

# Anticorrelated noise provably ensures stable convergence to flatter minima

Yongjoo Baek<sup>1</sup>, Minseok Kim<sup>1</sup>, Youngkyoung Bae<sup>1</sup>, Euijoon Kwon<sup>2</sup>, Yunsik Choe<sup>1</sup>, Yongjae Oh<sup>1</sup>

<sup>1</sup>Seoul National University, Gwanak-gu, South Korea, <sup>2</sup>Korean Institute for Advanced Study, Dongdaemun-gu, South Korea

While the implicit regularization of Stochastic Gradient Descent (SGD) is often attributed to the spatial covariance of gradient noise, the role of its temporal correlations remains insufficiently understood. In this work, we investigate how temporally structured noise affects optimization dynamics and generalization, and propose controllable Anti-correlated Noise Injection (cANI), a simple framework that introduces tunable temporal anti-correlation into gradient updates. By explicitly controlling the correlation time, cANI allows us to isolate the effect of temporal noise structure beyond conventional white-noise perturbations.

Working in the over-parameterized regime and at a finite learning rate, we derive an effective loss that governs the mean trajectory of the parameters. Our analysis shows that cANI induces a Hessian-dependent implicit regularizer that penalizes sharp directions and actively drives the optimizer toward flatter minima. At the same time, anti-correlated perturbations suppress long-range trajectory diffusion by confining parameter fluctuations. These two effects jointly give rise to a fundamental stability-regularization trade-off governed by the correlation time  $\tau$ : strong anti-correlation improves stability by reducing stochastic fluctuations, while excessively strong confinement weakens the effective drift toward flatter regions. As a result, the bias toward flat minima is maximized not at the strongest anti-correlation, but at an intermediate and optimal timescale.

Our theory further suggests that this trade-off becomes particularly important in over-parameterized models, where many nearly flat directions amplify residual fluctuations and can destabilize optimization. In this sense, cANI provides a mechanism not only for favoring flatter solutions, but also for preventing the instability often associated with over-parameterization by appropriately confining the optimization path. Thus, temporal anti-correlation acts as a tunable control knob for balancing exploration, regularization, and stability.

We validate our predictions through theory and experiments on the widening-valley model, matrix sensing, CIFAR-10 classification, and noisy-label benchmarks including CIFAR-10N and CIFAR-100N. Across these settings, we observe a consistent non-monotonic dependence on  $\tau$ , confirming the predicted trade-off and the existence of an optimal anti-correlation level. Empirically, cANI yields flatter solutions and improved generalization compared with standard SGD and PGD, while showing particular robustness in noisy and highly over-parameterized settings. Our results highlight temporal noise correlation as a previously underexplored but practically important design axis for understanding and improving optimization in modern machine learning.

