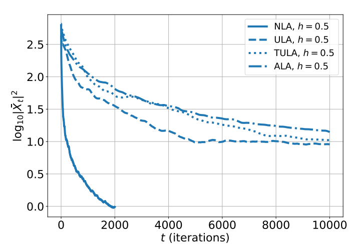
Main remarks

1. **An analysis focused on the continuous time process**
   Several referees have raised the question of the analysis of the discretized algorithm. We have deliberately avoided this question in order to deliver a crisp and clear message regarding interplay between functional inequalities and the chi-squared divergence.
   The study of discrete-time algorithms brings significant additional details and require additional assumptions that would distract the reader from the main message. In fact, our arXiv preprint has already prompted a successful follow-up study[1] of discretized Langevin using the chi-squared divergence, but again, at the cost of ad-hoc assumptions that may or may not be definitive.
2. **Some rates of convergence are artifacts of continuous time analysis**
   It is true that for an isotropic Gaussian target, NLD is simply a sped-up vanilla LD, but this is no longer the case for other target distributions where NLD has a real effect. In fact, for non-isotropic targets, our experiments demonstrate convincingly that NLD is not just a time-reparametrization of ULA, and that NLD is indeed superior.
3. **There should be comparisons with more algorithms**
   Given the diversity of modifications of ULA, we had decided to add only TULA as a comparator: the improvement of NLA over these algorithms is several orders of magnitude better than the variability within the cluster of ULA modifications. Comparison with Underdamped/Accelerated Langevin (ALA) is presented in the attached figure: it exhibits a behaviour similar to (T)ULA in the anisotropic Gaussian case ($d = 20$). HMC belongs to a different family of algorithms.

Specific comments

Reviewer 2

Instability of the "Newton scheme". As in any problem in optimization or sampling, we do not advocate for a one-size-fits-all algorithm and, in many examples, additional structure may be leveraged to improve performance. NLA displays generally better behaviour than competitors in this study and, in that sense, is a good off-the-shelf algorithm. Nevertheless, we describe in the appendix an example where the the flexibility of the mirror perspective can be useful to better exploit the structure of the problem.



TULA vs. ULA. For the same step size ULA indeed typically outperforms TULA in our experiments but the two have essentially the same behavior. We believe that the range of step sizes in our experiments is sufficient to demonstrate qualitatively the relative behaviour of various algorithms.

Reviewer 3

Using the chi-squared divergence is standard. We completely agree with reviewer on this point and acknowledge inspiration from the Markov semigroup perspective in the text. In fact, for the analysis of vanilla Langevin, the use of the chi-squared divergence is almost a tautology as indicated by our short proof based on semigroups; this fact had somehow been elusive in the sampling literature despite intense activity over the past few years.

The lack of dimension dependence is not surprising. For non-strongly-log-concave potentials, whether or not the Poincaré constant depends on the dimension is actually the object of the KLS conjecture (see Sections 3.2 and 4.1). Note that whether this conjecture is true or not, the rate for NLD does not even depend on the Poincaré constant.

Reviewer 4

Assumptions. In the theory of sampling algorithms, we produce some of the weakest conditions amenable to polynomial sampling algorithms.

Comparison with Zha+20 in terms of technical tools. In fact, Zha+20 fails to show convergence of the Mirror Langevin Algorithms even for vanishing step size. It also uses some assumptions that are stronger than ours, others that are incomparable, and finally some assumptions that are inherent to their discrete-time analysis.

[1] M. A. Erdogu and R. Hosseinhazadeh, (2020). A brief note on the convergence of Langevin Monte Carlo in chi-square divergence, arXiv:2007.11612.