Nonlinear Optimization II (553.762)

Lecture Notes - Spring 2023

Benjamin Grimmer
grimmer@jhu.edu
N419 Wyman Park Hall

[The notes corresponding to each lecture will be posted as we go along.]

Logistics: MW 1:30-2:45 in Hackerman B17 (In-Person!)

TA: Hongyu Cheng (hongyucheng@jhu.edu), Danlin Li (dli91@jhu.edu)

Office Hours: TBD

Prerequisite Skills: Linear Algebra
Real Analysis
Minor Programming background
(any language will suffice)

Not Prerequisite: Nonlinear Optimization I (553.761)

(761 covers unconstrained optimization by a variety of methods (linesearch vs trust region, 1^{st} order vs second order, etc.)

762 covers constrained optimization by a different variety of methods (topics outlined on next page).)

Nonlinear Optimization Topics (and very fuzzy time estimates)

1. Linear Programming (~2.5 weeks)
   (Polyhedral Geometry, Duality certifying Optimality, Simplex Method)
2. Nonlinear Programming Models (~2 weeks)
   (Quadratic/Second-order-cone/semidefinite programming, Geometry, Simple Optimality Conditions, Simple Algorithms)
3. Duality Theory (~3 weeks)
   (KKT conditions, Lagrange + Fenchel duality, Optimality Certificates)

Midterm and then Spring Break roughly fit here. →

4. First-Order Algorithms (~3 weeks)
   (Switching subgradient method, penalized methods, augmented Lagrangian methods, minimax optimization methods, bundle methods)
5. Second-Order Algorithms (~2.5 weeks)
   (Interior Point Methods, Sequential Quadratic Programming, etc.)

Course Evaluations/Grading

We will have 5 homeworks (approximately every two weeks) that will roughly consist of 3 proof-based questions and 1 programming question (do-able in any language).

We will have a take home midterm: March 15^{th} - 17^{th} (3pm to 3pm).
(covering 1-3 above)

We will have a take home final: TBD.
(covering 1-5 above)

A small part of your grade may be participation (in lecture, office hours, or emails).

Optimal Grading Rubric Selection

For each student, I will use the "best feasible" rubric to compute your course score. To formalize this, denote their four component scores as

$$\begin{cases} C_H = \text{Homework Score} & \text{(out of 100)} \\ C_M = \text{Midterm Score} & \text{(out of 100)} \\ C_F = \text{Final Score} & \text{(out of 100)} \\ C_P = \text{Participation Score} & \text{(out of 100)}. \end{cases}$$

A rubric is given by selecting component weights $(H, M, F, P) \in \mathbb{R}^4$

$$\begin{cases} H = \text{Percentage of Weight on Homework} & \text{(out of 100)} \\ M = \text{"} \quad \text{" Midterm} & \text{"} \quad \text{"} \\ F = \text{"} \quad \text{" Final} & \text{"} \quad \text{"} \\ P = \text{"} \quad \text{" Participation} & \text{"} \quad \text{"}. \\ \ \ = 100 - H - M - F & \end{cases}$$

Given a rubric and a students grades, their overall score is

$$\frac{1}{100}\left(C_H H + C_M M + C_F F + C_P P\right).$$

I am willing to grade each of you with any rubric satisfying the following rules

$$(*) \quad \begin{cases} H + M + F \leq 100 & \text{(participation is nonnegative)} \\ H \geq 15 & \\ M \geq 15 & \text{(H and M matter)} \\ F \geq M & \text{(F matters more than M)} \\ M + F \geq 50 & \text{(Exams matter)} \\ M + F \leq 80 & \text{(Exams don't matter too much)} \\ H + M + F \geq 90 & \text{(Participation is bounded)} \end{cases}$$

(Note P shows up nowhere as it is determined by $100 - H - M - F = P$.)

Hence computing the maximum course score amounts to

$$(\text{LP}) \quad \begin{cases} \max \ \frac{1}{100}\left(c_H H + c_M M + c_F F + c_P(100 - H - M - F)\right) \\ \text{subject to } \ (H, M, F) \in \mathbb{R}^3 \text{ satisfy } (*). \end{cases}$$

This is a "<u>linear program</u>":

it is maximizing (or minimizing) a linear function over a set defined by linear inequality constraints.

The set of allowed solutions (called the "<u>feasible region</u>") in this case lives in three-dimensions, so we can visualize it.

- 3D printed copies of this set were handed out in class.
- See "grading.pdf" posted for more formal documentation

<u>Sample Grading for Two Students</u>

Example 1: Suppose Alice skips every lecture, homework and midterm, but aces the final. That is, she has

$$(c_H, c_M, c_F, c_P) = (0, 0, 100, 0).$$

What is her maximum possible score? $\boxed{65\,\%}$ (by (20, 15, 65, 0))

How can we show this is the best?

Every rubric has $M + F \leq 80$ and $M \geq 15$.

Hence
$$\begin{array}{rl} & 1\cdot(M+F \leq 80) \\ + & -1\cdot(M \geq 15) \\ \hline & F \leq 65. \quad \square \end{array}$$

Example 2: Suppose Bob does half the midterm and homework (and still aces the final). Then he has

$$(c_H, c_M, c_F, c_P) = (50, 50, 100, 0).$$

What is his maximum possible score? $\boxed{82.5\%}$
by $(20, 15, 65, 0)$

How can we show this is the best possible?

Every rubric has $H+M+F \le 100$, $M \ge 15$, $M+F \le 80$.

Hence

$$\begin{aligned} &\tfrac{1}{2} \cdot (H+M+F \le 100) \\ -&\tfrac{1}{2} \cdot (M \ge 15) \\ +\ &\tfrac{1}{2} \cdot (M+F \le 80 \\ \hline &\tfrac{1}{2}H + \tfrac{1}{2}M + F \le 82.5 \end{aligned}$$

for every rubric. □

There is a pattern here!! To convincingly say what a student's maximum score is we need to things:

- A proposed rubric (H, M, F) [Called a <u>primal solution</u>]
- A clever set of weights to add up constraints in (*) showing no feasible rubric does better. [Called a <u>dual solution</u>]

Let's formalize this. For our seven constraints in (*), denote weights $\lambda_1, \dots, \lambda_7$. Then our general arguement looks like:

$$\begin{aligned} &\lambda_1(H + M + F \le 100) \\ &\lambda_2(H \ge 15) \\ &\lambda_3(M \ge 15) \\ &\lambda_4(-M + F \ge 0) \\ &\lambda_5(M + F \ge 50) \quad \{\lambda_6(M + F \le 80) \\ +\ &\lambda_7(H + M + F \ge 90) \\ \hline &\dots \end{aligned}$$

Warning! We cannot add $\leq$ with $\geq$ and conclude anything. For each "$\leq$" constraint we need $\lambda_i \geq 0$ and each "$\geq$" constraint needs $\lambda_i \leq 0$.

Summing up our generically weighted constraints, we have

$$(\lambda_1+\lambda_2+\lambda_7)H + (\lambda_1+\lambda_3-\lambda_4+\lambda_5+\lambda_6+\lambda_7)M + (\lambda_1+\lambda_4+\lambda_5+\lambda_6+\lambda_7)F$$
$$\leq 100\lambda_1 + 15\lambda_2 + 15\lambda_3 + 0\cdot\lambda_4 + 50\lambda_5 + 80\lambda_6 + 90\lambda_7.$$

Recall our objective in grading looks like (just rearranging the first line of (LP))

$$\left(\frac{C_H-C_P}{100}\right)H + \left(\frac{C_M-C_P}{100}\right)M + \left(\frac{C_F-C_P}{100}\right)F \quad + C_P.$$

So to give a meaningful upperbound on a students score, we need to pick our weights $\lambda_1, \dots, \lambda_7$ to get these three coefficients matching:

$$(**) \quad \begin{cases} \lambda_1 + \lambda_2 \qquad\qquad\qquad\qquad + \lambda_7 = \dfrac{C_H-C_P}{100} \\ \lambda_1 \qquad + \lambda_3 - \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 = \dfrac{C_M-C_P}{100} \\ \lambda_1 \qquad\qquad\quad + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 = \dfrac{C_F-C_P}{100} \\ \lambda_1 \geq 0,\ \lambda_2 \leq 0,\ \lambda_3 \leq 0,\ \lambda_4 \leq 0,\ \lambda_5 \leq 0,\ \lambda_6 \geq 0,\ \lambda_7 \leq 0. \end{cases}$$

Give any such weightings, we know the student cannot do better than

$$100\lambda_1 + 15\lambda_2 + 15\lambda_3 + 0\cdot\lambda_4 + 50\lambda_5 + 80\lambda_6 + 90\lambda_7 \quad + C_P.$$

But wait! That's a linear function (of the λ's) and (**) is a bunch of linear constraints. We can optimize to find the smallest upper bound!

$$\text{(Dual-LP)} \begin{cases} \min & 100\lambda_1 + 15\lambda_2 + 15\lambda_3 + 0\cdot\lambda_4 + 50\lambda_5 + 80\lambda_6 + 90\lambda_7 \\ \text{subject to} & (\lambda_1, \dots, \lambda_7) \in \mathbb{R}^7 \text{ satisfies } (**). \end{cases}$$

Natural Question Can we always find weights $\lambda_1, \dots, \lambda_7$ that give an upper bound equal to the actual maximum grade?

By construction, we have "weak duality"

$$\text{(LP)'s maximum value} \leq \text{(Dual-LP)'s minimum value}.$$

Fantastically, these two are actually always equal!
This is known as "strong duality" and will be our first major result.