Langevin Monte Carlo for Degenerately Convex Potentials

Murat A. Erdogdu*

April 23, 2020

Abstract

We study sampling from a target distribution $\nu_* = e^{-f}$ using the Langevin Monte Carlo (LMC) algorithm. For any potential function f that is degenerately convex at infinity, i.e., asymptotically behaves like $f(x) \sim \|x\|^\alpha$ for $\alpha \in [1, 2)$, with β-Hölder continuous gradient, we prove that $\tilde{\mathcal{O}}\big(d^{\frac{2}{\alpha}(1+\frac{1}{\beta})-1}/\epsilon^{\frac{1}{\beta}}\big)$ steps are sufficient to reach ϵ-neighborhood of a d-dimensional target ν_* in KL-divergence. This convergence rate, in terms of ϵ-dependence, is not directly influenced by the asymptotic growth rate α of the potential function as long as the growth is at least linear, and it only relies on the order of smoothness β. One notable consequence of this result is that for potentials with Lipschitz gradient, i.e. $\beta = 1$, the above rate recovers the best known rate $\tilde{\mathcal{O}}(d/\epsilon)$ established for strongly convex potentials with quadratic growth in ϵ-dependence, but our result implies that the same rate is achievable for a much wider class of degenerately convex potentials that even grow linearly. The growth rate α starts to have an effect on the convergence rate in high dimensions when d is large, where the above rate recovers the best-known dimension dependence in this setup as $\alpha \to 2$, when the growth of the potential becomes asymptotically quadratic.

We establish the convergence rate of LMC by first proving a moment dependent modified log-Sobolev inequality with explicit constants for a class of target distributions with degenerately convex potentials. Then, by establishing a linearly diverging estimate for any order moment of the Markov chain defined by the LMC algorithm, we prove the above convergence rate in KL-divergence, a result that may seem counterintuitive at first. Our framework also allows for finite perturbations and any order of smoothness $\beta \in (0, 1]$, consequently our results are applicable to a wide class of non-convex and non-smooth potentials.

*Department of Computer Science and Department of Statistical Sciences at University of Toronto, and Vector Institute `erdogdu@cs.toronto.edu`