Characterizing Dependence of Samples along the Langevin Dynamics and Algorithms via Contraction of Φ-Mutual Information

Jiaming Liang JIAMING.LIANG@ROCHESTER.EDU
Goergen Institute for Data Science & Department of Computer Science, University of Rochester

Siddharth Mitra SIDDHARTH.MITRA@YALE.EDU
Andre Wibisono ANDRE.WIBISONO@YALE.EDU
Department of Computer Science, Yale University

Editors: Nika Haghtalab and Ankur Moitra

Abstract

The mixing time of a Markov chain determines how fast the iterates of the Markov chain converge to the stationary distribution; however, it does not control the dependencies between samples along the Markov chain. In this paper, we study the question of how fast the samples become approximately independent along popular Markov chains for continuous-space sampling: the Langevin dynamics in continuous time (Villani, 2009; Chafaï, 2004), and the Unadjusted Langevin Algorithm (Dalalyan, 2017; Cheng and Bartlett, 2018; Vempala and Wibisono, 2019) and the Proximal Sampler (Lee et al., 2021; Chen et al., 2022; Mitra and Wibisono, 2025) in discrete time. We measure the dependence between samples via Φ-mutual information (Polyanskiy and Wu, 2025; Raginsky, 2016), which is a broad generalization of the standard mutual information, and which is equal to 0 if and only if the the samples are independent. We show that along these Markov chains, the Φ-mutual information between the first and the k-th iterate decreases to 0 exponentially fast in k when the target distribution is strongly log-concave. Our proof technique is based on showing the Strong Data Processing Inequalities (SDPIs) hold along the Markov chains (Polyanskiy and Wu, 2025). To prove fast mixing of the Markov chains, we only need to show the SDPIs hold for the stationary distribution. In contrast, to prove the contraction of Φ-mutual information, we need to show the SDPIs hold along the entire trajectories of the Markov chains; we prove this when the iterates along the Markov chains satisfy the corresponding Φ-Sobolev inequality, which is implied by the strong log-concavity of the target distribution.

Keywords: Markov chain, Langevin dynamics, unadjusted Langevin algorithm, proximal sampler, Φ-mutual information, Φ-Sobolev inequality, strong data processing inequality

1. Extended abstract. Full version appears as [arXiv:2402.17067v3].

References

Djalil Chafaï. Entropies, convexity, and functional inequalities, On Φ-entropies and Φ-Sobolev inequalities. *Journal of Mathematics of Kyoto University*, 2004.

Yongxin Chen, Sinho Chewi, Adil Salim, and Andre Wibisono. Improved analysis for a proximal algorithm for sampling. In *Conference on Learning Theory*, pages 2984–3014. PMLR, 2022.

Xiang Cheng and Peter Bartlett. Convergence of Langevin MCMC in KL-divergence. In *Algorithmic Learning Theory*, pages 186–211. PMLR, 2018.

Arnak Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In *Conference on Learning Theory*, pages 678–689. PMLR, 2017.

Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Structured logconcave sampling with a restricted Gaussian oracle. In *Conference on Learning Theory*, pages 2993–3050. PMLR, 2021.

Siddharth Mitra and Andre Wibisono. Fast Convergence of Φ-Divergence Along the Unadjusted Langevin Algorithm and Proximal Sampler. In *36th International Conference on Algorithmic Learning Theory*, 2025.

Yury Polyanskiy and Yihong Wu. *Information theory: From coding to learning*. Cambridge university press, 2025.

Maxim Raginsky. Strong data processing inequalities and Φ-Sobolev inequalities for discrete channels. *IEEE Transactions on Information Theory*, 62(6):3355–3389, 2016.

Santosh Vempala and Andre Wibisono. Rapid convergence of the Unadjusted Langevin Algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009.