We thank the reviewers for their useful feedback. We now address some of the issues/questions raised, starting with the high-level ones.

The contributions of this paper are (1) convergence of Langevin Dynamics in KL divergence under log-Sobolev (our most important result, since this is the first result assuming only L-smoothness and isoperimetry), (2) generalization of KL divergence to Renyi divergence (see below for more on this) and (3) relaxation of log-Sobolev to Poincare, again a fundamental and natural extension, as the latter isoperimetric condition is weaker and more elementary.

Examples of nonconvex functions with LSI/Poincare are quite generic, e.g., 1- Take a standard Gaussian and subtract from it the sum of a set of Gaussians with means centered on the integer points of the standard lattice and tiny variance. The resulting set is not logconcave but is isoperimetric. 2- Take a convex body and add small disjoint cones on any subset of its boundary so it looks like a prickly pear. The set is not convex, but isoperimetric. Such perturbations clearly destroy convexity (and in particular the Hessian having all nonnegative eigenvalues is not true at many points), but they maintain isoperimetry approximately, both Cheeger and log-Sobolev — the measure of the boundary of a subset is lower bounded by an increasing function of the measure of the subset. Such deviations from convexity can be generically expected, possibly due to noise but also because convexity everywhereappers to be too rigid and unrealistic. On the other hand, isoperimetry is a much weaker and more robust assumption.

Previous work applies to a significantly more restricted class of nonconvex functions. Relaxing convexity comes at the cost of strong assumptions either on the support of the "nonconvexity" or on higher derivatives of the density function. Our finding truly requires only isoperimetry and Lipschitz gradient, and the rate of convergence is *polynomial* in all parameters (the log-Sobolev or Cheeger constant, the Lipshitz constant of the gradient, the dimension and the desired error). With the numbering of R3, [2] (which we cite for the novel analysis of optimization via sampling) gives a guarantee for optimization, not sampling. Moreover, it assumes a strong "dissipativity" condition which (roughly) effectively says that outside a bounded region, the function behaves like a strongly convex function. This is in addition to a log-Sobolev inequality. [3] assumes that the Hessian is Lipshitz. (Unfortunately, the proof included in the arXiv version is only for the case of strongly convex outside a ball, and is missing some details; the general case claimed is not substantiated in proof). [4,5], which we will cite (thank you!) provide new analysis techniques for convergence and both apply to limited classes of functions as they require strong assumptions not implied by isoperimetry (e.g., log-Sobolev) or even convexity. [6] crucially assumes strong convexity.

In some applications, the Renyi divergence is the natural measure of convergence and more suitable than KL. In differential privacy, the widely used Exponential Mechanism is not known to be efficient in many cases. To preserve privacy, it is not sufficient to argue that a sampling algorithm is close in TV or KL. One ideally wants closeness in D_∞, but closeness in D_α for some reasonably large α will (roughly) imply (α, ϵ)-Renyi DP, which can be translated to an (ϵ', δ')-DP guarantee. See https://arxiv.org/abs/1702.07476 for a discussion. of Renyi divergence as a privacy notion and its relation to (ϵ, δ)-DP. In the paper, we mention several other motivating applications for Renyi divergence.

Regarding experiments, we agree that these results are closely connected to practice. Indeed variants of LD are used routinely, and typically for nonconvex functions. We see our analysis as providing rigorous backing for this approach rather than a new algorithm that should be empirically tested. Our bounds are the first polynomial bounds in this generality. We agree that it would be very interesting to determine the precise polynomial dependencies on all the parameters. Empirical tightness is potentially a harder question as the bounds are worst case.

We are grateful to the reviewers for the suggestions for improvement in the presentation and will incorporate them, including the following specific points:
– Regarding the Poincare inequality, we show in the appendix that it suffices to recover the convergence in Renyi divergence,
– For all the theorems and proofs of the discretized ULA, we explicitly have the assumption that the gradient of the potential is L-Lipschitz (as we say in the abstract etc.). For the continuous dynamics statements, the specific bound on the Lipschitz constant does not affect convergence; However, to be clear we should (and will) add the Lipschitz assumption explicitly.
– We note that the growth function is only used to bound the initial distance (in Renyi divergence) to the biased limit of the process. It is not used in the analysis of convergence.
– Our results only require starting at an approximate local optimum. We say local optimum for convenience to avoid adding another parameter to the analysis (distance to the closest local optimum).
– A deterministic starting point is a familiar problem in continuous state space Markov chains — we take the distribution after one step from the deterministic starting point as the starting distribution.

Finally, we are grateful for the insightful comment about contraction — LSI implies exponential decay of W_2 distance only between solution and the target, not between any two solutions (i.e., contraction, which is equivalent to strong logconcavity as noted by R3).