ISyE 8803: Advanced Tools and Special Topics in Mathematical Data Science

Spring 2026

Instructor: Dmitrii M. Ostrovskii – ostrov@gatech.edu – Skiles 263

Schedule & location: M/W 11:00–12:15, **Clough 146**

Canvas: https://gatech.instructure.com/courses/496224

Disclaimer. This course is targeted primarily at graduate and senior undergraduate students whose research interests are steering towards the interfaces between statistical theory, mathematics of data, information theory and/or optimization theory. Being more of a "problem-solver" than a "theory-builder," I do not intend to give a comprehensive overview of these areas. Instead, we shall first get equipped with a set of techniques and ideas that prove to be useful for solving interesting and challenging problems to optimality or near-optimality. We shall then explore several such problems, each concerning one or more of the following topics:

- self-concordant functions and local Gaussian approximations;
- information-theoretic tools, e.g. the Gibbs duality and f-divergences;
- estimation under structural constraints, such as shift- and affine-invariance;
- solving variational problems via the method of extreme points.

Some open problems shall be discussed along the way, along with promising approaches to them. In a nutshell, the goal of this course is to put one at the cutting edge of research in some of the most interesting, deep, and important topics in the mathematical foundations of data science.

Prerequisites. Multivariate calculus at the level of Taylor expansions in $\mathbb{R}^d$ (MATH 2550 or 1712); linear algebra (MATH 1554 or 1564); differential equations (MATH 2552 or 2562; only homogeneous linear equations with constant coefficients); introductory probability and statistics (MATH 3215, but preferably 3225, 3235, 3236, or 3670). Convex optimization basics (ISyE 3133) are desirable, but a crash course will be given.
You may lack some of these credentials, especially the latter ones, if you are highly motivated.

Contact and office hours. The best way to contact me is by email (please add the class name ISyE 8803 to the subject line) or in person in the office hours (Mon 1:30–3pm, Wed 1:30–3pm).

Outline of topics. The course is taught for the first time, so the list of topics might change.

I. **Toolkit: basic-level recap (Probability, Statistics, Optimization)**

- Probability: common distributions; concentration inequalities (Hoeffding, Bernstein).
- Statistics: central limit theorem; least squares; asymptotic theory of maximum likelihood estimators; exponential families; generalized linear models; minimax framework.
- Optimization: convexity; subgradients; first-order optimality conditions; Fenchel dual.

II. **Toolkit: advanced (Information theory, Analysis, Optimization)**

- Bregman and f-divergences, with some applications (hypotheses testing).
- Exponential families and the Gibbs duality. Generalized linear models.
- Extreme points & related theorems (Krein-Milman, Bauer, Dubins) with applications.
- Self-concordant barriers: key properties and geometric characterization.

III. **Applications of self-concordance; Gaussian approximation**

1. Sharp finite-sample analysis of logistic regression with random design.
2. Optimal and tractable online logistic regression.
3. Optimal and tractable online portfolio selection.
4. Extension to quantum tomography.
5. D-minimax experiment design in generalized linear models.

(?) Open problem discussion: near-optimal self-concordant barriers.

IV. **Super-resolution and estimation under shift-invariance**

1. Warm-up: sparse recovery via ℓ_0 and ℓ_1 penalization.
2. "Classical" super-resolution à-la Candès and Recht.
3. Reproducing filters and connection with Lagrange interpolation.
4. Near-optimal and tractable estimation under shift-invariance.

(?) Open problem discussion: super-resolution beyond the Abbe limit.

Literature. The first two thematic blocks consist of classic material; detailed lecture notes will be provided. The other blocks incorporate modern results; some of these have only been taught once (in Spring 2025), and the others have never been taught (these are to be scribbled).

Assignments and evaluation. There will be 2-3 large homework assignments with multi-step problems, designed to complement and broaden the lecture material. Some problems are multi-step and guided (here is a sample homework assignment for a related course). This will be complemented with an informal discussion of a classical paper pertaining to one of the application topics listed above. Grading is generally expected to be rather soft: the main goal is to encourage your research curiosity. On the other hand, depending on the enrollment, I anticipate giving favorable recommendations for a postdoc (or PhD program, if appropriate) to the top students, upon their request. Some advice:

- Start to work on homework assignments as early as possible.
- It is OK to work in a small group. However, always try to solve the problem yourself first. When blocked, use office hours to ask me the question, but be ready to show some work.
- Write solutions on your own; suspected plagiatrism will receive zero credit, and there will be further sanctions in case of a relapse. If you found a solution online – this is possible in some cases – study it carefully, rephrase it, and be ready to explain it if asked.
- It is not penalized, and to some extent even encouraged, to interact with a large language model (GPT, DeepSeek, Gemini, etc.) when preparing a question. Clearly, there is a tradeoff: on the one hand, the more you've managed to do hands-on, the more you have learned, and the problems that you actually solved yourselves will stay with you for years to come; on the other hand, AI-assisted mathematics research is becoming increasingly common, and you are in a good position to enroll in it not too late (if you have not yet).

Grading amd attendance. 66% homework assignments; 24% final presentation; 10% lecture scribbling. In general, I do not keep track of attendance. However, if attendance drops below a certain point, I reserve the right to start controlling it with modest bonuses and penalties. Scribbling will be done on the rolling basis; you should be present in class when it is your turn.