Applied Math 210:
Algebraic Fundamentals of Representing Data

Anna Seigal

Fall 2025

Outline

Algebra gives mathematical abstractions that allow us to process information. Many optimization problems in data and learning are built on algebraic ideas. For example, principal component analysis finds a low-rank approximation of a matrix, a problem central to linear algebra. This course builds from this example to study the algebraic fundamentals of optimization problems to find representations of data. The course is in four parts.

1. The course begins with *factorizations for data analysis*. Factorizations break an object into building blocks that we can understand and interpret. This process underpins classical paradigms for data analysis, including for dimensionality reduction and blind source separation. We will examine matrix factorizations and the models they describe.

2. We then move from the linear algebra of matrices to the multilinear algebra of *tensors*. Tensors encode multi-dimensional data, which arise when studying variables across spatial or temporal contexts and when studying higher-order interactions between variables. We examine the algebra and geometry of low-rank tensors and connect to examples concerning biological data.

3. Next we study *graphical models and causal inference*, and the question of determining how a collection of variables relate to one another. We see the factorizations appearing in graphical models and structural causal models, which contain as special cases the matrix and tensor decompositions encountered up to this point. We study the combinatorial optimization problems of learning causal structures.

4. Finally, we consider the algebraic fundamentals behind the multi-layer architectures of *deep learning models*. We will examine expressivity of networks, factorizations of functions, and what we can gain from deep linear and polynomial networks.

This is a graduate course in applied algebra. The course combines mathematical theory, computational and numerical experiments, and exploration of real world data. The focus is on current research developments and connections to open problems.

Prerequisites. Recommended preparation for the course is familiarity with proofs in linear algebra at the level of two semesters of Math 22A/B or Math 25A or Math 121 as well as programming experience at the level of AM 120.

Assessment. Class attendance (5%), four problem sets (20%), two midterms (40%) a research-oriented final project (35%).

Learning outcomes. Students will have a unified algebraic toolbox to understand existing methods, to design new models, and to prove results on their theoretical underpinnings across the areas of dimensionality reduction, tensor decomposition, causal inference, and deep learning.

Syllabus and References. See `https://tinyurl.com/yc5uf9nk`.