Lecture 1

Instructor: Mateo Diaz (mateodd@jhu.edu)

Office Hours: Th 3:00 - 4:30 pm
Wyman S429

TAs:

- Thabo Samakhona (tsamakh1@jhu.edu)
  OH: - am
- Pedro Izquierdo (pizquie1@jh.edu)
  OH: Tu 9:15 - 10:00 pm

Resources

- Canvas
- Website (mateodd25.github.io/nonlinear2
- Piazza
- Gradescope

Ask your questions here

Delete this 2 for resources from Nonlinear 1.

All submissions

Agenda

- Syllabus
- Motivation
- Overview

Syllabus

Four components:

- Homework — 5 total
- Midterm — Takehome (-)
- Final — Takehome (-)

3 q's about theory
1 code question
(Python please)

Might change

- Participation
  - Engaging in class, OH, Piazza.

Grading System

Let C_H, C_M, C_F, C_P denote your normalize grades $(0 - 1)$.

Let H, M, F be variable weights for each component.

Your grade will be the optimal value of $(C_H - C_p)H + (C_M - C_p)M + (C_F - C_p)F$

'' $+ C_F$

$$
p^* = \left\{
\begin{aligned}
\max \quad & C_H \cdot H + C_M \cdot M + C_F \cdot F \\
& \quad + C_p \cdot \underbrace{(1 - H - M - F)}_{= P} \\
\text{s.t.} \quad & (H, M, F) \in \mathbb{R}^3 \\
& H + M + F \leq 100 \\
& H, M \geq 15 \\
& F \geq M \\
& M + F \leq 80 \\
& M + F \geq 50 \\
& H + M + F \geq 90
\end{aligned}
\right. \qquad (P)
$$

Textbook

We will not follow any particular textbook. See website for suggested references.

Motivation

The goal of this class is to study problems of the form

$$\min_{x \in C} f(x)$$

unlike in nonlinear 1 we will assume $C \subsetneq \mathbb{R}^d$. This adds additional complications but gives rise to beautiful theory.

We consider two types of constrained sets:

- **Inequality constrains**

  $$C = \{ x \in \mathbb{R}^d \mid g_i(x) \leq 0 \}$$

  with $g_i : \mathbb{R}^d \to \mathbb{R}$ differentiable.

- **Structured convex**

  We let C be a "structured" convex set. Structured comes in different flavors, i.e.,

  a) We can project to C.

b) We have a "nice" description of C.

Examples

- **The grading scheme**

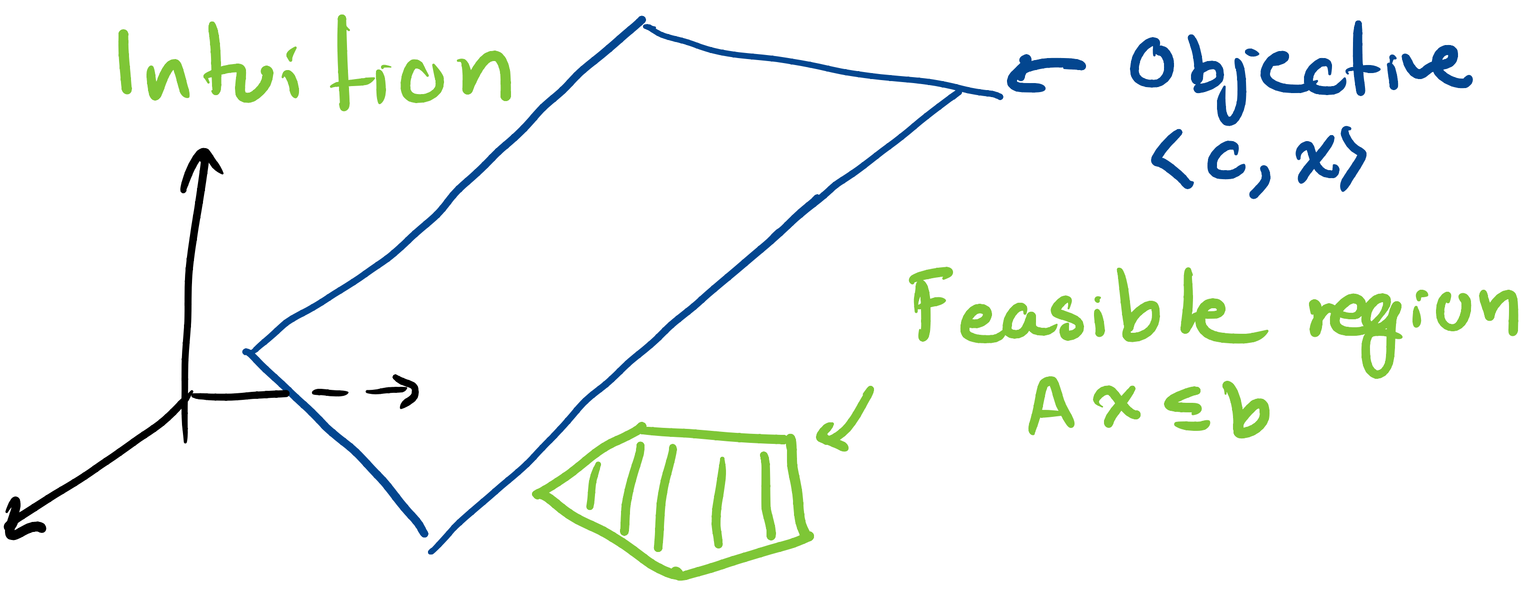
The problem we use to compute grades (P) is part of an important problem class called Linear Programing problems:

$$\min_{x \in \mathbb{R}^d} \langle c, x \rangle \qquad c^T x = \sum c_i x_i$$

$$Ax \leq b$$

$A \in \mathbb{R}^{m \times d}$ $\qquad \leq$ component-wise.

Intuition



History Aside: These problems and algorithms to solve them were one of the first applications of computers in the 1940s.

Question: How can we certify that a solution is optimal?

Let's consider two students

- Student 1:

Suppose Arisu obtained

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

Bad Arisu :(

I claim that the grade Arisu should get is

65 out 100

given by weights

$(H, M, F, P) = (20, 15, 65, 0)$

How can I show this is the best?

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

For every rubric $\langle (H, M, F, P), (C_H, C_M, C_F, C_P) \rangle$

Normalized

▷ Student 2

Suppose Boris got

$(C_H, C_M, C_F, C_P) = (50, 50, 100, 0)/100$

Seems like nobody participates!

I claim Boris would get

82.5 out of 100 by (20, 15, 65, 0)

How to certify this?

$$\begin{array}{rl} & \tfrac{1}{2} \cdot (H + M + F \leq 100) \\ & -\tfrac{1}{2} \cdot (M \geq 15) \\ + & \tfrac{1}{2} \cdot (M + F \leq 80) \\ \hline \end{array}$$

$$\frac{1}{2} H + \frac{1}{2} M + F \leq 82.5$$

← once again for every rubric.

It seems like this goes beyond these two students, we could have $\lambda_1, \dots, \lambda_7 \in \mathbb{R}$

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

We want $(c_H - c_p) H + (c_M - c_p) M + (c_F - c_p) F$

In order to have a valid bound we need two things:

a) The coefficients in front of H, M and F have to match the grades.

b) The λ's have to satisfy
For each "$\leq$" constraint we need $\lambda_i \geq 0$
For each "$\geq$" " " " $\lambda_i \leq 0$.

This leads to the problem

$$
d^* \begin{cases}
\min \; 100\lambda_1 + 15\lambda_2 + 15\lambda_3 + 80\lambda_5 + 80\lambda_6 + 90\lambda_7 \\
\text{s.t.} \quad \lambda_1 + \lambda_2 + \lambda_7 = c_H - c_P \\
\qquad \lambda_1 - \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 = c_M - c_P \\
\qquad \lambda_1 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 = c_F - c_P \\
\lambda_1 \geq 0,\ \lambda_2 \leq 0,\ \lambda_3 \leq 0,\ \lambda_4 \leq 0,\ \lambda_5 \geq 0 \\
\lambda_6 \leq 0,\ \lambda_7 \leq 0.
\end{cases}
$$

This is another LP called

the dual!

We have established that the $p^* \le d^*$. But are they equal?
Indeed, they are; this is known as "strong duality."

Overview

We will cover:

- Fundamentels
  - Background in Convexity
  - Optimality conditions
- Duality
  - Lagrange duality
  - Fenchel duality
- Algorithms
  - Classical methods
  - Splitting methods
  - Other first order methods

- Variational Analysis
  - Nonconvex calculus
  - Inverse problems & metric regularity

Time permitting.