



Dissertation Defense:

**Adela DePavia**

Computational and Applied Mathematics  
University of Chicago

**“Leveraging Structure for Learning and Optimization”**

**Monday, April 13, 2026, at 1:00 PM**

**Location: Data Science Institute  
5460 S University Ave, Room 105**

ABSTRACT

As practitioners advance machine learning, they target new kinds of structure in training data, propose new models to fit this data, and formulate new learning problems to more accurately model real-world applications. Each of these advances demands new algorithms to solve the resultant novel learning problems. In this thesis, we consider three classes of modern machine learning problems, featuring interactions of novel models, novel data acquisition methods, and novel data distributions. For each of these problems, we analyze the relevant geometric structures that govern the difficulty of the learning or optimization problem. In doing so, we produce new insights that lead to novel algorithms, thereby enabling new applications and advancing machine learning practice.

First, in Chapter 2 we consider the development of new models required to capture more complex structure in data. Many datasets exhibit higher-order structure, such as group memberships or multi-way interactions, that standard graph-based models cannot represent. As a result, recent research has sought to replace graphs with hypergraphs in several key applications. Unfortunately, the deployment of hypergraph-based methods is hindered by a lack of efficient algorithmic primitives for core optimization problems, such as solving hypergraph Laplacian systems. We address this gap by developing algorithms to efficiently solve key optimization problems, including hypergraph Laplacian systems, enabling hypergraph-based methods for semi-supervised learning and clustering that scale to massive datasets.

Rather than modeling new structure in existing, up-front data, Chapters 3 and 4 ask how learners can adapt when data acquisition becomes part of the learning problem. We consider two distinct settings: in Chapter 3, the learner is tasked with finding a hidden goal in an unknown environment. The learner must explore the environment to gather data relating to the goal’s location, and the learner’s aim is to minimize the total travel cost required to find the goal. In Chapter 4, the learner is tasked with recovering an unknown partition of a set by submitting queries to an oracle, modeling setups like crowd-sourced data labeling. The learner’s aim is to minimize the total number of queries submitted. In both settings, the gathered data may contain errors. We develop algorithms that learn efficiently while still remaining robust to adversarial errors. In both chapters, we prove matching lower bounds, confirming that the cost incurred by our algorithms is optimal or nearly optimal.

Chapters 5 and 6 turn to the third force: how contemporary optimization algorithms interact with geometric structure in loss landscapes and data. These chapters focus on adaptive gradient algorithms that use diagonal preconditioner matrices, such as the default implementations of Adam, Adagrad, and their

variants, which are now de facto methods for training massive deep learning models. However, these algorithms are not rotationally equivariant, and rotations in both data and parameter space can dramatically affect their performance. In Chapter 5, we focus on rotations in parameter space, and propose a data-driven method for computing an orthonormal transformation that can improve adaptive algorithms' convergence. We relate the impact of this reparameterization to approximate low-rank structure, an empirically prominent geometry that existing methods often fail to exploit. We then evaluate this reparameterization and the ubiquity of approximate low-rank structure on real-world datasets and architectures. In Chapter 6, we focus on rotations in data space and their impact on generalization. We prove that even arbitrarily small rotations of the training data can drastically impact adaptive algorithms' generalization. We then show how to remedy this sensitivity by proving that the reparameterization method proposed in Chapter 5 makes adaptive gradient methods robust to data rotations.

As datasets, models, problem formulations, and optimizers in machine learning continually evolve, one aim of this thesis is to serve as a stepping stone for future work. In Chapter 7, we highlight avenues for related future research, building on the novel algorithms and analysis developed herein. For each open problem addressed in this thesis, we discuss promising research directions and comment on related ongoing research.