Math 484: Topics Covered on the Final Exam

Misha Lavrov

Fall 2018

The final exam will be cumulative, but with added focus on the material covered between Monday, November 12 and Monday, December 10 (which we haven't had an exam on yet). In this document, I go through the new material and point out the things you should know from it.

The final exam will be closed-note, closed-book, closed-smartphone, and so on, but I will include a formula sheet for the final exam. A rough draft of it is also included in this document. Suggestions for what to add to the formula sheet are welcome, but will not necessarily be followed.

(In general, I don't consider definitions and statements of theorems to be things that should go on this sheet, but if there's a formula that you would want to look up even if you understand the theorem as a whole, that formula would belong there.)

To study for the exam, I suggest the following resources:

- Review the solution to the homework assignments. Moodle (`https://learn.illinois.edu/`) will have solution sets for each assignment on the day it is returned. It's worth reading over them even if you solved the problems correctly to see if you missed shortcuts or alternative approaches.
- When reviewing material from a specific section, look at the examples in the textbook, which are labeled "***(x.y.z)* Example.**", and try to work through or understand them on your own before checking against the book's solution.
- The exercises at the end of each chapter are pretty good too. Some of them develop extra theory that wasn't covered in that chapter, which occasionally means they get deeper into the material than we do in this class. If you have questions about how to solve any of the exercises, I'm happy to answer them.

6 Penalty Methods

6.2 The Penalty Method

You should still remember the conditions that guarantee that a solution obtained by the penalty method is optimal. (As stated in Lecture 30, the first lecture from 6.2 of the textbook, the nontrivial requirements here are that either f is bounded below and P is feasible, or else that we separately check that the point we get is feasible.)

The new topic is coercive functions. You should keep in mind the guarantees we get when the objective function in the penalty method is coercive, even though they're not as useful when solving specific problems.

You should also know the criteria for proving that a function is coercive, which we discussed in the next lecture. (The lecture notes are a better reference here than the textbook is.)

6.3 Applications of the Penalty Function Method to Convex Programs

You should review this section both for the slightly-more-general statement about when strong duality holds, and to see more instances of us working with the KKT dual.

Dealing with Equality Constraints

The material in this lecture did not come from a specific section of the textbook, but it is a rough substitute for spending more time on Chapter 7.

You should be prepared to face equality constraints in the KKT theorem (where, if the Slater condition is going to apply, they will only be linear constraints), the penalty method, and geometric programming.

3 Iterative Methods for Unconstrained Optimization

For all of these methods, don't worry about heavy numerical computation appearing on the final. You should be able to do the calculations in principle, assuming the numbers are not too bad.

3.1 Newton's method

You should be able to use all versions of Newton's method and understand their geometric interpretations (as finding roots of linear approximations, or critical points of quadratic approximations, to the original function). Be aware of cases where Newton's method fails and what that looks like.

You should understand Theorems 3.1.4 and 3.1.5, although Theorem 3.1.5 didn't come up for us until we started discussing the method of steepest descent.

You don't need to worry about the discussion of how to solve systems of linear equations efficiently.

3.2 The method of steepest descent

You should be able to use the method of steepest descent. Theorems 3.2.5 and 3.2.6 are important to know; theorem 3.2.3 is kind of cool.

3.3 Beyond steepest descent

You should know the criteria 1–4 for Wolfe's theorem (Theorem 3.3.1), as well as the theorem itself, and be able to do the following things:

1. Find a descent direction $\mathbf{p}^{(k)}$ of the form $(Hf(\mathbf{x}^{(k)}) + \mu_k I)\nabla f(\mathbf{x}^{(k)})$.
2. Test which values t_k in the iterative step $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t_k \mathbf{p}^{(k)}$ satisfy Criteria 3 and 4 of Wolfe's theorem.

3.4 Broyden's method

You should be able to use Broyden's method (3.4.1) and understand the geometric intuition behind the update rule.

This section is heavy on linear algebra. Most of it steps from the idea that the matrix $M = \frac{1}{\|\mathbf{u}\|^2}\mathbf{v}^\mathsf{T}\mathbf{u}$ is the matrix for the linear transformation that sends $\mathbf{u}$ to $\mathbf{v}$ and sends all vectors orthogonal to $\mathbf{u}$ to $\mathbf{0}$. (This is building on some ideas we also saw in Chapter 4.)

I'm not holding you responsible for the "bonus material" regarding the Sherman–Morrison formula, which you'll find in the lecture notes. But it's part of the motivation for using Broyden's method over another method.

Formula Sheet

Jensen's inequality: $f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n)$ whenever f is a convex function and $\lambda_1, \ldots, \lambda_n$ are nonnegative weights satisfying $\lambda_1 + \cdots + \lambda_n = 1$.

Newton's method, solving a system of equations: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \nabla \mathbf{g}(\mathbf{x}^{(k)})^{-1} \mathbf{g}(\mathbf{x}^{(k)})$.

Newton's method, finding a critical point: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - Hf(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$.

Wolfe's theorem criteria (where $\phi(t) = f(\mathbf{x}^{(k)} + t\mathbf{p}^{(k)})$):

1. $\phi(t_k) < \phi(0)$.
2. $\phi'(0) < 0$.
3. $\phi'(t_k) > \beta \phi'(0)$.
4. $\phi(t_k) < \phi(0) + (\alpha \phi'(0)) t_k$.

Broyden's method: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - D_k^{-1} \mathbf{g}(\mathbf{x}^{(k)})$, $D_{k+1} = D_k + \frac{\mathbf{g}(\mathbf{x}^{(k)})(\mathbf{b}^{(k)})^\mathsf{T}}{\mathbf{b}^{(k)} \cdot \mathbf{b}^{(k)}}$, where $\mathbf{b}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$.

Least-squared error solution to $A\mathbf{x} = \mathbf{b}$: $\mathbf{x}^* = (A^\mathsf{T} A)^{-1} A^\mathsf{T} \mathbf{b}$.

Minimum norm solution to $A\mathbf{x} = \mathbf{b}$: $\mathbf{x}^* = A^\mathsf{T} \mathbf{w}$, where $AA^\mathsf{T} \mathbf{w} = \mathbf{b}$.

Minimum H-norm solution: $\mathbf{x}^* = H^{-1} A^\mathsf{T} \mathbf{w}$, where $AH^{-1} A^\mathsf{T} \mathbf{w} = \mathbf{b}$.

Obtuse angle criterion: $\langle \mathbf{z} - \mathbf{x}^*, \mathbf{x} - \mathbf{x}^* \rangle \leq 0$ (where $\mathbf{z}$ is the point outside C, $\mathbf{x}^*$ is the point closest to it, and $\mathbf{x}$ is some other point in C)

Lagrangian function: $L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum\limits_{i=1}^{m} \lambda_i g_i(\mathbf{x})$.

Courant–Beltrami objective function: $F_k(\mathbf{x}) = f(\mathbf{x}) + k \sum_{i=1}^{m} (\max\{0, g_i(\mathbf{x})\})^2$.