

From this result together with the upper bounds given in Lemma 19, (39), (41), (42) and (43), it follows for the discretization error (12) that

$$\begin{aligned}&E_1(f,g,K,h)\\ &\leq\sum_{k=1}^{K}\mathcal{O}\Big(\exp\Big(-\mathcal{B}(T-t_k)e^{-h\beta(T)}+e^{-h\beta(T)}+h\beta(T)\mathcal{B}(T)\Big)\Big)e^{\mathcal{B}(T-t_k)}\\ &\qquad\cdot\left(L_1h\big(1+\mathcal{O}(\|X_0\|_{L_2})+\mathcal{O}(\|X_0\|_{L_2}+\sqrt{d})\big)\Big(e^{\beta(T)h}-1\Big)\right.\\ &\qquad\qquad\left.+\sqrt{h}\mathcal{O}\Big(h\beta(T)\big(\|X_0\|_{L_2}+\sqrt{d}+T\big)\Big)\Big(\mathcal{O}\Big(\beta(T)(e^{h\beta(T)}-1)\Big)\Big)^{\frac{1}{2}}\right).\\ &=\sum_{k=1}^{K}\mathcal{O}\Big(e^{\mathcal{B}(T-t_k)(1-e^{-h\beta(T)})}\cdot\exp\Big(-e^{-h\beta(T)}\Big)\cdot e^{\mathcal{B}(T)h\beta(T)}\Big)\\ &\qquad\cdot\left(\mathcal{O}\Big(h(\|X_0\|_{L_2}+\sqrt{d})(e^{h\beta(T)}-1)\Big)\right.\\ &\qquad\qquad\left.+(h\beta(T))^{3/2}\mathcal{O}\Big(\big(\|X_0\|_{L_2}+\sqrt{d}+T\big)\Big)\Big(\mathcal{O}\Big(e^{h\beta(T)}-1\Big)\Big)^{\frac{1}{2}}\right)\\ &\leq\mathcal{O}\Big(K\cdot e^{\mathcal{B}(T)(1-e^{-h\beta(T)})}\cdot\exp\left(-e^{-h\beta(T)}\right)\cdot e^{\mathcal{B}(T)h\beta(T)}\Big)\\ &\qquad\cdot\left(\mathcal{O}\Big(h(\|X_0\|_{L_2}+\sqrt{d})(e^{h\beta(T)}-1)\Big)\right.\\ &\qquad\qquad\left.+(h\beta(T))^{3/2}\mathcal{O}\Big(\big(\|X_0\|_{L_2}+\sqrt{d}+T\big)\Big)\Big(\mathcal{O}\Big(e^{h\beta(T)}-1\Big)\Big)^{\frac{1}{2}}\right).\end{aligned}$$

The fact that $\mathcal{O}(e^x-1)=\mathcal{O}(x)$ for small x, $h\beta(T)=o(1)$, $\beta(T)=\Omega(1)$, $\mathcal{B}(T)\leq T\beta(T)$ further simplifies the expression on the right-hand side, finally yielding

$$\begin{aligned}E_1(f,g,K,h)&\leq\mathcal{O}\Big(K\cdot e^{T\beta^2(T)h}\cdot e^{T\beta^2(T)h}\Big)\cdot\mathcal{O}\Big(h^2\beta(T)^2\big(\|X_0\|_{L_2}+\sqrt{d}+T\big)\Big))\\ &=\mathcal{O}\Big(e^{T\beta^2(T)h}T\beta(T)^2h\Big(\|X_0\|_{L_2}+\sqrt{d}+T\Big)\Big).\end{aligned}$$

Similarly, we get for the propagated score matching error (13)

$$\begin{aligned}E_2(f,g,K,h,M)&=\mathcal{O}\Big(K\cdot e^{\mathcal{B}(T)(1-e^{-h\beta(T)})}\cdot\exp\left(-e^{-h\beta(T)}\right)\cdot e^{\mathcal{B}(T)h\beta(T)}\Big)\mathcal{E}\Big(e^{\beta(T)h}-1\Big)\\ &=\mathcal{O}\Big(e^{T\beta^2(T)h}T\beta(T)\mathcal{E}\Big).\end{aligned}$$

□

If $\beta(t) = (a + bt)^\rho$, and if we assume $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$, the above errors become

$$\begin{aligned} E_0 &\leq \mathcal{O}\left(e^{-T^{\rho+1}} \|X_0\|_{L_2}\right) \\ E_1 &\leq \mathcal{O}\left(e^{T^{2\rho+1}h}\, T^{2\rho+1}\, h\, (\sqrt{d} + T)\right) \\ E_2 &\leq \mathcal{O}\left(e^{T^{2\rho+1}h} T^{\rho+1} \mathcal{E}\right). \end{aligned} \tag{46}$$

From these upper bounds, we can derive heuristics on how T, h, and $\mathcal{E}$ need to be chosen in order to achieve a small sampling error, that is $\mathcal{W}_2(\mathcal{L}(\widehat{Z}_T), p_0) \leq \mathcal{O}(\varepsilon) = o(1)$. From the initialization error E_0, we get

$$T \geq \mathcal{O}\left(\log \frac{\sqrt{d}}{\varepsilon}\right)^{\frac{1}{\rho+1}}.$$

For the above choice of T, the upper bound for E_1 implies

$$h \leq \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d} T^{2\rho+1}}\right) = \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\left(\log \frac{\sqrt{d}}{\varepsilon}\right)^{\frac{2\rho+1}{\rho+1}}}\right).$$

This matches the bound from Gao & Zhu (2025, Corollary D.6) up to a negligible log-factor, that also appears in their proof. Finally, from the bound for E_2, it follows that

$$\mathcal{E} \leq \mathcal{O}\left(\frac{\varepsilon}{\log \frac{\sqrt{d}}{\varepsilon}}\right).$$

Next, we compare these rates for the ODE model with analogous results for SDE-based models, taken from Table 2 in Gao et al. (2025). In the VP case, for polynomial $f(t) = (b + at)^\rho/2$, one has the requirement (see Corollary 18 and its proof, in particular p. 52, in the paper)

$$T = Kh \geq \mathcal{O}\left(\log \frac{\sqrt{d}}{\varepsilon}\right)^{\frac{1}{\rho+1}}, \quad h = \mathcal{O}\left(\frac{\varepsilon^2}{d}\right).$$

It follows that

$$\sqrt{d} e^{T^{\rho+1}} h = \sqrt{d}\, \mathcal{O}\left(\frac{\sqrt{d}}{\varepsilon}\right) \cdot \mathcal{O}\left(\frac{\varepsilon^2}{d}\right) = \mathcal{O}(\varepsilon),$$

so that, in order to achieve $o(1)$ error one needs to take

$$h = o\left(\frac{e^{-T^{\rho+1}}}{\sqrt{d}}\right).$$

D Interpretation of the main result

As $\gamma_{k,h}$ plays the role of a contraction rate for the discretization and propagated score matching error, i.e. the L_2-distance between Y_{t_k} and $\widehat{Z}_{t_k}$ (see Proposition 21), it is crucial to investigate whether or when it lies between 0 and 1. The following proposition establishes a regime shifting result (similar to Proposition 3) for this contraction rate.

Proposition 18 (Regime shift for $\gamma_{k,h}$)**.** *Assuming*

$$h < \bar{h} := \min\left\{\frac{\log(2)}{\max_{0\leq t\leq T} f(t)}, \min_{t > \tau(\alpha_0, M_0)}\left\{\frac{\frac{1}{4} g^2(t)(\alpha(t) - M(t))}{\frac{1}{8} g^4(t) L^2(t) + \frac{1}{2} L_1 g^2(t)}\right\}\right\},$$

we have

$$\begin{cases} \gamma_{k,h} \in (0,1), & k \in \left\{1, 2, \ldots, \left\lfloor K - \frac{\tau(\alpha_0, M_0)}{h} \right\rfloor\right\} \\ \gamma_{k,h} > 1, & k \in \left\{\left\lceil K - \frac{\tau(\alpha_0, M_0)}{h} \right\rceil + 1, \ldots, K-1, K\right\}. \end{cases}$$

Moreover, it holds that for $\tilde{T} = (K + \ell)h$, $\tilde{\gamma}_{k+\ell,h} = \gamma_{k,h}$.

Proof of Proposition 18. To simplify notation, we write

$$\gamma_{k,h} = 1 - \int_{T-t_k}^{T-t_{k-1}} \delta_k(t)\,\mathrm{d}t + \frac{1}{2}L_1 h \int_{T-t_k}^{T-t_{k-1}} g^2(t)\,\mathrm{d}t$$

and

$$\delta_k(t) = \frac{1}{2}e^{-\int_t^{T-t_{k-1}} f(s)\,\mathrm{d}s} g^2(t)\big(\alpha(t) - M(t)\big) - \frac{1}{8}hg^4(t)L^2(t).$$

By definition of the regime shift, if $t < \tau(\alpha_0, M_0)$ then $\alpha(t) - M(t) < 0$, and hence $\delta_k(t) < 0$. It follows that $\gamma_{k,h} > 1$ for all k with $T - t_{k-1} \le \tau(\alpha_0, M_0)$, i.e. $k \ge K + 1 - h^{-1}\tau(\alpha_0, M_0)$.

On the other hand, assume that $k \le K - h^{-1}\tau(\alpha_0, M_0)$ and thus $T - t_k \ge \tau(\alpha_0, M_0)$. Note that $h \le \bar{h}$ implies that $e^{-h\max_{0\le t\le T} f(t)} > \frac{1}{2}$ and thus

$$h < \min_{t>\tau(\alpha_0,M_0)} \left\{ \frac{\frac{1}{2}e^{-h\max_{0\le t\le T} f(t)} g^2(t)(\alpha(t) - M(t))}{\frac{1}{8}g^4(t)L^2(t) + \frac{1}{2}L_1 g^2(t)} \right\}.$$

It follows that for $t > \tau(\alpha_0, M_0)$, in particular $t \in [T - t_k, T - t_{k-1}]$, it holds

$$\delta_k(t) > h \cdot \left(\frac{1}{8}g^4(t)L^2(t) + \frac{1}{2}L_1 g^2(t)\right) - \frac{1}{8}hg^4(t)L^2(t) = \frac{1}{2}L_1 h g^2(t),$$

and hence $\gamma_{k,h} < 1$. Moreover, inequality (60) in the proof of Proposition 21 implies that

$$1 - \int_{t_{k-1}}^{t_k} \delta_k(T - t)\,\mathrm{d}t \ge 0.$$

Consequently, $\gamma_{k,h} > 0$ since we assumed g to be positive for all $t > 0$.

Now, let $\tilde{T} = (K + \ell)h$. Then we have $\tilde{T} - t_{k+\ell} = T - t_k$ and hence

$$\tilde{\gamma}_{k+\ell,h} = 1 - \int_{T-t_k}^{T-t_{k-1}} \tilde{\delta}_{k+\ell}(t)\,\mathrm{d}t + \frac{1}{2}L_1 h \int_{T-t_k}^{T-t_{k-1}} g^2(t)\,\mathrm{d}t,$$

and

$$\tilde{\delta}_{k+\ell}(t) = \frac{1}{2}e^{-\int_t^{\tilde{T}-t_{k+\ell-1}} f(s)\,\mathrm{d}s} g^2(t)\big(\alpha(t) - M(t)\big) - \frac{1}{8}hg^4(t)L^2(t) = \delta_k(t),$$

which completes the proof. □

Note that, if $\tau(\alpha_0, M_0)$ is not evenly divisible by h, it is not clear whether $\gamma_{k,h}$ will be less or greater than one for $k = \left\lceil K - \frac{\tau(\alpha_0,M_0)}{h} \right\rceil$. The second part of Proposition 18 means that, when increasing $T = Kh$ to $\tilde{T} = (K + \ell)h$ for some integer $\ell \ge 1$, we have

$$\prod_{k=1}^{K+\ell} \tilde{\gamma}_{k,h} < \prod_{k=1}^{K} \gamma_{k,h},$$

which lies at the core of the discretization error $E_1(f, g, K, h)$ defined in (12).

The following lemma provides an upper bound for $\theta(T)$, another term involved in the discretization error E_1. It is used when comparing our error bound to the strongly log-concave case in Section 4.3 and to establish the more accessible error bound for the VP-SDE in the proof of Theorem 4.

Lemma 19 (Upper bound for $\theta(T)$)**.** *It holds that*

$$\theta(T) \le \sqrt{C(\alpha_0, M_0)}\|X_0\|_{L_2},$$

where $C(\alpha_0, M_0)$ *is defined in* (14)*.*

Proof of Lemma 19. By the definition of $\theta(T)$ in (21) and the non-negativity of $f(t)$, we have

$$\begin{aligned}\theta(T) &= \sup_{0\le t\le T} e^{-\frac{1}{2}\int_0^t g^2(T-s)(\alpha(T-s)-M(T-s))\,\mathrm{d}s} e^{-\int_t^T f(T-s)\,\mathrm{d}s}\|X_0\|_{L_2} \\ &\le \sup_{0\le t\le T} e^{-\frac{1}{2}\int_t^T g^2(s)(\alpha(s)-M(s))\,\mathrm{d}s}\|X_0\|_{L_2}.\end{aligned}$$

Since $\alpha(s) - M(s) > 0$ for any $s > \tau(\alpha_0, M_0)$, it follows that

$$e^{-\frac{1}{2}\int_t^T g^2(s)(\alpha(s)-M(s))\,\mathrm{d}s} < 1$$

for all $t > \tau(\alpha_0, M_0)$, and therefore

$$\begin{aligned}
\theta(T) &\leq \max\left\{1, \sup_{0\leq t\leq \tau(\alpha_0,M_0)} e^{-\frac{1}{2}\int_t^T g^2(s)(\alpha(s)-M(s))\,\mathrm{d}s}\right\}\|X_0\|_{L_2} \\
&\leq \max\left\{1, \sup_{0\leq t\leq \tau(\alpha_0,M_0)} e^{-\frac{1}{2}\int_t^{\tau(\alpha_0,M_0)} g^2(s)(\alpha(s)-M(s))\,\mathrm{d}s}\right\}\|X_0\|_{L_2} \\
&= \max\left\{1, \sup_{0\leq t\leq \tau(\alpha_0,M_0)} e^{\frac{1}{2}\int_t^{\tau(\alpha_0,M_0)} g^2(s)|\alpha(s)-M(s)|\,\mathrm{d}s}\right\}\|X_0\|_{L_2} \\
&= e^{\frac{1}{2}\int_0^{\tau(\alpha_0,M_0)} g^2(s)|\alpha(s)-M(s)|\,\mathrm{d}s}\|X_0\|_{L_2}.
\end{aligned}$$

Using Lemma 15, we can further say that

$$\theta(T) \leq e^{\frac{1}{2}\frac{|\alpha_0-M_0|}{\alpha_0^2\wedge 1}\xi(\tau(\alpha_0,M_0))\int_0^{\tau(\alpha_0,M_0)} g^2(s)\,\mathrm{d}s}\|X_0\|_{L_2} = \sqrt{C(\alpha_0,M_0)}\|X_0\|_{L_2}. \qquad \square$$

D.1 Comparison to the strongly log-concave case

As mentioned in Section 4.3, our Theorem 5 matches the one by Gao & Zhu (2025, Theorem 3.4), if p_0 is in fact strongly log-concave, i.e. $M_0 = 0$. To see that, note that our result differs from Gao and Zhu's in the following ways:

1. In the initialization error, we have the additional coefficient $C(\alpha_0, M_0)$ as well as $|\alpha(t) - M(t)|$ instead of $a(t)$ in the exponent. If p_0 is strongly log-concave, then $\tau(\alpha_0, M_0) = 0$ and thus $\xi(\tau(\alpha_0, M_0)) = 0$ implying that $C(\alpha_0, M_0) = 1$. Moreover, from the definitions in Proposition 2, it can be seen that, if $M_0 = 0$, then $M(t) = 0$ and $\alpha(t)$ equals $a(t)$ defined in Gao & Zhu (2025, equation (49)) which is positive for all $t \in [0, T]$.
2. In $\delta_k(T-t)$ and $\theta(T)$, the strong log-concavity parameter $a(T-t)$ of p_{T-t} is naturally replaced by the weak log-concavity parameter $(\alpha(T-t)-M(T-t))$. As explained above, we have $\alpha(T-t) = a(T-t)$ and $M(T-t) = 0$ in case p_0 is strongly log-concave.
3. The definition of the Lipschitz constant $L(t)$ of p_t in Proposition 16 resembles the one in Gao & Zhu (2025, equation (27)) but involves the additional term $-\big(\alpha(t) - M(t)\big)$. If p_0 is strongly log-concave, we have $\tau(\alpha_0, M_0) = 0$ and thus $\alpha(t) - M(t) > 0$ for all $t \in [0, T]$. Since the minimum in the definition (17) of $L(t)$ is always non-negative, the additional term can be disregarded and the two definitions coincide.
4. The coefficient in front of the second summand of $\delta_k(T-t)$ is $\frac{1}{8}$ instead of $\frac{1}{4}$. Note that this is better in the sense that it yields a tighter error bound.
5. The definition of $\nu_{k,h}$ involves the coefficient $T + h$ instead of T. We believe that the same should apply to Gao and Zhu's result, correcting Gao & Zhu (2025, equation (72)) as illustrated in equation (57) in the proof of Lemma 24 in Appendix E.
6. In the first summand of the discretization error $E_1(f, g, K, h)$, the coefficient $\|X_0\|_{L_2}$ is replaced by $\theta(T)$. According to Lemma 19 in Appendix D, it holds that

   $$\theta(T) \leq \sqrt{C(\alpha_0, M_0)}\|X_0\|_{L_2}^2.$$

   In the strongly log-concave case, we have $C(\alpha_0, M_0) = 1$ as explained under point 1. Hence, in this case, $\theta(T) \leq \|X_0\|_{L_2}$, which is used in Gao & Zhu (2025).

Next, we provide the proof of Proposition 6, establishing the remarkable finding that the asymptotics of our error bound given in Theorem 5 are the same as under the stricter assumption of strong log-concavity.

Proof of Proposition 6. To analyze the differences in the asymptotics with respect to T, h, and $\mathcal{E}$ of the bound in Theorem 5 if p_0 is only weakly log-concave compared to the strongly log-concave case analyzed in Gao & Zhu (2025, Theorem 3.4), we just need to consider the consequences of the differences in the error bounds as listed above. We discuss the effect of each difference point by point.

1. The constant $C(\alpha_0, M_0)$ does not influence the asymptotics. For the exponential term in the initialization error, we have

$$e^{-\int_0^T g^2(t)|\alpha(t)-M(t)|\,\mathrm{d}t} \leq e^{-\int_0^T g^2(t)\alpha(t)\,\mathrm{d}t} \cdot e^{\int_0^T g^2(t)M(t)\,\mathrm{d}t}.$$

So, in order to identify the difference to the strongly log-concave case, we need to analyze the second coefficient involving $M(t)$. For a VE-SDE, i.e. $f(t) \equiv 0$, we have

$$\begin{aligned}\int_0^T g^2(t)M(t)\,\mathrm{d}t &= \int_0^T \frac{M_0 g^2(t)}{\left(1+\alpha_0\int_0^t g^2(s)\,\mathrm{d}s\right)^2}\,\mathrm{d}t = -\frac{M_0}{2\alpha_0}\left(\frac{1}{1+\alpha_0\int_0^T g^2(s)\,\mathrm{d}s}-1\right)\\ &= \mathcal{O}\left(1-\frac{1}{\int_0^T g^2(s)\,\mathrm{d}s}\right) = \mathcal{O}(1),\end{aligned} \tag{47}$$

where we used the fact that g is positive and T diverges. In the VP case, i.e. $f(t) = \frac{1}{2}\beta(t)$ and $g(t) = \sqrt{\beta(t)}$, on the other hand, it follows from the substitution $B(t) = 1 + \alpha_0(e^{\mathcal{B}(t)} - 1)$ that

$$\begin{aligned}\int_0^T g^2(t)M(t)\,\mathrm{d}t &= \int_0^T \frac{M_0\beta(t)e^{\mathcal{B}(t)}}{\left(1+\alpha_0(e^{\mathcal{B}(t)}-1)\right)^2}\,\mathrm{d}t = \frac{M_0}{\alpha_0}\int_0^T \frac{B'(t)}{B^2(t)}\,\mathrm{d}t\\ &= -\frac{M_0}{\alpha_0}\left(\frac{1}{B(T)}-\frac{1}{B(0)}\right)\mathrm{d}t = -\frac{M_0}{\alpha_0}\left(\frac{1}{1+\alpha_0(e^{\mathcal{B}(T)}-1)}-1\right)\\ &= \mathcal{O}\left(1-e^{-\mathcal{B}(T)}\right) = \mathcal{O}(1),\end{aligned} \tag{48}$$

where we reused the definition $\mathcal{B}(t) = \int_0^t \beta(s)\,\mathrm{d}s$ and the value of $M(t)$ given in (36) from Example 3. In both cases, there is no change in the asymptotics.

2. To determine how the change in $\delta_k(T-t)$ influences the limit behavior, we recapitulate how it is analyzed in Gao & Zhu (2025, Corollaries 3.8–3.11). The coefficient $\prod_{j=k+1}^K \gamma_{j,h}$ in the error bound given in Theorem 5 is upper bounded using the fact that $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$. Accordingly, we have

$$\begin{aligned}\prod_{j=k+1}^K \gamma_{j,h} &= \prod_{j=k+1}^K 1 - \int_{t_{j-1}}^{t_j} \delta_j(T-t)\,\mathrm{d}t + \frac{1}{2}L_1 h\int_{t_{j-1}}^{t_j} g^2(T-t)\,\mathrm{d}t\\ &\leq \exp\left(-\sum_{j=k+1}^K \int_{t_{j-1}}^{t_j}\delta_j(T-t)\,\mathrm{d}t + \frac{1}{2}L_1 h\int_{t_k}^T g^2(T-t)\,\mathrm{d}t\right).\end{aligned}$$

The only new term in the above display emerging in the weak log-concave case is

$$\begin{aligned}&\exp\left(\sum_{j=k+1}^K \int_{t_{j-1}}^{t_j}\frac{1}{2}e^{-\int_{t_{j-1}}^t f(T-s)\,\mathrm{d}s} g^2(T-t)M(T-t)\,\mathrm{d}t\right)\\ &\leq \exp\left(\frac{1}{2}\int_{t_k}^T g^2(T-t)M(T-t)\,\mathrm{d}t\right) \leq \exp\left(\frac{1}{2}\int_0^T g^2(t)M(t)\,\mathrm{d}t\right),\end{aligned}$$

where we used the non-negativity of $f(t)$ and $M(t)$. Both, in the VE and VP case, the term on the far right-hand side is in $\mathcal{O}(1)$ as shown in (47) and (48). Thus, the asymptotic behavior remains unchanged. For a discussion of $\theta(T)$, see point 6.

3. When analyzing the limit behavior in Gao & Zhu (2025, Corollaries 3.8–3.11), the time-dependent Lipschitz constant is dealt with by finding an upper bound for $L(t)$ in the VP case and $g^2(t)L(t)$ in the VE case. Denote the upper bound for the Lipschitz constant $L^{GZ}(t)$ in Gao and Zhu's paper by $\bar{L}^{GZ}$. Note that we have
$$L(t) = \max\left\{L^{GZ}(t), -(\alpha(t) - M(t))\right\},$$
so it suffices to show that $-(\alpha(t) - M(t))$ is appropriately bounded. By Lemma 15, we have
$$\begin{aligned}-(\alpha(t) - M(t)) &\leq \frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \min\left\{e^{2\int_0^t f(s)ds}, \frac{e^{2\int_0^t f(s)ds}}{(\int_0^t e^{2\int_0^s f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s)^2}\right\} \\ &\leq \frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \min\left\{e^{2\int_0^t f(s)ds}, \frac{1}{\int_0^t e^{-2\int_s^t f(v)dv} g^2(s)ds}\right\},\end{aligned}$$
where the last inequality follows from the arguments in Lemma 17. Since
$$\bar{L}^{GZ} \geq L^{GZ}(t) \geq (L_0 \wedge 1) \min\left\{e^{2\int_0^t f(s)ds}, \frac{1}{\int_0^t e^{-2\int_s^t f(v)dv} g^2(s)ds}\right\},$$
it follows that
$$-(\alpha(t) - M(t)) \leq \frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \frac{1}{L_0 \wedge 1} \bar{L}^{GZ},$$
and hence
$$L(t) \leq \max\left\{1, -\frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1} \frac{1}{L_0 \wedge 1}\right\} \bar{L}^{GZ}.$$
Similar arguments lead to an upper bound for $g^2(t)L(t)$. As the bound only differs by some coefficient that is independent of T, h, and $\mathcal{E}$, the asymptotics are not affected.

4. The difference of the coefficients does not have any effect on the asymptotics.

5. Since $h = o(1)$, we have $\mathcal{O}(T + h) = \mathcal{O}(T)$.

6. The constant coefficient $\sqrt{C(\alpha_0, M_0)}$ does not influence the asymptotics. □

The asymptotics remaining the same means that we can reuse the results from Gao & Zhu (2025, Section 3.3), deriving heuristics for how to choose the hyperparameters T, h, and $\mathcal{E}$ in order for the sampling error to be appropriately bounded. Note that, in these derivations, Gao and Zhu make use of the fact that $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$, which may not always apply when p_0 is only assumed to be weakly log-concave. Consequently, our bounds will involve an additional dependency on this term (as in Theorem 4). However, it seems natural to assume that the L_2-norm of X_0 scales with the dimension in this way as
$$\|X_0\|_{L_2}^2 = \mathbb{E}\left[\|X_0\|^2\right] = \|\mu_0\|^2 + \mathrm{tr}(\Sigma_0) = \sum_{i=1}^d \left(\mu_0^{(i)}\right)^2 + \sum_{i=1}^d \Sigma_0^{(i,i)},$$
where $\mu_0 = (\mu_0^{(1)}, \ldots, \mu_0^{(d)})^\top \in \mathbb{R}^d$ and $\Sigma_0 = (\Sigma_0^{(i,j)})_{i,j=1}^d \in \mathbb{R}^{d\times d}$ denote the mean and covariance matrix corresponding to p_0. Accordingly, $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$ holds if the entries of μ_0 and Σ_0 do not scale with the dimension d.

Table 2 presents the heuristics for how to choose the time scale T, step size h, and acceptable score-matching error $\mathcal{E}$ in order to guarantee the error to be bounded by some small $\varepsilon > 0$. It was directly derived from Gao & Zhu (2025, Table 1), translating the bounds for the number of steps K to bounds for T. Note that we assume that $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$ for the table to be applicable. We want to emphasize that this is not a limiting assumption as we can derive analogous results in case this condition is not met. Similar to the bounds for the VP-SDE, given in Section 4.1, this would entail the term $\|X_0\|_{L_2}$ arising in the heuristics for T and h. To keep the results simple, and because the assumption seems natural as argued above, we decided to not explicitly state this dependence in the table. For a derivation of the heuristics in Table 2, we refer to Gao & Zhu (2025, Corollaries 3.8–3.11). Here, we only want to remark that the proof techniques are similar as for the VP-SDE, unveiled in Appendix C, and do not change in our case as revealed in Proposition 6.

f	g	T	h	$\mathcal{E}$
0	ae^{bt}	$\mathcal{O}\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)$	$\mathcal{O}\left(\frac{\varepsilon^3}{d^{\frac{3}{2}}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^2}{\sqrt{d}}\right)$
0	$(b+at)^c$	$\mathcal{O}\left(\left(\frac{d}{\varepsilon^2}\right)^{\frac{1}{2c+1}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^3}{d^{\frac{3}{2}}}\right)$	$\mathcal{O}\left(\frac{\varepsilon^2}{\sqrt{d}}\right)$
$\frac{b}{2}$	$\sqrt{b}$	$\mathcal{O}\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$
$\frac{b+at}{2}$	$\sqrt{b+at}$	$\mathcal{O}\left(\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)^{\frac{1}{2}}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$
$\frac{(b+at)^\rho}{2}$	$(b+at)^{\frac{\rho}{2}}$	$\mathcal{O}\left(\left(\log\left(\frac{\sqrt{d}}{\varepsilon}\right)\right)^{\frac{1}{\rho+1}}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$	$\mathcal{O}\left(\frac{\varepsilon}{\log\left(\frac{\sqrt{d}}{\varepsilon}\right)}\right)$

Table 2: Heuristics for the choice of the time horizon T, the step size h, and the acceptable score-matching error $\mathcal{E}$ in order for the 2-Wasserstein distance between the generated distribution $\mathcal{L}(\widehat{Z}_{t_k})$ and the true data distribution p_0 to be less than or equal to $\varepsilon = o(1)$. Different choices for f and g are considered. The table is split into VE and VP SDEs, and it is assumed that $\|X_0\|_{L_2} = \mathcal{O}(\sqrt{d})$.

D.2 The price of weak log-concavity

Next, we present the proof of Proposition 7, showing that the exponential dependence on the weak log-concavity parameters arising in our upper bounds is unavoidable.

Proof of Proposition 7. The main idea of the proof is to find one specific candidate $p_0^* \in \mathcal{P}(\alpha_0, M_0)$ that attains the lower bound. For that, we define the distribution of the first coordinate by

$$p_0^{(1)} = \frac{1}{2}\mathcal{N}\left(0, \frac{1}{\alpha_0}\right) + \frac{1}{2}\mathcal{N}\left(\frac{2\sqrt{M_0}}{\alpha_0}, \frac{1}{\alpha_0}\right),$$

and extend it to a d-dimensional distribution with independent coordinates via

$$p_0^*(x) = p_0^{(1)}(x_1)\prod_{i=2}^{d} p_0^{(d)}(x_i) \quad \text{with} \quad p_0^{(d)} = \mathcal{N}\left(0, \frac{1}{(d-1)\alpha_0}\right).$$

To show that p_0^* is indeed (α_0, M_0)-weakly log-concave, we first analyze $p_0^{(1)}$ and then transfer the result to p_0^*. Note that $p_0^{(1)}$ is a translation of the symmetric Gaussian mixture

$$q_0 = \frac{1}{2}\mathcal{N}\left(-\frac{\sqrt{M_0}}{\alpha_0}, \frac{1}{\alpha_0}\right) + \frac{1}{2}\mathcal{N}\left(\frac{\sqrt{M_0}}{\alpha_0}, \frac{1}{\alpha_0}\right)$$

and thus exhibits the same concavity properties. We define $U(x) = -\log q_0(x)$. Algebraic calculations yield

$$U'(x) = \alpha_0 x - \sqrt{M_0}\tanh\left(\sqrt{M_0}x\right).$$

For $x - y = r > 0$, we have

$$\frac{U'(x) - U'(y)}{x - y} = \alpha_0 - \frac{\sqrt{M_0}}{r}\left[\tanh\left(\sqrt{M_0}x\right) - \tanh\left(\sqrt{M_0}y\right)\right].$$

For fixed r, the difference of the two hyperbolic tangents is maximal when the interval $[y, x]$ is centered at zero. Hence, the weak convexity profile corresponding to $p_0^{(1)}$ is given by

$$\kappa_U(r) = \alpha_0 - \frac{2\sqrt{M_0}}{r}\tanh\left(\frac{\sqrt{M_0}r}{2}\right).$$

Further, we have

$$\nabla(-\log p_0^*(x)) = \left(U'(x_1), (d-1)\alpha_0 x_2, \ldots, (d-1)\alpha_0 x_d\right)^T,$$

and thus (with $s = |x_1 - y_1| \leq \|x - y\| = r$)

$$\begin{aligned}\kappa_{-\log p_0^*}(r) &= \inf_{0\leq s\leq r} \frac{s^2}{r^2}\kappa_U(s) + (d-1)\alpha_0\left(1 - \frac{s^2}{r^2}\right) \\ &= \alpha_0(d-1) - \sup_{0\leq s\leq r} \frac{s^2}{r^2}(d-2)\alpha_0 - \frac{2\sqrt{M_0}s}{r^2}\tanh\left(\frac{\sqrt{M_0}s}{2}\right) \\ &= \alpha_0 - \frac{2\sqrt{M_0}}{r}\tanh\left(\frac{\sqrt{M_0}r}{2}\right).\end{aligned}$$

Next, we show that p_0^* attains the lower bound in Proposition 7. Because p_0^* has independent components, so do p_t and $\mathcal{L}(Y_t)$, for every $t > 0$. Consequently, it holds that

$$\mathcal{W}_2^2(p_0^*, \mathcal{L}(Y_T)) = \sum_{i=1}^d \mathcal{W}_2^2\big(p_{0,i}^*, \mathcal{L}(Y_{T,i})\big) \geq \mathcal{W}_2^2\Big(p_0^{(1)}, \mathcal{L}(Y_{T,1})\Big),$$

so that any lower bound for the error in the first component also applies to the overall error and we can reduce our analysis to the first dimension. Let F_t denote the (one-dimensional) distribution function of the first component of p_t, Q_{Y_T} denote the quantile function of $Y_{T,1}$, and Φ, φ be the distribution and density function of a standard normal distribution. The one-dimensional quantile representation of $\mathcal{W}_2$ yields

$$\mathcal{W}_2^2\Big(\mathcal{L}(Y_{T,1}), p_0^{(1)}\Big) = \int_0^1 \big(Q_{Y_T}(u) - F_0^{-1}(u)\big)^2 \,\mathrm{d}u\,.$$

We will derive Q_{Y_T} next. Note that in one dimension, the probability flow map Ψ_T from time T back to time 0 is increasing and transports $p_{T,1}$ exactly to $p_0^{(1)}$. Accordingly, if $X \sim p_{T,1}$, then $\Psi_T(X) \sim p_0^{(1)}$. It follows that

$$F_0(\Psi_T(x)) = \mathbb{P}(\Psi_T(X) \leq \Psi_T(x)) = \mathbb{P}(X \leq x) = F_T(x)$$

and thus $\Psi_T = F_0^{-1} \circ F_T$. Now, we want to consider the perturbed initialization $Y_{0,1} \sim \hat{p}_{T,1} = \mathcal{N}(0, s_T)$ with $s_T = \sqrt{1 - e^{-\mathcal{B}(T)}}$, $\mathcal{B}(T) \coloneqq \int_0^T \beta(t)\,\mathrm{d}t$, and its image $Y_{T,1} = \Psi_T(Y_{0,1})$. For the distribution function of $Y_{T,1}$, we get

$$F_{Y_T}(x) = \mathbb{P}(\Psi_T(Y_{0,1}) \leq x) = \mathbb{P}(Y_{0,1} \leq F_T^{-1}F_0(x)) = \Phi(s_T^{-1}F_T^{-1}(F_0(x)))\,,$$

and hence

$$Q_{Y_T}(u) = F_0^{-1}(F_T(s_T\Phi^{-1}(u)))\,.$$

Defining

$$\mathcal{T}(z) \coloneqq F_0^{-1}(\Phi(z))$$

and applying a change of variable $u = \Phi(z)$, it follows that

$$\begin{aligned}\mathcal{W}_2^2\Big(\mathcal{L}(Y_{T,1}), p_0^{(1)}\Big) &= \int_0^1 \big(F_0^{-1}(F_T(s_T\Phi^{-1}(u))) - F_0^{-1}(u)\big)^2 \,\mathrm{d}u \\ &= \int_{\mathbb{R}} (\mathcal{T}(\Phi^{-1}(F_T(s_T x))) - \mathcal{T}(z))^2 \varphi(z)\,\mathrm{d}z\,.\end{aligned}$$

Using this representation and the Dominated Convergence Theorem, we next want to show that

$$\lim_{T\to\infty} e^{\mathcal{B}(T)}\mathcal{W}_2^2\Big(\mathcal{L}(Y_{T,1}), p_0^{(1)}\Big) = \frac{M_0}{\alpha_0^2}A^2 \tag{49}$$

where

$$A^2 \coloneqq \int_{\mathbb{R}} \big(\mathcal{T}'(z)\big)^2 \varphi(z)\,\mathrm{d}z\,.$$

For that, we define

$$\varepsilon \coloneqq e^{-\frac{1}{2}\mathcal{B}(T)}, \quad H_\varepsilon(z) \coloneqq F_T(s_T z), \quad k_\varepsilon(z) \coloneqq \Phi^{-1}(H_\varepsilon(z))$$

and write

$$e^{\mathcal{B}(T)}\mathcal{W}_2^2\Big(\mathcal{L}(Y_{T,1}), p_0^{(1)}\Big) = \int_{\mathbb{R}} \left(\frac{\mathcal{T}(k_\varepsilon(z)) - \mathcal{T}(z)}{\varepsilon}\right)^2 \varphi(z)\,\mathrm{d}z\,.$$

We first establish the pointwise convergence of the integrand. Since the solution to the VP SDE can be reprented as

$$X_T = e^{-\frac{1}{2}\mathcal{B}(T)}X_0 + \sqrt{1 - e^{-\mathcal{B}(T)}}Z = \varepsilon X_0 + s_T Z, \quad Z \sim \mathcal{N}(0, I_d)$$

we have

$$\frac{X_{T,1}}{s_T} = \frac{\varepsilon}{s_T}X_{0,1} + Z_1 \sim \frac{1}{2}\mathcal{N}\big(0, r_\varepsilon^2\big) + \frac{1}{2}\mathcal{N}\left(\frac{2\sqrt{M_0}}{\alpha_0}t_\varepsilon, r_\varepsilon^2\right),$$

where

$$t_\varepsilon \coloneqq \frac{\varepsilon}{s_T}, \quad r_\varepsilon \coloneqq \sqrt{1 + \frac{t_\varepsilon^2}{\alpha_0}}\,.$$

Therefore,

$$H_\varepsilon(z) = \frac{1}{2}\Phi\left(\frac{z}{r_\varepsilon}\right) + \frac{1}{2}\Phi\left(\frac{z - \frac{2\sqrt{M_0}}{\alpha_0}t_\varepsilon}{r_\varepsilon}\right).$$

It follows that $H_0(z) = \Phi(z)$, $k_0(z) = z$, and hence

$$\frac{\mathcal{T}(k_\varepsilon(z)) - \mathcal{T}(z)}{\varepsilon} \to \left.\frac{\partial \mathcal{T}(k_\varepsilon(z))}{\partial \varepsilon}\right|_{\varepsilon=0} = \left.\frac{\partial k_\varepsilon(z)}{\partial \varepsilon}\right|_{\varepsilon=0} \cdot \mathcal{T}'(z)\,.$$

The inverse function theorem yields

$$\frac{\partial k_\varepsilon(z)}{\partial \varepsilon} = \big(\Phi^{-1}\big)'(H_\varepsilon(z)) \cdot \frac{\partial \mathcal{H}_\varepsilon(z)}{\partial \varepsilon} = \frac{1}{\varphi(\Phi^{-1}(H_\varepsilon(z)))} \cdot \frac{\partial \mathcal{H}_\varepsilon(z)}{\partial \varepsilon}\,.$$

Since, at $\varepsilon = 0$, it holds that

$$t_0 = 0, \quad r_0 = 1, \quad \left.\frac{\mathrm{d}t_\varepsilon}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 1, \quad \left.\frac{\mathrm{d}r_\varepsilon}{\mathrm{d}}\right|_{\varepsilon=0} = 0\,,$$

we have

$$\frac{\partial \mathcal{H}_\varepsilon(z)}{\partial \varepsilon} = -\frac{\sqrt{M_0}}{\alpha_0}\varphi(z)\,.$$

Altogether, this yields

$$\frac{\mathcal{T}(k_\varepsilon(z)) - \mathcal{T}(z)}{\varepsilon} \longrightarrow -\frac{\sqrt{M_0}}{\alpha_0}\mathcal{T}'(z) \quad \text{as } \varepsilon \to 0\,.$$

We next establish an integrable dominating function justifying the transition from pointwise convergence to that of the integral. For that, first note that the mean value theorem gives

$$\left|\frac{\mathcal{T}(k_\varepsilon(z)) - \mathcal{T}(z)}{\varepsilon}\right| = \left|\frac{\mathcal{T}(k_\varepsilon(z)) - \mathcal{T}(z)}{k_\varepsilon(z) - z}\right| \cdot \frac{|k_\varepsilon(z) - z|}{\varepsilon} \le \sup_{x\in\mathbb{R}}|\mathcal{T}'(x)| \cdot \frac{|k_\varepsilon(z) - z|}{\varepsilon}\,.$$

We will first bound the second factor and then the derivative of $\mathcal{T}$. From the definition of $H_\varepsilon(z)$ and the monotonicity of Φ, it follows that

$$\Phi\left(\frac{z - \frac{2\sqrt{M_0}}{\alpha_0}t_\varepsilon}{r_\varepsilon}\right) \le H_\varepsilon(z) \le \Phi\left(\frac{z}{r_\varepsilon}\right).$$

Applying the increasing function Φ^{-1} gives

$$\frac{z - \frac{2\sqrt{M_0}}{\alpha_0}t_\varepsilon}{r_\varepsilon} \le k_\varepsilon(z) \le \frac{z}{r_\varepsilon}\,,$$

and consequently

$$|k_\varepsilon(z) - z| \le \left(1 - \frac{1}{r_\varepsilon}\right)|z| + \frac{2\sqrt{M_0}t_\varepsilon}{\alpha_0 r_\varepsilon}\,.$$

We can further bound this by noting that

$$1 - \frac{1}{r_\varepsilon} = \frac{r_\varepsilon^2 - 1}{r_\varepsilon(r_\varepsilon + 1)} = \frac{t_\varepsilon^2}{\alpha_0 r_\varepsilon(r_\varepsilon + 1)} \le \frac{t_\varepsilon^2}{2\alpha_0},$$

and, for $0 < \varepsilon \le 1/2$,

$$\frac{t_\varepsilon}{\varepsilon} = \frac{1}{\sqrt{1-\varepsilon^2}} \le \frac{2}{\sqrt{3}}, \qquad \frac{t_\varepsilon^2}{\varepsilon} = \frac{\varepsilon}{(1-\varepsilon^2)} \le \frac{2}{3}.$$

It follows that

$$\frac{|k_\varepsilon(z) - z|}{\varepsilon} \le \frac{1}{3\alpha_0}|z| + \frac{4\sqrt{M_0}}{\sqrt{3}\,\alpha_0} \le \tilde{C}(\alpha_0, M_0)(1 + |z|)$$

where $\tilde{C}(\alpha_0, M_0) < \infty$ is independent of ε.

It remains to control the derivative

$$\mathcal{T}'(z) = \frac{\varphi(z)}{p_0^{(1)}(F_0^{-1}(\Phi(z)))} = \frac{\varphi(z)}{p_0^{(1)}(\mathcal{T}(z))}.$$

Since $\mathcal{T}'$ is continuous, it is bounded on $\mathbb{R}$ for every fixed α_0 and M_0, so that we only need to consider its limiting behavior. For $z \to -\infty$, we have $\mathcal{T}(z) \to -\infty$. Further, on the left tail, the Gaussian mixture is dominated by the first component, that is

$$p_0^{(1)}(x) \approx \frac{\sqrt{\alpha_0}}{2}\varphi(\sqrt{\alpha_0}x) \quad \text{and} \quad F_0(x) \approx \frac{1}{2}\Phi(\sqrt{\alpha_0}x) \qquad \text{for } x \to -\infty.$$

Consequently,

$$\mathcal{T}'(z) \approx \frac{\varphi(z)}{\frac{\sqrt{\alpha_0}}{2}\varphi\big(\sqrt{\alpha_0}\mathcal{T}(z)\big)} \quad \text{and} \quad \Phi(z) = F_0(\mathcal{T}(z)) \approx \frac{1}{2}\Phi(\sqrt{\alpha_0}\mathcal{T}(z)) \qquad \text{for } z \to -\infty.$$

By an application of Mills' ratio, that is $\Phi(t) \approx \varphi(t)/(-t)$ as $t \to -\infty$, it follows that

$$\mathcal{T}'(z) \approx \frac{z}{\alpha_0 \mathcal{T}(z)}.$$

Since

$$\frac{\Phi(z)}{\Phi(\sqrt{\alpha_0}\mathcal{T}(z))} \longrightarrow 1/2 \quad \text{as } z \to -\infty$$

implies that

$$\frac{z}{\sqrt{\alpha_0}\mathcal{T}(z)} \longrightarrow 1 \quad \text{as } z \to -\infty,$$

we have

$$\mathcal{T}'(z) \longrightarrow \frac{1}{\sqrt{\alpha_0}} \quad \text{as } z \to -\infty.$$

Similar arguments lead to the same limit for $z \to \infty$. Hence,

$$\tilde{L}(\alpha_0, M_0) \coloneqq \sup_{z \in \mathbb{R}} |\mathcal{T}'(z)| < \infty.$$

Altogether, this yields

$$\left|\frac{\mathcal{T}(k_\varepsilon(z)) - \mathcal{T}(z)}{\varepsilon}\right|^2 \varphi(z) \le \tilde{L}^2(\alpha_0, M_0)\tilde{C}^2(\alpha_0, M_0)(1 + |z|)^2\phi(z).$$

The right-hand side is integrable over $\mathbb{R}$ and independent of ε, justifying the application of the Dominated Convergence Theorem to conclude that (49) holds.

Next, we establish a lower bound for A. A change of variable using $z(x) = \Phi^{-1}(F_0(x))$ yields

$$A^2 = \int_{\mathbb{R}} \left(\frac{\varphi(z)}{p_0^{(1)}(F_0^{-1}(\Phi(z)))}\right)^2 \varphi(z)\,\mathrm{d}z = \int_{\mathbb{R}} \frac{\big(\varphi(\Phi^{-1}(F_0(x)))\big)^2}{p_0^{(1)}(x)}\,\mathrm{d}x.$$

Consider the interval

$$I \coloneqq \left[a - \frac{1}{\sqrt{M_0}}, a + \frac{1}{\sqrt{M_0}}\right] \quad \text{with} \quad a \coloneqq \frac{\sqrt{M_0}}{\alpha_0},$$

and note that for $x = a + r$ simple algebraic calculations yield

$$p_0^{(1)}(a+r) = \sqrt{\frac{\alpha_0}{2\pi}} \exp\left(-\frac{\alpha_0(a^2+r^2)}{2}\right) \cosh\left(\sqrt{M_0} r\right),$$

such that for $x \in I$, we have

$$p_0^{(1)}(x) \leq \cosh(1) p_0^{(1)}(a) = \cosh(1)\sqrt{\frac{\alpha_0}{2\pi}} \exp\left(-\frac{M_0}{2\alpha_0}\right).$$

The symmetry of the Gaussian mixture around a implies that $F_0(a) = 1/2$. Moreover, for $x \in I$ and $M_0 \geq \alpha_0$, we get

$$\left|F_0(x) - \frac{1}{2}\right| \leq \int_a^{a+\frac{1}{\sqrt{M_0}}} p_0^{(1)}(y)\,\mathrm{d}y \leq \frac{\cosh(1)}{\sqrt{2\pi}}\sqrt{\frac{\alpha_0}{M_0}} \exp\left(-\frac{M_0}{2\alpha_0}\right) \leq \frac{\cosh(1)}{\sqrt{2\pi}} \exp(-1) \leq \frac{1}{4}.$$

It follows that $\Phi^{-1}(F_0(x))$ remains in a fixed compact interval for $x \in I$. Hence, there exists a universal constant $c_\varphi > 0$ such that $\varphi(\Phi^{-1}(F_0(x))) \geq c_\varphi$ for all $x \in I$. Altogether, this yields

$$A^2 \geq \int_I \frac{(\varphi(\Phi^{-1}(F_0(x))))^2}{p_0^{(1)}(x)}\,\mathrm{d}x \geq \frac{2\sqrt{2\pi}c_\varphi^2}{\cosh(1)} \frac{1}{\sqrt{\alpha_0 M_0}} \exp\left(\frac{M_0}{2\alpha_0}\right),$$

and thus

$$\begin{aligned}
\liminf_{T\to\infty} e^{\frac{1}{2}\mathcal{B}(T)} \mathcal{W}_2(\mathcal{L}(Y_T), p_0^*) &\geq \liminf_{T\to\infty} e^{\frac{1}{2}\mathcal{B}(T)} \mathcal{W}_2\Big(\mathcal{L}(Y_{T,1}), p_0^{(1)}\Big) \\
&= \lim_{T\to\infty} e^{\frac{1}{2}\mathcal{B}(T)} \mathcal{W}_2\Big(\mathcal{L}(Y_{T,1}), p_0^{(1)}\Big) = a \cdot A \\
&\geq \cdot \frac{a\tilde{c}}{(\alpha_0 M_0)^{\frac{1}{4}}} \exp\left(\frac{M_0}{4\alpha_0}\right)
\end{aligned}$$

for a universal constant $\tilde{c} > 0$.

Finally, from the definition of p_0^*, it can be easily seen that

$$\mathbb{E}\big[X_0^2\big] = \mathbb{E}\big[X_{0,1}^2\big] + (d-1)\mathbb{E}\big[X_{0,2}^2\big] = \frac{2}{\alpha_0} + \frac{2M_0}{\alpha_0^2},$$

and consequently, for $M_0 \geq \alpha_0$,

$$\frac{a}{\|X_0\|_{L_2}} = \frac{1}{\sqrt{\frac{2\alpha_0}{M_0} + 2}} \geq \frac{1}{\sqrt{4}}.$$

In summary, it follows that

$$\liminf_{T\to\infty} \sup_{p_0 \in \mathcal{P}(\alpha_0, M_0)} \frac{\mathcal{W}_2(\mathcal{L}(Y_T), p_0)}{e^{-\frac{1}{2}\mathcal{B}(T)}\|X_0\|_{L_2}} \geq \liminf_{T\to\infty} \frac{\mathcal{W}_2(\mathcal{L}(Y_T), p_0^*)}{e^{-\frac{1}{2}\mathcal{B}(T)}\|X_0\|_{L_2}} \geq \frac{c}{(\alpha_0 M_0)^{\frac{1}{4}}} \exp\left(\frac{M_0}{4\alpha_0}\right)$$

with $c = \tilde{c}/\sqrt{4}$. □

Proposition 7 implies that any estimate of the form

$$\mathcal{W}_2(\mathcal{L}(Y_T), p_0) \leq C_{\mathrm{init}}(\alpha_0, M_0) e^{-\frac{1}{2}\int_0^T \beta(t)\,\mathrm{d}t} \|X_0\|_{L_2},$$

which is to hold uniformly over (α_0, M_0)-weakly log-concave targets p_0 for all sufficiently large T, necessarily satisfies

$$C_{\mathrm{init}}(\alpha_0, M_0) \geq \frac{c}{(\alpha_0 M_0)^{\frac{1}{4}}} \exp\left(\frac{M_0}{4\alpha_0}\right).$$

E PROOF OF THE MAIN RESULT

The proof of Theorem 5 relies on two propositions that are listed in the following and control the initialization error and the combined discretization and propagated score-matching error, respectively. Their proofs are given in the following subsections. The first one is a generalization of Gao & Zhu (2025, Proposition 4.1) to our setting. It establishes a control on the initialization error caused by replacing the unknown $\tilde{X}_0 \sim p_T$ by $Y_0 \sim \hat{p}_T$ in the reverse flow.

Proposition 20 (Initialization error)**.** *Under Assumption 1,*

$$\mathcal{W}_2(\mathcal{L}(Y_T), p_0) \leq C(\alpha_0, M_0) e^{-\frac{1}{2}\int_0^T g^2(t)|\alpha(t)-M(t)|\,\mathrm{d}t} \|X_0\|_{L_2}.$$

where $C(\alpha_0, M_0)$ is defined in (14)*.*

The quantity $C(\alpha_0, M_0)$ measures the increased cost caused by the lack of regularity of p_0. If p_0 is strongly log-concave, then $C(\alpha_0, M_0) = 1$, as $\tau(\alpha_0, M_0) = 0$. Note that the initialization error will decrease exponentially in T no matter whether p_0 is strongly or weakly log-concave. Next, we consider the discretization and propagated score-matching error. The following result is a generalization of Gao & Zhu (2025, Proposition 4.2).

Proposition 21 (Discretization and propagated score matching error)**.** *Under Assumptions 1, 2, and 3, it holds for any $k \in \{1, \dots, K\}$ that*

$$\begin{aligned}\left\|Y_{t_k} - \widehat{Z}_{t_k}\right\|_{L_2} \leq \Bigg(1 - \int_{t_{k-1}}^{t_k} \delta_k(T-t)\,\mathrm{d}t + \frac{1}{2}L_1 h \int_{t_{k-1}}^{t_k} g^2(T-t)\,\mathrm{d}t\Bigg) \\ \cdot e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\left\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\right\|_{L_2} \\ + \frac{1}{2}L_1 h(1+\theta(T)+\omega(T))\phi_{k,h} + \frac{1}{2}\mathcal{E}\phi_{k,h} + \frac{1}{2}\sqrt{h}\nu_{k,h}\sqrt{\psi_{k,h}},\end{aligned}$$

where $\delta_k(T-t)$, $\theta(T)$, $\omega(T)$, $\nu_{k,h}$, $\phi_{k,h}$, and $\psi_{k,h}$ are defined in (20)–(25) *in Appendix A.*

Now, we are ready to prove Theorem 5.

Proof of Theorem 5. By the triangle inequality for the 2-Wasserstein distance, we have

$$\mathcal{W}_2\Big(\mathcal{L}\Big(\widehat{Z}_T\Big), p_0\Big) \leq \mathcal{W}_2\Big(\mathcal{L}\Big(\widehat{Z}_T\Big), \mathcal{L}(Y_T)\Big) + \mathcal{W}_2(\mathcal{L}(Y_T), p_0). \tag{50}$$

To establish a bound for the first term, we will use Proposition 21. To simplify notation, define

$$\beta_{k,h} \coloneqq \frac{1}{2}L_1 h(1+\theta(T)+\omega(T))\phi_{k,h} + \frac{1}{2}\mathcal{E}\phi_{k,h} + \frac{1}{2}\sqrt{h}\nu_{k,h}\sqrt{\psi_{k,h}},$$

and recall the definition of $\gamma_{k,h}$ from (19). Then, Proposition 21 states that for $k \in \{1, \dots, K\}$

$$\left\|Y_{t_k} - \widehat{Z}_{t_k}\right\|_{L_2} \leq \gamma_{k,h} e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\left\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\right\|_{L_2} + \beta_{k,h}. \tag{51}$$

If we pick a coupling between Y_t and $\widehat{Z}_t$ such that $Y_0 = \widehat{Z}_0$ a.s., then by recalling that $T = t_K$ and applying (51) recursively, we get

$$\mathcal{W}_2\Big(\mathcal{L}\Big(\widehat{Z}_T\Big), \mathcal{L}(Y_T)\Big) \leq \left\|\widehat{Z}_{t_K} - Y_{t_K}\right\|_{L_2} = \sum_{k=1}^{K}\left(\prod_{j=k+1}^{K}\gamma_{j,h}\right) e^{\int_{t_k}^{T} f(T-t)\,\mathrm{d}t}\beta_{k,h}.$$

Together with Proposition 20, bounding the second term in (50), it follows that

$$\begin{aligned}\mathcal{W}_2\Big(\mathcal{L}\Big(\widehat{Z}_T\Big), p_0\Big) \leq C(\alpha_0, M_0) e^{-\int_0^T g^2(t)|\alpha(t)-M(t)|\,\mathrm{d}t}\|X_0\|_{L_2} \\ + \sum_{k=1}^{K}\left(\prod_{j=k+1}^{K}\gamma_{j,h}\right) e^{\int_{t_k}^{T} f(T-t)\,\mathrm{d}t}\beta_{k,h}.\end{aligned}$$

The definitions of E_0, E_1, and E_2 complete the proof. □

E.1 Proof of Proposition 20

We start by analyzing the initialization error. Recall the following result from Gao & Zhu (2025, Lemma B.3).

Lemma 22. *It holds that*

$$\mathcal{W}_2\big(p_T, \hat{p}_T\big) \;\le\; e^{-\int_0^T f(s)\,\mathrm{d}s}\,\|X_0\|_{L_2}.$$

Proof of Proposition 20. The result is a consequence of the propagation over time of the weak log-concavity, combined with the regime change results from Section 3. We start by following the steps in Gao & Zhu (2025, Proposition 4.1). Let

$$m(t) = -2f(t) + g^2(t)(\alpha(t) - M(t)).$$

By computing the derivative, using (6) and (7), and by Proposition 2, we get

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\tilde{X}_t - Y_t\|^2\, e^{\int_0^t m(T-s)\,\mathrm{d}s}\right)\\
&\quad= m(T-t)\, e^{\int_0^t m(T-s)\,\mathrm{d}s}\,\|\tilde{X}_t - Y_t\|^2 \;+ 2\, e^{\int_0^t m(T-s)\,\mathrm{d}s}\left\langle \tilde{X}_t - Y_t,\; \frac{\mathrm{d}}{\mathrm{d}t}\tilde{X}_t - \frac{\mathrm{d}}{\mathrm{d}t}Y_t\right\rangle\\
&\quad= m(T-t)\, e^{\int_0^t m(T-s)\,\mathrm{d}s}\,\|\tilde{X}_t - Y_t\|^2 + 2\, e^{\int_0^t m(T-s)\,\mathrm{d}s}\left\langle \tilde{X}_t - Y_t,\; f(T-t)\big(\tilde{X}_t - Y_t\big)\right\rangle\\
&\qquad+ 2\, e^{\int_0^t m(T-s)\,\mathrm{d}s}\left\langle \tilde{X}_t - Y_t,\; \tfrac{1}{2}g^2(T-t)\big(\nabla \log p_{T-t}(\tilde{X}_t) - \nabla \log p_{T-t}(Y_t)\big)\right\rangle\\
&\quad\le e^{\int_0^t m(T-s)\,\mathrm{d}s}\,\|\tilde{X}_t - Y_t\|^2\left[m(T-t) + 2f(T-t) - g^2(T-t)(\alpha(T-t) - M(T-t))\right]\\
&\quad= 0.
\end{aligned}$$

Hence, for any $t \in [0,T]$,

$$\|\tilde{X}_t - Y_t\|^2 e^{\int_0^t m(T-s)\,ds} \le \|\tilde{X}_0 - Y_0\|^2, \tag{52}$$

so that

$$\mathbb{E}\|\tilde{X}_T - Y_T\|^2 \le e^{-\int_0^T m(T-s)\,ds}\mathbb{E}\|\tilde{X}_0 - Y_0\|^2.$$

Next, consider a coupling of $(\tilde{X}_0, Y_0)$ such that $\tilde{X}_0 \sim p_T$, $Y_0 \sim \hat{p}_T$, and $\mathbb{E}\|\tilde{X}_0 - Y_0\|^2 = \mathcal{W}_2^2(p_T, \hat{p}_T)$. By combining the previous result with Lemma 22 and by the definition of the Wasserstein distance, we have

$$\begin{aligned}
\mathcal{W}_2^2(\mathcal{L}(Y_T), p_0) = W_2^2(\mathcal{L}(Y_T), \mathcal{L}(\tilde{X}_T)) &\le \mathbb{E}\|\tilde{X}_T - Y_T\|^2 \le e^{-\int_0^T m(T-s)\;\mathrm{d}s}\mathcal{W}_2^2(p_T, \hat{p}_T)\\
&\le e^{-\int_0^T m(s)\,\mathrm{d}s} e^{-2\int_0^T f(s)\,\mathrm{d}s}\|X_0\|_{L_2}^2 = e^{-\int_0^T g^2(t)(\alpha(t)-M(t))\,\mathrm{d}t}\|X_0\|_{L_2}^2.
\end{aligned} \tag{53}$$

By the regime shift result from Proposition 3 and Lemma 15, we get

$$\begin{aligned}
&\exp\left(-\int_0^T g^2(t)(\alpha(t) - M(t))\,\mathrm{d}t\right)\\
&\quad= \exp\left(-\int_0^T g^2(t)|\alpha(t) - M(t)|\,\mathrm{d}t + 2\int_0^{\tau(\alpha_0, M_0)\wedge T} g^2(t)|\alpha(t) - M(t)|\,\mathrm{d}t\right)\\
&\quad\le \exp\left(-\int_0^T g^2(t)|\alpha(t) - M(t)|\,\mathrm{d}t + 2\sup_{0\le t\le \tau(\alpha_0,M_0)}|\alpha(t) - M(t)|\int_0^{\tau(\alpha_0,M_0)} g^2(t)\,\mathrm{d}t\right)\\
&\quad= \exp\left(-\int_0^T g^2(t)|\alpha(t) - M(t)|\,\mathrm{d}t - 2\inf_{0\le t\le \tau(\alpha_0,M_0)} K(t)\int_0^{\tau(\alpha_0,M_0)} g^2(t)\,\mathrm{d}t\right)\\
&\quad\le \exp\left(-\int_0^T g^2(t)|\alpha(t) - M(t)|\,\mathrm{d}t + 2\frac{|\alpha_0 - M_0|}{\alpha_0^2 \wedge 1}\xi(\tau(\alpha_0, M_0))\int_0^{\tau(\alpha_0,M_0)} g^2(t)\,\mathrm{d}t\right)\\
&\quad= C^2(\alpha_0, M_0)\exp\left(-\int_0^T g^2(t)|\alpha(t) - M(t)|\,\mathrm{d}t\right).
\end{aligned}$$

Together with (53), it follows that

$$\mathcal{W}_2^2(\mathcal{L}(Y_T), p_0) \leq C^2(\alpha_0, M_0) \exp\left(-\int_0^T g^2(t)|\alpha(t) - M(t)| \,\mathrm{d}t\right) \|X_0\|_{L_2}^2.$$

We note that the quantity $C(\alpha_0, M_0)$ is always finite for any positive α_0 and M_0, since g is continuous and $\tau(\alpha_0, M_0)$ is finite. □

E.2 Proof of Proposition 21

Next, we examine the discretization and propagated score-matching error. For that, we need two technical lemmas.

Lemma 23. *With $\omega(T)$ defined in* (22)*, it holds that*

$$\sup_{0 \leq t \leq T} \|X_t\|_{L_2} = \omega(T).$$

Proof of Lemma 23. Using the explicit formula for X_t given in (2) and the distribution of the stochastic integral therein as well as its independence of X_0, we get

$$\begin{aligned}
\|X_t\|_{L_2}^2 &= \mathbb{E}\left[\left\| e^{-\int_0^t f(s)\,\mathrm{d}s} X_0 + \int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s \right\|^2\right] \\
&= \mathbb{E}\left[\left\| e^{-\int_0^t f(s)\,\mathrm{d}s} X_0 \right\|^2\right] + \mathbb{E}\left[\left\| \int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s \right\|^2\right] \\
&= e^{-2\int_0^t f(s)\,\mathrm{d}s} \|X_0\|_{L_2}^2 + \mathrm{tr}\left(\mathrm{Var}\left(\int_0^t e^{-\int_s^t f(v)\,\mathrm{d}v} g(s)\,\mathrm{d}B_s\right)\right) \\
&= e^{-2\int_0^t f(s)\,\mathrm{d}s} \|X_0\|_{L_2}^2 + d \cdot \int_0^t e^{-2\int_s^t f(v)\,\mathrm{d}v} g^2(s)\,\mathrm{d}s.
\end{aligned}$$

□

Lemma 24. *With $\nu_{k,h}$ defined in* (23)*, it holds for any $k \in \{1, \ldots, K\}$ that*

$$\sup_{t_{k-1} \leq t \leq t_k} \left\|Y_t - Y_{t_{k-1}}\right\|_{L_2} \leq \nu_{k,h}.$$

Proof of Lemma 24. From (6) and (7), it follows that for any $t \in [t_{k-1}, t_k]$

$$\begin{aligned}
\tilde{X}_t &= \tilde{X}_{t_{k-1}} + \int_{t_{k-1}}^t \left[f(T-s)\tilde{X}_s + \frac{1}{2} g^2(T-s) \nabla \log p_{T-s}(\tilde{X}_s)\right] \mathrm{d}s \qquad (54) \\
Y_t &= Y_{t_{k-1}} + \int_{t_{k-1}}^t \left[f(T-s)Y_s + \frac{1}{2} g^2(T-s) \nabla \log p_{T-s}(Y_s)\right] \mathrm{d}s, ,
\end{aligned}$$

so that, by an application of the triangle inequality,

$$\begin{aligned}
\left\|Y_t - Y_{t_{k-1}}\right\|_{L_2} \leq{}& \left\|\tilde{X}_t - \tilde{X}_{t_{k-1}}\right\|_{L_2} + \int_{t_{k-1}}^t f(T-s) \left\|Y_s - \tilde{X}_s\right\|_{L_2} \mathrm{d}s \\
&+ \int_{t_{k-1}}^t \frac{1}{2} g^2(T-s) \left\|\nabla \log p_{T-s}(Y_s) - \nabla \log p_{T-s}(\tilde{X}_s)\right\|_{L_2} \mathrm{d}s.
\end{aligned}$$

An application of Proposition 16 further yields

$$\begin{aligned}
\left\|Y_t - Y_{t_{k-1}}\right\|_{L_2} \leq{}& \left\|\tilde{X}_t - \tilde{X}_{t_{k-1}}\right\|_{L_2} \\
&+ \int_{t_{k-1}}^t \left[f(T-s) + \frac{1}{2} g^2(T-s) L(T-s)\right] \left\|Y_s - \tilde{X}_s\right\|_{L_2} \mathrm{d}s.
\end{aligned}$$

From the proof of Proposition 20, specifically (52) and the lines thereafter, we have

$$\left\|Y_t - \tilde{X}_t\right\|_{L_2} \le e^{-\frac{1}{2}\int_0^t m(T-s)\,\mathrm{d}s}\left\|Y_0 - \tilde{X}_0\right\|_{L_2} \le e^{-\frac{1}{2}\int_0^t m(T-s)\,\mathrm{d}s} e^{-\int_0^T f(s)\,\mathrm{d}s}\|X_0\|_{L_2} \le \theta(T), \tag{55}$$

and therefore

$$\left\|Y_t - Y_{t_{k-1}}\right\|_{L_2} \le \left\|X_t - X_{t_{k-1}}\right\|_{L_2} + \theta(T)\int_{t_{k-1}}^t \left[f(T-s) + \frac{1}{2}g^2(T-s)L(T-s)\right]\mathrm{d}s. \tag{56}$$

Next, (54) implies that

$$\left\|\tilde{X}_t - \tilde{X}_{t_{k-1}}\right\|_{L_2} \le \int_{t_{k-1}}^t \left[f(T-s)\left\|\tilde{X}_s\right\|_{L_2} + \frac{1}{2}g^2(T-s)\left\|\nabla \log p_{T-s}(\tilde{X}_s)\right\|_{L_2}\right]\mathrm{d}s.$$

Another application of Proposition 16 and the fact that $p_{T-s}(\mathbf{0})$ is determistic yields

$$\begin{aligned}\left\|\nabla \log p_{T-s}(\tilde{X}_s)\right\|_{L_2} &\le \left\|\nabla \log p_{T-s}(\tilde{X}_s) - \nabla \log p_{T-s}(\mathbf{0})\right\|_{L_2} + \|\nabla \log p_{T-s}(\mathbf{0})\|_{L_2} \\ &\le L(T-s)\left\|\tilde{X}_s\right\|_{L_2} + \|\nabla \log p_{T-s}(\mathbf{0})\|\end{aligned}$$

Moreover, since $s \in [t_{k-1}, t_k]$ and $t_K = Kh = T$, it follows from Assumption 2 that

$$\begin{aligned}\|\nabla \log p_{T-s}(\mathbf{0})\| &\le \left\|\nabla \log p_{T-s}(\mathbf{0}) - \nabla \log p_{T-t_{k-1}}(\mathbf{0})\right\| \\ &\quad + \sum_{j=k}^K \left\|\nabla \log p_{T-t_j}(\mathbf{0}) - \nabla \log p_{T-t_{j-1}}(\mathbf{0})\right\| + \|\nabla \log p_0(\mathbf{0})\| \\ &\le (K-k+2)L_1 h + \|\nabla \log p_0(\mathbf{0})\| \\ &\le (K+1)L_1 h + \|\nabla \log p_0(\mathbf{0})\| \\ &\le (T+h)L_1 + \|\nabla \log p_0(\mathbf{0})\|. \end{aligned} \tag{57}$$

In summary, we conclude that

$$\begin{aligned}\left\|\tilde{X}_t - \tilde{X}_{t_{k-1}}\right\|_{L_2} &\le \int_{t_{k-1}}^t \left[f(T-s) + \frac{1}{2}g^2(T-s)L(T-s)\right]\left\|\tilde{X}_s\right\|_{L_2}\mathrm{d}s \\ &\quad + \frac{1}{2}\Big((T+h)L_1 + \|\nabla \log p_0(\mathbf{0})\|\Big)\int_{t_{k-1}}^t g^2(T-s)\,\mathrm{d}s. \end{aligned} \tag{58}$$

The final result follows from a combination of (56) and (58) together with the observation that

$$\sup_{0\le s\le T}\left\|\tilde{X}_s\right\|_{L_2} = \sup_{0\le s\le T}\|X_{T-s}\|_{L_2} = \omega(T),$$

where the first equality holds because $\tilde{X}_s = X_{T-s}$ in distribution and the second equality is verified in Lemma 23. □

Now, we are ready to prove Proposition 21.

Proof of Proposition 21. We follow the steps in Gao & Zhu (2025, Proposition 4.2). Specifically, we split the distance between Y_{t_k} and $\widehat{Z}_{t_k}$ into several parts and derive upper bounds for each one of them separately, repeatedly making use of the propagation of weak log-concavity and Lipschitz-smoothness from p_0 to p_t as established in Proposition 2 and 16.

By the definition of Y_t and $\widehat{Z}_t$ given in (7) and (9), we have for any $t \in [t_{k-1}, t_k]$

$$\begin{aligned}Y_t &= Y_{t_{k-1}} + \int_{t_{k-1}}^t \left[f(T-s)Y_s + \frac{1}{2}g^2(T-s)\nabla \log p_{T-s}(Y_s)\right]\mathrm{d}s, \\ \widehat{Z}_t &= \widehat{Z}_{t_{k-1}} + \int_{t_{k-1}}^t \left[f(T-s)\widehat{Z}_s + \frac{1}{2}g^2(T-s)s_\theta\Big(\widehat{Z}_{t_{k-1}}, T-t_{k-1}\Big)\right]\mathrm{d}s,\end{aligned}$$

which yields the solutions

$$Y_{t_k} = e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t} Y_{t_{k-1}} + \frac{1}{2}\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\nabla \log p_{T-t}(Y_t)\,\mathrm{d}t,$$
$$\widehat{Z}_{t_k} = e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t} \widehat{Z}_{t_{k-1}} + \frac{1}{2}\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) s_\theta\Big(\widehat{Z}_{t_{k-1}}, T-t_{k-1}\Big)\,\mathrm{d}t.$$

By adding and subtracting some additional terms as well as several applications of the triangle inequality, it follows that

$$\begin{aligned}
&\Big\|Y_{t_k} - \widehat{Z}_{t_k}\Big\|_{L_2} \\
&\quad\le \Bigg\| e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big(Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big) \\
&\qquad + \frac{1}{2}\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big(\nabla \log p_{T-t}(Y_{t_{k-1}}) - \nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}})\Big)\,\mathrm{d}t\Bigg\|_{L_2} \\
&\qquad + \frac{1}{2}\Bigg\|\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\big(\nabla \log p_{T-t}(Y_t) - \nabla \log p_{T-t}(Y_{t_{k-1}})\big)\,\mathrm{d}t\Bigg\|_{L_2} \\
&\qquad + \frac{1}{2}\Bigg\|\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big(\nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}}) - \nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}})\Big)\,\mathrm{d}t\Bigg\|_{L_2} \\
&\qquad + \frac{1}{2}\Bigg\|\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) \\
&\qquad\qquad \cdot \Big(\nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}}) - s_\theta\Big(\widehat{Z}_{t_{k-1}}, T-t_{k-1}\Big)\Big)\,\mathrm{d}t\Bigg\|_{L_2} \\
&\quad=: \|S_1(k,h)\|_{L_2} + \frac{1}{2}\|S_2(k,h)\|_{L_2} + \frac{1}{2}\|S_3(k,h)\|_{L_2} + \frac{1}{2}\|S_4(k,h)\|_{L_2}.
\end{aligned} \tag{59}$$

Next, we derive upper bounds for the four summands $\|S_i(k,h)\|_{L_2}$, $i \in \{1,\dots,4\}$, that appear in (59). For $S_1(k,h)$ and $S_2(k,h)$, we first derive an upper bound for the Euclidean norm and then deduct one for the L_2-norm.

For the first term, we get

$$
\begin{aligned}
&\|S_1(k,h)\|^2 \\
&\quad = e^{2\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2 \\
&\qquad + \left\|\frac{1}{2}\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big(\nabla \log p_{T-t}(Y_{t_{k-1}}) - \nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}})\Big)\,\mathrm{d}t\right\|^2 \\
&\qquad + 2\left\langle e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big(Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big),\right. \\
&\qquad\qquad \left.\frac{1}{2}\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big(\nabla \log p_{T-t}(Y_{t_{k-1}}) - \nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}})\Big)\,\mathrm{d}t\right\rangle \\
&\quad \le e^{2\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2 \\
&\qquad + \frac{1}{4}\left(\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big\|\nabla \log p_{T-t}(Y_{t_{k-1}}) - \nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}})\Big\|\,\mathrm{d}t\right)^2 \\
&\qquad + e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) \\
&\qquad\qquad \cdot \left\langle Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}, \nabla \log p_{T-t}(Y_{t_{k-1}}) - \nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}})\right\rangle\,\mathrm{d}t.
\end{aligned}
$$

From the weak concavity and Lipschitz continuity of $\nabla \log p_{T-t}$, established in Proposition 2 and 16, respectively, it follows that

$$
\begin{aligned}
\|S_1(k,h)\|^2 &\le e^{2\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2 \\
&\quad + \frac{1}{4}\left(\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)L(T-t)\,\mathrm{d}t\right)^2\Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2 \\
&\quad - e^{2\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\int_{t_{k-1}}^{t_k} e^{\int_{t_{k-1}}^{t} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big(\alpha(T-t) - M(T-t)\Big)\,\mathrm{d}t \\
&\quad \cdot \Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2.
\end{aligned}
$$

By Cauchy-Schwartz inequality, it holds that

$$
\begin{aligned}
&\left(\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)L(T-t)\,\mathrm{d}t\right)^2 \\
&\qquad \le h e^{2\int_{t_{k-1}}^{t_k} f(T-s)\,\mathrm{d}s}\cdot\int_{t_{k-1}}^{t_k} g^4(T-t)L^2(T-t)\,\mathrm{d}t,
\end{aligned}
$$

which further yields

$$
\begin{aligned}
&\|S_1(k,h)\|^2 \\
&\quad \le \left(1 - \int_{t_{k-1}}^{t_k}\left[e^{\int_{t_{k-1}}^{t} f(T-s)\,\mathrm{d}s} g^2(T-t)\Big(\alpha(T-t) - M(T-t)\Big)\right.\right. \\
&\qquad\qquad \left.\left. - \frac{1}{4}hg^4(T-t)L^2(T-t)\right]\mathrm{d}t\right)\cdot e^{2\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2 \\
&\quad = \left(1 - \int_{t_{k-1}}^{t_k} 2\delta_k(T-t)\,\mathrm{d}t\right)e^{2\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t}\Big\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\Big\|^2.
\end{aligned}
$$

Note that, since the left-hand side of this inequality is non-negative, the right-hand side is guaranteed to be non-negative as well. Hence, using the inequality $\sqrt{1-x} \leq 1 - \frac{x}{2}$, which holds for any $x \leq 1$, we conclude that

$$\|S_1(k,h)\|_{L_2} \leq \left(1 - \int_{t_{k-1}}^{t_k} \delta_k(T-t)\,\mathrm{d}t\right) e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t} \left\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\right\|_{L_2}. \tag{60}$$

By Proposition 16, we get for the second term that

$$\begin{aligned}\|S_2(k,h)\|^2 &= \left\| \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\big(\nabla \log p_{T-t}(Y_t) - \nabla \log p_{T-t}(Y_{t_{k-1}})\big)\,\mathrm{d}t \right\|^2 \\ &\leq \left(\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L(T-t) \big\|Y_t - Y_{t_{k-1}}\big\|\,\mathrm{d}t \right)^2.\end{aligned}$$

An application of Cauchy-Schwartz inequality further yields

$$\begin{aligned}\|S_2(k,h)\|^2 &\leq \int_{t_{k-1}}^{t_k} 1^2\,\mathrm{d}t \cdot \int_{t_{k-1}}^{t_k} \left(e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L(T-t) \big\|Y_t - Y_{t_{k-1}}\big\|\right)^2 \mathrm{d}t \\ &\leq h \int_{t_{k-1}}^{t_k} \left[e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L(T-t)\right]^2 \mathrm{d}t \sup_{t_{k-1} \leq t \leq t_k} \big\|Y_t - Y_{t_{k-1}}\big\|^2.\end{aligned}$$

It follows that

$$\begin{aligned}\|S_2(k,h)\|_{L_2} &\leq \sqrt{h} \left(\int_{t_{k-1}}^{t_k} \left[e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L(T-t)\right]^2 \mathrm{d}t \right)^{\frac{1}{2}} \sup_{t_{k-1} \leq t \leq t_k} \big\|Y_t - Y_{t_{k-1}}\big\|_{L_2} \\ &\leq \sqrt{h}\,\nu_{k,h} \left(\int_{t_{k-1}}^{t_k} \left[e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L(T-t)\right]^2 \mathrm{d}t \right)^{\frac{1}{2}},\end{aligned}$$

where for the last inequality, we used Lemma 24.

For the third term, Assumption 2 implies that

$$\begin{aligned}\|S_3(k,h)\|_{L_2} &\leq \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) \left\|\nabla \log p_{T-t}(\widehat{Z}_{t_{k-1}}) - \nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}})\right\|_{L_2} \mathrm{d}t \\ &\leq \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L_1 h \left(1 + \left\|\widehat{Z}_{t_{k-1}}\right\|_{L_2}\right) \mathrm{d}t \\ &\leq L_1 h \left(1 + \left\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\right\|_{L_2} + \left\|Y_{t_{k-1}} - \tilde{X}_{t_{k-1}}\right\|_{L_2} + \left\|\tilde{X}_{t_{k-1}}\right\|_{L_2}\right) \\ &\quad \cdot \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t.\end{aligned}$$

By (55), we have

$$\left\|Y_{t_{k-1}} - \tilde{X}_{t_{k-1}}\right\|_{L_2} \leq e^{-\frac{1}{2}\int_{t=0}^{t_{k-1}} m(T-s)\,\mathrm{d}s} e^{-\int_0^T f(s)ds} \left\|\tilde{X}_0\right\|_{L_2} \leq \theta(T).$$

Moreover, since $\tilde{X}_t = X_{T-t}$ in distribution for any $t \in [0,T]$, Lemma 23 implies that

$$\left\|\tilde{X}_{t_{k-1}}\right\|_{L_2} = \left\|X_{T-t_{k-1}}\right\|_{L_2} \leq \sup_{t \in [0,T]} \|X_t\|_{L_2} = \omega(T).$$

In summary, this yields

$$\|S_3(k,h)\|_{L_2} \leq L_1 h \left(1 + \theta(T) + \omega(T) + \left\|Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}}\right\|_{L_2}\right) \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t.$$

The fourth term can be easily bounded by Assumption 3. In particular, we have

$$\begin{aligned}
&\|S_4(k,h)\|_{L_2} \\
&\quad\le \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) \left\| \nabla \log p_{T-t_{k-1}}(\widehat{Z}_{t_{k-1}}) - s_\theta\Big(\widehat{Z}_{t_{k-1}}, T-t_{k-1}\Big) \right\|_{L_2} \mathrm{d}t \\
&\quad\le \mathcal{E} \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t.
\end{aligned}$$

Combining the bounds for all four summands in (59), we conclude that

$$\begin{aligned}
\left\| Y_{t_k} - \widehat{Z}_{t_k} \right\|_{L_2} \le{}& \left(1 - \int_{t_{k-1}}^{t_k} \delta_k(T-t)\,\mathrm{d}t\right) e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t} \left\| Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}} \right\|_{L_2} \\
&+ \frac{1}{2}\sqrt{h}\,\nu_{k,h} \left(\int_{t_{k-1}}^{t_k} \left[e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t) L(T-t) \right]^2 \mathrm{d}t \right)^{\frac{1}{2}} \\
&+ \frac{1}{2} L_1 h \left(1 + \theta(T) + \omega(T) + \left\| Y_{t_{k-1}} - \widehat{Z}_{t_{k-1}} \right\|_{L_2} \right) \\
&\quad \cdot \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t \\
&+ \frac{1}{2}\mathcal{E} \int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t.
\end{aligned}$$

Using the fact that

$$\int_{t_{k-1}}^{t_k} e^{\int_t^{t_k} f(T-s)\,\mathrm{d}s} g^2(T-t)\,\mathrm{d}t \le e^{\int_{t_{k-1}}^{t_k} f(T-t)\,\mathrm{d}t} \int_{t_{k-1}}^{t_k} g^2(T-t)\,\mathrm{d}t$$

and slightly rearranging the terms finally completes the proof. □