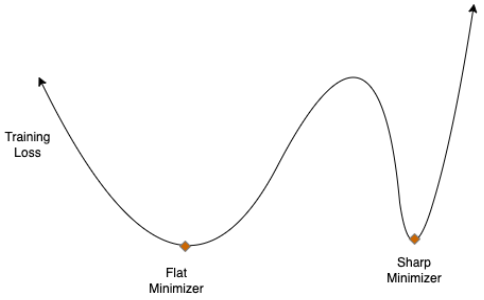
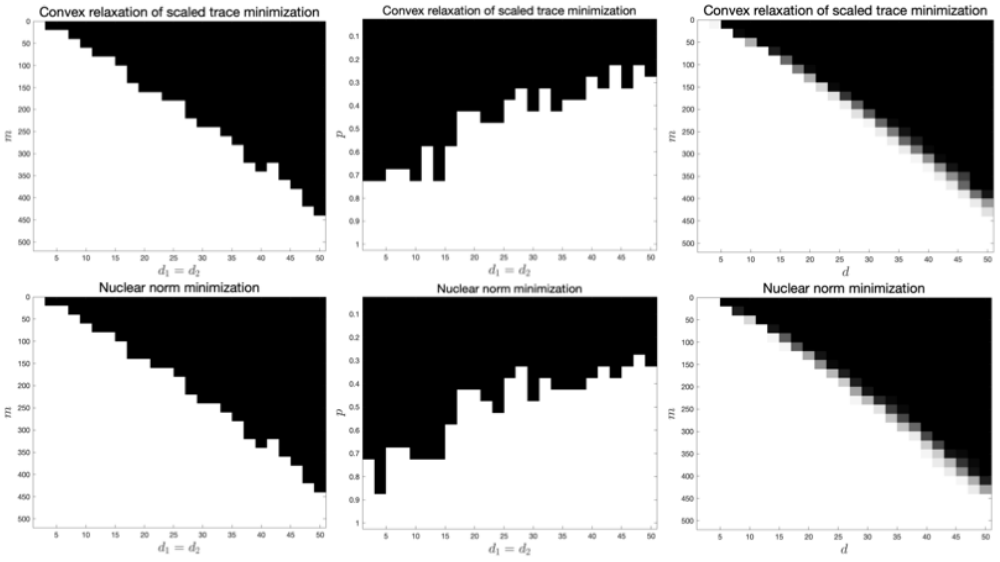
Flat Minima Generalize for Low-rank Matrix Recovery

Many behaviors empirically observed in deep neural networks still lack satisfactory explanation. For example, a fundamental question is: How does an overparameterized neural network avoid overfitting to its training data, and generalize to unseen data? Characterizing the double-descent property for these networks in recent years has shed some light on this question. However, empirical evidence suggests that generalization depends on *which* zero-loss local minimum is attained during training. The shape of the training loss around a local minimum seems to strongly impact the model's performance: “Flat” local minima---around which the loss grows slowly—appear to generalize well. Clarifying this phenomenon can help explain generalization properties, which still largely remain a mystery.



We took steps in this direction by focusing on the simplest class of overparameterized nonlinear models, those arising in *low-rank matrix recovery*. We analyze the following key models: (i) overparametrized matrix sensing and bilinear sensing, (ii) robust Principal Component Analysis, (iii) covariance matrix estimation, (iv) single hidden layer neural networks with quadratic activation functions, and (v) matrix completion. We prove that flat minima (measured by the trace of the Hessian, a notion of average curvature), *exactly recover* the ground truth under standard statistical assumptions, for the first four models. For matrix completion, we obtain weak recovery guarantees, but in simulation always observe exact recovery as well. These results extend to the case where the given information or measurements are noisy.



(a) Matrix Sensing (b) Matrix Completion (c) NNQA

From a broader practical perspective, these results suggest (1) a theoretical basis for favoring methods that bias iterates towards flat solutions, (2) use of Hessian trace as a reasonable regularizer for some learning tasks. This work opens up new research questions; e.g., the landscape properties we examined are algorithm-agnostic, a future direction is to pair these findings with analysis of common deep net training algorithms to understand the interplay between the loss landscape and algorithmic implicit bias.

Credits. Work supported in part by NSF TRIPODS II DMS award 2023166, NSF CCF 2212261.