Resolving the Mixing Time of the Langevin Algorithm to its Stationary Distribution for Log-Concave Sampling

Jason M. Altschuler JA4775@NYU.EDU
NYU

Kunal Talwar KTALWAR@APPLE.COM
Apple

Editors: Gergely Neu and Lorenzo Rosasco

Sampling from a high-dimensional distribution is a fundamental task in statistics, engineering, and the sciences. A canonical approach is the Langevin Algorithm, i.e., the Markov chain for the discretized Langevin Diffusion. This is the sampling analog of Gradient Descent. However, despite being studied for several decades in multiple communities, tight mixing time bounds (a.k.a., convergence analyses) for this algorithm remain unresolved even in the seemingly simple setting of log-concave distributions over a bounded domain.

Our main result characterizes the mixing time of the Langevin Algorithm to its stationary distribution in this setting (and others). This mixing result can be combined with any bound on the discretization bias in order to sample from the stationary distribution of the Langevin Diffusion. In this way, we disentangle the study of the mixing and bias of the Langevin Algorithm.

Our key technical insight is to use a new Lyapunov function to analyze the convergence of the Langevin Algorithm. Specifically, the Lyapunov function we propose is the shifted divergence—a quantity studied in the differential privacy literature (Feldman et al., 2018; Altschuler and Talwar, 2022). Briefly, this Lyapunov function is a version of the Rényi divergence that is smoothed in optimal transport distance, and we use the amount of smoothing to measure the progress of the Langevin Algorithm. In addition to giving a short, simple proof of optimal mixing bounds, this analysis approach also has several additional appealing properties:

- Our approach removes all unnecessary assumptions required by other sampling analyses (e.g., warm starts, curvature assumptions on the set, etc.).
- Our approach unifies many settings: it gives tight bounds if the Langevin Algorithm uses projections, mini-batch gradients, or strongly convex potentials (whereby our mixing time improves exponentially).
- Our approach unifies many metrics: it proves mixing in the stringent notion of Rényi divergence, which implies mixing in all common metrics via standard comparison inequalities.
- Our approach exploits convexity only through the contractivity of a gradient step—reminiscent of how convexity is used in textbook proofs of Gradient Descent.

. Extended abstract. Full version appears as arXiv:2210.08448, v2.

References

Jason M Altschuler and Kunal Talwar. Privacy of Noisy Stochastic Gradient Descent: More iterations without more privacy loss. In *Advances in Neural Information Processing Systems*, 2022.

Vitaly Feldman, Ilya Mironov, Kunal Talwar, and Abhradeep Thakurta. Privacy amplification by iteration. In *Symposium on Foundations of Computer Science*, pages 521–532. IEEE, 2018.