Curvature Controlled Mixing on the Interpolation Manifold: Sharp Cutoffs in Late-Phase Training Dynamics

Anonymous authors
Paper under double-blind review

Abstract

After a network interpolates its training data, weight decay drives a slow norm-minimising flow on the interpolation manifold. We study the stochastic version of this late phase and show that its approach to equilibrium is not gradual but abrupt.

We analyse the effective diffusion on the interpolation manifold $M \subset \mathbb{R}^n$, whose invariant measure carries the Fixman $\frac{1}{2}\log\det$ correction produced by integrating out normal fluctuations. Our main result is a lower bound on the Bakry–Émery Ricci curvature of this measure, $\mathrm{Ric}_\phi \succeq \kappa g$ with $\kappa = \frac{2\alpha}{\sigma^2}(1 - o(1))$ for weight decay α and noise scale σ, in which every correction — the intrinsic Ricci curvature via the Gauss equation, the extrinsic shape-operator coupling, and the Hessian of the $\log\det$ term — is controlled by quantities measurable on a trained network: the model Hessian norm, its third derivative, the tangential Hessian rank, and the spectrum of the empirical NTK Gram matrix. We also record what the hypotheses force: strong convexity of the log-density makes M complete, noncompact and contractible, so the theorem describes mixing within a single well-conditioned basin rather than across the permutation-symmetric zero-loss set.

Positive curvature has two consequences. Existing results on cutoff for non-negatively curved diffusions then give a total-variation cutoff, and we locate it: $t_c = \frac{1}{2\alpha}\log(d_0^2/\xi^2)$ with window $O(1/\alpha)$, where d_0 is the initial displacement and $\xi = \kappa^{-1/2}$ the equilibrium correlation length. The clock is set by weight decay alone, independent of noise temperature at leading order. We identify the equilibrium as the minimum-norm interpolator measure, concentrated to $O(\sigma/\sqrt{\alpha})$ in parameter norm, recovering deterministic norm-minimisation results as its zero-temperature mode. The resulting prediction — delay logarithmic in the ratio of initial displacement to equilibrium fluctuation scale — is distinct from existing norm-ratio delay laws. In simulations of constrained diffusions where every assumption holds by construction, the predicted location holds with regression slope 1.003 ($R^2 = 0.99$) across independent variation of dimension, weight decay, temperature and initial displacement, and the total-variation profiles collapse under the predicted rescaling.

Placeholder submitted at the abstract-registration deadline. The full paper will be uploaded before the paper deadline.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.