While other Neurips authors complain massively on social networks about the quality of the reports on their papers, we consider ourselves privileged having received 4 fair and constructive reports. We would like to wholeheartedly thank the reviewers for their suggestions and thought provoking remarks. We really want to present our work at NeurIPS and hope that our responses below will convince the reviewers that this submission is worth being accepted.

Answers to Reviewer # 1

1. [. . .] authors do not carry over the analysis of the discretization error. We find it impossible to add the results on discretization to this paper without sacrificing the clarity and the precision, and respecting the 9-pages limitation. There are many different discretization schemes: Euler, Ozaki, randomized mid-point, *etc*. Each of these schemes has its strengths and weaknesses. We are preparing a separate paper analyzing and comparing these discretization schemes along with two novel variants. Note that the randomized mid-point discretization alone was the object of a full NeurIPS paper [2] with 18 pages supplementary material. We believe that an in-depth study of continuous-time dynamics is of interest on its own (this opinion seems to be shared by the authors of [3,4]).
2. [. . .] this extension does not seem to be very challenging and the proof is classical.

   We disagree with the reviewer on this assessment. As kindly mentioned by the reviewer in other comments, the paper is written in a mathematically detailed manner and every passage has been meticulously polished. This is perhaps the main reason why the proof, as it is currently presented, seems not challenging and classical. On the other hand, the term "classical" is not appropriate here, since we are not aware of any other paper where such a proof is used.

Answers to Reviewer # 2

3. [. . .] Does diffusion process end up closest to minimizer that is closest to the origin? Yes, this is easy to prove. We will mention this point along with a sketch of proof in the revised version.
4. [. . .] In the setting $m = 0$, do you think a discretization of this diffusion process will be faster, in terms of number of iterations as compared to simple LMC with a fixed quadratic penalty added? According to our preliminary rough computations, using a clever discretization leads to rates that improve on results in [1]. Let us also mention here that an obvious advantage of the time-varying penalty is the adaptivity to the time horizon.

Answers to Reviewer # 3

5. They obtain an "optimal" rate of convergence by choosing a particular alpha that decreases in t; however, choosing alpha=0 gives a faster rate of convergence, suggesting that their upper bound is loose.

   The upper bound depends of course on t, but also on other parameters such as the dimension and the condition number. The best known [4] exponential bound is $W_2^2(\nu_t, \pi) \leq 2\mathcal{C}_P e^{-2t/\mathcal{C}_P} \chi^2(\nu_0, \pi)$. The χ^2-divergence therein might be very large (recall that for m-strongly convex potentials the μ_2 is of order p/m while the χ^2-divergence is $(M/m)^p$ [Lemma 5, arXiv:1412.7392]). The currently available upper bounds on $\mathcal{C}_P$ depend also badly on p. All in all, the dependence of our bound on p and the condition number is much better than that of the foregoing exponential-in-t bound. This makes it advantageous to use our bound in the high-dimensional or badly conditioned settings. We will further elaborate on this in the revised version.
6. The assumption A, although discussed in the paper, remains a bit mysterious to me for the interesting case of $q < 1$. We agree and intend to further discuss this assumption in the revised version.
7. [. . .] do not explicitely compare their results. Even though rates in continuous time do not tell much about discretize schemes rates, it would still have been interesting to compare the rates of convergence of the different time-continuous variants of gradient flow (such as Nesterov accelartion) for the optimization and variants of Langevin Dynamics for the sampling. Thank you for this suggestion. We will revise the paper accordingly if accepted.

Answers to Reviewer # 4

8. [. . .] it looks like the bound on the Wasserstein-2 metric is $O(t^{-1/2})$ for $m = 0$, while it appears that exponentially decreasing bounds exist (Chewi et al. 2020). Please see our response to Q5 above (raised by Rev #3).
9. The authors explain in the appendix that their bound doesn't require unknown constants (e.g. the Poincare constant) but it does require knowing $\mu_2(\pi)$ [. . .]. Estimating the second-order moment (SOM) is arguably qualitatively simpler than estimating the Poincaré constant (PC). But this is not the main advantage of our result. For many concrete weakly convex potentials, one can easily compute (see Prop 2-4 in [1]) upper bounds on the SOM in which the dependence on the dimension is tight and polynomial. This is not the case of the PC. Finally, our result leads to an upper bound which is smaller than the so-far-known exponential-in-t bound for a large range of values of t.
10. [. . .] do not really discuss other work on temperature-related augmentations to the dynamics. I am a bit concerned that similar results may already exist. We would be grateful if the reviewer could provide some references.

References

[1] Dalalyan, Karagulyan and Riou-Durand (2020) arXiv:1906.08530.

[2] Shen and Lee (2019) NeurIPS and arXiv:1909.05503.

[3] Su, Boyd and Candes (2014) NIPS and arXiv:1503.01243.

[4] Chewi, S., Gouic, T. L., Lu, C., Maunu, T., Rigollet, P., and Stromme, A. (2020) arXiv:2005.09669.