Proceedings of Machine Learning Research vol 291:1–2, 2025 38th Annual Conference on Learning Theory

Mixing Time of the Proximal Sampler in Relative Fisher Information via Strong Data Processing Inequality (Extended Abstract)

Andre Wibisono ANDRE.WIBISONO@YALE.EDU
Department of Computer Science, Yale University, New Haven, CT, USA

Editors: Nika Haghtalab and Ankur Moitra

Abstract

We study sampling from a probability distribution $\nu \propto e^{-f}$ on $\mathbb{R}^d$, from the perspective of minimizing relative entropy (KL divergence) $H_\nu(\rho) = \mathsf{KL}(\rho \,\|\, \nu) = \mathbb{E}_\rho[\log \frac{\rho}{\nu}]$ on the space of probability distributions $\mathcal{P}(\mathbb{R}^d)$ with the Wasserstein geometry: $\nu = \arg\min_{\rho \in \mathcal{P}(\mathbb{R}^d)} H_\nu(\rho)$. The Langevin dynamics is a continuous-time stochastic process in $\mathbb{R}^d$ that implements the Wasserstein gradient flow for minimizing H_ν in $\mathcal{P}(\mathbb{R}^d)$. The relative Fisher information $\mathsf{FI}(\rho \,\|\, \nu) = \mathbb{E}_\rho[\|\nabla \log \frac{\rho}{\nu}\|^2]$ has the geometric meaning as the squared Wasserstein gradient of relative entropy H_ν. When ν is strongly log-concave ($-\log \nu$ is a strongly convex function on $\mathbb{R}^d$), relative entropy H_ν is a strongly convex function on $\mathcal{P}(\mathbb{R}^d)$, and Langevin dynamics has fast convergence guarantees including in relative Fisher information: $\mathsf{FI}(\rho_t \,\|\, \nu) \le e^{-2\alpha t} \mathsf{FI}(\rho_0 \,\|\, \nu)$. In discrete time, we study the Proximal Sampler (Lee et al., 2021), a two-step Gibbs sampling algorithm to sample from an auxiliary joint distribution which has the original target distribution as the x-marginal. The Proximal Sampler can be seen as an approximate proximal discretization of the Langevin dynamics, and it has matching convergence rates with the continuous-time Langevin dynamics in many settings, for example an exponential convergence rate in KL divergence under log-Sobolev inequality (Chen et al., 2022).

In this work, we show that when ν is α-strongly log-concave, Proximal Sampler also has an exponential convergence in relative Fisher information: $\mathsf{FI}(\rho_k \,\|\, \nu) \le (1+\alpha\eta)^{-2k} \cdot \mathsf{FI}(\rho_0 \,\|\, \nu)$. This matches the convergence guarantee from the Langevin dynamics in continuous time, and from the proximal gradient algorithm in optimization. If ν is α-strongly log-concave and L-log-smooth, then using standard rejection sampling implementation of Proximal Sampler, we conclude its iteration complexity is $k = \tilde{O}(\frac{dL}{\alpha} \log \frac{\mathsf{FI}(\rho_0 \,\|\, \nu)}{\varepsilon})$ to reach $\mathsf{FI}(\rho_k \,\|\, \nu) \le \varepsilon$.

Our analysis proceeds via establishing the *strong data processing inequality (SDPI)* for a family of Fokker-Planck channels driven by diffusion processes, including the Gaussian channel, the Ornstein-Uhlenbeck (OU) channel, the Langevin dynamics, and the reverse Gaussian channel. We show that even along the Gaussian channel, data processing inequality in relative Fisher information may not hold when the second distribution is arbitrary. We also show along the Gaussian channel, (S)DPI in relative Fisher information holds when the second distribution is (strongly) log-concave; we also show SDPI in relative Fisher information eventually holds when the second distribution is a log-Lipschitz perturbation of a strongly log-concave distribution. Along the Ornstein-Uhlenbeck channel, we show that SDPI in relative Fisher information eventually holds when the second distribution is strongly log-concave, and exhibit an example where DPI initially does not hold even when both input distributions are Gaussian. For our algorithmic result, we can write the Proximal Sampler as a composition of the Gaussian and reverse Gaussian channels. Then we can combine the SDPI for the Gaussian channel under SLC and the DPI for the reverse Gaussian channel to show that relative Fisher information converges exponentially fast along the Proximal Sampler.[1]

Keywords: Mixing time, Fisher information, Langevin dynamics, data processing inequality

Acknowledgments: The author is supported by NSF Award CCF-2403391 and CCF-2443097.

1. Extended abstract. Full version appears as (Wibisono, 2025): `https://arxiv.org/abs/2502.05623v2`.

References

Yongxin Chen, Sinho Chewi, Adil Salim, and Andre Wibisono. Improved analysis for a proximal algorithm for sampling. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 2984–3014. PMLR, 02–05 Jul 2022.

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Structured logconcave sampling with a restricted Gaussian oracle. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 2993–3050. PMLR, 8 2021.

Andre Wibisono. Mixing time of the proximal sampler in relative Fisher information via strong data processing inequality, 2025. URL https://arxiv.org/abs/2502.05623v2.