From Global to Tail Convergence: Dynamics of α-Stable Score-Based Generative Models

Ianis Hammani[1,2] Claire Boyer[2,3]

[1]Ecole Normale Supérieur Paris-Saclay, France
[2]Université Paris-Saclay, CNRS, Inria, Laboratoire de mathématiques d'Orsay, 91405 Orsay, France
[3]Institut Universitaire de France

Correspondence: ianis.hammani@ens-paris-saclay.fr

ABSTRACT

Score-based generative models driven by α-stable Lévy processes have been proposed as an alternative to Brownian diffusion models, with empirical evidence suggesting improved sampling and better modeling of heavy-tailed data. Yet theoretical guarantees for these models remain limited. We develop a non-asymptotic convergence analysis for scale-preserving α-stable generative dynamics, covering both compactly supported and power-law targets. In p-Wasserstein distance, we quantify the effects of early stopping, initialization, score approximation, and time discretization, and show exponential decay of the initialization error. A comparison with Brownian dynamics reveals that global Wasserstein convergence alone does not fundamentally distinguish the two models: both initialization errors decay exponentially with the forward horizon on the common class of heavy-tailed targets. A sharp separation emerges instead in the tails. At any positive generation time, stable dynamics produces power-law tails with index α, whereas Brownian dynamics remains light-tailed. For a target with tail index ν, asymptotically correct extreme-quantile scaling is possible precisely when $\alpha = \nu$. These results identify tail-index matching as a principled mechanism behind the ability of stable generative models to capture extreme events.

Keywords Score-based generative models · α-stable Lévy processes · Heavy-tailed distributions · Non-asymptotic convergence · Wasserstein bound · Tail-index matching

1 INTRODUCTION

Score-based generative models learn to reverse a noising process that maps the data distribution to a tractable reference law (Song et al., 2021). Standard constructions use Brownian motion, based on Gaussian noising transitions. However, its continuous, light-tailed paths evolve through local increments, making distant or rare modes harder to explore and recover (Yoon et al., 2023; Paquet et al., 2024). Similarly, Brownian score-based models may also underrepresent power-law tails (Pandey et al., 2025) and be inadequately initialized for heavy-tailed targets (Fassina et al., 2026). These limitations motivate considering non-Gaussian noising processes (Nachmani et al., 2021; Deasy et al., 2021).

Isotropic α-stable noise represents a heavy-tailed alternative, where the parameter α controls the tails and jump scaling. In particular, $\alpha = 2$ recovers the Gaussian case, while $\alpha \in (1, 2)$ yields heavy-tailed noise. Yoon et al. (2023) use an α-stable noise in the continuous scale-preserving Lévy–Itô model (LIM), with explicit forward transitions and fractional score. LIM and the discrete model of Shariatian et al. (2025) have been reported to accelerate and diversify sampling, improve tail and rare-mode coverage, and better explore high-dimensional spaces (Yoon et al., 2023; Paquet et al., 2024; Shariatian et al., 2025; Popov et al., 2025). Why stable noise helps remains only partially understood theoretically: for the discrete model in the finite-sample regime, Cherkaoui et al. (2026) show that its initialization gain can be offset by a harder estimation problem in the reverse dynamics.

In continuous time, the main difficulty lies in constructing a tractable backward dynamics. For jump processes, the exact time reversal is generally no longer an additive Lévy SDE (Conforti and Léonard, 2022; Qi et al., 2026), making it impractical for generative modeling. Popov et al. (2025) circumvent this difficulty by introducing a family of backward SDEs with additive Lévy noise that, while not pathwise reversals, exactly reproduce the forward marginal flow in reverse.

Non-asymptotic convergence guarantees are well developed for Brownian diffusion models (Conforti et al., 2025; Strasman et al., 2025; Gao et al., 2025), while general Lévy-type denoising Markov models have recently been studied through a unified generator framework (Ren et al., 2026). In contrast, quantitative guarantees for continuous α-stable score-based models remain largely unexplored, particularly for heavy-tailed data.

Contributions. We develop a convergence theory for continuous α-stable generative models and compare their global and tail behavior with Brownian dynamics. Our main contributions are:

- Building on a fractional Tweedie identity, we relate reverse-drift regularity to the Lipschitz modulus of the optimal denoiser and establish the required control for both compactly supported and power-law targets.
- For $1 \leq p < \alpha$, we establish a non-asymptotic W_p bound for the practical α-stable sampler, separating early stopping, initialization, score approximation, and discretization errors. For power-law targets, we further show that the Brownian initialization error also decays exponentially despite the loss of the contractive coupling available in the stable case. Thus, global Wasserstein convergence, established so far, alone does not fundamentally separate the two dynamics.
- A sharp distinction emerges in the tails. For exact-score dynamics initialized at the reference laws, stable outputs have power-law tails with index α, whereas Brownian outputs are light-tailed at every positive generation time. For a target with tail index ν, the stable extreme-quantile ratio vanishes for $\alpha > \nu$, diverges for $\alpha < \nu$, and converges to a finite nonzero limit precisely when $\alpha = \nu$. Hence, tail-index matching is the unique regime allowing asymptotically correct extreme-quantile scaling.

Complete proofs and the protocols for numerical experiments are deferred to the appendices.

2 α-STABLE DYNAMICS

In this section, we introduce the scale-preserving forward process and its marginally exact additive reverse dynamics, then we characterize its regularity and dissipativity through the denoiser representation of the fractional score.

2.1 Forward and Backward Dynamics

For any $\alpha \in (1,2]$, an isotropic α-stable Lévy process $(L_t^\alpha)_{t\geq 0}$ on $\mathbb{R}^d$ has stationary independent increments and characteristic function $\mathbb{E}[e^{i\xi\cdot L_t^\alpha}] = e^{-t\|\xi\|^\alpha}$ for $\xi \in \mathbb{R}^d$. It satisfies the scaling relation $L_t^\alpha \overset{\mathcal{L}}{=} t^{1/\alpha} L_1^\alpha$. Since $L_t^2 \overset{\mathcal{L}}{=} \sqrt{2} W_t$ for a standard Brownian motion W, we henceforth reserve "α-stable" for $\alpha \in (1,2)$ and treat the conventionally normalized Brownian case separately.

Denote the standard isotropic α-stable law of L_1^α by $\mathrm{S}\alpha\mathrm{S}_d(1)$, with density φ_α, which is smooth and satisfies $\varphi_\alpha(z) \asymp \|z\|^{-(d+\alpha)}$ as $\|z\| \to \infty$ (Applebaum, 2009; Samorodnitsky and Taqqu, 1994). Following Yoon et al. (2023), we consider the scale-preserving stochastic differential equation (SP-SDE)

$$\mathrm{d}X_t = -\frac{\beta(t)}{\alpha} X_t \,\mathrm{d}t + \beta(t)^{1/\alpha}\,\mathrm{d}L_t^\alpha, \qquad X_0 \sim \pi_{\mathrm{data}}, \tag{2.1}$$

where $\beta > 0$ is measurable and locally bounded, and X_0 is independent of L^α. For $t > 0$, write $\pi_t = \mathrm{Law}(X_t)$ and use the same symbol for its density. Setting

$$a(t) := \exp\left(-\frac{1}{\alpha}\int_0^t \beta(s)\,\mathrm{d}s\right), \qquad \gamma(t)^\alpha := 1 - a(t)^\alpha,$$

its solution satisfies

$$X_t \overset{\mathcal{L}}{=} a(t)X_0 + \gamma(t)Z, \qquad Z \sim \mathrm{S}\alpha\mathrm{S}_d(1),$$

with Z independent of X_0. In particular, the SP-SDE preserves the isotropic stable law $\mathrm{S}\alpha\mathrm{S}_d(1)$, and the conditional density of X_t given $X_0 = x_0$ takes the explicit form $\pi_t(x \mid x_0) = \gamma(t)^{-d}\varphi_\alpha\big((x - a(t)x_0)/\gamma(t)\big)$.

The reverse dynamics involves the fractional score

$$S_t^{(\alpha)}(x) := \frac{\Delta^{(\alpha-2)/2}\nabla \pi_t(x)}{\pi_t(x)}, \tag{2.2}$$

where $\Delta^{(\alpha-2)/2}\nabla$ has Fourier multiplier $-i\xi\|\xi\|^{\alpha-2}$ under the convention $\widehat{f}(\xi) = \int f(x)e^{i\xi\cdot x}\,\mathrm{d}x$; for $\alpha = 2$, this reduces to the classical score $\nabla \log \pi_t$. Since the exact time reversal of (2.1) has a state-dependent jump kernel (Conforti and Léonard, 2022), we instead use the marginally exact additive family of Popov et al. (2025, Theorem 1).

Proposition 1 (Marginally exact reverse dynamics, Popov et al. (2025))**.** *Fix $\eta > 0$ and an early-stopping time $0 < \varepsilon < T$, write $\bar{\beta}(\tau) := \beta(T-\tau)$, and assume $\mathbb{E}\,\|X_0\| < \infty$. Let*

$$\bar{b}(\tau, x) := \frac{\bar{\beta}(\tau)}{\alpha} x + (1+\eta)\bar{\beta}(\tau) S_{T-\tau}^{(\alpha)}(x). \tag{2.3}$$

Let $\overleftarrow{L}^\alpha$ be an isotropic α-stable process independent of $\overleftarrow{Z}_0$. Under the well-posedness and Fokker–Planck uniqueness conditions of Appendix A.3, the solution of

$$\mathrm{d}\overleftarrow{Z}_\tau = \bar{b}(\tau, \overleftarrow{Z}_\tau)\,\mathrm{d}\tau + \big(\eta\bar{\beta}(\tau)\big)^{1/\alpha}\mathrm{d}\overleftarrow{L}_\tau^\alpha, \qquad \overleftarrow{Z}_0 \sim \pi_T, \tag{2.4}$$

satisfies $\mathrm{Law}(\overleftarrow{Z}_\tau) = \pi_{T-\tau}$ for $0 \leq \tau \leq T - \varepsilon$.

The parameter $\eta > 0$ is a genuine degree of freedom in the reverse sampler. Every choice preserves the same marginal evolution, but changes the backward noise intensity and drift. The family interpolates between deterministic and stochastic sampling: $\eta \downarrow 0$ yields the probability-flow ODE, while $\eta = 1$ matches the forward noise intensity, paralleling the probability-flow ODE and reverse-time SDE of Brownian variance-preserving (VP) models (Song et al., 2021).

Equation (2.4) defines the ideal reverse dynamics. In practice, three approximations are required to obtain a tractable sampler: (i) the unknown fractional score $S_t^{(\alpha)}$ is replaced by a learned approximation s_θ; (ii) the inaccessible initialization π_T is replaced by the reference distribution $\pi_\infty^\alpha = \mathrm{S}\alpha\mathrm{S}_d(1)$; and (iii) the continuous-time dynamics is approximated by an exponential integrator on a uniform grid of N steps, with early stopping at time $\varepsilon > 0$. We denote by $\hat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}$ the law generated by the resulting algorithm, which we study in the next sections.

2.2 Fractional Tweedie Formula

As in Gaussian score-based models, the fractional score admits a Tweedie-type representation in terms of the optimal denoiser. The following identity, essentially due to Yoon et al. (2023), extends the classical Tweedie formula (Efron, 2011) to the α-stable setting.

Proposition 2 (Fractional Tweedie formula, Yoon et al. (2023))**.** *For the SP-SDE (2.1) with any initial distribution, the denoiser and fractional score are related, at every $t > 0$, by*

$$m_t(x) = \frac{x + \alpha\gamma(t)^\alpha S_t^{(\alpha)}(x)}{a(t)}, \tag{2.5}$$

where $m_t(x) := \mathbb{E}[X_0 \mid X_t = x]$ is the conditional mean, which is MSE-optimal when $X_0 \in L^2$.

In what follows, we therefore quantify the regularity of m_t through its global Lipschitz constant

$$C_t := \sup_{x\in\mathbb{R}^d} \|\nabla m_t(x)\|_{\mathrm{op}}.$$

2.3 Dissipativity

The denoiser representation (2.5) yields dissipativity estimates for the fractional score and one-sided Lipschitz bounds for the reverse drift, which determine the contractivity and stability of the sampling dynamics.

Proposition 3 (Dissipativity and one-sided Lipschitz estimates)**.** *Let $t \in (0, T]$, set $\tau = T - t$, and suppose $C_t < \infty$. Then, for all $x, y \in \mathbb{R}^d$, the score and reverse drift satisfy*

$$\begin{aligned} \langle S_t^{(\alpha)}(y) - S_t^{(\alpha)}(x), x - y\rangle &\geq C_t^{(\alpha)} \|x - y\|^2, \\ \langle \bar{b}(\tau, x) - \bar{b}(\tau, y), x - y\rangle &\leq \bar{L}(\tau) \|x - y\|^2, \end{aligned} \tag{2.6}$$

$$\begin{aligned} \textit{where} \quad C_t^{(\alpha)} &:= \frac{1 - a(t)C_t}{\alpha\gamma(t)^\alpha}, \\ \bar{L}(\tau) &:= \beta(T-\tau)\Big(\tfrac{1}{\alpha} - (1+\eta)\, C_{T-\tau}^{(\alpha)}\Big). \end{aligned}$$

The reverse drift is therefore contractive whenever $C_t^{(\alpha)} > 1/(\alpha(1+\eta))$. We note here that increasing η facilitates contractivity.

3 COMPACTLY SUPPORTED TARGETS

3.1 Non-asymptotic W_p bound

We bound

$$W_p\left(\pi_{\mathrm{data}}, \hat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}\right)$$

for the exponential-integrator (EI) sampler defined in Appendix B.1, Equation (B.1). It integrates the dissipative linear part in (3.3) and the stochastic convolution exactly, while freezing the score residual over each step.

Backward processes. Choose a uniform grid of $[0, T-\varepsilon]$ as $\tau_k = kh$, for $k = 0, \dots, N$, of constant time step $h = (T-\varepsilon)/N$. To disentangle the different sources of approximation error, we introduce three backward processes: (i) the exact process $\overleftarrow{Z}$ solves (2.4) with the exact score and is initialized from $\overleftarrow{Z}_0 \sim \pi_T$; (ii) the ideal process $\overleftarrow{Y}$ also solves (2.4) but is initialized from $\overleftarrow{Y}_0 \sim \pi_\infty^\alpha$ instead; and (iii) the (piecewise-constant) numerical process $\bar{X}^{\theta,h}$ is the EI approximation driven by the learned score s_θ, initialized from the same distribution π_∞^α. All three processes use the same backward Lévy process $\overleftarrow{L}^\alpha$; the EI noise is its exponentially weighted stochastic integral in (B.1). Consequently, the only differences between the three dynamics arise from the initialization, the score approximation, and the time discretization.

Standing assumptions. The analysis rests on the following assumptions, bearing on the noise schedule, score approximation, temporal score regularity, and spatial denoiser regularity after early stopping.

Assumption A1 (Noise schedule)**.** The schedule $\beta : [0, \infty) \to (0, \infty)$ is continuous, nondecreasing, and satisfies $B(T) = \int_0^T \beta(t)\,\mathrm{d}t \to \infty$ as $T \to \infty$; for a given horizon it is used restricted to $[0, T]$.

Assumption A2 (Score approximation error)**.** The spatial modeling error of the network s_θ is uniformly bounded in L^p over the backward horizon: there exists $\mathrm{err} \geq 0$ with

$$\sup_{0 \leq k \leq N-1} \left\| s_\theta(\tau_k, \bar{X}^{\theta,h}_{\tau_k}) - S^{(\alpha)}_{T-\tau_k}(\bar{X}^{\theta,h}_{\tau_k}) \right\|_{L^p(\Omega)} \leq \mathrm{err}.$$

Assumption A3 (Temporal regularity of the score)**.** For the early-stopping threshold $\varepsilon \in (0, T)$ there is $L^{\mathrm{time}}_\varepsilon > 0$, possibly depending on T, such that, for all $x \in \mathbb{R}^d$ and all $s, t \in [\varepsilon, T]$,

$$\left\| S^{(\alpha)}_t(x) - S^{(\alpha)}_s(x) \right\| \leq L^{\mathrm{time}}_\varepsilon\, |t-s|\, (1 + \|x\|).$$

Assumption A4 (Uniform denoiser regularity)**.** For every early-stopping threshold $\varepsilon > 0$, $\bar{C}_\varepsilon := \sup_{t \geq \varepsilon} C_t < \infty$.

The assumptions play distinct roles. Assumption A1 fixes the time scale, while Assumptions A2–A3 control the two approximations made by the numerical sampler. They are stable analogues of standard Brownian SGM conditions (Conforti et al., 2025; Strasman et al., 2025). Finally, Assumption A4 controls the spatial regularity of the score after early stopping; it is target-dependent and is verified separately in the regimes studied.

Error decomposition. Since $\mathrm{Law}(\overleftarrow{Z}_{T-\varepsilon}) = \pi_\varepsilon$, one has

$$W_p\left(\pi_{\mathrm{data}}, \hat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}\right) \leq \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{E}_2, \tag{3.1}$$

$$\text{where} \quad \begin{cases} \mathcal{E}_0 = W_p(\pi_{\mathrm{data}}, \pi_\varepsilon), \\ \mathcal{E}_1 = W_p(\mathrm{Law}(\overleftarrow{Z}_{T-\varepsilon}), \mathrm{Law}(\overleftarrow{Y}_{T-\varepsilon})), \\ \mathcal{E}_2 = W_p(\mathrm{Law}(\overleftarrow{Y}_{T-\varepsilon}), \mathrm{Law}(\bar{X}^{\theta,h}_{T-\varepsilon})), \end{cases}$$

respectively correspond to the stopping, initialization, and score+discretization errors.

For clarity of exposition, we state hereafter the result for a constant noise schedule. See Appendix B for a general version, allowing time-dependent schedules.

Theorem 4 (Constant schedule)**.** *Suppose π_{data} is supported in $\bar{B}(0,R)$. Fix $\eta > 0$, choose $p \in [1,\alpha)$, a constant noise schedule $\beta \equiv \beta_0$, and arbitrary times $0 < \varepsilon < T$. Under Assumptions A1–A3, for h small enough there exist $\mathsf{c}_0, \dots, \mathsf{c}_4 > 0$ depending only on $(\alpha, \eta, d, p, \beta_0, R, \varepsilon, L_\varepsilon^{\text{time}})$ such that*

$$W_p(\pi_{\text{data}}, \hat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}) \le \mathsf{c}_0 \varepsilon^{1/\alpha} + \mathsf{c}_1 e^{-\rho_\star (T-\varepsilon)} + \mathsf{c}_2 (T-\varepsilon)\mathrm{err} + \mathsf{c}_3 \tfrac{h}{\varepsilon} + \mathsf{c}_4 (1 + T + \log(1+\tfrac{1}{\varepsilon}))h^{1/\alpha}, \tag{3.2}$$

with $\rho_\star = \beta_0 \frac{1+\eta}{\alpha}$.

Following the Brownian analysis of Conforti et al. (2025), the proof first recasts the backward drift so that its linear part becomes dissipative. The stationary fractional score is linear, $S_\infty^{(\alpha)}(x) = -x/\alpha$, and subtracting it yields

$$\bar{b}(\tau, x) = -\tfrac{\eta}{\alpha}\beta(T-\tau)\,x + (1+\eta)\beta(T-\tau)\big[S_{T-\tau}^{(\alpha)}(x) - S_\infty^{(\alpha)}(x)\big]. \tag{3.3}$$

Equation (2.4) is thus a perturbation of a stable Ornstein–Uhlenbeck process of rate $\eta\beta(T-\tau)/\alpha$ by a term vanishing at the reference, a structure essential for the analysis of $\mathcal{E}_2$. The same reformulation of the Brownian reverse SDE gives a rate $\beta(T-\tau)/2$, exactly recovered here at $\alpha = 2$ and $\eta = 1$.

The five terms of (3.2) have distinct origins. The first is the early-stopping bias, the price of not integrating the dynamics down to $t = 0$; it decreases with ε. The second is the initialization error of the backward dynamics, which decays exponentially in the horizon T at the rate $\rho_\star$, itself increasing in η. The next two control score approximation and drift discretization. The last discretization term, of order $h^{1/\alpha}$, reflects stable fluctuations over an EI step, where the nonlinear drift in (3.3) is frozen; its bound also absorbs drift and temporal-freezing contributions. This matches the usual $h^{1/2}$ term in the W_2 analysis of the Brownian dynamics by Strasman et al. (2025).

Decreasing ε therefore reduces the first term but inflates the last three, through the singular behavior of the score near forward time $t = 0$. The restriction $p < \alpha$ is required for the stable stationary law to have a finite p-th moment.

3.2 Denoiser regularity

The proof of Theorem 4 uses Assumption A4, which the following proposition verifies for compactly supported targets. It also shows that the denoiser Lipschitz constant vanishes asymptotically.

Proposition 5 (Decay of C_t)**.** *If π_{data} is supported in $\bar{B}(0,R)$, then with $\Lambda := \sup_u \|\nabla^2 \log \varphi_\alpha(u)\| < \infty$,*

$$C_t \le \frac{d\,\Lambda\,R^2\,a(t)}{\gamma(t)^2} \xrightarrow[t\to\infty]{} 0. \tag{3.4}$$

Consequently, Assumption A4 is satisfied for compactly supported targets. Compact support also ensures the well-posedness conditions required by Proposition 1, so that the convergence guarantee of Theorem 4 applies for any early-stopping time $0 < \varepsilon < T$. Moreover, since $C_t \to 0$, the dissipativity constant satisfies $C_t^{(\alpha)} \longrightarrow \frac{1}{\alpha}$. The reverse drift is therefore contractive whenever the corresponding forward time is sufficiently large. Any possible lack of contractivity is thus confined to a finite transient regime, whose contribution can be absorbed into the constants of the convergence bound; see Appendix B. Section 4 shows that the same asymptotic decay persists beyond compact support, for a class of heavy-tailed target distributions.

4 HEAVY-TAILED TARGETS

We now turn to heavy-tailed target distributions, where compact-support arguments no longer apply. Such distributions naturally arise in applications where rare but extreme events carry a significant fraction of the information, including financial and insurance losses, network traffic, and climate records (Pandey et al., 2025; Resnick, 2007). Our next result shows that the convergence guarantees established in the previous section extend beyond compactly supported targets to a broad class of power-law distributions, as specified by the following assumption.

Assumption A5 (Heavy-tailed target)**.** The target π_{data} admits a bounded, positive and continuous density with an asymptotically radial power-law tail of index $\nu > 1$:

$$\pi_{\text{data}}(x) = \left(c_{\pi_{\text{data}}} + o(1)\right) \|x\|^{-(d+\nu)} \tag{4.1}$$

when $\|x\| \to \infty$ uniformly in the direction $x/\|x\|$.

In order to extend the convergence analysis to heavy-tailed targets, the main additional difficulty is to control the denoiser regularity without relying on bounded support. Under Assumption A5, its value at the origin is controlled by the finite first moment,

$$\|m_t(0)\| \leq M_1 := \mathbb{E}[\|X_0\|] < \infty.$$

The next result characterizes the denoiser regularity for heavy-tailed targets.

Proposition 6 (Denoiser regularity)**.** *Under Assumptions A1 and A5, the following hold.*

(i) If $1 < \nu < \alpha$, then $C_t \geq a(t)^{-1} \xrightarrow[t\to\infty]{} \infty$.

(ii) If $\alpha \leq \nu < 1 + \alpha$, Assumption A5 alone does not ensure $C_t < \infty$: for each such ν, there is a smooth, positive density in dimension one satisfying this assumption for which $C_t = \infty$ for every $t > 0$.

(iii) If $\nu \geq 1 + \alpha$, then for every $\varepsilon > 0$ there exists $c_\varepsilon < \infty$, independent of t, such that $C_t \leq c_\varepsilon a(t)$ for all $t \geq \varepsilon$. Consequently, $C_t \to 0$ and Assumption A4 holds.

For $\nu \geq 1 + \alpha$, we can therefore apply the convergence analysis of Theorem 4 without imposing any additional regularity assumption on the denoiser.

Theorem 7 (W_p-bound for heavy-tailed targets)**.** *Let π_{data} satisfy Assumption A5 with $\nu \geq 1 + \alpha$. Under the remaining hypotheses of Theorem 4, the same bound holds. In the drift estimates, the support-radius control of $\|m_t(0)\|$ is replaced by M_1. The resulting constants may additionally depend on $\mu_p = \mathbb{E}[\|X_0\|^p]^{1/p}$ and on the quantitative constants in Assumption A5.*

5 COMPARISON WITH BROWNIAN DYNAMICS

We now compare α-stable and Brownian dynamics on heavy-tailed targets from two complementary perspectives: global convergence of the generated distribution (which primarily reflects reconstruction of the bulk) and preservation of its tail behavior.

5.1 Global comparison

Our goal is to determine whether α-stable and Brownian dynamics exhibit a meaningful separation at the level of the generated distribution as a whole, in W_p. For the Brownian model, we consider a VP-SDE with a general noise schedule and its associated reverse-time dynamics. We denote its forward marginals by $\pi_t^{(2)}$ and its stationary law by $\pi_\infty^{(2)} = \mathcal{N}(0, I_d)$. Since the stable and Brownian formulations share the same algebraic structure, we use $a(t)$, $\beta(t)$, $\gamma(t)$, and C_t for the corresponding quantities in both settings. When a distinction is needed, we make it explicit, e.g. writing $\mathcal{E}_1^{(\alpha)}$ for the stable initialization error and $\mathcal{E}_1^{(2)}$ for its Brownian counterpart.

Following the error decomposition of Section 3, we discuss the two dynamics in terms of denoiser regularity, mixing of the reverse dynamics, discretization, and early-stopping error.

Denoiser regularity. We first compare the regularity of the denoiser, as measured by its Lipschitz constant C_t. Under stable noise, Propositions 5 and 6 give $C_t \to 0$ for compactly supported targets and power-law targets with $\nu \geq 1 + \alpha$. In the Brownian case, the behavior actually depends on the target type.

Proposition 8 (Brownian denoiser regularity)**.** *Under Assumption A1, the Brownian VP denoiser satisfies:*

- *If π_{data} is supported on $\bar{B}(0, R)$, then $C_t \leq a(t)R^2/\gamma(t)^2 \to 0$ as $t \to \infty$.*
- *Under Assumption A5, $\nabla m_t(x) \to a(t)^{-1} I_d$ as $\|x\| \to \infty$ for every $t > 0$; hence $C_t \geq a(t)^{-1} \to \infty$ as $t \to \infty$.*

Thus, for heavy-tailed targets with $\nu \geq 1 + \alpha$, denoiser regularity separates the two models: stable noise asymptotically flattens the denoiser, yielding eventual contractivity, whereas Brownian noise produces the opposite behavior.

Figure 1 illustrates this dichotomy on a Student-t_4 target with $\beta \equiv 1$ in both models through $\widehat{C}_t(40)$, the maximum numerically evaluated $|m_t'|$ on a grid of $[-40, 40]$: it decreases under stable noise ($\alpha = 1.5$) and grows under Brownian noise, consistently with the lower bound $C_t \geq 1/a(t) = e^{t/2}$; see Appendix F.2 for details.

Crucially, this regularity gap does not imply a separation in global convergence. As we show next, the loss of Brownian contractivity makes the synchronous-coupling argument ineffective, but does not prevent the Brownian reverse dynamics from mixing.

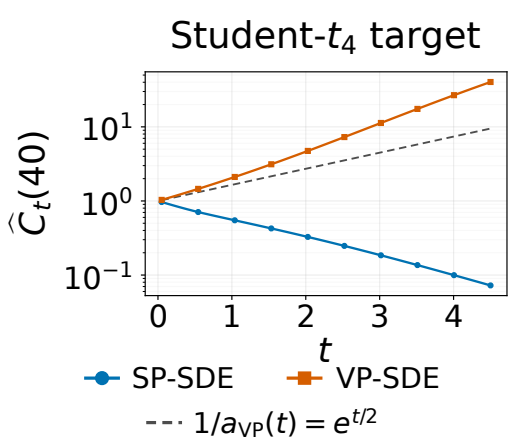

Figure 1: Illustration of $\widehat{C}_t(40)$ on a Student-t_4 target for the stable ($\alpha = 1.5$) and VP models.

Mixing. Here, mixing error denotes the propagated initialization error. For stable dynamics, the one-sided Lipschitz bound yields a contractive synchronous coupling, as $\mathcal{E}_1^{(\alpha)} \leq W_p(\pi_{\text{data}}, \pi_\infty^\alpha) e^{-\Phi^{(\alpha)}(T)}$ where

$$\Phi^{(\alpha)}(T) := \frac{1+\eta}{\alpha} \int_\varepsilon^T \frac{\beta(t)\big(1 - a(t)C_t\big)}{\gamma(t)^\alpha} \, \mathrm{d}t.$$

Combined with the decay of C_t, this gives an explicit exponential rate for the initialization error.

Corollary 9 (Stable mixing error)**.** *Under Assumptions A1 and A5 with $\nu \geq 1+\alpha$, for every $p \in [1, \alpha)$ stable noise mixes exponentially in B,*

$$\mathcal{E}_1^{(\alpha)} \leq C\, W_p(\pi_{\text{data}}, \pi_\infty^\alpha)\, e^{-\frac{1+\eta}{\alpha}(B(T)-B(\varepsilon))} \xrightarrow[T\to\infty]{} 0,$$

where $C < \infty$ is independent of T.

For Brownian dynamics, Proposition 8 gives $a(t)C_t \geq 1$, so the analogous synchronous-coupling bound is vacuous. Nevertheless, a direct analysis of the initialization error still yields exponential mixing.

Theorem 10 (Brownian mixing error)**.** *Under Assumption A1, suppose π_{data} satisfies Assumption A5 with $\nu > 2$. Let $\mu_{T,\varepsilon}^{(2)}$ denote the law at reverse time $T-\varepsilon$ of the Brownian reverse SDE with the exact score, initialized at the reference $\pi_\infty^{(2)}$ in place of the forward marginal $\pi_T^{(2)}$. For $p \in [1, \nu)$ and any $q \in (p, \nu)$ there is $C_q < \infty$, independent of ε, such that for T large enough*

$$\mathcal{E}_1^{(2)} = W_p(\mu_{T,\varepsilon}^{(2)}, \pi_\varepsilon^{(2)}) \leq C_q\, e^{-\frac{q-p}{2pq}B(T)} \xrightarrow[T\to\infty]{} 0.$$

Figure 2 illustrates the decay of the initialization error on a Student-t_4 target. For each model, we compare samplers initialized from the forward marginal π_T and from the stationary reference, using the same numerically evaluated score, step size $h = 0.0125$, and stopping time $\varepsilon = 0.05$. The resulting $W_{1.1}$ and $W_{1.4}$ discrepancies decrease with T for both stable and Brownian dynamics, eventually approaching the finite-sample baseline obtained from two independent samples from π_ε. This is consistent with the predicted loss of initialization dependence in both models, including the non-contractive Brownian regime; see Appendix F.3.

Thus, on the common class of targets with $\nu \geq 1+\alpha$, both initialization errors decay exponentially in W_p, despite fundamentally different mechanisms: contractive coupling for stable dynamics and a direct mixing argument for Brownian dynamics. Global initialization error therefore does not separate the two models.

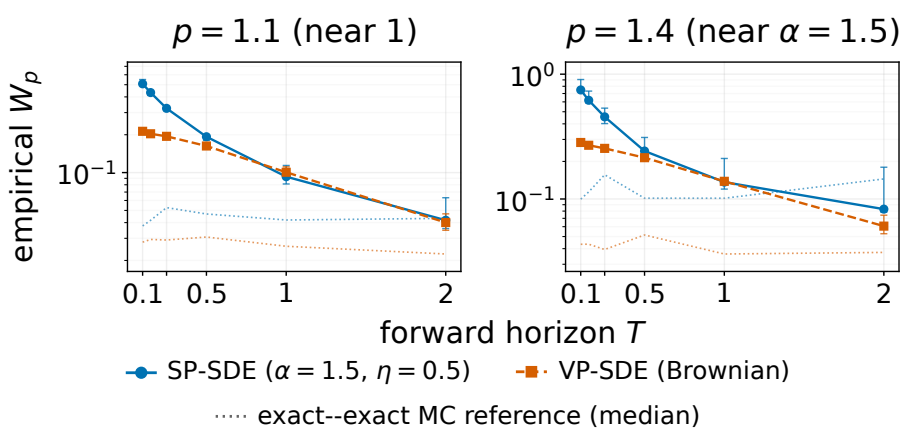

Figure 2: Empirical $W_{1.1}$ and $W_{1.4}$ discrepancies between samplers initialized from the stationary reference and from the exact forward marginal π_T, for stable and Brownian dynamics on a Student-t_4 target. Both discrepancies decrease with T toward the finite-sample baseline (dotted), estimated from independent samples from π_ε. Points and bars report medians and 90% bootstrap intervals over 12 runs; see Appendix F.3.

Early stopping and discretization. Having found no qualitative separation in the initialization error, we finally compare the scaling of the remaining two terms. For the stable sampler, both directly

reflect the self-similarity of the driving process: $\gamma(\varepsilon) \asymp (\beta_0\varepsilon)^{1/\alpha}$ and $L_h^\alpha \overset{\mathcal{L}}{=} h^{1/\alpha} L_1^\alpha$ yield the respective upper bounds $\varepsilon^{1/\alpha}$ and $h^{1/\alpha}$. For $\alpha < 2$, L^α is a pure-jump process and $1/\alpha > 1/2$ is its self-similarity index. It expresses superdiffusive scaling, in a Lévy-flight sense: heavy-tailed jumps permit occasional long-range moves and thereby provide a mechanism for broader exploration of a large space (Viswanathan et al., 2008). At $\alpha = 2$, both exponents reduce to the Brownian scaling $1/2$.

A rigorous comparison of these bounds with Brownian dynamics is not currently available under the heavy-tailed assumptions considered here. In particular, Strasman et al. (2025) obtain an $h^{1/2}$ term for the same exponential-integrator construction, but under global score regularity and strong log-concavity and without the early-stopping error. The shared construction nevertheless suggests $\varepsilon^{1/2}$ and $h^{1/2}$ as the corresponding Brownian scalings. If these rates extend to the present heavy-tailed setting, the stable bounds would exhibit more favorable exponents for $\alpha < 2$. This comparison is therefore suggestive rather than conclusive: it identifies a possible scaling mechanism consistent with the empirical sampling gains reported for Lévy-based generative models (Yoon et al., 2023; Paquet et al., 2024).

5.2 Tail behavior

Global convergence in W_p does not by itself guarantee that the generated distribution reproduces the extreme behavior of the target. In particular, it does not determine whether the generated law has the correct tail index or assigns the correct mass to extreme events. We therefore complement the Wasserstein analysis with two tail-sensitive quantities motivated by classical Extreme Value Theory (EVT, Resnick, 2007; de Haan and Ferreira, 2006; Ghosh and Resnick, 2010). The first, based on the mean-excess function, identifies the asymptotic tail regime; the second compares generated and target quantiles at increasingly extreme probability levels.

Tail regime. A target-independent criterion, called limiting relative mean excess, distinguishes power-law tails from lighter tails, such as Gaussian tails.

Definition 11 (Limiting relative mean excess)**.** For a probability measure μ on $\mathbb{R}^d$, let $F_\mu(r) = \mu(\{x : \|x\| \le r\})$ denote its radial CDF. For a level $\kappa \in (0,1)$, let $q_\kappa(\mu) = \inf\{r : F_\mu(r) \ge \kappa\}$ be its radial quantile of level κ. Then, for u such that $1 - F_\mu(u) > 0$, introduce the mean-excess function $e_\mu(u) = \mathbb{E}_\mu[\|X\| - u \mid \|X\| > u]$ so that the *limiting relative mean excess* is given by

$$\Gamma(\mu) := \lim_{u\to\infty} \frac{e_\mu(u)}{u},$$

whenever the limit exists.

Note that if $q_\kappa(\mu) \to \infty$ as $\kappa \to 1$, then the limiting relative mean excess can be re-expressed as $\Gamma(\mu) = \lim_{\kappa\to 1} e_\mu(q_\kappa(\mu))/q_\kappa(\mu)$. This condition is satisfied by the distributions considered here.

We also introduce $\overline{G}_\mu(\cdot) := 1 - F_\mu(\cdot)$, the radial survival function of μ. In particular, Assumption A5 implies that $\overline{G}_{\pi_{\text{data}}}(r) \sim c'_{\pi_{\text{data}}} r^{-\nu}$ as $r \to \infty$, with $c'_{\pi_{\text{data}}} = \omega_{d-1} c_{\pi_{\text{data}}}/\nu$, where ω_{d-1} is the surface area of the unit sphere.

Proposition 12 (Tail-regime dichotomy)**.**
(i) If $\overline{G}_\mu(r) \sim c_\mu r^{-\vartheta}$ for some $c_\mu > 0$ and $\vartheta > 1$, then $e_\mu(u) \sim u/(\vartheta - 1)$ and

$$\Gamma(\mu) = \frac{1}{\vartheta - 1} > 0.$$

(ii) If $X \sim \mu$ decomposes as $X = G + R$ with G a nondegenerate isotropic Gaussian vector and $\|R\| \le K$ almost surely, then $e_\mu(u) = \mathcal{O}(1)$ and $\Gamma(\mu) = 0$.

Thus, $\Gamma(\mu)$ identifies a qualitative difference in tail regime, but does not measure whether the generated distribution reproduces the target at a given extreme probability level. To capture this second aspect, we compare their extreme quantiles directly.

Extreme-quantile matching. For a level $\kappa \in (0,1)$, define the *extreme-quantile ratio* of μ with respect to a target reference π by

$$\mathcal{M}_\mu(\kappa) := \frac{q_\kappa(\mu)}{q_\kappa(\pi)}.$$

Thus, $\mathcal{M}_\mu(\kappa) \approx 1$ means that generated and target distributions reach comparable radial scales at probability level κ. For power-law tails, its asymptotic behavior follows directly from the corresponding tail indices and constants.

Lemma 13 (Extreme-quantile ratio under power-law tails)**.** *Consider a distribution μ such that $\overline{G}_\mu(r) \sim c_\mu r^{-\alpha}$ for $c_\mu, \alpha > 0$. Then, its quantile function satisfies $q_\kappa(\mu) \sim c_\mu^{1/\alpha}(1-\kappa)^{-1/\alpha}$ as $\kappa \uparrow 1$. Consider in addition a target distribution π such that $\overline{G}_\pi(r) \sim c_\pi r^{-\nu}$ for $c_\pi, \nu > 0$. Then, the extreme-quantile ratio of μ w.r.t. π satisfies*

$$\mathcal{M}_\mu(\kappa) \sim c_\mu^{1/\alpha}\, c_\pi^{-1/\nu}\, (1-\kappa)^{\frac{1}{\nu}-\frac{1}{\alpha}} \qquad (\kappa \uparrow 1). \tag{5.1}$$

Lemma 13 separates tail mismatch as follows. A discrepancy between the tail indices produces a polynomial divergence or collapse of $\mathcal{M}_\mu(\kappa)$ as $\kappa \uparrow 1$. When the indices match ($\alpha = \nu$) this polynomial discrepancy disappears and $\mathcal{M}_\mu(\kappa) \longrightarrow (c_\mu/c_\pi)^{1/\alpha}$.

Stable/Brownian separation in the tails. Fix $0 < \varepsilon < T < \infty$. Write $\mu_\tau^{(\alpha)} := \mathrm{Law}(\overleftarrow{Y}_\tau)$ for the stable exact-score reverse law initialized at π_∞^α, and define $\mu_\tau^{(2)}$ analogously for the Brownian reverse SDE initialized at $\pi_\infty^{(2)}$. The symbol μ denotes a generated law at reverse time $\tau \in (0, T-\varepsilon]$, whereas π denotes a forward marginal. In particular, $\mu_{T-\varepsilon}^{(2)} = \mu_{T,\varepsilon}^{(2)}$ in Theorem 10.

Proposition 14 (Stable/Brownian tail separation)**.** *Suppose Assumption A5 holds, so that $\overline{G}_{\pi_{\mathrm{data}}}(r) \sim c'_{\pi_{\mathrm{data}}} r^{-\nu}$. At every generation time $\tau \in (0, T-\varepsilon]$, the Brownian output $\mu_\tau^{(2)}$ has sub-Gaussian radial tails, whereas the α-stable output satisfies*

$$\overline{G}_{\mu_\tau^{(\alpha)}}(r) \sim c_\tau r^{-\alpha}, \qquad c_\tau \in (0,\infty).$$

In terms of relative mean excess, this gives

$$\Gamma(\mu_\tau^{(\alpha)}) = \frac{1}{\alpha-1} \;>\; 0 = \Gamma(\mu_\tau^{(2)}).$$

For the extreme-quantile ratio, $\mathcal{M}_{\mu_\tau^{(2)}}(\kappa) \xrightarrow[\kappa\to 1]{} 0$, while

$$\mathcal{M}_{\mu_\tau^{(\alpha)}}(\kappa) \underset{\kappa\to 1}{\sim} c_\tau^{1/\alpha}(c'_{\pi_{\mathrm{data}}})^{-1/\nu}(1-\kappa)^{\frac{1}{\nu}-\frac{1}{\alpha}}.$$

The proof decomposes the stable output into an exactly α-stable leading term and a uniformly sublinear remainder, which is negligible in the tail. The tail constant is $c_\tau = C_{d,\alpha}\rho_\tau^\alpha$, where $C_{d,\alpha}$ is the radial tail constant of $\mathrm{S}\alpha\mathrm{S}_d(1)$ and ρ_τ is the scale of the leading stable component (see Appendix E for details).

The two diagnostics reveal complementary aspects of tail reconstruction: Γ distinguishes light from power-law tails, whereas $\mathcal{M}_\mu$ measures the scale of extreme quantiles. Their joint behavior is summarized in Table 1.

Thus, producing heavy tails is not sufficient for correctly calibrated extremes. Indeed, if $\alpha > \nu$, the generated tail is lighter than the target one and the extreme-quantile ratio vanishes. If $\alpha < \nu$, the generated tail is heavier so that the extreme-quantile ratio diverges. Tail-index matching, $\alpha = \nu$,

Table 1: Tail regimes under Assumption A5, at fixed $0 < \varepsilon < T < \infty$, as in Proposition 14.

	$\Gamma(\mu_\tau)$	$\mathcal{M}_{\mu_\tau}(\kappa)$ as $\kappa \uparrow 1$
Brownian	0	0
Stable, $\alpha > \nu$	$\frac{1}{\alpha-1}$	0
Stable, $\alpha = \nu$	$\frac{1}{\nu-1}$	$\left(c_\tau / c'_{\pi_{\text{data}}}\right)^{1/\nu}$
Stable, $\alpha < \nu$	$\frac{1}{\alpha-1}$	$+\infty$

is therefore the unique stable regime allowing asymptotically correct extreme-quantile scaling; exact matching further requires the tail constants to agree.

Figure 3 illustrates the three tail-reconstruction regimes predicted by Proposition 14 for Student-t_ν targets. With stable noise ($\alpha = 1.5$), as $\kappa \uparrow 1$ the extreme-quantile ratio decreases for $\nu < \alpha$, remains approximately constant for $\nu = \alpha$, and increases for $\nu > \alpha$, consistently with the predicted scaling $(1-\kappa)^{1/\nu - 1/\alpha}$. In the index-matched case $\nu = \alpha = 1.5$, the Brownian ratio instead decreases toward zero, illustrating its failure to reproduce the target tail scale. The off-match experiments reproduce the predicted qualitative trends, although the available sample size does not reliably resolve their asymptotic exponents.

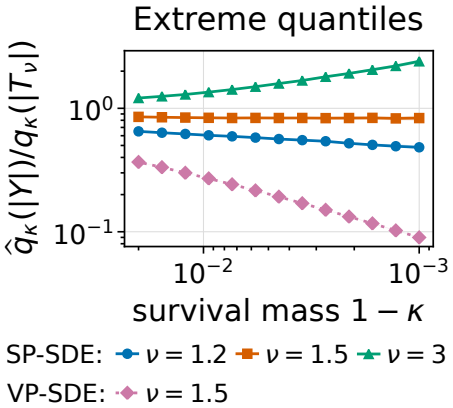

Figure 3: Generated-to-target extreme-quantile ratios for Student-t_ν targets. Stable samplers use $(\alpha, \eta) = (1.5, 0.5)$; the VP baseline uses $\nu = 1.5$. Curves show medians over 12 runs; bands give the inter-seed 16–84% range. See Appendix F.4.

Overall, the exact-score dynamics initialized at the reference laws differ in their tails: Brownian outputs remain light-tailed at finite generation time, while stable outputs inherit the noise tail index. This distinction is not determined by global W_p convergence. Tail-index matching therefore provides a possible mechanism behind the improved tail coverage observed empirically with Lévy-based generative models (Yoon et al., 2023; Paquet et al., 2024; Pandey et al., 2025; Shariatian et al., 2025).

CONCLUSION

Our results distinguish global accuracy from tail reconstruction in α-stable generative models. Our tail analysis includes the index-matched regime $\nu = \alpha$, whereas our global convergence guarantees require stronger tail assumptions. Tail-index matching explains when stable dynamics can reproduce the correct asymptotic scale of extreme events, yet the role of tail constants, anisotropy, and multivariate dependence remains largely unexplored. It is also unclear whether Lévy noise provides advantages

beyond tail reconstruction, for instance through faster exploration of multimodal landscapes or improved finite-step sampling. Finally, extending the tail analysis to learned scores is essential: estimation errors in low-density regions may interact strongly with the jump structure precisely where the stable dynamics is meant to improve tail coverage.

AI Use Statement

In this work, we used generative AI tools, including ChatGPT, to assist with English writing and editing, literature checks, mathematical proofreading, and the implementation and verification of some numerical experiments. Generative AI tools were not used to define the research questions, formulate the main methodological ideas, or determine the scientific conclusions of the work.

All AI-assisted mathematical arguments were independently checked by the authors, and no proof or theoretical claim was accepted without verification. AI-assisted code and numerical experiments were reviewed, tested, and validated by the authors, including by checking their consistency with the underlying mathematical formulation and the reported results. Literature-related suggestions were verified against the original sources before being used or cited.

We take full responsibility for the final content of this work, including its text, mathematical claims, numerical results, and conclusions.

References

Yang Song, Jascha Sohl-Dickstein, Diederik Pieter Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Eunbi Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim. Score-based generative models with Lévy processes. In *Advances in Neural Information Processing Systems*, volume 36, pages 40694–40707, 2023. doi: 10.52202/075280-1772. URL https://doi.org/10.52202/075280-1772.

Eric Paquet, Farzan Soleymani, Herna Lydia Viktor, and Wojtek Michalowski. Annealed fractional Lévy–Itô diffusion models for protein generation. *Computational and Structural Biotechnology Journal*, 23:1641–1653, 2024. doi: 10.1016/j.csbj.2024.04.009.

Kushagra Pandey, Jaideep Pathak, Yilun Xu, Stephan Mandt, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Heavy-tailed diffusion models. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.14171.

Tiziano Fassina, Gabriel Cardoso, Sylvain Le Corff, and Thomas Romary. Generalizing score-based generative models for heavy-tailed distributions, 2026. URL https://arxiv.org/abs/2603.00772.

Eliya Nachmani, Robin San Roman, and Lior Wolf. Non-gaussian denoising diffusion models, 2021. URL https://arxiv.org/abs/2106.07582.

Jacob Deasy, Nikola Simidjievski, and Pietro Liò. Heavy-tailed denoising score matching, 2021. URL https://arxiv.org/abs/2112.09788.

Dario Shariatian, Umut Şimşekli, and Alain Durmus. Heavy-tailed diffusion with denoising Lévy probabilistic models. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2407.18609.

Vadim Popov, Assel Yermekova, Tasnima Sadekova, Artem Khrapov, and Mikhail Kudinov. Improved sampling algorithms for Lévy–Itô diffusion models. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=XxCgeWSTNp.

Hamza Cherkaoui, Hélène Halconruy, and Antonio Ocello. Do heavy tails help diffusion? on the subtle trade-off between initialization and training, 2026. URL https://arxiv.org/abs/2605.13175.

Giovanni Conforti and Christian Léonard. Time reversal of markov processes with jumps under a finite entropy condition. *Stochastic Processes and their Applications*, 144:85–124, 2022. doi: 10.1016/j.spa.2021.10.002.

Tianfu Qi, Jun Wang, and Jun Zhang. Generator-guided inverse sampling for Lévy-driven generative models, 2026. URL https://arxiv.org/abs/2608.10384.

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, 2025. doi: 10.1137/23M1613670.

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

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. URL https://www.jmlr.org/papers/v26/24-0902.html.

Yinuo Ren, Grant M. Rotskoff, and Lexing Ying. A unified approach to analysis and design of denoising markov models. *Journal of Machine Learning Research*, 27(105):1–69, 2026. URL https://www.jmlr.org/papers/v27/25-0693.html.

David Applebaum. *Lévy Processes and Stochastic Calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2 edition, 2009. doi: 10.1017/CBO9780511809781.

Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, 1994.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181.

Sidney I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York, 2007. doi: 10.1007/978-0-387-45024-7.

Gandhimohan M. Viswanathan, Ernesto Pessoa Raposo, and Marcos Gomes Eleutério da Luz. Lévy flights and superdiffusion in the context of biological encounters and random searches. *Physics of Life Reviews*, 5(3):133–150, 2008. doi: 10.1016/j.plrev.2008.03.002. URL https://doi.org/10.1016/j.plrev.2008.03.002.

Laurens de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006. doi: 10.1007/0-387-34471-3.

Souvik Ghosh and Sidney I. Resnick. A discussion on mean excess plots. *Stochastic Processes and their Applications*, 120(8):1492–1517, 2010. doi: 10.1016/j.spa.2010.04.002.

Michael Röckner, Longjie Xie, and Xicheng Zhang. Superposition principle for non-local Fokker–Planck–Kolmogorov operators. *Probability Theory and Related Fields*, 178(3–4):699–733, 2020. doi: 10.1007/s00440-020-00985-8. URL https://arxiv.org/abs/1910.11022v1.

Krzysztof Bogdan, Tomasz Grzywny, and Michał Ryznar. Density and tails of unimodal convolution semigroups. *Journal of Functional Analysis*, 266(6):3543–3571, 2014. URL https://arxiv.org/abs/1305.0976v2.

Kamil Kaleta and Paweł Sztonyk. Estimates of transition densities and their derivatives for jump Lévy processes. *Journal of Mathematical Analysis and Applications*, 431(1):260–282, 2015. doi: 10.1016/j.jmaa.2015.05.061. URL https://arxiv.org/abs/1307.1302v3.

Henrik Hult and Filip Lindskog. Extremal behavior of stochastic integrals driven by regularly varying Lévy processes. *The Annals of Probability*, 35(1):309–339, 2007. doi: 10.1214/009117906000000548. URL https://arxiv.org/abs/math/0703802.

Nicholas Hugh Bingham, Charles M. Goldie, and Edward Omey. Regularly varying probability densities. *Publications de l'Institut Mathématique*, 80(94):47–57, 2006. doi: 10.2298/PIM0694047B. URL https://www.ma.imperial.ac.uk/~bin06/Papers/bgo.pdf.

Michel Denuit, Patricia Ortega-Jiménez, and Christian-Yann Robert. Conditional expectations given the sum of independent random variables with regularly varying densities. *ASTIN Bulletin*, 55(2):449–485, 2025. doi: 10.1017/asb.2025.11.

Alex Dytso, Harold Vincent Poor, and Shlomo Shamai. A general derivative identity for the conditional mean estimator in Gaussian noise and some applications, 2021. URL https://arxiv.org/abs/2104.01883v1.

Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 59(4):1844–1881, 2023. doi: 10.1214/22-AIHP1320. URL https://arxiv.org/abs/2104.07708v3.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, Cham, 2014. doi: 10.1007/978-3-319-00227-9.

Péter Kevei. Regularly varying functions. Lecture notes, University of Szeged, n.d. URL https://www.math.u-szeged.hu/~kevei/tanitas/jegyzet/RegVar_notes.pdf.

National Institute of Standards and Technology. NIST Digital Library of Mathematical Functions, n.d. URL https://dlmf.nist.gov/8.11#E2. Section 8.11(i), equation 8.11.2; consulted on 10 September 2026.

René L. Schilling. An introduction to Lévy and Feller processes. Lecture notes, 2016. URL https://arxiv.org/abs/1603.00251.

Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975. doi: 10.1214/aos/1176343247.

Sergey Foss, Dmitry Korshunov, and Stan Zachary. Convolutions of long-tailed and subexponential distributions. *Journal of Applied Probability*, 46(3):756–767, 2009. doi: 10.1239/jap/1253279850.

John M. Chambers, Colin Lingwood Mallows, and Bart W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976. doi: 10.1080/01621459.1976.10480344.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL https://github.com/jax-ml/jax. Version 0.7.2.

Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL https://doi.org/10.1038/s41592-019-0686-2.

Appendix

Table of Contents

Notation **18**

A Proofs for Section 2: Stable Dynamics **20**
A.1 Strong solution and scale preservation . . . 20
A.2 Proof of Proposition 2: Fractional Tweedie formula . . . 20
A.3 Proof of Proposition 1: Matching marginals . . . 21
A.4 Proof of Proposition 3: Dissipativity . . . 22
A.5 Measurability and transient anti-contractivity . . . 22

B Proof of Theorem 4: The Non-Asymptotic Bound **23**
B.1 Dissipative reformulation and the EI scheme . . . 23
B.2 The early-stopping bias . . . 23
B.3 Exponential contraction of the initialization error . . . 23
B.4 Score and EI discretization error . . . 24
B.5 Assembly: the explicit bound . . . 27
B.6 Schedule corollaries and proof of Theorem 4 . . . 27

C Proof of Proposition 5: Denoiser Regularity under Compact Support **29**

D Proofs for Sections 4 and 5.1: Heavy-Tailed Targets **29**
D.1 Proof of Theorem 7: Heavy-tailed bound . . . 30
D.2 Proofs of Propositions 6 and 8: Denoiser regularity . . . 30
D.3 Proof of Corollary 9: Stable mixing . . . 34
D.4 Proof of Theorem 10: Brownian initialization error . . . 34

E Proofs for Section 5.2: Tail Regimes and Extreme Quantiles **36**
E.1 Proof of Proposition 12: Tail-regime dichotomy . . . 37
E.2 Brownian tail bounds for Proposition 14 . . . 37
E.3 Stable tail bounds for Proposition 14 . . . 38
E.4 Proofs of Lemma 13 and Proposition 14 . . . 39

F Numerical Protocol for Figures 1–4 **40**
F.1 Common setup and score evaluation . . . 40
F.2 Figure 1: Deterministic denoiser modulus . . . 41
F.3 Figure 2: Initialization error . . . 41
F.4 Figures 3 and 4: Tail coverage . . . 42
F.5 Implementation and numerical controls . . . 44

Notation

The following notation is used throughout the supplement. Vectors live in $\mathbb{R}^d$ and all stable laws are isotropic; $\alpha \in (1,2)$ and $\nu > 1$ are fixed, and subscripts t (resp. τ) denote forward (resp. reverse) time.

Table 2: Notation used throughout the supplement.

Symbol	Meaning	Symbol	Meaning
Indices, horizon, and schedule			
$\alpha \in (1,2)$	stability index of the driving noise	ν	target tail index, $\nu > 1$ (stable W_p bounds use $\nu \geq 1+\alpha$)
$d,\ p$	dimension; Wasserstein / moment order, $p \in [1,\alpha)$	$T,\ \varepsilon$	terminal time and early-stopping threshold
$\beta(t)$	noise schedule; $\beta_{\min} = \beta(\varepsilon)$ and $\beta(T)$ is its maximum on $[\varepsilon, T]$	$B(t)$	integrated schedule, $\int_0^t \beta(s)\,\mathrm{d}s$; $B_\varepsilon = B(\varepsilon)$, $B_T = B(T)$
$\bar{\beta}(\tau)$	reversed schedule, $\beta(T-\tau)$	$a(t)$	signal scale, $e^{-B(t)/\alpha}$ (in VP, $a_t = e^{-B(t)/2}$)
$\gamma(t)$	noise scale, $\gamma(t)^\alpha = 1 - a(t)^\alpha$ (in VP, $\gamma_t^2 = 1 - a_t^2$)	η/α	coefficient of the linear dissipative term
$\widetilde{m}(\tau)$	exponential attenuator, $\exp(-\frac{\eta}{\alpha}\int_0^\tau \bar{\beta})$	$h,\ N,\ \tau_k$	step size, number of steps, and grid points; $h = (T-\varepsilon)/N$
ΔB	per-step injected noise, $\Delta B \leq h\beta(T)$		
Processes and laws			
$L_t^\alpha,\ \overleftarrow{L}_\tau^\alpha$	forward and backward isotropic α-stable driving processes	ν_α	Lévy measure, $\nu_\alpha(\lVert z\rVert > \rho) = C_{d,\alpha}\rho^{-\alpha}$
X_t	forward process, $X_0 \sim \pi_{\text{data}}$	$\overleftarrow{Y}_\tau$	exact-score reverse process initialized at π_∞^α
$\bar{X}_{\tau_k}^{\theta,h}$	exponential-integrator (EI) scheme	$\pi_t,\ \pi_t^{(2)}$	stable and Brownian forward marginals and densities
$\mu_\tau^{(\alpha)},\ \mu_\tau^{(2)}$	stable and Brownian exact-score outputs under reference initialization; μ_τ locally	$\widehat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}$	sampled law after N steps
π_∞^α	stationary stable law, $\mathrm{S}\alpha\mathrm{S}_d(1)$	$\pi_\infty^{(2)}$	stationary VP law, $\mathcal{N}(0, I_d)$
$\mathrm{Law}(\cdot),\ \overset{\mathrm{d}}{=}$	law of a variable; equality in law	$\overleftarrow{Z}_\tau$	exact-score reverse process initialized at π_T, with law $\pi_{T-\tau}$
Fractional score, denoiser, and drift			
$\Delta^{(\alpha-2)/2}\nabla$	fractional gradient, multiplier $-i\xi\lVert\xi\rVert^{\alpha-2}$	$S_t^{(\alpha)}$	marginal fractional score
$S_t^{(\alpha)}(\cdot \mid x_0)$	conditional fractional score	$m_t(x)$	posterior barycenter; MSE-optimal denoiser when $X_0 \in L^2$
C_t	global denoiser modulus, $\sup_x \lVert\nabla m_t(x)\rVert_{\mathrm{op}}$	$\bar{C}_\varepsilon$	uniform modulus, $\sup_{t\geq\varepsilon} C_t$, in Assumption A4
$C_t^{(\alpha)}$	score dissipativity constant, $(1 - a(t)C_t)/(\alpha\gamma(t)^\alpha)$; $C_t^{(2)}$ is its VP analogue	$\bar{b}(\tau, x),\ \bar{L}(\tau)$	backward drift and one-sided Lipschitz constant
$S_t^{(\alpha),\mathrm{res}}$	stationary-score residual, $x + \alpha S_t^{(\alpha)}(x) = \alpha(S_t^{(\alpha)} - S_\infty^{(\alpha)})$	$L_t^{\mathrm{res}},\ C_t^{\mathrm{res}}$	Lipschitz and one-sided Lipschitz constants of the residual
$s = \nabla \log \varphi_\alpha$	score of the standard stable density	$s_\theta,\ h_\theta$	learned score and learned residual map $x + \alpha s_\theta$
χ_t	direct-time contraction coefficient, $(1+\eta)C_t^{(\alpha)} - 1/\alpha$	$\mathfrak{A}_\varepsilon$	cumulative non-contractivity after early stopping

Symbol	Meaning	Symbol	Meaning
$\mathfrak{D}_\varepsilon^{\mathrm{mix}}$	cumulative deficit from the asymptotic mixing rate	$\mathcal{G}_{\varepsilon,h}$	uniform bound on products of EI one-step factors
Constants of the standard stable law			
φ_α	density of $\mathrm{S}\alpha\mathrm{S}_d(1)$	$k_{d,\alpha}$	density tail constant, $\varphi_\alpha(z) \sim k_{d,\alpha}\|z\|^{-(d+\alpha)}$
$C_{d,\alpha}$	radial tail constant, $\omega_{d-1}k_{d,\alpha}/\alpha$	$c_{d,\alpha,p}$	p-moment constant, $\mathbb{E}[\|Z\|^p]^{1/p}$, finite iff $p<\alpha$
$c_\varphi,\ C_\varphi$	two-sided envelope constants of φ_α	$\Lambda_1,\ \Lambda_2$	$\sup\|s\|$ and $\sup\|\nabla s\|$
Λ	$\sup\|\nabla^2 \log\varphi_\alpha\|$		
Target distribution and its tail			
π_{data} (π locally)	target law and density, represented by the same symbol	$\overline{G}_{\pi_{\mathrm{data}}}(r)$	radial survival, $\mathbb{P}(\|X_0\|>r)$
$c_{\pi,-},\ c_{\pi,+},\ R_\pi$	two-sided density-envelope constants and threshold	$M_1,\ M_2$	$\mathbb{E}\|X_0\|$ and $\mathbb{E}\|X_0\|^2$
$S_{\varepsilon,T}$	uniform bound on the smoothed Brownian score for $s\in[\varepsilon,T]$	$M_*^{(p)}$	uniform conditional p-th moment
$B_\varepsilon^{(p)}$	uniform moment bound for the ideal reverse process	R	support radius in the compact case
Effective tail scales of the generated law			
ρ_τ ($\rho_{\nu,\tau}$ locally)	scale of the exactly stable leading term $\mathcal{W}_\tau$	$E(\tau,s),\ E_\nu(\tau,s)$	gains of the corresponding linear flows from s to τ
c_τ	tail constant, $\overline{G}_{\mu_\tau^{(\alpha)}}(r)\sim c_\tau r^{-\alpha}$	$\sigma_\tau^2,\ K_\tau$	Brownian variance and bounded-remainder radius
Tail-regime and extreme-quantile functionals			
$e_\mu(u)$	radial mean excess, $\mathbb{E}_\mu[\|X\|-u \mid \|X\|>u]$	$\Gamma(\mu)$	limiting relative mean excess, $\lim_{u\to\infty} e_\mu(u)/u$
$q_\kappa(\mu)$	radial κ-quantile	$\mathcal{M}_\mu(\kappa)$	generated-to-target extreme-quantile ratio
$\mathrm{osc}_S(f)$	oscillation, $\sup_S f-\inf_S f$		
Error terms and metrics			
$\mathcal{E}_0,\ \mathcal{E}_1,\ \mathcal{E}_2$	early-stopping, mixing, and score-plus-discretization errors	err	score-approximation error in Assumption A2
$L_\varepsilon^{\mathrm{time}}$	time-regularity constant in Assumption A3	W_p	order-p Wasserstein distance
$\|\cdot\|_{\mathrm{TV}}$	total-variation distance	$\mathrm{KL}(\cdot\,\|\,\cdot)$	Kullback–Leibler divergence
Miscellaneous			
I_d	$d\times d$ identity matrix	$\varphi_\gamma,\ \varphi$	Brownian density of scale γ; standard Brownian density
ω_{d-1}	surface area of $\mathbb{S}^{d-1}$	$\bar{B}(0,R)$	closed Euclidean ball of radius R
$\langle\cdot,\cdot\rangle,\ \|\cdot\|$	Euclidean inner product and norm	$(x)_+$	positive part, $\max(x,0)$

The constants $k_{d,\alpha}$, $C_{d,\alpha}$, and $c_{d,\alpha,p}$ are, respectively, density-tail, radial-tail, and moment constants; despite their similar notation, they are not interchangeable.

A Proofs for Section 2: Stable Dynamics

The explicit stable Ornstein–Uhlenbeck transition law gives both scale preservation and the conditional score. The Fourier convention stated below (2.2) then identifies the marginal score and the reverse marginals.

A.1 Strong solution and scale preservation

Lemma 15 (Strong solution)**.** *The SP-SDE* (2.1) *admits the unique strong solution*

$$X_t = a(t)X_0 + \int_0^t \frac{a(t)}{a(s)}\,\beta(s)^{1/\alpha}\,\mathrm{d}L_s^\alpha, \tag{A.1}$$

and $X_t \overset{\mathcal{L}}{=} a(t)X_0 + \gamma(t)Z$ *with* $Z \sim \mathrm{S}\alpha\mathrm{S}_d(1)$ *independent of* X_0*. If* $X_0 \sim \mathrm{S}\alpha\mathrm{S}_d(1)$*, then* $X_t \sim \mathrm{S}\alpha\mathrm{S}_d(1)$ *for all* t*.*

Proof. Variation of constants gives (A.1) and uniqueness. For a deterministic scalar $k \in L^\alpha$, independence of stable increments gives $\mathbb{E}e^{i\xi\cdot\int k\,\mathrm{d}L^\alpha} = \exp(-\|\xi\|^\alpha \int |k|^\alpha)$, first for simple k and then by approximation. Here the scale to the power α is

$$\int_0^t \left(\frac{a(t)}{a(s)}\right)^\alpha \beta(s)\,\mathrm{d}s = \int_0^t \beta(s)e^{-[B(t)-B(s)]}\,\mathrm{d}s = 1 - e^{-B(t)} = \gamma(t)^\alpha.$$

The integral is independent of X_0, and scale preservation follows from $a(t)^\alpha + \gamma(t)^\alpha = 1$. □

A.2 Proof of Proposition 2: Fractional Tweedie formula

The following conditional-score identity is standard for the Lévy–Itô model (Yoon et al., 2023). We include its short derivation to fix our Fourier and stable-scale conventions.

Lemma 16 (Conditional score)**.** *The conditional fractional score is exactly linear:*

$$S_t^{(\alpha)}(x \mid x_0) = -\frac{1}{\alpha\gamma(t)^\alpha}(x - a(t)x_0). \tag{A.2}$$

Proof. This is the conditional fractional-score formula of Yoon et al. (2023, Theorem 4.4), with our stable-scale convention. Indeed, for $f(x) = \pi_t(x \mid x_0)$, $a = a(t)$ and $\gamma = \gamma(t)$,

$$\widehat{f}(\xi) = e^{iax_0\cdot\xi - \gamma^\alpha\|\xi\|^\alpha}, \qquad \mathcal{F}[(x - ax_0)f](\xi) = -\alpha\gamma^\alpha(-i\xi\|\xi\|^{\alpha-2})\widehat{f}(\xi).$$

Fourier inversion and division by $f > 0$ give (A.2). □

Proof of Proposition 2. For $t > 0$, the envelope $(1+\|\xi\|^{\alpha-1})e^{-\gamma(t)^\alpha\|\xi\|^\alpha}$ is integrable and independent of x_0. Fourier inversion and Fubini therefore allow $\Delta^{(\alpha-2)/2}\nabla$ to commute with the mixing integral. For fixed (t, x), the map $x_0 \mapsto \|x_0\|\varphi_\alpha((x - a(t)x_0)/\gamma(t))$ is bounded by the stable density envelope. Thus the conditional first moment is finite even without an unconditional moment assumption. Integrating (A.2) against the conditional law yields

$$S_t^{(\alpha)}(x) = \int S_t^{(\alpha)}(x \mid x_0)\,\mathbb{P}(X_0 \in \mathrm{d}x_0 \mid X_t = x) = \frac{a(t)m_t(x) - x}{\alpha\gamma(t)^\alpha}.$$

□

A.3 Proof of Proposition 1: Matching marginals

The exact time reversal of (2.1) is a jump process whose jump kernel is the stable Lévy measure tilted by a ratio of marginals, $\bar{\beta}(\tau)\nu_\alpha(\mathrm{d}z)\,\pi_{T-\tau}(x+z)/\pi_{T-\tau}(x)$ (Conforti and Léonard, 2022). It is not driven by a pure Lévy process and is not used here. We justify instead the matching-marginals family (2.4) in two steps.

Set $H := T-\varepsilon$ and $\bar{\beta}(\tau) := \beta(T-\tau)$. For $r \geq 1$, write $\mathcal{P}_r(\mathbb{R}^d)$ for the probability measures with finite r-th moment. For a compact interval I, let $\mathcal{C}_1(I)$ denote the narrowly continuous curves $(\mu_t)_{t\in I}$ in $\mathcal{P}(\mathbb{R}^d)$ such that $\int_I\int_{\mathbb{R}^d}\|x\|\,\mu_t(\mathrm{d}x)\,\mathrm{d}t < \infty$. Here $\Delta^{\alpha/2} = -(-\Delta)^{\alpha/2}$ denotes the generator with Fourier multiplier $-\|\xi\|^\alpha$. In the weak sense, the forward Fokker–Planck Cauchy problem on $[\varepsilon, T]$ is

$$\int \varphi\,\mathrm{d}\nu_t - \int \varphi\,\mathrm{d}\pi_\varepsilon = \int_\varepsilon^t \int \Big[-\frac{\beta(s)}{\alpha}x\cdot\nabla\varphi(x) + \beta(s)\Delta^{\alpha/2}\varphi(x)\Big]\,\nu_s(\mathrm{d}x)\,\mathrm{d}s, \tag{A.3}$$

for $t \in [\varepsilon, T]$ and every $\varphi \in C_c^2(\mathbb{R}^d)$. The corresponding reverse problem on $[0, H]$ is

$$\begin{aligned}\int \varphi\,\mathrm{d}\mu_\tau - \int \varphi\,\mathrm{d}\pi_T = &\int_0^\tau \int \bar{b}(u,x)\cdot\nabla\varphi(x)\,\mu_u(\mathrm{d}x)\,\mathrm{d}u \\ &+ \eta\int_0^\tau \bar{\beta}(u)\int \Delta^{\alpha/2}\varphi(x)\,\mu_u(\mathrm{d}x)\,\mathrm{d}u.\end{aligned} \tag{A.4}$$

Proposition 17 (Specialization of Popov et al.'s theorem)**.** *Assume $\pi_{\mathrm{data}} \in \mathcal{P}_1(\mathbb{R}^d)$ and the following sufficient weak formulation of the well-posedness requirements underlying Popov et al. (2025, Theorem 1 and Appendix A):*

(P1) *the forward SDE (2.1), restricted to $[\varepsilon, T]$ with initial law π_ε, and the reverse SDE (2.4) admit weak solutions whose marginal curves belong, respectively, to $\mathcal{C}_1([\varepsilon, T])$ and $\mathcal{C}_1([0, H])$;*

(P2) *$(\pi_t)_{t\in[\varepsilon,T]}$ is the unique solution of (A.3) in $\mathcal{C}_1([\varepsilon, T])$, and (A.4) has a unique solution in $\mathcal{C}_1([0, H])$.*

Then every weak solution of (2.4) whose marginal curve belongs to $\mathcal{C}_1([0, H])$ satisfies $\mathrm{Law}(\overleftarrow{Z}_\tau) = \pi_{T-\tau}, \qquad 0 \leq \tau \leq H$.

Proof. Specialize the Fokker–Planck calculation of Popov et al. (2025, Theorem 1 and Appendix A) to $\mu(x,t) = -\beta(t)x/\alpha$, $\sigma_t = \beta(t)^{1/\alpha}$ and $\eta_t = \eta$. For $\mu_\tau = \pi_{T-\tau}$, the identity $\mathrm{div}(\mu_\tau S^{(\alpha)}_{T-\tau}) = \Delta^{\alpha/2}\mu_\tau$ gives, in the weak sense,

$$-\,\mathrm{div}(\bar{b}\,\mu_\tau) + \eta\bar{\beta}(\tau)\Delta^{\alpha/2}\mu_\tau = -\frac{\bar{\beta}(\tau)}{\alpha}\,\mathrm{div}(x\mu_\tau) - \bar{\beta}(\tau)\Delta^{\alpha/2}\mu_\tau = \partial_\tau\mu_\tau.$$

This curve belongs to $\mathcal{C}_1([0, H])$ and starts from π_T. Under (P1) the reverse marginals solve the same weak problem; uniqueness in (P2) identifies them with μ_τ. □

Proposition 18 (Verification of the Popov–et–al. well-posedness hypothesis)**.** *Assume $\pi_{\mathrm{data}} \in \mathcal{P}_1(\mathbb{R}^d)$ and that the drift (2.3) is Borel measurable, globally Lipschitz in x uniformly in $\tau \in [0, H]$, and satisfies $\sup_{\tau\in[0,H]}\|\bar{b}(\tau, 0)\| < \infty$. Then conditions (P1)–(P2) of Proposition 17 hold. Consequently, Proposition 1 holds.*

Moreover, the stated regularity conditions on $\bar{b}$ follow from

$$\bar{C}_{\varepsilon,T} := \sup_{t\in[\varepsilon,T]} C_t < \infty.$$

In particular, since $\bar{C}_{\varepsilon,T} \leq \bar{C}_\varepsilon$, Assumption A4 together with $\pi_{\mathrm{data}} \in \mathcal{P}_1(\mathbb{R}^d)$ verifies the conditions of Proposition 1. In that case,

$$\sup_{\tau\in[0,H]} \mathrm{Lip}_x(\bar{b}(\tau,\cdot)) \leq \|\beta\|_{L^\infty([\varepsilon,T])}\Big(\tfrac{1}{\alpha} + (1+\eta)\tfrac{1+\bar{C}_{\varepsilon,T}}{\alpha\gamma(\varepsilon)^\alpha}\Big). \tag{A.5}$$

Proof. Subtracting the additive stochastic integral reduces either SDE to a pathwise Carathéodory equation with a Lipschitz vector field of linear growth. This gives a non-explosive strong solution and uniqueness from every starting time. The additive stable integral is a martingale with finite r-th moments for $1 < r < \alpha$; Doob's L^r inequality gives a finite maximal first moment on $[0, H]$. Grönwall therefore gives (P1). The jump measures are $\beta(t)\nu_\alpha$ and $\eta\bar{\beta}(\tau)\nu_\alpha$, with bounded multipliers. Their required characteristics satisfy

$$\int_{\|z\|\le 1} \|z\|^2 \nu_\alpha(\mathrm{d}z) < \infty, \qquad \int_{\|z\|>1} \log(1+\|z\|)\nu_\alpha(\mathrm{d}z) < \infty.$$

Together with linear drift growth, these bounds verify condition (1.17) of Röckner et al. (2020, arXiv version 1). Extend the coefficients past the fixed interval with the same bounds. Strong well-posedness gives well-posed martingale problems; their Corollary 1.8 then gives (P2). Finally, (2.5) gives $\mathrm{Lip}(S_s^{(\alpha)}) \le (1+a(s)C_s)/(\alpha\gamma(s)^\alpha)$, hence (A.5). In the integral formula for $m_s(0)$ the numerator is uniformly bounded by $C\mathbb{E}\|X_0\|$ and the positive continuous denominator is bounded away from zero on $[\varepsilon, T]$. This also bounds $\bar{b}(\tau, 0)$; the same formula gives its measurability. □

A.4 Proof of Proposition 3: Dissipativity

Proof of Proposition 3. By (2.5) and the definition of C_t,

$$\langle S_t^{(\alpha)}(x) - S_t^{(\alpha)}(y), x-y\rangle \le -\frac{1-a(t)C_t}{\alpha\gamma(t)^\alpha}\|x-y\|^2.$$

This proves the score estimate. Substitution into (2.3) gives (2.6), with $\bar{L}(\tau) = \bar{\beta}(\tau)(1/\alpha - (1+\eta)C_{T-\tau}^{(\alpha)})$. □

A.5 Measurability and transient anti-contractivity

Lemma 19 (Measurability of the denoiser modulus)**.** *Assume $\pi_{\mathrm{data}} \in \mathcal{P}_1(\mathbb{R}^d)$. Then $(t,x) \mapsto \nabla m_t(x)$ is continuous on $(0,\infty)\times\mathbb{R}^d$ and $t \mapsto C_t$ is Borel measurable, with values in $[0,\infty]$.*

Proof. Write $m = N/D$ using the conditional numerator and normalizer. Boundedness of $\varphi_\alpha, \nabla\varphi_\alpha$, the first moment of π, and $\inf_{[\varepsilon,T]}\gamma > 0$ justify differentiation and joint dominated convergence. Since $D > 0$, ∇m is jointly continuous. Consequently $C_t = \sup_{x\in\mathbb{Q}^d}\|\nabla m_t(x)\|_{\mathrm{op}}$ is a countable supremum of continuous functions, hence Borel measurable. □

Define the direct-time contraction coefficient

$$\chi_t := (1+\eta)C_t^{(\alpha)} - \frac{1}{\alpha}, \qquad \bar{L}(\tau) = -\bar{\beta}(\tau)\chi_{T-\tau}, \tag{A.6}$$

and, for a fixed early-stopping time $\varepsilon > 0$, the cumulative non-contractivity $\mathfrak{A}_\varepsilon := \int_\varepsilon^\infty \beta(t)(\chi_t)_-\,\mathrm{d}t$. By Lemma 19, this integral is well defined.

Lemma 20 (Finite transient amplification)**.** *Under Assumptions A1 and A4, and $\pi_{\mathrm{data}} \in \mathcal{P}_1(\mathbb{R}^d)$, one has $\chi_t \to \eta/\alpha$. In particular, there exists $t_+ \ge \varepsilon$, independent of the horizon T, such that*

$$\chi_t \ge \frac{\eta}{2\alpha}, \qquad t \ge t_+, \tag{A.7}$$

and $\mathfrak{A}_\varepsilon < \infty$. More precisely,

$$\mathfrak{A}_\varepsilon \le K_\varepsilon(B(t_+) - B(\varepsilon)), \qquad K_\varepsilon := \frac{1}{\alpha} + \frac{1+\eta}{\alpha\gamma(\varepsilon)^\alpha}(a(\varepsilon)\bar{C}_\varepsilon - 1)_+. \tag{A.8}$$

Proof. The bound $a(t)C_t \le a(t)\bar{C}_\varepsilon \to 0$ gives $C_t^{(\alpha)} \to 1/\alpha$ and $\chi_t \to \eta/\alpha$. Thus (A.7) holds after a finite t_+ independent of T. Moreover, for $t \ge \varepsilon$,

$$(\chi_t)_- \le \frac{1}{\alpha} + \frac{1+\eta}{\alpha\gamma(\varepsilon)^\alpha}(a(\varepsilon)\bar{C}_\varepsilon - 1)_+ = K_\varepsilon.$$

The negative part vanishes after t_+; integration gives (A.8). □

B Proof of Theorem 4: The Non-Asymptotic Bound

We follow the error-propagation structure of Strasman et al. (2025, Appendix C.2), adapting it to stable noise in L^p, $p < \alpha$, and allowing a finite non-contractive phase.

We prove the bound under Assumptions A1–A3 and A4 for a target supported in $\bar{B}(0, R)$. Appendix D treats the heavy-tailed case. Propositions 5 and 6, respectively, verify Assumption A4 for these two target classes. Write $\bar{\beta}(\tau) = \beta(T - \tau)$.

B.1 Dissipative reformulation and the EI scheme

Since $S_\infty^{(\alpha)}(x) = -x/\alpha$, define the stationary-score residual map $S_t^{(\alpha),\mathrm{res}}(x) := x + \alpha S_t^{(\alpha)}(x) = \alpha(S_t^{(\alpha)}(x) - S_\infty^{(\alpha)}(x))$. Then $\bar{b}(\tau, x) = -\frac{\eta}{\alpha}\,\bar{\beta}(\tau)\,x + \frac{1+\eta}{\alpha}\,\bar{\beta}(\tau)\,S_{T-\tau}^{(\alpha),\mathrm{res}}(x)$, and $S_\infty^{(\alpha),\mathrm{res}} \equiv 0$. On the uniform grid $0 = \tau_0 < \cdots < \tau_N = T - \varepsilon$, the EI scheme is

$$\bar{X}_{\tau_{k+1}}^{\theta,h} = \widetilde{m}_k\,\bar{X}_{\tau_k}^{\theta,h} + \frac{1+\eta}{\alpha}\int_{\tau_k}^{\tau_{k+1}} \frac{\widetilde{m}(\tau_{k+1})}{\widetilde{m}(u)}\,\bar{\beta}(u)\,h_\theta(\tau_k, \bar{X}_{\tau_k}^{\theta,h})\,\mathrm{d}u + \int_{\tau_k}^{\tau_{k+1}} \frac{\widetilde{m}(\tau_{k+1})}{\widetilde{m}(u)}\,(\eta\,\bar{\beta}(u))^{1/\alpha}\,\mathrm{d}\overleftarrow{L}_u^\alpha, \tag{B.1}$$

where $h_\theta(t, x) := x + \alpha s_\theta(t, x)$ and the exponential attenuator and its one-step ratio are $\widetilde{m}(\tau) := \exp\Big(-\frac{\eta}{\alpha}\int_0^\tau \bar{\beta}(v)\,\mathrm{d}v\Big)$ and $\widetilde{m}_k := \frac{\widetilde{m}(\tau_{k+1})}{\widetilde{m}(\tau_k)}$. The attenuator is decreasing. Define $L_t^{\mathrm{res}} := 1 + \frac{1+a(t)C_t}{\gamma(t)^\alpha}$ and $C_t^{\mathrm{res}} := 1 - \alpha C_t^{(\alpha)}$, which respectively bound the Lipschitz and one-sided Lipschitz constants of $S_t^{(\alpha),\mathrm{res}}$.

B.2 The early-stopping bias

Lemma 21 (Early-stopping bias)**.** *For every $\varepsilon \in (0, T)$ and $p \in [1, \alpha)$,*

$$\mathcal{E}_0 = W_p(\pi_{\mathrm{data}}, \pi_\varepsilon) \;\le\; (1 - a(\varepsilon))\,\mathbb{E}[\|X_0\|^p]^{1/p} + \gamma(\varepsilon)\,c_{d,\alpha,p}.$$

Proof. By Lemma 15, $X_\varepsilon \overset{\mathcal{L}}{=} a(\varepsilon)X_0 + \gamma(\varepsilon)Z_\alpha$ with $Z_\alpha \sim \mathrm{S\alpha S}_d(1)$ independent of X_0. Coupling π_{data} and π_ε through $(X_0,\, a(\varepsilon)X_0 + \gamma(\varepsilon)Z_\alpha)$ and applying Minkowski's inequality, $W_p(\pi_{\mathrm{data}}, \pi_\varepsilon) \le \|X_0 - a(\varepsilon)X_0 - \gamma(\varepsilon)Z_\alpha\|_{L^p} \le (1 - a(\varepsilon))\,\|X_0\|_{L^p} + \gamma(\varepsilon)\,\|Z_\alpha\|_{L^p}$. □

B.3 Exponential contraction of the initialization error

Lemma 22 (Forward contraction)**.** *For every $p \in [1, \alpha)$, $W_p(\pi_T, \pi_\infty^\alpha) \le a(T)\,W_p(\pi_{\mathrm{data}}, \pi_\infty^\alpha)$.*

Proof. For any coupling of the two initial laws, use the same forward noise. Equation (A.1) gives $X_T - X_T' = a(T)(X_0 - X_0')$. Taking L^p norms and the infimum over initial couplings proves the claim, since the second marginal is invariant. □

Lemma 23 (Backward propagation of the mixing error)**.** *For every $p \in [1, \alpha)$ and $\varepsilon \in (0, T)$,*

$$\mathcal{E}_1 \;\le\; W_p(\pi_{\mathrm{data}}, \pi_\infty^\alpha)\,\exp\Big(-\frac{1+\eta}{\alpha}\int_\varepsilon^T \beta(t)\,\frac{1 - a(t)C_t}{\gamma(t)^\alpha}\,\mathrm{d}t\Big). \tag{B.2}$$

If $\mathfrak{A}_\varepsilon < \infty$, then

$$\mathcal{E}_1 \le e^{\mathfrak{A}_\varepsilon} W_p(\pi_{\text{data}}, \pi_\infty^\alpha) e^{-\frac{1}{\alpha}(B(T)-B(\varepsilon))}. \tag{B.3}$$

Moreover, define $\mathfrak{D}_\varepsilon^{\text{mix}} := \int_\varepsilon^\infty \beta(t) \left(\frac{\eta}{\alpha} - \chi_t\right)_+ \mathrm{d}t$. If $\mathfrak{D}_\varepsilon^{\text{mix}} < \infty$, then

$$\mathcal{E}_1 \le e^{\mathfrak{D}_\varepsilon^{\text{mix}}} W_p(\pi_{\text{data}}, \pi_\infty^\alpha) e^{-\frac{1+\eta}{\alpha}(B(T)-B(\varepsilon))}. \tag{B.4}$$

Proof. Realize the two well-posed backward SDEs with common noise and any coupling of $(\pi_\infty^\alpha, \pi_T)$, independent of that noise. Their difference D is absolutely continuous, and (2.6) gives $\frac{\mathrm{d}}{\mathrm{d}\tau}\|D_\tau\|^2 \le 2\bar{L}(\tau)\|D_\tau\|^2$. Grönwall, the infimum over couplings, the marginal identity, and Lemma 22 yield

$$\mathcal{E}_1 \le W_p(\pi_{\text{data}}, \pi_\infty^\alpha) a(T) \exp\left(-\int_\varepsilon^T \beta(t)\chi_t \,\mathrm{d}t\right).$$

The logarithm of this last factor equals $-B(\varepsilon)/\alpha - (1+\eta)\int_\varepsilon^T \beta(t) C_t^{(\alpha)} \,\mathrm{d}t$. Dropping $-B(\varepsilon)/\alpha$ gives (B.2). Writing its exponent as $-[B(T)-B(\varepsilon)]/\alpha - \int_\varepsilon^T \beta\chi$ or as $-(1+\eta)[B(T)-B(\varepsilon)]/\alpha - \int_\varepsilon^T \beta(\chi - \eta/\alpha)$ then gives (B.3) and (B.4), respectively. □

B.4 Score and EI discretization error

We first record a uniform moment bound for the ideal process, used to close the recursion.

Proposition 24 (Uniform moment bound)**.** *Under Assumptions A1 and A4, for $p \in [1, \alpha)$ and $\pi_{\text{data}} \in \mathcal{P}_p(\mathbb{R}^d)$,*

$$\sup_{\tau \in [0, T-\varepsilon]} \mathbb{E}\left[\left\|\overleftarrow{Y}_\tau\right\|^p\right]^{1/p} \le (1 + e^{\mathfrak{A}_\varepsilon})\left(\mathbb{E}[\|X_0\|^p]^{1/p} + c_{d,\alpha,p}\right) =: B_\varepsilon^{(p)}. \tag{B.5}$$

Proof. Lemma 20 gives $\mathfrak{A}_\varepsilon < \infty$. Take any $\zeta \in \Pi(\pi_\infty^\alpha, \pi_T)$ and drive $\overleftarrow{Y}$ and the matching-marginals process $\overleftarrow{Z}$ by the same stable noise, independent of their initial pair of law ζ. Proposition 18 identifies $\mathrm{Law}(\overleftarrow{Z}_\tau) = \pi_{T-\tau}$, so Minkowski and the forward representation give $\|\overleftarrow{Z}_\tau\|_{L^p} \le \|X_0\|_{L^p} + c_{d,\alpha,p}$. The synchronous gap bound and the triangle inequality yield

$$\|\overleftarrow{Y}_\tau\|_{L^p} \le \|X_0\|_{L^p} + c_{d,\alpha,p} + e^{\mathfrak{A}_\varepsilon}\left(\int \|y-z\|^p \,\zeta(\mathrm{d}y, \mathrm{d}z)\right)^{1/p}.$$

The left-hand side is independent of ζ. Taking the infimum gives $W_p(\pi_\infty^\alpha, \pi_T)$ in the last term. Forward contraction (Lemma 22) and an independent coupling then bound it by $W_p(\pi_\infty^\alpha, \pi_{\text{data}}) \le \|X_0\|_{L^p} + c_{d,\alpha,p}$, proving (B.5). □

We now bound $\mathcal{E}_2$, tracking the discretization gap $D_k := \bar{X}_{\tau_k}^{\theta,h} - \overleftarrow{Y}_{\tau_k}$ through a one-step contraction analysis and a telescoping sum.

Proposition 25 (Cumulative bound)**.** *Under Assumptions A1–A3 and A4, for a target either supported in $\bar{B}(0,R)$ or satisfying Assumption A5 with $\nu \ge 1+\alpha$ (in which case R is replaced by M_1), fix arbitrary $0 < \varepsilon < T$ and let $h \le h_0(\varepsilon, T)$ be small enough. Then*

$$\mathcal{E}_2 \le \mathcal{G}_{\varepsilon,h}\left\{(1+\eta)\left[\mathrm{err} + L_\varepsilon^{\text{time}} h\,(1+B_\varepsilon^{(p)})\right](B(T)-B(\varepsilon)) + \frac{1+\eta}{\alpha}\sum_{k=0}^{N-1}\left(\int_{\tau_k}^{\tau_{k+1}} \bar{\beta}(u)\, L_{T-u}^{\text{res}} \,\mathrm{d}u\right)\Theta_k\right\} \tag{B.6}$$

with $\Theta_k := (B_\varepsilon^{(p)} + R)\left[\frac{\Delta B_k}{\alpha} + \frac{1+\eta}{\alpha}\int_{\tau_k}^{\tau_{k+1}} \bar{\beta}(v)\,\frac{1+a(T-v)C_{T-v}}{\gamma(T-v)^\alpha}\,\mathrm{d}v\right] + c_{d,\alpha,p}\,\eta^{1/\alpha}\,\Delta B_k^{1/\alpha}$, where $\Delta B_k := \int_{\tau_k}^{\tau_{k+1}} \bar{\beta} \le h\beta(T)$ and $\mathcal{G}_{\varepsilon,h}$ is defined in (B.15) below.

Proof. Both target classes belong to $\mathcal{P}_1(\mathbb{R}^d)$, so Assumption A4 and Proposition 18 verify Proposition 1. Lemma 20 also provides t_+ and $\mathfrak{A}_\varepsilon < \infty$. Set $r_k(u) := \widetilde{m}(\tau_{k+1})/\widetilde{m}(u) \le 1$ for $u \in [\tau_k, \tau_{k+1}]$. Write $D_k := \bar{X}^{\theta,h}_{\tau_k} - \overleftarrow{Y}_{\tau_k}$ for the discretization gap. Subtracting the integral form of the exact SDE for $\overleftarrow{Y}$ from the scheme (B.1) (the shared jump integral cancels) and inserting $\pm S^{(\alpha),\mathrm{res}}_{T-u}(\bar{X}^{\theta,h}_{\tau_k})$ and $\pm S^{(\alpha),\mathrm{res}}_{T-u}(\overleftarrow{Y}_{\tau_k})$ splits the one-step gap as $D_{k+1} = A_k + R^{(1)}_k + R^{(2)}_k$, with

$$A_k := \widetilde{m}_k D_k + \tfrac{1+\eta}{\alpha} \int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u)\big[S^{(\alpha),\mathrm{res}}_{T-u}(\bar{X}^{\theta,h}_{\tau_k}) - S^{(\alpha),\mathrm{res}}_{T-u}(\overleftarrow{Y}_{\tau_k})\big]\,\mathrm{d}u, \tag{B.7}$$

$$R^{(1)}_k := \tfrac{1+\eta}{\alpha} \int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u)\big[S^{(\alpha),\mathrm{res}}_{T-u}(\overleftarrow{Y}_{\tau_k}) - S^{(\alpha),\mathrm{res}}_{T-u}(\overleftarrow{Y}_{u})\big]\,\mathrm{d}u, \tag{B.8}$$

$$R^{(2)}_k := \tfrac{1+\eta}{\alpha} \int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u)\big[h_\theta(\tau_k, \bar{X}^{\theta,h}_{\tau_k}) - S^{(\alpha),\mathrm{res}}_{T-u}(\bar{X}^{\theta,h}_{\tau_k})\big]\,\mathrm{d}u, \tag{B.9}$$

The three terms are, respectively, a contraction of the previous gap (A_k), the local fluctuation of the ideal process ($R^{(1)}_k$), and the modeling-plus-time-freezing residual ($R^{(2)}_k$). We bound each in L^p.

Step 1 (one-step contraction of A_k). Expanding $\|A_k\|^2 \le \delta_k^2 \|D_k\|^2$ pathwise from (B.7), the deterministic factor δ_k is

$$\delta_k^2 = \underbrace{\widetilde{m}_k^2}_{(a)} + \underbrace{2\widetilde{m}_k \tfrac{1+\eta}{\alpha} \int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u) C^{\mathrm{res}}_{T-u}\,\mathrm{d}u}_{(b)} + \underbrace{\Big(\tfrac{1+\eta}{\alpha} \int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u) L^{\mathrm{res}}_{T-u}\,\mathrm{d}u\Big)^2}_{(c)}. \tag{B.10}$$

We bound the three terms without expanding $\bar{\beta}$ or the score constants at the left endpoint, neither of which is assumed Lipschitz in time. Write $I_k := \int_{\tau_k}^{\tau_{k+1}} \bar{\beta}$, so that $\widetilde{m}_k = e^{-\frac{\eta}{\alpha} I_k}$ and $r_k(u) = \exp(-\frac{\eta}{\alpha}\int_u^{\tau_{k+1}} \bar{\beta}) \in (0,1]$. The substitution $v = \int_u^{\tau_{k+1}} \bar{\beta}$, for which $\mathrm{d}v = -\bar{\beta}(u)\,\mathrm{d}u$, gives the *exact* identity

$$\int_{\tau_k}^{\tau_{k+1}} r_k(u)\,\bar{\beta}(u)\,\mathrm{d}u = \int_0^{I_k} e^{-\frac{\eta}{\alpha}v}\,\mathrm{d}v = \frac{\alpha}{\eta}(1-\widetilde{m}_k). \tag{B.11}$$

Set $J_k := \int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u)\chi_{T-u}\,\mathrm{d}u$, and $Q_k := \frac{1+\eta}{\alpha}\int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u)L^{\mathrm{res}}_{T-u}\,\mathrm{d}u$. Since $\frac{1+\eta}{\alpha}C^{\mathrm{res}}_t = \frac{\eta}{\alpha} - \chi_t$ and $\frac{\eta}{\alpha}\int_{\tau_k}^{\tau_{k+1}} r_k(u)\bar{\beta}(u)\,\mathrm{d}u = 1 - \widetilde{m}_k$, the factor in (B.10) satisfies the exact identity

$$\delta_k^2 = 1 - (1-\widetilde{m}_k)^2 - 2\widetilde{m}_k J_k + Q_k^2. \tag{B.12}$$

Furthermore,

$$L^{\mathrm{res}}_t \le L^{\mathrm{res}}_\varepsilon := 1 + \frac{1 + a(\varepsilon)\bar{C}_\varepsilon}{\gamma(\varepsilon)^\alpha}, \qquad Q_k \le C_L I_k, \qquad C_L := \frac{1+\eta}{\alpha} L^{\mathrm{res}}_\varepsilon.$$

For every step, (B.12) gives

$$\delta_k \le 1 + (J_k)_- + \frac{1}{2}C_L^2 I_k^2, \tag{B.13}$$

because $\delta_k^2 \le 1 + 2(J_k)_- + C_L^2 I_k^2$ and $\sqrt{1+y} \le 1 + y/2$ for $y \ge 0$.

If the direct-time cell $\mathcal{I}_k := [T - \tau_{k+1}, T - \tau_k]$ is contained in $[t_+, \infty)$, then (A.7) and (B.11) give $J_k \ge (1-\widetilde{m}_k)/2$. Hence $\delta_k^2 \le \widetilde{m}_k + C_L^2 I_k^2$. If $(\eta/\alpha)I_k \le 1$, then $1 - \widetilde{m}_k \ge (\eta/(2\alpha))I_k$, so $\delta_k^2 \le 1 - (\eta/(2\alpha))I_k + C_L^2 I_k^2$. Since the right-hand side is at least $1/2$ under the stated condition, $\sqrt{1+y} \le 1 + y/2$ for $y \ge -1$ gives

$$\delta_k \le 1 - \frac{\eta}{4\alpha} I_k + \frac{1}{2}C_L^2 I_k^2. \tag{B.14}$$

As δ_k is deterministic, taking L^p norms in (B.7) gives $\|A_k\|_{L^p} \le \delta_k \|D_k\|_{L^p}$.

Step 2 (the residual $R_k^{(2)}$ and closing the recursion). At $\bar{X}^{\theta,h}_{\tau_k}$, split the bracket in (B.9) into a modeling part and a time-freezing part,

$$h_\theta(\tau_k,\cdot) - S^{(\alpha),\mathrm{res}}_{T-u}(\cdot) = \underbrace{\alpha\big(s_\theta(\tau_k,\cdot) - S^{(\alpha)}_{T-\tau_k}(\cdot)\big)}_{\text{Assumption A2: } \|\cdot\|_{L^p}\le \alpha\,\mathrm{err}} + \underbrace{\alpha\big(S^{(\alpha)}_{T-\tau_k} - S^{(\alpha)}_{T-u}\big)(\cdot)}_{\text{Assumption A3}}.$$

Assumption A3 bounds the second part by $\alpha L^{\mathrm{time}}_\varepsilon (u-\tau_k)(1+\big\|\bar{X}^{\theta,h}_{\tau_k}\big\|)$, and $\big\|\bar{X}^{\theta,h}_{\tau_k}\big\|_{L^p} \le \|D_k\|_{L^p} + B^{(p)}_\varepsilon$ by Proposition 24. Since $u-\tau_k \le h \le I_k/\beta_{\min}$ and $r_k(u)\le 1$, the part proportional to $\|D_k\|_{L^p}$ is at most $C_2 I_k^2 \|D_k\|_{L^p}$ with $C_2 := (1+\eta)L^{\mathrm{time}}_\varepsilon/\beta_{\min}$ (the outer weight $\frac{1+\eta}{\alpha}$ and the factor α of h_θ combining into $1+\eta$); the remaining, D_k-independent term is $\mathrm{Source}_k := (1+\eta)\int_{\tau_k}^{\tau_{k+1}} \bar\beta(u)\big[\mathrm{err}+L^{\mathrm{time}}_\varepsilon(u-\tau_k)(1+B^{(p)}_\varepsilon)\big]\,\mathrm{d}u$. Set $\widetilde{\delta}_k := \delta_k + C_2 I_k^2$ and $K_1 := \frac{1}{2}C_L^2 + C_2$. For every step, (B.13) gives $\widetilde{\delta}_k \le \exp\big((J_k)_- + K_1 I_k^2\big)$. For a cell contained in $[t_+,\infty)$, (B.14) instead gives $\widetilde{\delta}_k \le 1$ provided $I_k \le I_\star := \min\Big\{\frac{\alpha}{\eta}, \frac{\eta}{4\alpha K_1}\Big\}$.

Let $\mathcal{K}_{\mathrm{tr}}$ contain the cells not entirely included in $[t_+,\infty)$. Since the grid is uniform in direct time, for $h\le 1$ their union is contained in $[\varepsilon, t_+ + 1]$. Consequently, $\sum_{k\in\mathcal{K}_{\mathrm{tr}}}(J_k)_- \le \mathfrak{A}_\varepsilon$, and $\sum_{k\in\mathcal{K}_{\mathrm{tr}}} I_k^2 \le h\,\Xi_\varepsilon$ with $\Xi_\varepsilon := \beta(t_+ + 1)\big(B(t_+ + 1) - B(\varepsilon)\big)$. Here we used $(J_k)_- \le \int_{\mathcal{I}_k} \beta(t)(\chi_t)_-\,\mathrm{d}t$, $I_k \le h\beta(t_+ + 1)$ on transient cells, and disjointness of the cells. Thus every partial product of the one-step factors satisfies

$$\prod_j \widetilde{\delta}_j \le \mathcal{G}_{\varepsilon,h} := \exp\big(\mathfrak{A}_\varepsilon + K_1 h \Xi_\varepsilon\big). \tag{B.15}$$

Since $I_k \le h\beta(T)$, it is enough to impose

$$h \le h_0(\varepsilon,T) := \min\Big\{1, \frac{I_\star}{\beta(T)}, \frac{1}{K_1\Xi_\varepsilon}\Big\}, \tag{B.16}$$

with the convention that the last term is $+\infty$ if $K_1\Xi_\varepsilon = 0$. Then $\mathcal{G}_{\varepsilon,h} \le e^{\mathfrak{A}_\varepsilon + 1}$, independently of T.

Step 3 (the fluctuation $R_k^{(1)}$). Integrating the exact backward SDE over one step,

$$\overleftarrow{Y}_u - \overleftarrow{Y}_{\tau_k} = \int_{\tau_k}^u \bar\beta(v)\Big[\tfrac{1}{\alpha}\overleftarrow{Y}_v + (1+\eta)S^{(\alpha)}_{T-v}(\overleftarrow{Y}_v)\Big]\,\mathrm{d}v + \int_{\tau_k}^u \big(\eta\bar\beta(v)\big)^{1/\alpha}\,\mathrm{d}\overleftarrow{L}^\alpha_v. \tag{B.17}$$

The score–denoiser identity bounds the score at the origin, $\big\|S^{(\alpha)}_t(0)\big\| = \frac{a}{\alpha\gamma^\alpha}\|m_t(0)\| \le \frac{a}{\alpha\gamma^\alpha}R \le \frac{1+aC_t}{\alpha\gamma^\alpha}R$ (as $a\le 1$), so its at-most-linear growth gives $\big\|S^{(\alpha)}_{T-v}(\overleftarrow{Y}_v)\big\|_{L^p} \le \frac{1+a(T-v)C_{T-v}}{\alpha\gamma(T-v)^\alpha}(B^{(p)}_\varepsilon + R)$. The drift term of (B.17) is therefore bounded in L^p by $\int_{\tau_k}^u \bar\beta(v)\big[\frac{B^{(p)}_\varepsilon}{\alpha} + \frac{1+\eta}{\alpha}\frac{1+a(T-v)C_{T-v}}{\gamma(T-v)^\alpha}(B^{(p)}_\varepsilon + R)\big]\,\mathrm{d}v \le (B^{(p)}_\varepsilon + R)\big[\frac{\Delta B_k}{\alpha} + \frac{1+\eta}{\alpha}\int_{\tau_k}^{\tau_{k+1}} \bar\beta(v)\frac{1+a(T-v)C_{T-v}}{\gamma(T-v)^\alpha}\,\mathrm{d}v\big]$; the stochastic integral in (B.17) is $\mathrm{S}\alpha\mathrm{S}_d$ with scale $(\eta\int_{\tau_k}^u \bar\beta)^{1/\alpha} \le (\eta\,\Delta B_k)^{1/\alpha}$, contributing $c_{d,\alpha,p}\,\eta^{1/\alpha}\Delta B_k^{1/\alpha}$. Hence $\big\|\overleftarrow{Y}_u - \overleftarrow{Y}_{\tau_k}\big\|_{L^p} \le \Theta_k$, and by (B.8),

$$\big\|R_k^{(1)}\big\|_{L^p} \le \frac{1+\eta}{\alpha}\Big(\int_{\tau_k}^{\tau_{k+1}} \bar\beta(u)\,L^{\mathrm{res}}_{T-u}\,\mathrm{d}u\Big)\Theta_k. \tag{B.18}$$

Step 4 (telescoping). Combining Steps 1–3, $\|D_{k+1}\|_{L^p} \le \widetilde{\delta}_k\|D_k\|_{L^p} + \big\|R_k^{(1)}\big\|_{L^p} + \mathrm{Source}_k$. Starting from $\|D_0\|_{L^p} = 0$ (both processes start at π^α_∞, synchronously coupled) and using (B.15),

$$\|D_N\|_{L^p} \le \mathcal{G}_{\varepsilon,h}\Big(\sum_{k=0}^{N-1}\mathrm{Source}_k + \sum_{k=0}^{N-1}\big\|R_k^{(1)}\big\|_{L^p}\Big).$$

In the source sum, factor err and $L^{\mathrm{time}}_\varepsilon h(1+B^{(p)}_\varepsilon)$ out of $\sum_k\int_{\tau_k}^{\tau_{k+1}}\bar\beta = \int_0^{T-\varepsilon}\bar\beta = B(T)-B(\varepsilon)$; this is the first term inside braces in (B.6). The second sum, bounded by (B.18), is the second term. Finally $\mathcal{E}_2 \le \|D_N\|_{L^p}$ since W_p is an infimum over couplings. □

B.5 Assembly: the explicit bound

Theorem 26 (Explicit non-asymptotic bound)**.** *Under Assumptions A1–A3 and A4, for a target supported in $\bar{B}(0,R)$, $p \in [1,\alpha)$, arbitrary $0 < \varepsilon < T$, and $h \le h_0(\varepsilon, T)$,*

$$
\begin{aligned}
W_p(\pi_{\text{data}}, \hat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}) &\le \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{E}_2, \\
\mathcal{E}_0 &\le (1 - a(\varepsilon))\mathbb{E}[\|X_0\|^p]^{1/p} + \gamma(\varepsilon) c_{d,\alpha,p}, \\
\mathcal{E}_1 &\le W_p(\pi_{\text{data}}, \pi_\infty^\alpha) \\
&\quad \times \exp\Big(-\tfrac{1+\eta}{\alpha} \int_\varepsilon^T \beta(t) \tfrac{1-a(t)C_t}{\gamma(t)^\alpha}\, \mathrm{d}t\Big), \\
\mathcal{E}_2 &\le \mathcal{G}_{\varepsilon,h} \Bigg\{ (1+\eta)\big[\text{err} + L_\varepsilon^{\text{time}} h (1 + B_\varepsilon^{(p)})\big](B(T) - B(\varepsilon)) \\
&\quad + \frac{1+\eta}{\alpha} \sum_{k=0}^{N-1} \Big(\int_{\tau_k}^{\tau_{k+1}} \bar{\beta}(u) L_{T-u}^{\text{res}}\, \mathrm{d}u \Big) \Theta_k \Bigg\}.
\end{aligned}
\tag{B.19}
$$

Here Θ_k, $B_\varepsilon^{(p)}$ and $\mathcal{G}_{\varepsilon,h}$ are as in Proposition 25.

Proof. Combine the triangle decomposition (3.1) with Lemmas 21 and 23 and Proposition 25. □

B.6 Schedule corollaries and proof of Theorem 4

We express (B.19) through $B(t) = \int_0^t \beta$, with $B_\varepsilon := B(\varepsilon)$, $B_T := B(T)$, the per-step injected noise $\Delta B := \max_k \int_{\tau_k}^{\tau_{k+1}} \bar{\beta} \le h\beta(T)$, and $\mu_p := \mathbb{E}[\|X_0\|^p]^{1/p}$. The change of variable $\mathrm{d}B = \beta(t)\,\mathrm{d}t$, $\gamma(t)^\alpha = 1 - e^{-B(t)}$, gives

$$
\frac{\beta(t)}{\gamma(t)^\alpha}\,\mathrm{d}t = \frac{\mathrm{d}B}{1 - e^{-B}}. \tag{B.20}
$$

Corollary 27 (General schedule, intrinsic scale)**.** *Under the assumptions of Theorem 26 and the resolution $\Delta B \le \min\{B_\varepsilon, 1\}$,*

$$
\begin{aligned}
W_p(\pi_{\text{data}}, \hat{\pi}_N^{(\beta,\theta,\eta,\varepsilon)}) \le\ & c_0' B_\varepsilon^{1/\alpha} \\
& + c_1 e^{-\rho_\star (B_T - B_\varepsilon)} + c_2 (B_T - B_\varepsilon)\text{err} \\
& + c_3 \frac{\Delta B}{B_\varepsilon} \\
& + c_4 \left(B_T - B_\varepsilon + \log\left(1 + \frac{1}{B_\varepsilon}\right)\right) \Delta B^{1/\alpha},
\end{aligned}
\tag{B.21}
$$

where, writing $w := \frac{1+\eta}{\alpha}$, $\bar{w} := \max(1, w)$ and $\mathcal{G}_\varepsilon := e^{\mathfrak{A}_\varepsilon + 1}$,

$$
\rho_\star = \frac{1+\eta}{\alpha}, \qquad c_0' = \mu_p + c_{d,\alpha,p}, \qquad c_1 = e^{\mathfrak{D}_\varepsilon^{\text{mix}}}(\mu_p + c_{d,\alpha,p}), \qquad c_2 = \mathcal{G}_\varepsilon(1+\eta),
$$

$c_3 = 8\,\mathcal{G}_\varepsilon w\,\bar{w}\,(B_\varepsilon^{(p)} + R)(2 + \bar{C}_\varepsilon)^2$,

$$
\begin{aligned}
c_4 = \mathcal{G}_\varepsilon \Big[& w\,\eta^{1/\alpha} c_{d,\alpha,p}(2 + \bar{C}_\varepsilon) + 2w\,\bar{w}\,(B_\varepsilon^{(p)} + R)(2 + \bar{C}_\varepsilon)^2 \\
& + (1+\eta) L_\varepsilon^{\text{time}} (1 + B_\varepsilon^{(p)})/\beta_{\min} \Big].
\end{aligned}
$$

Proof. Bias. The global inequalities

$$1 - a(\varepsilon) \le \min\{B_\varepsilon/\alpha, 1\} \le B_\varepsilon^{1/\alpha}, \qquad \gamma(\varepsilon) \le \min\{B_\varepsilon, 1\}^{1/\alpha} \le B_\varepsilon^{1/\alpha}$$

and Lemma 21 yield $\mathcal{E}_0 \le c_0' B_\varepsilon^{1/\alpha}$.

Mixing. By (A.6), $\left(\frac{\eta}{\alpha} - \chi_t\right)_+ = \frac{1+\eta}{\alpha}\left(\frac{a(t)C_t - a(t)^\alpha}{\gamma(t)^\alpha}\right)_+ \le \frac{1+\eta}{\alpha}\frac{a(t)C_t}{\gamma(t)^\alpha}$. Assumption A4, the monotonicity of γ, and $\int_\varepsilon^\infty \beta(t)a(t)\,\mathrm{d}t = \alpha a(\varepsilon)$ yield

$$\mathfrak{D}_\varepsilon^{\mathrm{mix}} \le (1+\eta)\bar{C}_\varepsilon \frac{a(\varepsilon)}{\gamma(\varepsilon)^\alpha} < \infty. \tag{B.22}$$

Lemma 23 gives the mixing term in (B.21), while the independent coupling gives $W_p(\pi_{\mathrm{data}}, \pi_\infty^\alpha) \le \mu_p + c_{d,\alpha,p}$.

Score. The first term inside braces in (B.6) is $(1+\eta)[\mathrm{err} + L_\varepsilon^{\mathrm{time}} h(1+B_\varepsilon^{(p)})](B_T - B_\varepsilon)$. After multiplication by $\mathcal{G}_{\varepsilon,h} \le \mathcal{G}_\varepsilon$, the modeling part gives the coefficient c_2. The temporal-regularity part, bounded via $h \le \Delta B/\beta_{\min}$ and $\Delta B \le \Delta B^{1/\alpha}$ because $\Delta B \le 1$, is absorbed into the $c_4(B_T - B_\varepsilon)\Delta B^{1/\alpha}$ term below.

Discretization. Under compact support, $a(t)C_t \le \bar{C}_\varepsilon$, so, using $\gamma(t)^\alpha \le 1$ and $\alpha > 1$, both $L_t^{\mathrm{res}} = 1 + \frac{1+a(t)C_t}{\gamma(t)^\alpha}$ and $\frac{1}{\alpha} + \frac{1+a(t)C_t}{\gamma(t)^\alpha}$ are at most $\frac{2+\bar{C}_\varepsilon}{\gamma(t)^\alpha}$. Passing to direct time $t = T - u$ and using (B.20), each one-step regularity factor is bounded by $(2+\bar{C}_\varepsilon)H_k$, where $H_k := \int_{B(t_{k+1})}^{B(t_k)} \frac{\mathrm{d}B}{1-e^{-B}}$, $t_k := T - \tau_k$; the integration intervals partition $[B_\varepsilon, B_T]$. Since $e^{B_\varepsilon} \ge 1 + B_\varepsilon$ implies $\frac{1}{1-e^{-B_\varepsilon}} \le 1 + \frac{1}{B_\varepsilon}$, the antiderivative $B + \log(1-e^{-B})$ gives the global estimate $\sum_k H_k = \int_{B_\varepsilon}^{B_T} \frac{\mathrm{d}B}{1-e^{-B}} \le B_T - B_\varepsilon + \log\left(1 + \frac{1}{B_\varepsilon}\right)$. Moreover $\Theta_k \le \bar{w}\,(B_\varepsilon^{(p)} + R)(2+\bar{C}_\varepsilon)H_k + c_{d,\alpha,p}\eta^{1/\alpha}\Delta B_k^{1/\alpha}$, since $\frac{\Delta B_k}{\alpha} + w\int_{\tau_k}^{\tau_{k+1}} \bar{\beta}\,\frac{1+aC}{\gamma^\alpha} \le \Delta B_k + w(1+\bar{C}_\varepsilon)H_k \le \bar{w}(2+\bar{C}_\varepsilon)H_k$ ($\Delta B_k \le H_k$, the integrand of H_k being ≥ 1, and $\frac{1}{\alpha} \le 1$); hence

$$w\sum_k \Big(\int \bar{\beta} L^{\mathrm{res}}\Big)\Theta_k \le w\,\bar{w}\,(B_\varepsilon^{(p)} + R)(2+\bar{C}_\varepsilon)^2 \sum_k H_k^2 + w\,\eta^{1/\alpha} c_{d,\alpha,p}(2+\bar{C}_\varepsilon)\sum_k H_k \Delta B_k^{1/\alpha}.$$

Drift part. The inequality $\frac{1}{1-e^{-B}} \le 1 + \frac{1}{B}$ gives $H_k \le \Delta B_k(1 + \frac{1}{B(t_{k+1})})$. The resolution $\Delta B \le B_\varepsilon$ entails $B(t_k) \le 2B(t_{k+1})$, so $\frac{1}{B(t_{k+1})^2} \le \frac{4}{B(t_k)^2}$, and the monotonicity of $B \mapsto B^{-2}$ gives $\frac{\Delta B_k}{B(t_k)^2} \le \int_{B(t_{k+1})}^{B(t_k)} \frac{\mathrm{d}B}{B^2}$. Thus $\sum_k H_k^2 \le 2\Delta B\Big[(B_T - B_\varepsilon) + 4\int_{B_\varepsilon}^{B_T} \frac{\mathrm{d}B}{B^2}\Big] \le 2\Delta B\,(B_T - B_\varepsilon) + \frac{8\Delta B}{B_\varepsilon}$. The first piece, at most $2\Delta B^{1/\alpha}(B_T - B_\varepsilon)$ because $\Delta B \le 1$, joins the noise term; the second gives $c_3 \Delta B/B_\varepsilon$. *Noise part.* The global integral estimate above gives $\sum_k H_k \Delta B_k^{1/\alpha} \le \Delta B^{1/\alpha}\left(B_T - B_\varepsilon + \log\left(1 + \frac{1}{B_\varepsilon}\right)\right)$. Summing all five contributions and multiplying the score–discretization terms by $\mathcal{G}_{\varepsilon,h} \le \mathcal{G}_\varepsilon$ gives (B.21) with the stated constants. □

Proof of Theorem 4. Set $\rho_\star = \beta_0 \frac{1+\eta}{\alpha}$, $\quad c_0 = \beta_0^{1/\alpha}(\mu_p + c_{d,\alpha,p})$. Since Theorem 4 only requires h to be small enough, decrease its admissible threshold so that $h \le \min\{h_0(\varepsilon, T), \varepsilon, \beta_0^{-1}\}$. Then $\Delta B = \beta_0 h \le \min\{B_\varepsilon, 1\}$, $\quad B_\varepsilon = \beta_0 \varepsilon$, so Corollary 27 applies with $B_T = \beta_0 T$ and $\beta_{\min} = \beta_0$. Its bias term is $c_0 \varepsilon^{1/\alpha}$. The mixing exponent is $\rho_\star(T - \varepsilon)$; the score coefficient is $\beta_0 c_2$, and the drift term is $c_3 \beta_0 h/(\beta_0 \varepsilon) = c_3 h/\varepsilon$. Finally, $\left(\beta_0(T-\varepsilon) + \log\left(1 + \frac{1}{\beta_0 \varepsilon}\right)\right)(\beta_0 h)^{1/\alpha} \le c_{\beta_0}\left(1 + T + \log\left(1 + \frac{1}{\varepsilon}\right)\right) h^{1/\alpha}$ for a constant $c_{\beta_0} < \infty$ depending only on β_0. After absorbing this factor into c_4, we obtain (3.2). The constants c_3, c_4 inherited from Corollary 27 depend on ε through $\bar{C}_\varepsilon$ and $L_\varepsilon^{\mathrm{time}}$, and h_0 of (B.16) depends on (ε, T), including any horizon dependence of $L_\varepsilon^{\mathrm{time}}$. At fixed (ε, T), the discretization contribution is $\mathcal{O}(h^{1/\alpha})$. A joint accuracy or step-complexity bound also requires balancing early stopping, initialization, and score error. □

C Proof of Proposition 5: Denoiser Regularity under Compact Support

The following properties of the isotropic stable kernel will also be used in the heavy-tailed analysis. Writing $\varphi_{\alpha,d} = \varphi_\alpha$, the stable specialization of Bogdan et al. (2014, Corollary 23) gives $\varphi_{\alpha,d}(x) \asymp (1+\|x\|)^{-d-\alpha}$. Moreover, Kaleta and Sztonyk (2015, Theorem 3), applied to the stable Lévy measure, gives $|\partial^j \varphi_{\alpha,d}(x)| \le C(1+\|x\|)^{-d-\alpha}$ for $|j| \le 2$. Thus, with operator norm for matrices,

$$s := \nabla \log \varphi_\alpha, \qquad \Lambda_1 := \sup_x \|s(x)\| < \infty, \qquad \Lambda_2 := \sup_x \|\nabla s(x)\| < \infty, \qquad \Lambda := \Lambda_2.$$

For the sharper derivative bound needed below, let S be positive $\alpha/2$-stable with $\mathbb{E}e^{-vS} = e^{-v^{\alpha/2}}$. Gaussian subordination yields

$$\varphi_{\alpha,d}(r) = \mathbb{E}\big[(4\pi S)^{-d/2} e^{-r^2/(4S)}\big], \qquad \varphi'_{\alpha,d}(r) = -2\pi r \varphi_{\alpha,d+2}(r), \quad r > 0.$$

Differentiation is justified by domination on compact subsets of $(0,\infty)$. The density bound in dimension $d+2$ gives $|\varphi'_{\alpha,d}(r)| \le Cr(1+r)^{-d-\alpha-2}$ and radial monotonicity. Finally, Hult and Lindskog (2007, Lemma 2.1) gives $\mathbb{P}(\|L_1^\alpha\| > r) \sim C_{d,\alpha} r^{-\alpha}$, where $C_{d,\alpha} := \nu_\alpha(\|z\| > 1)$. Radial monotonicity and comparison of annuli with radii r and λr, followed by $\lambda \downarrow 1$, give

$$\varphi_\alpha(re) \sim k_{d,\alpha} r^{-d-\alpha}, \qquad k_{d,\alpha} = \frac{\alpha C_{d,\alpha}}{\omega_{d-1}}, \qquad e \in \mathbb{S}^{d-1}, \quad \omega_{d-1} = |\mathbb{S}^{d-1}|.$$

All kernel constants depend only on d and α.

Assume now $R := \sup_{x_0 \in \operatorname{supp} \pi_{\text{data}}} \|x_0\| < \infty$.

Proof of Proposition 5. Let $\mu_{t,x}$ be the conditional law of X_0 given $X_t = x$ and set $a = a(t)$, $\gamma = \gamma(t)$, $U = (x - aX_0)/\gamma$ and $s = \nabla \log \varphi_\alpha$. Differentiating the conditional-mean quotient, justified by compact support, gives

$$\nabla m_t(x) = \gamma^{-1} \operatorname{Cov}_{\mu_{t,x}}(X_0, s(U)). \tag{C.1}$$

Subtract $s((x - am_t(x))/\gamma)$ from the second covariance argument. Since s is Λ-Lipschitz,

$$\|\nabla m_t(x)\|_{\text{op}} \le \frac{\Lambda a}{\gamma^2} \mathbb{E}_{\mu_{t,x}} \|X_0 - m_t(x)\|^2 \le \frac{\Lambda a R^2}{\gamma^2} \le \frac{d\Lambda a R^2}{\gamma^2}.$$

This implies (3.4). Since $a \to 0$ and $\gamma \to 1$, $C_t \to 0$, $C_t^{(\alpha)} \to 1/\alpha$, and Assumption A4 holds. Lemma 20 gives the remaining conclusion. □

D Proofs for Sections 4 and 5.1: Heavy-Tailed Targets

Assumption A5 implies that, for suitable constants $0 < c_{\pi,-} \le c_{\pi,+}$ and $R_\pi \ge 1$,

$$c_{\pi,-}\|x\|^{-(d+\nu)} \;\le\; \pi(x) \;\le\; c_{\pi,+}\|x\|^{-(d+\nu)}, \qquad \|x\| \ge R_\pi, \tag{A5$'$}$$

because the asymptotic relation (4.1) is uniform in direction. We use this density envelope throughout the appendix; the full asymptotic relation is invoked only where an exact limit is required.

Write π for the target density, μ_x for the conditional law of X_0 given $X_t = x$, and ν_x for its image under $u = (x - ax_0)/\gamma$. We use the notation s, Λ_1, Λ_2 and the stable-kernel estimates from Appendix C, in particular

$$c_\varphi(1+\|u\|)^{-(d+\alpha)} \;\le\; \varphi_\alpha(u) \;\le\; C_\varphi(1+\|u\|)^{-(d+\alpha)}.$$

D.1 Proof of Theorem 7: Heavy-tailed bound

The heavy-tailed transfer requires a bound on $\|m_t(0)\|$ together with the uniform regularity and decay estimates of Proposition 6.

Lemma 28 (Uniform bound of m_t at zero)**.** *Under Assumption A5,* $\|m_t(0)\| \le \mathbb{E}_\pi[\|X_0\|] =: M_1 < \infty$ *for all* t*.*

Proof. At $x = 0$ the conditional law is weighted by $w(R) = \varphi_\alpha(aR/\gamma)$, $R = \|X_0\|$, a nonincreasing function. For an independent copy R', the identity $2\mathrm{Cov}(R, w(R)) = \mathbb{E}[(R - R')(w(R) - w(R'))] \le 0$ gives $\|m_t(0)\| \le \mathbb{E}[Rw(R)]/\mathbb{E}[w(R)] \le \mathbb{E}R$. □

Proof of Theorem 7. Use Lemma 28 in Step 3 of Proposition 25. Since $p < \alpha < \nu$ also at $\nu = 1 + \alpha$, $\mu_p < \infty$ and controls the moment terms in the bound. Proposition 6 verifies Assumption A4, including at this boundary, and Lemma 20 gives $\mathfrak{A}_\varepsilon < \infty$ and an eventual-contractivity time t_+ independent of T, while (B.22) gives $\mathfrak{D}_\varepsilon^{\mathrm{mix}} < \infty$. Therefore Proposition 24, the cumulative bound and the refined mixing estimate (B.4) all apply, with R replaced by M_1. □

D.2 Proofs of Propositions 6 and 8: Denoiser regularity

Spatial asymptotics give lower bounds on C_t; global upper bounds require separate estimates of the Jacobian.

Asymptotics of the denoiser m_s. For $s \in [\varepsilon, T]$ let $f_s(u) := a(s)^{-d}\pi(u/a(s))$ and $g_s(v) := \gamma(s)^{-d}\varphi_\alpha(v/\gamma(s))$ denote the densities of $a(s)X_0$ and of $\gamma(s)Z$, and set

$$\mathsf{D}_s := f_s * g_s, \qquad \mathsf{H}_s(x) := \int_{\mathbb{R}^d} u\, f_s(u)\, g_s(x-u)\,\mathrm{d}u, \tag{D.1}$$

so that D_s is the density of X_s and, by Bayes' formula,

$$a(s)\, m_s(x) = \mathbb{E}[a(s)X_0 \mid X_s = x] = \frac{\mathsf{H}_s(x)}{\mathsf{D}_s(x)}. \tag{D.2}$$

Since $a(s)$ and $\gamma(s)$ range in compact subsets of $(0, 1]$ as s ranges over $[\varepsilon, T]$, (4.1) and $\varphi_\alpha(z) \sim k_{d,\alpha}\|z\|^{-(d+\alpha)}$ give, uniformly in $s \in [\varepsilon, T]$ and $e \in \mathbb{S}^{d-1}$,

$$f_s(re) = (c_\pi a(s)^\nu + o(1))r^{-(d+\nu)}, \qquad g_s(re) = (k_{d,\alpha}\gamma(s)^\alpha + o(1))r^{-(d+\alpha)}, \tag{D.3}$$

together with the uniform global envelopes

$$f_s(y) \le C(1 + \|y\|)^{-(d+\nu)}, \qquad g_s(y) \le C(1 + \|y\|)^{-(d+\alpha)}, \tag{D.4}$$

and, because $\nu > 1$, the uniform first moment $\sup_{s\in[\varepsilon,T]} \int_{\mathbb{R}^d} \|u\|\, f_s(u)\,\mathrm{d}u < \infty$.

The unweighted convolution principle is classical (Bingham et al., 2006, Theorem 1.1); related conditional-mean regimes for positive scalar variables are studied by Denuit et al. (2025, Section 3). We give the weighted $\mathbb{R}^d$ argument, including the uniformity in s needed below.

Lemma 29 (Two-centre convolution asymptotics)**.** *Uniformly in* $s \in [\varepsilon, T]$ *and* $e \in \mathbb{S}^{d-1}$*, as* $r \to \infty$*,*

$$\mathsf{D}_s(re) = f_s(re) + g_s(re) + o(f_s(re) + g_s(re)), \tag{D.5}$$

and, for the numerator H_s *of the denoiser, the following two alternatives hold, again uniformly in* s *and* e*:*

$$\mathsf{H}_s(re) = re\, f_s(re) + o(r f_s(re)), \qquad 1 < \nu < 1 + \alpha, \tag{D.6}$$

$$\|\mathsf{H}_s(re)\| = \mathcal{O}(g_s(re)), \qquad \nu \ge 1 + \alpha. \tag{D.7}$$

Proof. Fix $x = re$ and split the convolution into $A = \{\|u\| \le r/2\}$, $B = \{\|x-u\| \le r/2\}$ and their complement C; $A \cap B$ is Lebesgue-null. Uniformly in the time parameter and direction, the tail equivalents and envelopes in (D.3)–(D.4) imply

$$\int_A f_s(u) g_s(x-u)\,\mathrm{d}u \sim g_s(x), \qquad \int_B f_s(u) g_s(x-u)\,\mathrm{d}u \sim f_s(x).$$

Indeed, on A the ratio $g_s(x-u)/g_s(x)$ is uniformly bounded and tends to one for bounded u; the common integrable envelope for f_s gives uniform dominated convergence. The argument on B is identical after setting $v = x - u$. On C, scaling $u = rw$ gives the bound $Cr^{-d-\nu-\alpha}$, proving (D.5). For the weighted numerator the same argument on B gives $\int_B u f_s(u) g_s(x-u)\,\mathrm{d}u = x f_s(x) + o(r f_s(x))$, since $\|(x-v)/r\| \le 3/2$ there. The contribution from A is $O(g_s(x))$ by the common first-moment bound, and that from C is $O(r^{1-d-\nu-\alpha})$. For $\nu < 1+\alpha$ these two terms are $o(r f_s(x))$; for $\nu \ge 1+\alpha$, all three terms are $O(g_s(x))$. This proves (D.6)–(D.7), including the boundary $\nu = 1+\alpha$, with the required uniformity. □

Proposition 30 (Spatial asymptotic regimes of the denoiser)**.** *Under Assumption A5 with $\nu > 1$, the following hold uniformly in $s \in [\varepsilon, T]$ and $e \in \mathbb{S}^{d-1}$.*

(i) If $1 < \nu < \alpha$, then $m_s(re) = a(s)^{-1} re + o(r)$.

(ii) If $\nu = \alpha$, then $m_s(re) = \varkappa_{\alpha,s}\, re + o(r)$ with $\varkappa_{\alpha,s} = \dfrac{c_\pi a(s)^{\alpha-1}}{c_\pi a(s)^\alpha + k_{d,\alpha}\gamma(s)^\alpha}$.

(iii) If $\alpha < \nu < 1+\alpha$, then $m_s(re) = \Big(\dfrac{c_\pi a(s)^{\nu-1}}{k_{d,\alpha}\gamma(s)^\alpha} + o(1)\Big) r^{\alpha+1-\nu} e$; in particular m_s is unbounded, but sublinear since $\alpha + 1 - \nu \in (0,1)$.

(iv) If $\nu \ge 1+\alpha$, then $\sup_{s\in[\varepsilon,T]} \sup_x \|m_s(x)\| < \infty$; the boundary $\nu = 1+\alpha$ is included.

Consequently, setting

$$\varkappa_{\nu,s} := \begin{cases} a(s)^{-1}, & 1 < \nu < \alpha, \\ \dfrac{c_\pi a(s)^{\alpha-1}}{c_\pi a(s)^\alpha + k_{d,\alpha}\gamma(s)^\alpha}, & \nu = \alpha, \\ 0, & \nu > \alpha, \end{cases} \tag{D.8}$$

one has the unified decomposition $m_s(x) = \varkappa_{\nu,s}\, x + \mathsf{r}_s(x)$ with a uniformly sublinear *remainder,*

$$\lim_{R\to\infty} \sup_{s\in[\varepsilon,T]} \sup_{\|x\|\ge R} \frac{\|\mathsf{r}_s(x)\|}{\|x\|} = 0. \tag{D.9}$$

Moreover m_s is C^1 in x, locally Lipschitz uniformly on compact space–time cylinders, and of uniform linear growth, $\sup_{s\in[\varepsilon,T]} \|m_s(x)\| \le C(1+\|x\|)$.

Proof. By (D.2), $m_s = a(s)^{-1}\mathsf{H}_s/\mathsf{D}_s$. In Lemma 29, the denominator is asymptotic to f_s if $\nu < \alpha$, to $f_s + g_s$ if $\nu = \alpha$, and to g_s if $\nu > \alpha$. Substitution of (D.3) gives (i)–(iii); for $\nu \ge 1+\alpha$ the numerator bound gives (iv) outside a fixed ball. Lemma 19 gives joint continuity of m_s and ∇m_s, hence boundedness on compact cylinders and the remaining part of (iv). The uniform equivalents imply (D.9); together with continuity on the complementary compact set, they also give uniform linear growth. Joint continuity of the Jacobian gives the stated local Lipschitz property. □

Lemma 31 (Covariance identity)**.** *Under Assumption A5, for all x and $t > 0$, $m_t \in C^1$ and $\nabla m_t(x) = -\frac{1}{a}\operatorname{Cov}_{\nu_x}(U, s(U))$.*

Proof. The differentiation in (C.1) remains valid under the finite first moment, by boundedness of $\nabla\varphi_\alpha$. The affine relation $X_0 = (x - \gamma U)/a$ therefore gives $\nabla m_t(x) = -a^{-1}\mathrm{Cov}_{\nu_x}(U, s(U))$. □

Proposition 32 (Uniform conditional moment)**.** *Under Assumption A5, for* $1 \le p \le \nu - \alpha$ *and* $0 < \gamma_0 \le \gamma_1$, $M_*^{(p)} := \sup\{\mathbb{E}_{\mu_x}\|X_0\|^p : \ x \in \mathbb{R}^d, \ a \in (0,1), \ \gamma \in [\gamma_0, \gamma_1]\} < \infty$.

Proof. Let $K = d + \alpha$ and omit the common likelihood factor γ^{-d}. Uniformly in the stated parameters, the conditional normalizer satisfies

$$Z(x) = \int \pi(y)\varphi_\alpha((x - ay)/\gamma)\,\mathrm{d}y \ge c(1 + \|x\|)^{-K},$$

by restricting to $\|y\| \le 1$. The numerator is at most CM_p on bounded sets of x, where $M_p = \mathbb{E}_\pi\|X_0\|^p < \infty$. For $r = \|x\| \ge \max(2R_\pi, 2)$, split it at $\|ay\| = r/2$. The near contribution is at most $Cr^{-K}M_p$. On the far region the power-law envelope gives $\|y\|^p\pi(y) \le C(r/a)^{p-d-\nu}$; consequently

$$N_{p,\mathrm{far}}(x) \le C(r/a)^{p-d-\nu}\int_{\mathbb{R}^d}\varphi_\alpha((x - ay)/\gamma)\,\mathrm{d}y \le Ca^{\nu-p}r^{p-d-\nu}. \tag{D.10}$$

Dividing by $Z(x)$ bounds the conditional moment by $C(M_p + a^{\nu-p}r^{p+\alpha-\nu})$, uniformly bounded because $p + \alpha - \nu \le 0$. This includes $p = \nu - \alpha$. □

Proof of Proposition 6. *Step 1 (spatial lower bounds).* For $\nu < \alpha$, Proposition 30 gives

$$C_t \ge \frac{\|m_t(re) - m_t(0)\|}{r} \longrightarrow a(t)^{-1},$$

which proves (i) without differentiating a spatial equivalent. The same argument gives $C_t \ge \varkappa_{\alpha,t}$ at $\nu = \alpha$.

Step 2 (lack of global regularity when $\alpha \le \nu < 1 + \alpha$). Fix such a ν, choose

$$\max\{0, 2 - \nu\} < \vartheta < 1 + \alpha - \nu, \qquad 0 < \lambda < \tfrac{1}{2},$$

and, in dimension one, let

$$\pi(y) = c(1 + y^2)^{-(1+\nu)/2}\left[1 + \lambda(1 + y^2)^{-\vartheta/2}\cos y\right], \tag{D.11}$$

where $c > 0$ normalizes the density. This density is smooth, positive, even, and satisfies Assumption A5 with $c_\pi = c$.

We first record an elementary convolution estimate. Let g be a one-dimensional symmetric stable density of index α, and let h satisfy

$$\int h = 0, \qquad \int |uh(u)|\,\mathrm{d}u < \infty, \qquad |h(u)| \le C(1 + |u|)^{-k}, \qquad 2 < k < 2 + \alpha.$$

If $h(x) = bx^{-k}\sin(\omega x) + o(x^{-k})$ as $x \to +\infty$, with $\omega > 0$, then

$$(h * g)(x) = b\widehat{g}(\omega)x^{-k}\sin(\omega x) + o(x^{-k}). \tag{D.12}$$

Indeed, on $|u| \le x/2$, subtract $g(x)$ and use $|g'(z)| \le C(1 + |z|)^{-2-\alpha}$, established for the stable kernel in Appendix C. The resulting integral is $\mathcal{O}(x^{-2-\alpha})$; the subtracted term is $\mathcal{O}(x^{-k-\alpha})$ because $\int h = 0$. On $|x - u| \le x/2$, set $v = x - u$. After multiplication by x^k, the integrand is bounded by $Cg(v)$, and the expansion of $h(x - v)$ gives $b\int \sin(\omega(x - v))g(v)\,\mathrm{d}v = b\widehat{g}(\omega)\sin(\omega x)$, with an error tending to zero by dominated convergence. On the remaining region $|u|, |x - u| > x/2$, the two envelopes give $\mathcal{O}(x^{-k-\alpha})$. These bounds prove (D.12).

Now fix any $t > 0$, abbreviate $a = a(t)$, $\gamma = \gamma(t)$, and use $f = f_t$, $g = g_t$, $\mathsf{D} = \mathsf{D}_t$, $\mathsf{H} = \mathsf{H}_t$ from (D.1). Set

$$A = ca^{\nu}, \qquad B = c\lambda a^{\nu+\vartheta}, \qquad G = k_{1,\alpha}\gamma^{\alpha}, \qquad \omega = a^{-1}, \qquad \phi = \widehat{g}(\omega) = e^{-\gamma^{\alpha}\omega^{\alpha}} > 0.$$

Direct differentiation of (D.11) after rescaling gives

$$f'(x) = -\omega B x^{-1-\nu-\vartheta}\sin(\omega x) + \mathcal{O}(x^{-2-\nu}), \qquad (xf(x))' = -\omega B x^{-\nu-\vartheta}\sin(\omega x) + \mathcal{O}(x^{-1-\nu}).$$

Both derivatives have integral zero and a finite absolute first moment, since $\nu + \vartheta > 2$. Moreover, $\vartheta < 1$ and $\nu + \vartheta < 1 + \alpha$, so (D.12) applies with $k = 1 + \nu + \vartheta$ and $k = \nu + \vartheta$, respectively. Differentiating the convolutions therefore yields

$$\begin{aligned}\mathsf{D}'(x) &= -\omega B\phi\, x^{-1-\nu-\vartheta}\sin(\omega x) + o(x^{-1-\nu-\vartheta}),\\ \mathsf{H}'(x) &= -\omega B\phi\, x^{-\nu-\vartheta}\sin(\omega x) + o(x^{-\nu-\vartheta}).\end{aligned}$$

Meanwhile, Lemma 29 gives

$$\mathsf{H}(x) \sim Ax^{-\nu}, \qquad \mathsf{D}(x) \sim \begin{cases}(A+G)x^{-1-\alpha}, & \nu = \alpha,\\ Gx^{-1-\alpha}, & \nu > \alpha.\end{cases}$$

Substitution into $m_t' = a^{-1}(\mathsf{H}'\mathsf{D} - \mathsf{H}\mathsf{D}')/\mathsf{D}^2$ gives

$$m_t'(x) = -K_t x^{1+\alpha-\nu-\vartheta}\sin(\omega x) + o(x^{1+\alpha-\nu-\vartheta}), \qquad K_t = \frac{\omega B\phi}{a}\begin{cases}G/(A+G)^2, & \nu = \alpha,\\ 1/G, & \nu > \alpha.\end{cases}$$

Since $K_t > 0$ and $1 + \alpha - \nu - \vartheta > 0$, evaluating along $\omega x = \pi/2 + 2\pi n$ proves $C_t = \infty$. The density (D.11) is independent of t, so this holds for every $t > 0$, proving (ii).

This failure is not forced by the matched index itself: if $X_0 \sim \mathrm{S}\alpha\mathrm{S}_1(\sigma)$, the fractional Tweedie formula gives

$$m_t(x) = \frac{\sigma^{\alpha}a(t)^{\alpha-1}}{\sigma^{\alpha}a(t)^{\alpha} + \gamma(t)^{\alpha}}\, x, \qquad C_t = \frac{\sigma^{\alpha}a(t)^{\alpha-1}}{\sigma^{\alpha}a(t)^{\alpha} + \gamma(t)^{\alpha}} \longrightarrow 0.$$

Step 3 (uniform decay for $\nu \geq 1 + \alpha$). Fix $\varepsilon > 0$. Proposition 32 at $p = 1$ gives $\sup_{t\geq\varepsilon,x}\mathbb{E}_{\mu_x}\|X_0\| \leq M_* < \infty$. Also $M_2 < \infty$, since $\nu \geq 1 + \alpha > 2$. With $a = a(t)$, $\gamma = \gamma(t)$ and conditional density w_x, the covariance identity gives

$$\nabla m_t(x) = \gamma^{-1}\int (y - m_t(x))[s((x-ay)/\gamma) - s(x/\gamma)]^{\top} w_x(y)\,\mathrm{d}y. \tag{D.13}$$

For bounded x, and also on $\|ay\| \leq \|x\|/2$ when $\|x\|$ is large, the stable envelopes imply $w_x(y) \leq C_\varepsilon\pi(y)$. The Lipschitz bound on s therefore bounds this contribution by $C_\varepsilon a(M_2 + M_*M_1)$. On the remaining region, $r = \|x\| \geq \max(2R_\pi, 2)$ and $\|ay\| > r/2$. Equation (D.10) with $p = 1$, divided by the normalizer, gives

$$\int_{\|ay\|>r/2}\|y\|w_x(y)\,\mathrm{d}y \leq C_\varepsilon a^{\nu-1}r^{1+\alpha-\nu}.$$

Since $\|y\| > 1$ there, boundedness of s bounds the remaining contribution to (D.13) by the same expression up to a constant. It is at most $C_\varepsilon a$ because $\nu - 1 > 1$ and $1 + \alpha - \nu \leq 0$. Taking the supremum over x proves (iii), including $\nu = 1 + \alpha$. □

Proof of Proposition 8. The Gaussian conditional-mean derivative identity (Dytso et al., 2021, Proposition 1), applied to $aX_0 + \gamma Z$, gives $\nabla m_t(x) = (a/\gamma^2)\mathrm{Cov}(X_0 \mid X_t = x)$. The conditional second moments are finite by Gaussian damping. For compact support, the covariance operator norm is at most R^2, giving $C_t \le aR^2/\gamma^2 \to 0$. Under Assumption A5, set $U = (aX_0 - x)/\gamma$. Its conditional density is proportional to

$$\rho_x(u)\varphi(u), \qquad \rho_x(u) = \frac{\pi((x+\gamma u)/a)}{\pi(x/a)},$$

where φ is the standard Gaussian density. As $\|x\| \to \infty$, $\rho_x(u) \to 1$ for bounded u. On $\|u\| \le \|x\|/(2\gamma)$ it is uniformly bounded by the power-law envelopes; on the complement it is at most $C\|x\|^{d+\nu}$. Dominated convergence on the first region and Gaussian tail decay on the second therefore give, for $w = 1, u, uu^\top$, $\int w(u)\rho_x(u)\varphi(u)\,\mathrm{d}u \to \int w(u)\varphi(u)\,\mathrm{d}u$. Thus $\mathrm{Cov}(U \mid X_t = x) \to I_d$ and $\nabla m_t(x) \to a^{-1}I_d$. It follows that $C_t \ge a(t)^{-1} \to \infty$ and $C_t^{(2)} \le 0$. All these spatial limits are uniform on compact positive-time intervals. Gaussian domination also gives joint continuity of the covariance there, hence its global spatial boundedness uniformly on each such interval. □

D.3 Proof of Corollary 9: Stable mixing

Proof of Corollary 9. Since $\nu \ge 1 + \alpha > \alpha$, both initial laws have finite p-th moments. Proposition 6 gives $C_t \le c_\varepsilon a(t)$ for $t \ge \varepsilon$, and hence

$$\mathfrak{D}_\varepsilon^{\mathrm{mix}} \le \frac{(1+\eta)c_\varepsilon}{\alpha\gamma(\varepsilon)^\alpha}\int_\varepsilon^\infty \beta(t)a(t)^2\,\mathrm{d}t = \frac{(1+\eta)c_\varepsilon a(\varepsilon)^2}{2\gamma(\varepsilon)^\alpha} < \infty,$$

using $\mathrm{d}(a^2) = -(2/\alpha)\beta a^2\,\mathrm{d}t$. Equation (B.4) proves the claim with $C = e^{\mathfrak{D}_\varepsilon^{\mathrm{mix}}}$, independent of T. □

Brownian synchronous coupling. The same argument for the VP SDE produces $\Phi^{(2)}(T) := \int_\varepsilon^T \beta(t)(1 - a_tC_t)/\gamma_t^2\,\mathrm{d}t$. Proposition 8 gives $a_tC_t \ge 1$, so $\Phi^{(2)}(T) \le 0$ and $e^{-\Phi^{(2)}(T)} \ge 1$. Thus this synchronous bound is non-contractive; it does not imply that Brownian mixing fails.

D.4 Proof of Theorem 10: Brownian initialization error

We prove the exponential W_p convergence of the Brownian model (Theorem 10), for the *reverse SDE*, so that the object compared with the stable sampler of Corollary 9 is of the same nature. We work in the constant-schedule VP model $\mathrm{d}X_t = -\frac{\beta_0}{2}X_t\,\mathrm{d}t + \sqrt{\beta_0}\,\mathrm{d}W_t$, with $a_t = e^{-\beta_0 t/2}$, $\gamma_t^2 = 1 - a_t^2$ and stationary law $\pi_\infty^{(2)} = \mathcal{N}(0, I_d)$; the general schedule follows by a time change. The standing hypotheses are those of Theorem 10:

$$\text{(H)} \qquad \pi_{\mathrm{data}} \text{ satisfies Assumption A5 with } \nu > 2. \tag{D.14}$$

In particular $M_2 := \mathbb{E}\,\|X_0\|^2 < \infty$.

The sampler analyzed is the Brownian reverse SDE

$$\mathrm{d}\overleftarrow{Y}_\tau = \Big[\tfrac{\bar\beta(\tau)}{2}\overleftarrow{Y}_\tau + \bar\beta(\tau)\nabla\log\pi_{T-\tau}(\overleftarrow{Y}_\tau)\Big]\,\mathrm{d}\tau + \sqrt{\bar\beta(\tau)}\,\mathrm{d}\overleftarrow{W}_\tau, \tag{D.15}$$

run over $\tau \in [0, T-\varepsilon]$. Lemmas 33 and 34 show that (D.15) is well posed from any initial law and that its transition kernel $\mathsf{K}_{T,\varepsilon}$ is the forward conditional kernel,

$$\mathsf{K}_{T,\varepsilon}(x,\cdot) = \mathrm{Law}\big(X_\varepsilon \mid X_T = x\big), \qquad \text{so that} \qquad \pi_T\mathsf{K}_{T,\varepsilon} = \pi_\varepsilon. \tag{D.16}$$

The generated law is then $\mu_{T,\varepsilon}^{(2)} := \pi_\infty^{(2)}\mathsf{K}_{T,\varepsilon}$; to lighten notation, we abbreviate it as μ_T throughout this subsection and write $\mathcal{E}_1 = W_p(\mu_T, \pi_\varepsilon)$. Everything else follows from three ingredients: a total-variation

bound of order a_T for $\left\|\pi_T - \pi_\infty^{(2)}\right\|_{\mathrm{TV}}$, a uniform moment bound for μ_T obtained from the likelihood ratio $\pi_\infty^{(2)}/\pi_T$, and an interpolation between total variation and W_p.

Lemma 33 (Score regularity and well-posedness)**.** *Under Assumption A5 and for $\varepsilon > 0$, write $\Sigma_t(y) := \mathrm{Cov}_{\mu_y}(X_0)$ for the conditional covariance entering the Jacobian of the denoiser m_t. Then $\pi_t \in C^\infty$, $\pi_t > 0$, and*

$$\nabla \log \pi_t(y) = \frac{a_t\, m_t(y) - y}{\gamma_t^2}, \qquad \nabla^2 \log \pi_t(y) = -\frac{1}{\gamma_t^2} I_d + \frac{a_t^2}{\gamma_t^4}\Sigma_t(y),$$

with $\Lambda_{\varepsilon,T} := \sup_{t\in[\varepsilon,T]} \sup_y \left\|\nabla^2 \log \pi_t(y)\right\| < \infty$. Consequently the drift of (D.15) is globally Lipschitz in y, uniformly in $\tau \in [0, T-\varepsilon]$, so (D.15) has a unique strong non-explosive solution from every initial law, and its transition kernel $\mathsf{K}_{T,\varepsilon}$ is well defined.

Proof. Gaussian Tweedie calculus and Dytso et al. (2021, Proposition 1) give

$$\nabla \log \pi_t(y) = \frac{a_t m_t(y) - y}{\gamma_t^2}, \qquad \nabla^2 \log \pi_t(y) = -\gamma_t^{-2} I_d + a_t^2 \gamma_t^{-4} \Sigma_t(y).$$

The proof of Proposition 8 bounds Σ_t uniformly on $[\varepsilon, T] \times \mathbb{R}^d$. Joint continuity also bounds $\nabla \log \pi_t(0)$ on that interval. The backward drift is therefore uniformly Lipschitz with linear growth; subtracting the additive Gaussian integral gives strong existence, pathwise uniqueness and non-explosion. □

Lemma 34 (The reverse kernel is the forward conditional kernel)**.** *Under (D.14), (D.16) holds for π_T-almost every x; in particular $\pi_T \mathsf{K}_{T,\varepsilon} = \pi_\varepsilon$, and $\mu_T = \pi_\infty^{(2)} \mathsf{K}_{T,\varepsilon}$ is the law of the output of (D.15) initialized at $\pi_\infty^{(2)}$.*

Proof. Let P be the forward Ornstein–Uhlenbeck path law and R the same process started from its Gaussian invariant law. Their conditional path laws given X_0 agree, so $H(P \mid R) = \mathrm{KL}(\pi_{\mathrm{data}} \mid \pi_\infty^{(2)}) < \infty$, by boundedness of π and its finite second moment. The stationary Gaussian reference satisfies Hypotheses 1.10 of Cattiaux et al. (2023, arXiv version 3); their Theorem 1.12 identifies the reverse drift as $\beta_0 y/2 + \beta_0 \nabla \log \pi_{T-\tau}(y)$. Restricting to $[0, T-\varepsilon]$ and using Lemma 33 identifies the reverse transition kernel with $\mathrm{Law}(X_\varepsilon \mid X_T = x)$, π_T-almost everywhere. Since π_T is positive, the identity also holds $\pi_\infty^{(2)}$-almost everywhere. Propagating the two initial laws gives the two asserted kernel identities. □

Lemma 35 (TV bound)**.** *Under (D.14), let $K_0 := \mathrm{KL}(\pi_{\mathrm{data}} \,\|\, \pi_\infty^{(2)}) < \infty$. Then*

$$\|\mu_T - \pi_\varepsilon\|_{\mathrm{TV}} \le \|\pi_T - \pi_\infty^{(2)}\|_{\mathrm{TV}} \le \sqrt{K_0/2}\, a_T.$$

Proof. The first inequality is contraction of total variation under $\mathsf{K}_{T,\varepsilon}$, by Lemma 34. Gaussian log-Sobolev entropy decay and Pinsker (Bakry et al., 2014, Theorem 5.2.1, Eq. (5.2.2), and Proposition 5.5.1) give the second inequality, with the VP normalization $\mathrm{KL}(\pi_T \,\|\, \pi_\infty^{(2)}) \le a_T^2 K_0$. Here $K_0 \le \log \|\pi\|_\infty + M_2/2 + (d/2)\log(2\pi) < \infty$. □

Lemma 36 (Density comparison)**.** *If $M_2 < \infty$, then $\pi_\infty^{(2)}(x) \le K_T e^{\lambda_T \|x\|^2} \pi_T(x)$ with $\lambda_T = \frac{a_T}{2(1-a_T)} \to 0$ and $K_T = \gamma_T^d\, e^{\lambda_T M_2} \to 1$.*

Proof. For $a = a_T \in (0,1)$, Young's inequality gives $\|x - aX_0\|^2 \le (1+a)(\|x\|^2 + a\|X_0\|^2)$. Jensen's inequality and $\gamma_T^2 = (1-a)(1+a)$ therefore yield

$$\pi_T(x) \ge (2\pi\gamma_T^2)^{-d/2} \exp\left[-\frac{\|x\|^2 + aM_2}{2(1-a)}\right].$$

Dividing the standard Gaussian density by this lower bound gives $K_T e^{\lambda_T \|x\|^2}$ without a centering assumption. □

Lemma 37 (Uniform moments of the generated law)**.** *Assume* (D.14) *and let* $q \in [1, \nu)$*. Choose* $r > 1$ *with* $qr < \nu$ *and set* $r' = r/(r-1)$*. Then for every* T *with* $r' a_T < 1$*,*

$$\mathbb{E}_{\mu_T} \|Y\|^q \;\le\; \left(\mathbb{E}\,\|X_\varepsilon\|^{qr}\right)^{1/r} \left[K_T^{r'-1}\big(1 - 2(r'-1)\lambda_T\big)^{-d/2}\right]^{1/r'},$$

with K_T, λ_T *as in Lemma 36. In particular* $\sup_{T \ge T_q,\, 0<\varepsilon<T} \mathbb{E}_{\mu_T} \|Y\|^q < \infty$ *for a threshold* T_q *determined by, for instance,* $r' a_{T_q} \le \frac{1}{2}$*, and the bound holds* for every early-stopping time $0 < \varepsilon < T$*, since* $a_\varepsilon, \gamma_\varepsilon \le 1$ *makes* $\mathbb{E}\,\|X_\varepsilon\|^{qr}$ *bounded independently of* ε*.*

Proof. Let $R_T = \pi_\infty^{(2)}/\pi_T$. The reverse-kernel identity and Hölder give

$$\mathbb{E}_{\mu_T}\|Y\|^q = \mathbb{E}[\|X_\varepsilon\|^q R_T(X_T)] \le (\mathbb{E}\|X_\varepsilon\|^{qr})^{1/r} (\mathbb{E} R_T(X_T)^{r'})^{1/r'}.$$

Lemma 36 and Gaussian integration give

$$\mathbb{E} R_T(X_T)^{r'} = \int R_T^{r'-1} \pi_\infty^{(2)} \le K_T^{r'-1} \left(1 - 2(r'-1)\lambda_T\right)^{-d/2},$$

provided $2(r'-1)\lambda_T < 1$, equivalently $r' a_T < 1$. The representation $X_\varepsilon = a_\varepsilon X_0 + \gamma_\varepsilon Z$ and $qr < \nu$ bound the first factor independently of ε. Taking $r' a_{T_q} \le 1/2$ bounds the second uniformly for $T \ge T_q$. □

Proposition 38 (Constant-schedule Brownian W_p convergence)**.** *Assume* (D.14)*. In the constant-schedule VP model, for* $p \in [1, \nu)$ *and any* $q \in (p, \nu)$ *there is* $K_{p,q} < \infty$*, independent of the early-stopping time* $0 < \varepsilon < T$*, such that for* T *large enough* $\mathcal{E}_1 = W_p(\mu_T, \pi_\varepsilon) \le K_{p,q}\, e^{-\frac{q-p}{2pq}\beta_0 T}$*.*

Proof. Couple the common part of two probability measures identically and their residual parts arbitrarily. Since $\|x-y\|^p \le 2^{p-1}(\|x\|^p + \|y\|^p)$, this gives $W_p(\mu,\nu)^p \le 2^{p-1} \int \|x\|^p \,\mathrm{d}|\mu - \nu|$. Hölder with respect to $|\mu - \nu| \le \mu + \nu$ then gives

$$W_p(\mu,\nu)^p \le 2^{p-1} \left(\int \|x\|^q \,\mathrm{d}(\mu+\nu)\right)^{p/q} (2\|\mu-\nu\|_{\mathrm{TV}})^{1-p/q}.$$

Apply this with $(\mu,\nu) = (\mu_T, \pi_\varepsilon)$. Lemmas 35 and 37 give $W_p \le K_{p,q} a_T^{1/p-1/q} = K_{p,q} e^{-(q-p)\beta_0 T/(2pq)}$, uniformly in ε. □

Lemma 39 (Reduction by time change)**.** *In the VP model the joint law of* (X_ε, X_T) *— hence the kernel* $\mathsf{K}_{T,\varepsilon}$*, the marginals* π_ε*,* π_T *and the generated law* μ_T *— depends on* (ε, T) *only through* $(B(\varepsilon), B(T))$*. Consequently the pair* (μ_T, π_ε) *for a schedule* β *is that of the unit schedule at horizon* $B(T)$ *and early-stopping time* $B(\varepsilon)$*.*

Proof. The Ornstein–Uhlenbeck transition from s to t has mean multiplier $e^{-[B(t)-B(s)]/2}$ and variance $1 - e^{-[B(t)-B(s)]}$. Consequently the joint law of (X_ε, X_T) is that of the unit schedule at $(B(\varepsilon), B(T))$. The same is true of its conditional kernel and the two propagated laws. □

Proof of Theorem 10. Apply Proposition 38 with unit schedule at $(B(\varepsilon), B(T))$, using Lemma 39. Its constants and threshold are uniform in the early-stopping time, and $B(T) \to \infty$. □

E Proofs for Section 5.2: Tail Regimes and Extreme Quantiles

The two consequences of Assumption A5 used in this appendix are the density envelope (A5′) and $\overline{G}_{\pi_{\mathrm{data}}}(r) \sim c'_\pi r^{-\nu}$, $\quad c'_\pi = \frac{c_\pi \omega_{d-1}}{\nu}$, where locally $c_\pi := c_{\pi_{\mathrm{data}}}$, obtained by integrating the uniform asymptotic relation (4.1) in polar coordinates. The former controls the denoiser estimates, while the latter identifies the radial tail index and constant. Both hold here with $\nu > 1$. Let $\mu_\tau = \mathrm{Law}(\overleftarrow{Y}_\tau)$ denote the output at $\tau \in (0, T-\varepsilon]$ from the appropriate reference law, and write $\mu_\tau^{(\alpha)}$ and $\mu_\tau^{(2)}$ when the stable and Brownian models must be distinguished. Both models are treated under Assumption A5, with $\nu > 1$ and fixed $0 < \varepsilon < T < \infty$.

E.1 Proof of Proposition 12: Tail-regime dichotomy

Proof of Proposition 12. For (i), the survival representation and Karamata's theorem (Kevei, n.d., Theorem 12(ii)) give

$$e_\mu(u) = \frac{\int_u^\infty \overline{G}_\mu(r)\,\mathrm{d}r}{\overline{G}_\mu(u)} \sim \frac{u}{\vartheta - 1}.$$

For (ii), put $V = \|G + R\|$ and $Q = \|G\|$. The bound $|V - Q| \le K$ implies

$$(V-u)^+ \le 2K\mathbf{1}_{\{V>u\}} + (Q-u-K)^+, \qquad \{Q > u+K\} \subseteq \{V > u\}.$$

Consequently $e_\mu(u) \le 2K + e_Q(u+K)$, without any independence assumption. For $G \sim \mathcal{N}(0, \sigma^2 I_d)$, the incomplete-gamma asymptotic (National Institute of Standards and Technology, n.d., Eq. 8.11.2) gives $\mathbb{P}(Q > v) \sim cv^{d-2}e^{-v^2/(2\sigma^2)}$ and $e_Q(v) \sim \sigma^2/v$. Thus $e_\mu(u) = \mathcal{O}(1)$ and $\Gamma(\mu) = 0$. □

E.2 Brownian tail bounds for Proposition 14

Lemma 40 (Bounded smoothed Brownian score)**.** *Under Assumption A5, for every $0 < \varepsilon < T < \infty$,*

$$S_{\varepsilon,T} := \sup_{s\in[\varepsilon,T],\, y\in\mathbb{R}^d} \|\nabla \log \pi_s^{(2)}(y)\| < \infty.$$

Proof. Set $U = (a_s X_0 - y)/\gamma_s$. The weighted Gaussian-convolution argument in the proof of Proposition 8 gives $\mathbb{E}[U \mid X_s = y] \to 0$ as $\|y\| \to \infty$, uniformly for $s \in [\varepsilon, T]$. Differentiating the Gaussian kernel yields

$$\nabla \log \pi_s^{(2)}(y) = \gamma_s^{-1}\mathbb{E}[U \mid X_s = y].$$

Since γ_s is bounded away from zero on this interval, the score tends to zero uniformly at spatial infinity. Gaussian convolution and positivity give joint continuity, hence boundedness on the remaining compact set. □

Lemma 41 (Brownian-plus-bounded decomposition)**.** *Under Assumption A5, fix $0 < \varepsilon < T < \infty$ and consider the Brownian backward SDE $\mathrm{d}\overleftarrow{Y}_u = [\frac{\bar\beta(u)}{2}\overleftarrow{Y}_u + \bar\beta(u)\nabla\log\pi^{(2)}_{T-u}(\overleftarrow{Y}_u)]\,\mathrm{d}u + \sqrt{\bar\beta(u)}\,\mathrm{d}W_u$, $\overleftarrow{Y}_0 \sim \mathcal{N}(0, I_d)$ independently of W. Then for each $\tau \in (0, T-\varepsilon]$,*

$$\overleftarrow{Y}_\tau = G_\tau + R_\tau, \qquad G_\tau \sim \mathcal{N}(0, \sigma_\tau^2 I_d), \quad \sigma_\tau^2 = 2e^{2H(\tau)} - 1, \quad \|R_\tau\| \le K_\tau := 2S_{\varepsilon,T}(e^{H(\tau)} - 1) \ \text{a.s.},$$

with $H(\tau) := \frac{1}{2}\int_0^\tau \bar\beta(v)\,\mathrm{d}v$ (and R_τ adapted, not independent of G_τ).

Proof. Variation of constants gives

$$G_\tau = e^{H(\tau)}\overleftarrow{Y}_0 + \int_0^\tau e^{H(\tau)-H(u)}\sqrt{\bar\beta(u)}\,\mathrm{d}W_u, \quad R_\tau = \int_0^\tau e^{H(\tau)-H(u)}\bar\beta(u)\nabla\log\pi^{(2)}_{T-u}(\overleftarrow{Y}_u)\,\mathrm{d}u.$$

The two terms defining G_τ are independent Gaussians; their summed per-coordinate variance is $e^{2H(\tau)} + (e^{2H(\tau)} - 1) = \sigma_\tau^2$. Lemma 40 gives $\|R_\tau\| \le S_{\varepsilon,T}\int_0^\tau e^{H(\tau)-H(u)}\bar\beta(u)\,\mathrm{d}u = 2S_{\varepsilon,T}(e^{H(\tau)} - 1) = K_\tau$. □

Corollary 42 (Brownian relative excess)**.** *Under Assumption A5, for every $\tau \in (0, T-\varepsilon]$ the Brownian generated law satisfies $e_{\mu_\tau^{(2)}}(u) = \mathcal{O}(1)$ and $\Gamma(\mu_\tau^{(2)}) = 0$, and its radial quantiles grow at most at square-root logarithmic order: $q_\kappa(\mu_\tau^{(2)}) = \mathcal{O}_d(\sigma_\tau\sqrt{\log\frac{1}{1-\kappa}}) + K_\tau$ for κ close to 1.*

Proof. Proposition 12(ii) applies to Lemma 41. Moreover, for $r > K_\tau$,

$$\mathbb{P}(\|\overleftarrow{Y}_\tau\| > r) \le 2^{d/2}e^{-(r-K_\tau)^2/(4\sigma_\tau^2)},$$

by $\|\overleftarrow{Y}_\tau\| \le \|G_\tau\| + K_\tau$ and $\mathbb{E}e^{\|G_\tau\|^2/(4\sigma_\tau^2)} = 2^{d/2}$. Inverting this upper bound gives the stated quantile estimate. □

E.3 Stable tail bounds for Proposition 14

Throughout this subsection $\mu_\tau = \mathrm{Law}(\overleftarrow{Y}_\tau)$ is the law at a fixed generation time $\tau \in (0, T-\varepsilon]$. The normalization of Appendix C gives $\nu_\alpha(\|z\| > r) = C_{d,\alpha} r^{-\alpha}$ and $\mathbb{P}(\|L_1^\alpha\| > r) \sim C_{d,\alpha} r^{-\alpha}$.

Scope. The fixed-time tail results in this subsection hold under Assumption A5 for every $\nu > 1$. Corollary 9 and the non-asymptotic generation bound require $\nu \geq 1+\alpha$, including equality. Proposition 6 gives the corresponding global regularity regimes.

The spatial asymptotics of m_s and its uniformly sublinear remainder were established in Proposition 30; we use that result below.

The next lemma requires only uniform sublinearity of the drift remainder, not independence from the stable leading term.

Lemma 43 (Tail stability under a uniformly sublinear drift)**.** *Fix $\tau > 0$ and consider*

$$\mathrm{d}\mathcal{X}_u = (\lambda(u)\mathcal{X}_u + \mathsf{r}(u, \mathcal{X}_u))\,\mathrm{d}u + \sigma(u)\,\mathrm{d}L_u^\alpha, \qquad \mathcal{X}_0 \sim \mathrm{S}\alpha\mathrm{S}_d(1) \text{ independent of } L^\alpha, \tag{E.1}$$

where λ, σ are bounded deterministic measurable scalar functions on $[0,\tau]$, and r is jointly measurable, locally Lipschitz in x uniformly on compact subsets of $[0,\tau] \times \mathbb{R}^d$, of linear growth, and uniformly sublinear: $\lim_{R\to\infty} \sup_{u\in[0,\tau]} \sup_{\|x\|\geq R} \|\mathsf{r}(u,x)\| / \|x\| = 0$. Then (E.1) has a unique non-explosive strong solution and, with $E(\tau, u) := \exp\Big(\int_u^\tau \lambda(v)\,\mathrm{d}v\Big)$, $\quad \rho_\tau^\alpha := E(\tau,0)^\alpha + \int_0^\tau E(\tau,u)^\alpha\,|\sigma(u)|^\alpha\,\mathrm{d}u$, one has $\mathbb{P}(\|\mathcal{X}_\tau\| > R) \sim C_{d,\alpha}\rho_\tau^\alpha R^{-\alpha}$ as $R \to \infty$.

Proof. Subtracting the additive noise integral reduces the equation to a Carathéodory equation; local Lipschitz continuity and linear growth give strong existence, uniqueness and non-explosion. On the same probability space set

$$\mathcal{W}_u = E(u,0)\mathcal{X}_0 + \int_0^u E(u,v)\sigma(v)\,\mathrm{d}L_v^\alpha.$$

The stable-integral characteristic function gives $\mathcal{W}_\tau \sim \mathrm{S}\alpha\mathrm{S}_d(\rho_\tau)$. The marginal tail constant agrees with that of the stable Lévy measure (Hult and Lindskog, 2007, Lemma 2.1), hence

$$\mathbb{P}(\|\mathcal{W}_\tau\| > R) \sim C_{d,\alpha}\rho_\tau^\alpha R^{-\alpha}. \tag{E.2}$$

Writing $\mathcal{W}_u = E(u,0)(\mathcal{X}_0 + M_u)$, the process $M_u = \int_0^u E(v,0)^{-1}\sigma(v)\,\mathrm{d}L_v^\alpha$ has the law of a stable process at the deterministic clock $\int_0^u |E(v,0)^{-1}\sigma(v)|^\alpha\,\mathrm{d}v$. The stable maximal bound (Schilling, 2016, Theorem 13.1) therefore gives

$$H_\tau := \limsup_{R\to\infty} R^\alpha \mathbb{P}\left(\sup_{u\leq\tau} \|\mathcal{W}_u\| > R\right) < \infty. \tag{E.3}$$

For any $\delta > 0$, uniform sublinearity gives $\|\mathsf{r}(u,x)\| \leq \delta\|x\| + K_\delta$. Variation of constants and Grönwall imply

$$\sup_{u\leq\tau} \|\mathcal{X}_u - \mathcal{W}_u\| \leq A_\delta \sup_{u\leq\tau} \|\mathcal{W}_u\| + B_\delta, \quad A_\delta = C_E\delta\tau e^{C_E\delta\tau}, \quad B_\delta = C_E K_\delta \tau e^{C_E\delta\tau},$$

where $C_E = \sup_{v\leq u\leq\tau} E(u,v) < \infty$. For every $\zeta > 0$, (E.3) bounds the limsup of $R^\alpha\mathbb{P}(\|\mathcal{X}_\tau - \mathcal{W}_\tau\| > \zeta R)$ by $H_\tau(A_\delta/\zeta)^\alpha$. Letting $\delta \downarrow 0$ yields

$$R^\alpha\mathbb{P}(\|\mathcal{X}_\tau - \mathcal{W}_\tau\| > \zeta R) \longrightarrow 0. \tag{E.4}$$

Sandwich $\mathbb{P}(\|\mathcal{X}_\tau\| > R)$ between the tails of $\|\mathcal{W}_\tau\|$ at $(1 \pm \zeta)R$, with the error in (E.4). Equation (E.2) and $\zeta \downarrow 0$ give the asserted constant. □

Lemma 44 (Linear-plus-sublinear decomposition of the backward drift)**.** *For $u \in [0, T-\varepsilon]$ set*

$$\ell_0(u) := \bar{\beta}(u)\Big(\tfrac{1}{\alpha} - \tfrac{1+\eta}{\alpha}\gamma(T-u)^{-\alpha}\Big) < 0, \qquad c(u) := \tfrac{1+\eta}{\alpha}\,\bar{\beta}(u)\,a(T-u)\,\gamma(T-u)^{-\alpha} > 0,$$

and, with $\varkappa_{\nu,s}$ as in (D.8),

$$\ell_\nu(u) := \ell_0(u) + c(u)\,\varkappa_{\nu,T-u} = \begin{cases} \dfrac{\bar{\beta}(u)}{\alpha}, & 1 < \nu < \alpha, \\ \bar{\beta}(u)\Big[\dfrac{1}{\alpha} - \dfrac{1+\eta}{\alpha}\cdot\dfrac{k_{d,\alpha}}{c_\pi a(T-u)^\alpha + k_{d,\alpha}\gamma(T-u)^\alpha}\Big], & \nu = \alpha, \\ \ell_0(u), & \nu > \alpha. \end{cases} \tag{E.5}$$

Then the backward drift of (2.4) splits as

$$\bar{b}(u,x) = \ell_\nu(u)\,x + \tilde{\mathsf{r}}(u,x), \qquad \tilde{\mathsf{r}}(u,x) := c(u)\,\mathsf{r}_{T-u}(x), \tag{E.6}$$

where ℓ_ν and c are bounded on $[0, T-\varepsilon]$ and $\tilde{\mathsf{r}}$ is locally Lipschitz, of linear growth, and uniformly sublinear in the sense of (D.9). Setting

$$E_\nu(\tau,u) := \exp\Big(\int_u^\tau \ell_\nu(v)\,\mathrm{d}v\Big), \qquad \rho_{\nu,\tau}^\alpha := E_\nu(\tau,0)^\alpha + \eta\int_0^\tau E_\nu(\tau,u)^\alpha\,\bar{\beta}(u)\,\mathrm{d}u \;\in (0,\infty), \tag{E.7}$$

the hypotheses of Lemma 43 hold with $\lambda = \ell_\nu$ and $\sigma(u) = \big(\eta\bar{\beta}(u)\big)^{1/\alpha}$.

Proof. The fractional Tweedie identity gives $\bar{b}(u,x) = \ell_0(u)x + c(u)m_{T-u}(x)$. Substitution of $m_s(x) = \varkappa_{\nu,s}x + \mathsf{r}_s(x)$ from Proposition 30 gives (E.6) and the three coefficients in (E.5). All coefficients are bounded on $[0, T-\varepsilon]$ since $\gamma(T-u) \geq \gamma(\varepsilon) > 0$; $\ell_0 < 0$ follows from $\gamma^\alpha < 1+\eta$. Multiplication by the bounded factor c preserves the local regularity, linear growth and uniform sublinearity of the remainder, so Lemma 43 applies with the announced coefficients. □

Theorem 45 (Exact tail of the stable generated law)**.** *Assume Assumption A5 for some $\nu > 1$. Then for every $\tau \in (0, T-\varepsilon]$,*

$$\overline{G}_{\mu_\tau}(r) \;\sim\; c_\tau\, r^{-\alpha} \quad (r \to \infty), \qquad c_\tau := C_{d,\alpha}\,\rho_{\nu,\tau}^\alpha \in (0,\infty), \tag{E.8}$$

with $\rho_{\nu,\tau}$ given by (E.7) (written ρ_τ in the main text). In particular the generated law is power-law of index α for every target index $\nu > 1$, *and, by Proposition 12,* $\Gamma(\mu_\tau^{(\alpha)}) = \frac{1}{\alpha-1}$.

Proof. Lemma 44 verifies every hypothesis of Lemma 43, with independent stable initialization. The latter gives (E.8) with $c_\tau = C_{d,\alpha}\rho_{\nu,\tau}^\alpha$ at each fixed generation time. Proposition 12(i) gives $\Gamma(\mu_\tau^{(\alpha)}) = 1/(\alpha-1)$. □

E.4 Proofs of Lemma 13 and Proposition 14

Proof of Lemma 13. Asymptotic inversion of regularly varying functions (Kevei, n.d., Theorem 16) gives $q_\kappa(\mu) \sim (c_\mu/(1-\kappa))^{1/\alpha}$. This applies to generalized quantiles as well; either inverse convention has the same equivalent. Apply the same result to π and take the ratio to obtain (5.1). □

Proof of Proposition 14. The tail regimes follow from Corollary 42 and Theorem 45. The Brownian quantile grows at most as $\sqrt{\log(1/(1-\kappa))}$, whereas the target quantile grows as $(c'_\pi/(1-\kappa))^{1/\nu}$, so their ratio vanishes. For the stable output, apply Lemma 13 with tail constants (c_τ, c'_π) and indices (α, ν). □

F Numerical Protocol for Figures 1–4

All reported experiments are one-dimensional, synthetic, and training-free. We set $\beta(t) \equiv 1$, and use $\alpha = 1.5$ and $\eta = 0.5$ for the stable backward runs. Scores are not learned: the model-specific continuous scores are evaluated numerically from exact forward characteristic functions. We distinguish deterministic numerical approximations, such as $\widehat{C}_t(40)$, from empirical Monte-Carlo quantities, such as $\widehat{W}_p$ and the tail estimators.

F.1 Common setup and score evaluation

On the backward grid $\tau_k = kh$, $t_k = T - \tau_k$, $h = (T - \varepsilon)/N$, the stable EI scheme is

$$Y_{k+1} = e^{-\lambda h} Y_k + \frac{1 - e^{-\lambda h}}{\lambda} F(Y_k, t_k) + (1 - e^{-\eta h})^{1/\alpha} Z_k, \qquad Z_k \sim \mathrm{S}\alpha\mathrm{S}_1(1), \tag{F.1}$$

where

$$\lambda = \frac{\eta}{\alpha}, \qquad F(x,t) = \frac{1+\eta}{\alpha}\big(x + \alpha S_t^{(\alpha)}(x)\big).$$

Thus the linear term and its stable stochastic convolution are integrated exactly, while F is frozen at the left endpoint. In particular, the noise scale in (F.1) uses the same η as the drift. The VP baseline uses the same splitting principle with $\lambda = 1/2$, $F_{\mathrm{VP}}(x,t) = x + s_t(x)$, and Gaussian noise of variance $1 - e^{-h}$. No trajectory is clipped or projected.

For a Student-t_ν target with characteristic function ϕ_ν, the forward characteristic functions are

$$\phi_t^{(\alpha)}(u) = \phi_\nu(e^{-t/\alpha}u)\exp\big(-(1-e^{-t})|u|^\alpha\big),$$
$$\phi_t^{(\mathrm{VP})}(u) = \phi_\nu(e^{-t/2}u)\exp\big(-\tfrac{1}{2}(1-e^{-t})u^2\big).$$

Fourier inversion of these exact expressions provides the density and the numerator of the corresponding score. Separate space–time tables are built for every pair of target and forward model and interpolated during propagation. Stable tables use 2^{19} Fourier modes and VP tables use 2^{18}; the main tables have 129×8193 points for the tail experiment and 257×4097 points for the initialization experiment. Independent evaluations and tables with doubled Fourier and interpolation resolutions are used as controls. Outside the tabulated spatial range, the known score asymptotics are used rather than truncating the state.

Table 3: Parameters of the three reported experiments. Stable backward runs use $\eta = 0.5$.

Experiment	Target and models	Time discretization	Sampling and displayed quantity
Denoiser modulus	Student-t_4; stable $\alpha = 1.5$, VP	64 times in $[0.05, 4.5]$; no backward dynamics	deterministic $\widehat{C}_t(40)$, no particles or seeds
Initialization	Student-t_4; stable $(\alpha, \eta) = (1.5, 0.5)$, VP	$\varepsilon = 0.05$, $h = 0.0125$, $T \in \{0.1, 0.15, 0.25, 0.5, 1, 2\}$	12 seeds, $n = 16384$; I_p for $p \in \{1.1, 1.4\}$
Tail coverage	Student-t_ν, $\nu \in \{1.2, 1.5, 3\}$; stable $\alpha = 1.5$, VP only for $\nu = 1.5$	$T = 2$, $\varepsilon = 0.5$, $N = 80$	12 seeds, $n = 262144$; Hill index, quantile ratio, and tail constant

F.2 Figure 1: Deterministic denoiser modulus

In this section, we describe how the Lipschitz modulus of the denoiser is computed in Figure 1, for Brownian and stable dynamics in the case of a Student-t_4 target. In the Student-t_4 case, $\phi_4(u) = 2u^2 K_2(2|u|)$ for $u \neq 0$, with $\phi_4(0) = 1$. Using the convention $\widehat{f}(u) = \int_{\mathbb{R}} f(x)e^{ixu}\,\mathrm{d}x$, the stable fractional-score numerator has multiplier $-iu|u|^{\alpha-2}$, while the ordinary VP-score numerator has multiplier $-iu$. A further spatial derivative contributes another factor $-iu$. Fourier inversion therefore gives p_t, g_t, p_t', g_t', from which

$$S_t'(x) = \frac{g_t'(x)}{p_t(x)} - \frac{g_t(x)p_t'(x)}{p_t(x)^2}.$$

The denoiser derivative is then evaluated from the Tweedie identities,

$$m_{t,\alpha}'(x) = \frac{1 + \alpha(1 - e^{-t})S_{t,\alpha}'(x)}{e^{-t/\alpha}}, \qquad m_{t,\mathrm{VP}}'(x) = \frac{1 + (1 - e^{-t})s_t'(x)}{e^{-t/2}}.$$

Figure 1 reports

$$\widehat{C}_t(40) = \max_{x \in \mathcal{G} \cap [-40,40]} |\widehat{m}_t'(x)|. \tag{F.2}$$

Here $\mathcal{G}$ is the evaluation grid and $\widehat{m}_t'$ is the numerically computed derivative. This is a truncated, discretized maximum, not a closed-form evaluation of $\sup_{\mathbb{R}} |m_t'|$.

The main Fourier grids use $(L, N_{\mathrm{FFT}}) = (8192, 2^{20})$ for the stable model and $(1024, 2^{19})$ for VP. At seven times, N_{FFT} is doubled at fixed L, leaving (F.2) stable at the displayed resolution. The maxima also agree on the nested domains $|x| \leq 8, 16, 24, 40$, with every detected maximizer lying in the innermost domain. Since the calculation does not involve random samples, Figure 1 shows no error bars.

F.3 Figure 2: Initialization error

For each T, model, and seed, two independent EI populations are propagated to $t = \varepsilon$. One starts from the model-specific stationary reference, $\mathrm{S}\alpha\mathrm{S}_1(1)$ or $\mathcal{N}(0,1)$; the other starts from an exact draw of the forward marginal

$$X_T \overset{\mathcal{L}}{=} e^{-T/\alpha} X_0 + (1 - e^{-T})^{1/\alpha} Z$$

on the stable side, with the Gaussian analogue on the VP side. The two arms use the same score table and step size but independent initial samples and innovations.

For equal-size one-dimensional samples X, Y, empirical Wasserstein distance is computed exactly by sorting,

$$\widehat{W}_p(X,Y) = \left(\frac{1}{n} \sum_{i=1}^{n} |X_{(i)} - Y_{(i)}|^p \right)^{1/p}.$$

The displayed initialization proxy is

$$I_p(T) = \widehat{W}_p(Y_{\varepsilon,h}^{\mathrm{stat}}, Y_{\varepsilon,h}^{\pi_T}), \qquad p \in \{1.1, 1.4\}.$$

The exact–exact reference

$$F_p(T) = \widehat{W}_p(X_\varepsilon^{(1)}, X_\varepsilon^{(2)})$$

uses two independent exact samples from the same π_ε. It measures finite-sample resolution, and is displayed on the final figure as an empirical sampling baseline to account for the quality of the statistical estimation in the experiment.

Figure 2 shows medians over 12 seeds with pointwise 90% percentile-bootstrap intervals for the median, obtained by resampling seed labels jointly across horizons. At $T = 2$, temporal refinements $N = 156, 312, 624$ and a doubled score table control the EI and interpolation errors. The choices $p = 1.1$ and $p = 1.4$ remain below $\alpha = 1.5$.

F.4 Figures 3 and 4: Tail coverage

Each Student-t_ν target determines its own forward marginal, score table, reverse drift, and target quantiles. Stable runs use $\nu \in \{1.2, 1.5, 3\}$, while the VP baseline is included only at $\nu = 1.5$. The main runs start from the model-specific stationary reference and use $T = 2$, $\varepsilon = 0.5$, and $N = 80$, with $n = 262144$ terminal samples per seed. Figure 4 collects three complementary diagnostics: the generated tail index, the scale of generated extreme quantiles relative to the target, and the tail constant relative to the EI prediction.

Tail index. Let $R_{(1)} \geq \cdots \geq R_{(n)}$ be the descending order statistics of the generated radii $R_i = |Y_i|$. Panel (a) reports the reciprocal Hill estimate (Hill, 1975)

$$\widehat{\alpha}_k = \left(\frac{1}{k} \sum_{j=1}^{k} \log \frac{R_{(j)}}{R_{(k+1)}} \right)^{-1},$$

which estimates the power-law exponent from logarithmic excesses above the threshold $R_{(k+1)}$. It is evaluated at the pre-specified fractions

$$k/n \in \{0.003, 0.004, 0.006, 0.008, 0.012, 0.018, 0.03\},$$

with principal fraction $k/n = 0.006$. Smaller fractions use fewer, more extreme observations; comparison across fractions assesses sensitivity to the threshold. The reference value is $\alpha = 1.5$ for all three stable outputs, irrespective of the target index ν.

Extreme-quantile ratio. Writing $s = 1 - \kappa$, panel (b) reports

$$\widehat{\mathcal{M}}_\nu(s) = \frac{\widehat{q}_{1-s}(|Y|)}{q_{1-s}(|T_\nu|)},$$

where $\widehat{q}$ is the empirical quantile of the generated radii and T_ν is a Student-t_ν random variable. The denominator is evaluated from the Student CDF F_ν through $q_{1-s}(|T_\nu|) = F_\nu^{-1}(1 - s/2)$, without a simulated target sample. The horizontal axis decreases to the right, so the rightmost points correspond to more extreme quantiles.

For the stable outputs, the predicted slope of $\log \mathcal{M}_\nu(s)$ against $\log s$ is $1/\nu - 1/\alpha$. Thus, as $s \downarrow 0$, the ratio should decrease for $\nu < \alpha$, approach a finite nonzero constant for $\nu = \alpha$, and increase for $\nu > \alpha$. At index match, this constant need not equal one: equality additionally requires matching tail constants. Slopes are fitted separately within each seed over the pre-specified extreme-quantile window, then aggregated across seeds.

Tail constant. For a putative α-tail, panel (c) uses

$$\widehat{c}_k = (k/n) R_{(k+1)}^{\alpha}.$$

This follows from $\mathbb{P}(|Y| > r) \sim c r^{-\alpha}$, replacing the survival probability at $R_{(k+1)}$ by k/n. The displayed quantity is $\widehat{c}_k / c_{N,h}^{\mathrm{EI}}$, where $c_{N,h}^{\mathrm{EI}}$ is the tail constant predicted for the actual finite-step EI chain.

More precisely, let G_k denote the deterministic part of one step of (F.1), and set $r_k = \lim_{x\to\infty} G_k(x)/x$. For these symmetric targets, the same asymptotic slope applies at both tails. If c_k is the two-sided α-tail constant before that step, asymptotic linearity and addition of independent regularly varying tails (Foss et al., 2009, Corollary 7) give

$$c_{k+1} = |r_k|^\alpha c_k + C_{1,\alpha}(1 - e^{-\eta h}), \qquad c_0 = C_{1,\alpha},$$

where $C_{1,\alpha} = \lim_{r\to\infty} r^\alpha \mathbb{P}(|Z| > r)$ for $Z \sim \mathrm{S}\alpha\mathrm{S}_1(1)$. We set $c^{\mathrm{EI}}_{N,h} = c_N$. The horizontal level one in panel (c) therefore tests agreement with the discretized chain's tail constant; it does not assert equality with the target tail constant or the continuous-time limit. Panels (a) and (c) concern only the stable samplers.

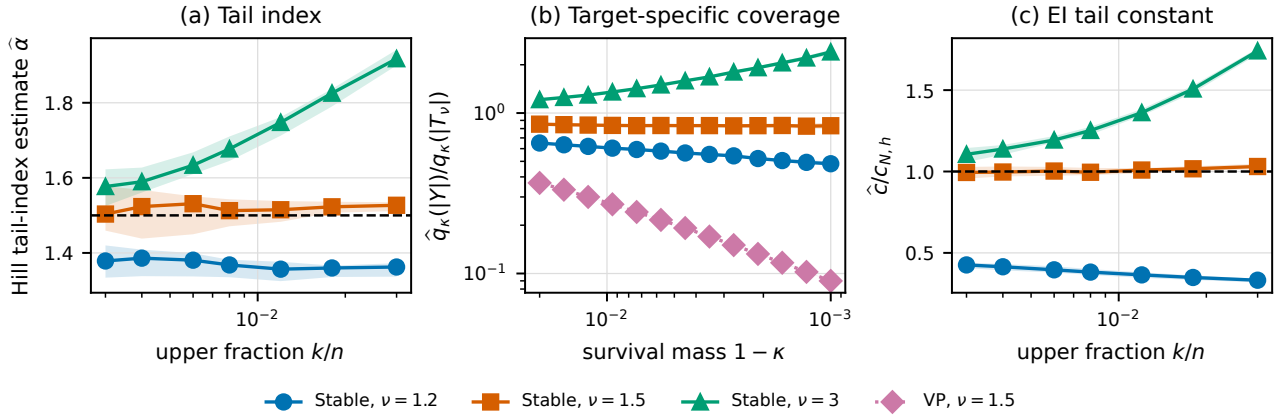

Figure 4: Tail diagnostics for Student-t_ν targets with stable $(\alpha, \eta) = (1.5, 0.5)$, $T = 2$, $\varepsilon = 0.5$, and $N = 80$. **(a)** Hill tail-index estimate; the dashed reference is $\alpha = 1.5$. **(b)** Generated-to-target extreme-quantile ratio; the Brownian VP baseline is included for $\nu = 1.5$. **(c)** Estimated tail constant normalized by the finite-step EI prediction; the dashed reference is one. Curves are medians over 12 independent seeds, and shaded bands show the inter-seed 16%–84% range.

Interpretation. At index match, $\nu = \alpha = 1.5$, the three diagnostics are mutually consistent: the Hill estimate remains close to 1.5, the extreme-quantile ratio is approximately constant, and the normalized tail constant remains close to one. The quantile plateau is nevertheless below one over the displayed range, so generated extreme quantiles remain smaller than their target counterparts. This illustrates the distinction between matching the tail index and reproducing the target's extreme scales: agreement with the EI tail constant in panel (c) does not remove the residual target discrepancy measured in panel (b).

Away from index match, panel (b) exhibits the predicted directions: the ratio decreases for $\nu = 1.2 < \alpha$ and increases for $\nu = 3 > \alpha$ as the quantiles become more extreme. The other diagnostics are less conclusive. For $\nu = 1.2$, the Hill estimates and normalized tail constants remain below their references; for $\nu = 3$, both drift upward as larger upper fractions are included. Thus the displayed thresholds do not yield joint stabilization of the tail-index and tail-constant estimates. These observations support the qualitative quantile comparison but do not resolve all asymptotic predictions at the available sample size.

For the Brownian baseline at $\nu = 1.5$, the quantile ratio decreases markedly as $s \downarrow 0$, whereas the stable ratio remains approximately flat. This illustrates increasingly undersized Brownian extreme quantiles, consistently with Proposition 14.

Numerical and statistical controls. Temporal refinements $N \in \{40, 80, 160\}$, independent high-resolution score tables, exact-π_T initialization, and direct SαS, Pareto, and Gaussian samples provide numerical and estimator controls. The displayed inter-seed bands describe variability across replications and are not confidence intervals; seed-level bootstrap diagnostics use 10000 resamples. At index match, the displayed diagnostics are qualitatively consistent with the predicted tail index, a constant quantile ratio, and the EI tail constant. Away from index match, threshold drift prevents an equally conclusive

joint interpretation. Pickands estimates, generalized Pareto fits, and mean-excess ratios are retained as supplementary checks.

F.5 Implementation and numerical controls

Stable variables use the convention $\mathbb{E}[e^{iuZ}] = e^{-|u|^\alpha}$ and are generated with the Chambers–Mallows–Stuck construction (Chambers et al., 1976). Random seeds are independent across reported replications; common random numbers are used only for paired resolution controls. Propagation costs $O(nN)$, and each one-dimensional Wasserstein evaluation costs $O(n \log n)$ because of sorting.

The reported propagation runs were produced with JAX (Bradbury et al., 2018) in CPU `float64`; the deterministic denoiser calculation uses NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020) FFTs in `float64`. The environment was Windows 11 with Python 3.12.10, JAX/JAXLIB 0.7.2, NumPy 2.3.3, SciPy 1.16.2, and 16 logical CPU cores. No GPU result is reported. Configurations, seeds, score-table hashes, raw seed-level summaries, and figure-generation commands are included in the supplementary code, available at `https://github.com/ainsi333/stable-sbgm-experiments`.

The project licenses for the stated versions are Apache-2.0 for JAX, BSD-3-Clause for NumPy, and BSD-3-Clause for SciPy. These identify the upstream project licenses; bundled third-party components retain their own applicable licenses and notices.