

# Lecture 1

Instructor: Mateo Diaz (mateodd@jhu.edu)

Office Hours: Mondays 4:00 - 6:00 pm  
Wyman S429

## TAs:

- Thabo Samakhona (tsamakh1@jhu.edu)  
OH: - am
- Pedro Izquierdo (pedroizq@jhu.edu)  
OH: - pm

## Resources

- Canvas
- Website (mateodd25.github.io/nonlinear2)
- Piazza
- Grade scope

Ask your questions here

All submissions

↑  
Delete thrs 2  
for resources from  
Nonlinear 1.

# Agenda

- ▷ Syllabus
- ▷ Motivation
- ▷ Overview

## Syllabus

Four components:

- Homework
- Midterm
- Final

- Participation

- Engaging in class, OH, Piazza

5 total  
Takehome ( - )  
Takehome ( - )

Might change

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

## 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  $(C_H - C_p)H + (C_M - C_p)M + (C_F - C_p)F$

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

## 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  $\mathbb{R}^d$  we will assume  $C \subseteq \mathbb{R}^d$ . This adds additional complications but gives rise to beautiful theory.

We consider two types of constrained sets:

- ▷ Inequality constraints

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

with  $g_i: \mathbb{R}^d \rightarrow \mathbb{R}$  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

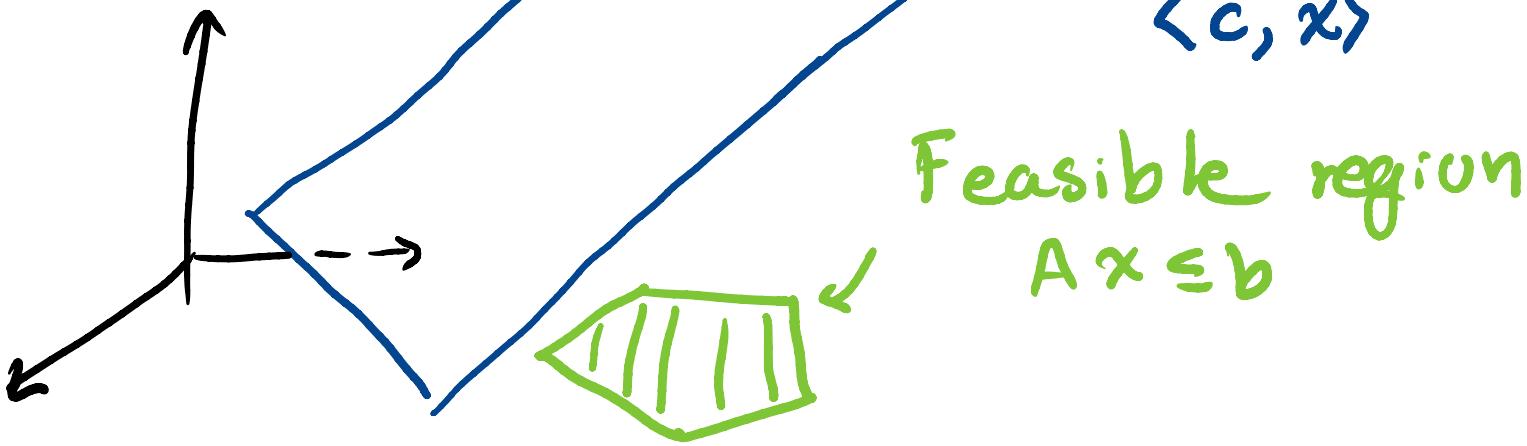
• The grading scheme

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

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

$$Ax \leq b \quad \leftarrow \text{component-wise.}$$

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

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

**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?

$$1 \cdot (M + F \leq 80)$$

$$+ -1 \cdot (M \geq 15)$$

For every rubric  $\Rightarrow F \leq 65$

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

► Student 2      <sup>↑ Normalized</sup>  
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

How to certify this?

$$\frac{1}{2} \cdot (H + M + F \leq 100)$$

$$- \frac{1}{2} \cdot (M \geq 15)$$

$$+ \frac{1}{2} \cdot (M + F \leq 80)$$

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

Once again for every rubric.

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

$$\lambda_1 \cdot (H + M + F) \leq 100$$

$$\lambda_2 \cdot (H) \geq 15$$

$$\lambda_3 \cdot (M) \geq 15$$

$$\lambda_4 \cdot (-M + F) \geq 0$$

$$\lambda_5 \cdot (M + F) \geq 50$$

$$\lambda_6 \cdot (M + F) \leq 80$$

$$+ \lambda_7 \cdot (H + M + F) \geq 90$$

$$(\lambda_1 + \lambda_2 + \lambda_7) H$$

$$+ (\lambda_1 + \lambda_3 - \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7) M = (C_H - C_P) H + (C_M - C_P) M$$

$$+ (\lambda_1 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7) F = (C_F - C_P) F$$

$$\leq 100\lambda_1 + 15\lambda_2 + 15\lambda_3 + 80\lambda_6 + 90\lambda_7.$$

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  $\lambda$ '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

$$\begin{aligned}
 & \max_{\lambda} 100\lambda_1 + 15\lambda_2 + 15\lambda_3 + 80\lambda_6 + 90\lambda_7 \\
 \text{s.t.} \\
 & \lambda_1 + \lambda_2 + \dots + \lambda_7 = C_H - C_P \\
 & \lambda_1 + \lambda_3 - \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 = C_M - C_P \\
 & \lambda_1 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 = C_F - C_P \\
 & \lambda_1 \geq 0, \lambda_2 \leq 0, \lambda_3 \leq 0, \lambda_4 \leq 0, \lambda_6 \geq 0, \lambda_7 \geq 0.
 \end{aligned}$$

This is another LP called

the dual!

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

## Overview

We will cover :

- ▷ Fundamentals
  - ▷ 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.