

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

Research

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:

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

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

$$\Pr [\|\hat{s}_t(x) - \nabla \log q_t(x)\| \geq u] \leq C \exp(-u/\sigma_{\text{se}}) \quad (2)$$

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

117 et al. [3] enables stronger (high-accuracy, dimension-free) guarantees.
 118 Sub-exponential and sub-Gaussian concentration is surveyed
 119 in [7].

120 121 5 CONCLUSION

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

131 132 REFERENCES

- | | |
|---|---|
| 133 [1] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. 2024.
134 Nearly d-Linear Convergence Bounds for Diffusion Models via Stochastic Localiza-
135 tion. <i>Proceedings of ICLR</i> (2024).
136 [2] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R Zhang.
137 2023. The Probability Flow ODE Is Provably Fast. <i>Advances in Neural Information
 138 Processing Systems</i> 36 (2023).
139 [3] Khashayar Gatmiry, Jonathan Kelner, and Holden Lee. 2026. High-accuracy and
140 Dimension-free Sampling with Diffusions. <i>arXiv preprint arXiv:2601.10708</i> (2026).
141 [4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic
142 Models. <i>Advances in Neural Information Processing Systems</i> 33 (2020), 6840–6851.
143 [5] Holden Lee, Jianfeng Lu, and Yixin Tan. 2023. Convergence of Score-Based
144 Generative Modeling for General Data Distributions. <i>Proceedings of ALT</i> (2023),
145 946–985.
146 [6] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano
147 Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic
148 Differential Equations. <i>Proceedings of ICLR</i> (2021).
149 [7] Roman Vershynin. 2018. High-Dimensional Probability: An Introduction with
150 Applications in Data Science. (2018). | 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 |
|---|---|