Broadening Target Distributions for Accelerated Diffusion Models via a Novel Analysis Approach

Yuchen Liang[†], Peizhong Ju[‡], Yingbin Liang[†], Ness Shroff[†]

[†]The Ohio State University [‡]University of Kentucky

Presented at ICLR 2025

Overview

Background on Diffusion Models

- Problem: Generate samples from a complex distribution (e.g., images)
- Denoising Diffusion Probabilistic Models (DDPMs) (Ho *et al.*, 2020)

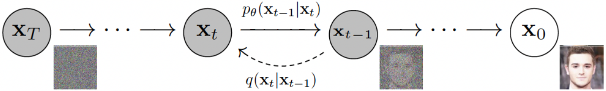


Figure 2: The directed graphical model considered in this work.

1. Forward process Q: $x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t} w_t$, for **training**
2. Reverse process $\widehat{P}$: $x_{t-1} = \widehat{\mu}_t(x_t) + \sqrt{\frac{1-\alpha_t}{\alpha_t}} z_t$, for **sampling**

- Train for posterior-mean $\widehat{\mu}_t(x_t)$
- Good sample quality, but slow...

Literature Review

Regular DDPMs:

- Best theoretical result: Q_0 finite var, $\mathrm{KL}(Q_0||\widehat{P}_0) \lesssim dT^{-1} + \varepsilon^2_{\mathsf{score}}$ (Chen *et al.*, 2023, Benton *et al.*, 2024)
- Existing Analysis: SDE-based analysis (Fokker-Planck equation, Girsanov change-of-measure) (Chen *et al.*, 2023a, Benton *et al.*, 2024a), typical-sets (Li *et al.*, 2024c)

Literature Review

Regular DDPMs:

- Best theoretical result: Q_0 finite var, $\mathrm{KL}(Q_0||\widehat{P}_0) \lesssim dT^{-1} + \varepsilon^2_{\text{score}}$ (Chen *et al*., 2023, Benton *et al*., 2024)
- Existing Analysis: SDE-based analysis (Fokker-Planck equation, Girsanov change-of-measure) (Chen *et al*., 2023a, Benton *et al*., 2024a), typical-sets (Li *et al*., 2024c)

Accelerated DDPMs:

- Existing Samplers: additional estimates (Li *et al*., 2024c), intermediate points (Li *et al*., 2024a), ODE (Chen *et al*., 2023c)...
- Challenge: Strong assumptions on Q
  1. **All** Q_t's have Lipschitz scores (hard to check)
  2. Q_0 **bounded support** (not finite var)

Main Questions

1. Can we obtain an accelerated convergence rate for a much **broader set of target distributions**?
   - Can we relax the Lipschitz smoothness (i.e., not on the entire path)?
   - Can we only require finite variance?
2. For existing cases, can we achieve a **lower dimensional dependency**?

New Accelerated Sampler

- Intuition: Approximate **posterior variance**
  - In comparison, the regular sampler only approximates posterior mean and assumes $\Sigma_t(x_t) \equiv \frac{1-\alpha_t}{\alpha_t} I_d$
- Define true post var $\Sigma_t(x_t) := \frac{1-\alpha_t}{\alpha_t}\left(I_d + (1-\alpha_t)\nabla^2 \log q_t(x_t)\right)$
  - Define $H_t(x_t)$ an estimate for $\nabla^2 \log q_t(x_t)$
  - We also provide examples how to estimate for H_t with theoretical guarantees (See Eqns (8) and (9) in paper)
- Define the **accelerated** reverse process: $x_{t-1} = \hat{\mu}_t(x_t) + \hat{\Sigma}_t^{\frac{1}{2}}(x_t) z_t$

Accelerated Convergence

- Noise schedule: $1-\alpha_t \lesssim \frac{\log T}{T}$ while $\bar{\alpha}_T := \prod_{t=1}^{T} \alpha_t = o\left(T^{-2}\right)$
- List of Assumptions:
  1. **Finite-var** Q_0
  2. Absolute continuity (when q_0 exists)
  3. Score and Hessian est error: $\varepsilon^2_{\text{score}} = \tilde{O}(T^{-2})$ and $\varepsilon^2_{\text{Hessian}} = \tilde{O}(T^{-1})$
  4. Regular partial derivatives

Accelerated Convergence

- Noise schedule: $1-\alpha_t \lesssim \frac{\log T}{T}$ while $\bar{\alpha}_T := \prod_{t=1}^{T} \alpha_t = o\left(T^{-2}\right)$
- List of Assumptions:
  1. **Finite-var** Q_0
  2. Absolute continuity (when q_0 exists)
  3. Score and Hessian est error: $\varepsilon^2_{\text{score}} = \tilde{O}(T^{-2})$ and $\varepsilon^2_{\text{Hessian}} = \tilde{O}(T^{-1})$
  4. Regular partial derivatives
- Theorem 1: Under these assumptions,

$$
\begin{aligned}
&\mathrm{KL}(Q_0 || \widehat{P}'_0) \\
&\lesssim (\log T)\varepsilon^2_{\text{score}} + \frac{\log^2 T}{T}\varepsilon^2_{\text{Hessian}} + \sum_{t=1}^{T}(1-\alpha_t)^3 \times \\
&\qquad \mathbb{E}_{X_t \sim Q_t} \sum_{i,j,k=1}^{d} \partial^3_{ijk} \log q_{t-1}(\mu_t(X_t)) \partial^3_{ijk} \log q_t(X_t) \\
&= \tilde{O}(T^{-2})
\end{aligned}
$$

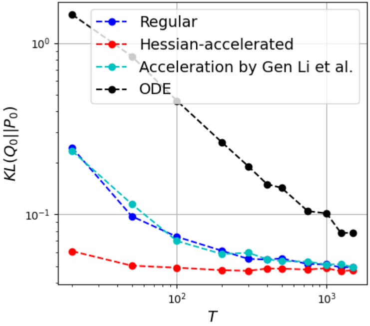
- Analysis not based on SDE techniques, but **tilting-factors**

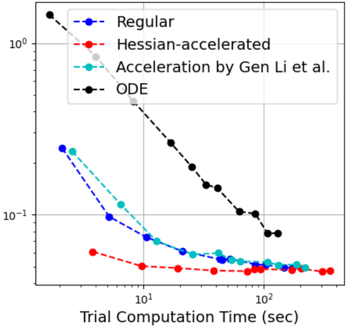
Example Q_0's with Explicit Parameter Dependency

Target distribution Q_0	Method	Num of steps	Results
$\nabla \log q_t, s_t$ L-Lips. $\forall t$	ODE-based	$\mathcal{O}\left(\frac{\sqrt{d}L^2}{\varepsilon}\right)$	(Chen et al., 2023c, Thm 3)
$\nabla \log q_t$ L-Lips. $\forall t$	DDPM accl.	$\mathcal{O}\left(\frac{\sqrt{d}L^2}{\varepsilon}\right)$	(Huang et al., 2024b, Thm 4.4)†
$\left\|\partial_{\boldsymbol{a}}^k s_t(x)\right\| \leq L\ \forall x, t, \boldsymbol{a}$ and $\forall k \leq p+1$, Q_0 Bounded Support	ODE	$\mathcal{O}\left(\frac{d^{\frac{p+1}{p}}}{\varepsilon^{\frac{1}{p}}}\right)^*$	(Huang et al., 2024a, Thm 3.10)†
$\nabla^2 \log q_0$ M-Lips.	DDPM accl.	$\mathcal{O}\left(\frac{d^{1.5}\log^{1.5} M}{\varepsilon}\right)$	(This paper, Thm 4)
Q_0 Gaussian Mixture	DDPM accl.	$\mathcal{O}\left(\frac{d^{1.5}N^{1.5}}{\varepsilon}\right)$	(This paper, Thm 2)
Q_0 Bounded Support	DDPM accl.	$\mathcal{O}\left(\frac{d^3}{\varepsilon}\right)^*$	(Li et al., 2024c, Thm 4) (Li et al., 2024a, Thm 2)†
	ODE	$\mathcal{O}\left(\frac{d^3}{\sqrt{\varepsilon}}\right)^*$	(Li et al., 2024c, Thm 2) (Li et al., 2024a, Thm 1)†
	ODE	$\mathcal{O}\left(\frac{d^2}{\varepsilon}\right)^*$	(Li et al., 2024c, Thm 1)
Q_0 Finite Variance	DDPM accl.	$\mathcal{O}\left(\frac{d^{1.5}}{\varepsilon}\right)^*$	(This paper, Thm 3)

- Broader set of Q_0
- Better dimensional dependencies (for finite var)

Numerical Performance





- Q_0: Gaussian mixture

Conclusion

- Accelerated convergence on broader set of target distributions
  - ...for both smooth Q_0 and general Q_0 with early-stopping
- New Hessian-based accelerated sampler
- Novel tilting-factor analysis technique