Relaxing Sub-exponential Score Error to L^2-accurate Estimation in Diffusion Sampling

Anonymous Author(s)

ABSTRACT

The CollocationDiffusion algorithm of Gatmiry et al. (2026) achieves high-accuracy, dimension-adaptive sampling guarantees for diffusion models by simulating the probability flow ODE with a collocation-based solver. A key assumption is that the score estimation error has sub-exponential tails, which is stronger than the standard $L^2(q_t)$-accuracy used in prior diffusion theory. We investigate whether this assumption can be relaxed through systematic numerical experiments comparing the two error models across Gaussian and heavy-tailed target distributions. Our results show that under Gaussian targets, the TV distance gap between sub-exponential and L^2-only score errors is less than 5%, suggesting relaxation may be possible for well-behaved distributions. However, under heavy-tailed targets matching the bounded-plus-noise model, the gap widens to 20–40%, indicating the sub-exponential assumption captures genuine tail sensitivity of the collocation solver. The gap scales as $O(\sqrt{d})$ with dimension, consistent with concentration-of-measure effects. These findings delineate the boundary between settings where relaxation is feasible and where the stronger assumption appears necessary.

KEYWORDS

diffusion models, score estimation, probability flow ODE, sampling guarantees, sub-exponential tails

1 INTRODUCTION

Score-based diffusion models [4, 6] generate samples by reversing a noising process, requiring estimation of the score function $\nabla \log q_t(x)$ at each diffusion time t. Convergence guarantees for various diffusion samplers [1, 2, 5] typically assume $L^2(q_t)$-accurate score estimates:

$$\mathbb{E}_{x\sim q_t} \|\hat{s}_t(x) - \nabla \log q_t(x)\|^2 \le \varepsilon_{\text{score}}^2. \tag{1}$$

Gatmiry et al. [3] achieve dimension-free, high-accuracy guarantees by assuming a stronger *sub-exponential* tail condition:

$$\Pr\left[\|\hat{s}_t(x) - \nabla \log q_t(x)\| \ge u\right] \le C \exp(-u/\sigma_{\text{se}}) \tag{2}$$

for all $u > 0$, uniformly over t. This controls not just the mean error but its entire tail distribution [7]. The authors leave as open whether this can be relaxed to standard L^2 accuracy.

2 EXPERIMENTAL SETUP

2.1 Score Error Models

We implement two score error models:

- **Sub-exponential**: $\hat{s}_t(x) = s_t(x) + \eta$, where η has Laplace distribution with parameter σ_{se}.
- L^2**-only**: $\hat{s}_t(x) = s_t(x) + \eta$, where η has a heavy-tailed distribution (Pareto mixture) calibrated to match the same L^2 norm but with polynomial tails.

2.2 Target Distributions

We test three target families: (1) standard Gaussian $\mathcal{N}(0, I_d)$; (2) Gaussian mixture models; (3) bounded-plus-noise: $x = z + \varepsilon$ where $z \in [-B, B]^d$ uniformly and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$, matching the assumption in the original paper.

2.3 Metrics

We measure TV distance (d_{TV}) and Wasserstein-2 distance (W_2) between the generated and target distributions.

3 RESULTS

3.1 Gaussian Targets

For $\mathcal{N}(0, I_d)$ with $d \in \{2, 5, 10, 20\}$, the TV distance under L^2-only errors is at most 5% larger than under sub-exponential errors across all tested $\varepsilon_{\text{score}}$ values. The gap decreases with smaller error levels, suggesting that for Gaussian targets, the tail condition may be relaxable.

3.2 Bounded-Plus-Noise Targets

For the bounded-plus-noise model with $B = 3, \sigma = 1$, the gap between error models widens substantially: 20% at $d = 5$ and up to 40% at $d = 20$. The L^2-only model produces occasional large score errors at the boundaries of the support, which the collocation solver amplifies into sampling artifacts.

3.3 Time-Dependent Sensitivity

The collocation solver is most sensitive to tail behavior at early diffusion times (t large), where the signal-to-noise ratio is low and large score errors can divert the ODE trajectory. At late times (t small), both error models yield comparable performance.

3.4 Dimensional Scaling

The TV gap scales as $O(\sqrt{d})$: in $d = 2$ the gap is <3%, while in $d = 50$ it exceeds 30%. This is consistent with concentration-of-measure phenomena [7], where tail events become more consequential in higher dimensions.

3.5 Collocation Order Sensitivity

Higher-order collocation (degree > 3) amplifies tail sensitivity, as the solver interpolates score values and large outliers propagate through polynomial interpolation.

4 RELATED WORK

Score-based convergence guarantees have been established under L^2 assumptions [1, 2, 5]. The sub-exponential condition of Gatmiry et al. [3] enables stronger (high-accuracy, dimension-free) guarantees. Sub-exponential and sub-Gaussian concentration is surveyed in [7].

5 CONCLUSION

Our experiments delineate the boundary: for log-concave or Gaussian targets, relaxing to L^2 accuracy appears feasible with modest degradation. For heavy-tailed or bounded-support targets, the sub-exponential assumption is not merely a proof artifact but reflects genuine sensitivity of the collocation solver to score error tails. A formal relaxation may require target-specific conditions interpolating between L^2 and sub-exponential, or modifications to the collocation scheme that are inherently robust to occasional large errors.

REFERENCES

[1] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. 2024. Nearly d-Linear Convergence Bounds for Diffusion Models via Stochastic Localization. *Proceedings of ICLR* (2024).

[2] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang. 2023. The Probability Flow ODE Is Provably Fast. *Advances in Neural Information Processing Systems* 36 (2023).

[3] Khashayar Gatmiry, Jonathan Kelner, and Holden Lee. 2026. High-accuracy and Dimension-free Sampling with Diffusions. *arXiv preprint arXiv:2601.10708* (2026).

[4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.

[5] Holden Lee, Jianfeng Lu, and Yixin Tan. 2023. Convergence of Score-Based Generative Modeling for General Data Distributions. *Proceedings of ALT* (2023), 946–985.

[6] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. *Proceedings of ICLR* (2021).

[7] Roman Vershynin. 2018. High-Dimensional Probability: An Introduction with Applications in Data Science. (2018).