OPTIMIZATION (SI 416) – LECTURE 1

Harsha Hutridurga

IIT Bombay

LET'S BEGIN

- My name is HARSHA HUTRIDURGA RAMAIAH
- I am the instructor for SI 416 during Spring 2025
- SI 416 is a DEPARTMENT CORE COURSE for
  - M.Sc. Statistics – First year
- It is also a DEPARTMENT ELECTIVE and an INSTITUTE ELECTIVE
- Learning materials include
  - **Lectures and Tutorial sessions**
  - **Reading material** – presentations, problem sheets and handouts
  - **Assessment** – Quizzes, Exams

LECTURES AND TUTORIAL SESSIONS

- We are assigned SLOT 6 for the lectures
  - WEDNESDAYS between 11.05 AM and 12.30 PM
  - FRIDAYS between 11.05 AM and 12.30 PM
- Lectures will be in LH 302
- As this is a six credit course, there is NO separate tutorial slot
- Some lectures will be replaced by tutorials in LH 302
- Office hours to discuss anything course related
  - WEDNESDAYS between 03.00 PM and 04.00 PM
  - Or by appointment (taken via email)
  - Office: 202-D, second floor, Department of Mathematics

ATTENDANCE

- Attendance monitored, for lectures and tutorials, via the SAFE app
- Misrepresentation of attendance on the SAFE app will have unpleasant consequences
- Registration on the ASC portal must have automatically enrolled you for a course called SI416-2025-JAN on the SAFE app

SAFE APP

- Download SAFE app on to your phone and register using LDAP
- During each lecture, you will be asked to take attendance on the SAFE app on your phone
- You have to bring your smartphone (with ample amount of charge) to the classroom

EXAMS AND QUIZZES

- This course will have
  - TWO quizzes
  - SOME surprise quizzes
  - a mid-semester exam (between 22 February and 02 March 2025)
  - an end-semester exam (between 21 April and 01 May 2025)
- Quiz 1 will be on
  - Saturday the 08 February 2025
- Quiz 2 will be on
  - Saturday the 05 April 2025

TUTORIAL SESSIONS

- KSHITIJ, SURAJIT, DEBAPRIYA are the TAs for this course
- A new problem sheet is posted every week
  - on the course webpage
    `https://hutridurga.wordpress.com/teaching/si-416/`
  - You are strongly encouraged to solve the problem sheets
  - Solutions will be discussed during tutorial sessions

WEIGHTAGE AND GRADING

- Here is the points distribution:
  - FIRST QUIZ: 12.5 POINTS
  - MID-SEMESTER EXAM: 25 POINTS
  - SECOND QUIZ: 12.5 POINTS
  - SURPRISE QUIZZES: 10 POINTS
  - END-SEMESTER EXAM: 40 POINTS
- We follow ABSOLUTE GRADING:
  - Grading slab for the letter grades will be disclosed on the last instruction day (17 April 2025)

GO-TO EXAMPLE

- The following example is a typical practical problem addressed in
  - Oil and chemical companies
  - Logistics companies
- Suppose
  - there are four factories B_1, B_2, B_3 and B_4 in Mumbai
  - there are ten retail outlets $S_1, \ldots, S_{10}$ that sell a certain product
  - Factory B_i can produce a_i tons of that product daily
  - Outlet S_j has a daily demand of b_j tons of that product
  - cost of shipping a ton of that product from B_i to S_j is c_{ij}
- How much to ship from each factory to each retail outlet so as to
  - satisfy all the requirements
  - minimize the cost
- Let x_{ij} denote the desired number of tons of the product to be shipped from factory B_i to the retail outlet S_j

GO-TO EXAMPLE (CONTD.)

- Our goal then is to find x_{ij} such that

$$\min \sum_{i,j} c_{ij} x_{ij}$$

subject to

$$\begin{cases} x_{ij} \geq 0 & \text{for } i = 1, \ldots, 4, \quad j = 1, \ldots, 10, \\ \displaystyle\sum_{j=1}^{10} x_{ij} \leq a_i & \text{for } i = 1, \ldots, 4, \\ \displaystyle\sum_{i=1}^{4} x_{ij} \geq b_j & \text{for } j = 1, \ldots, 10. \end{cases}$$

- This is a typical problem in LINEAR PROGRAMMING

- Given a set $\Omega \subseteq \mathbb{R}^n$ and a function $f : \Omega \to \mathbb{R}$, the goal is to

$$\text{find} \quad x_* \in \Omega \quad \text{such that} \quad f(x_*) \leq f(x) \quad \text{for all} \quad x \in \Omega$$

- f is referred to as the OBJECTIVE FUNCTION
- $\Omega \subseteq \mathbb{R}^n$ is usually defined by constraints, i.e. we are given

$$c_j : \mathbb{R}^n \to \mathbb{R} \qquad j = 1, \ldots, m_I + m_E$$

  and the problem is

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) & \\ c_j(x) \leq 0 & \text{for } j \in I \\ c_j(x) = 0 & \text{for } j \in E \end{cases}$$

  Here m_I is the cardinality of I and m_E is the cardinality of E
- We have inequality and equality constraints

BROAD CLASSIFICATIONS

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x) & \\ c_j(x) \le 0 & \text{for } j \in I \\ c_j(x) = 0 & \text{for } j \in E \end{cases}$$

♣ If $I = E = \emptyset$ then UNCONSTRAINED OPTIMIZATION

- Linear problems: f is linear, i.e.

$$f(x) := \langle \mathbf{b}, x\rangle + a \qquad \text{with } \mathbf{b} \in \mathbb{R}^n \text{ and } a \in \mathbb{R}.$$

  Note that the above f has a minimum if and only if $\mathbf{b} = \mathbf{0}$

- Quadratic problems: f is quadratic, i.e.

$$f(x) := \frac{1}{2}\langle Ax, x\rangle + \langle \mathbf{b}, x\rangle + a \quad \text{with } A \in \mathcal{M}_n(\mathbb{R}),\ \mathbf{b} \in \mathbb{R}^n \text{ and } a \in \mathbb{R}.$$

- Nonlinear problems: f is neither linear nor quadratic

BROAD CLASSIFICATIONS (CONTD.)

$$\begin{cases} \min_{x \in \mathbb{R}^n} f(x) & \\ c_j(x) \leq 0 & \text{for } j \in I \\ c_j(x) = 0 & \text{for } j \in E \end{cases}$$

♣ If I and E is (are) nonempty then CONSTRAINED OPTIMIZATION

- Linearly constrained problems (All c_j are linear)
  - ⋆ Problems with equality constraints ($I = \emptyset$)
    Linear-Quadratic problems: f quadratic
    Nonlinear problems: f is neither linear nor quadratic
  - ⋆ Problems with inequality constraints
    Linear programming: f linear
    Linear-Quadratic problems: f quadratic
    Linearly constrained nonlinear problems
- Nonlinear programming
  - ⋆ with equality constraints only
  - ⋆ general nonlinear programming

REFERENCES

[NW] J.NOCEDAL, S.J.WRIGHT

Numerical Optimization (2006).

[CZ] E.CHONG, S.ZAK

An introduction to optimization (2013).

[BR] V.S.BORKAR, K.S.MALLIKARJUNA RAO

Elementary convexity with optimization (2022).

[F] R.FLETCHER

Practical methods of optimization (2001).

[G] O.GÜLER

Foundations of optimization (2010).

Global versus Local

- Take a function $f : \mathbb{R}^n \to \mathbb{R}$
  - A point x_* is a GLOBAL MINIMIZER if

$$f(x_*) \leq f(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

  - A point x_* is a LOCAL MINIMIZER if there is an open set $B \ni x_*$ such that

$$f(x_*) \leq f(x) \qquad \text{for all } x \in B.$$

  - A point x_* is a STRICT LOCAL MINIMIZER if there is an open set $B \ni x_*$ such that

$$f(x_*) < f(x) \qquad \text{for all } x \in B \text{ with } x \neq x_*.$$

Theorem

Let m be a nonnegative integer and let $a \in \mathbb{R}$ be nonnegative.
Let $f : \mathbb{R} \to \mathbb{R}$ be $(m+1)-$times continuously differentiable on $\mathbb{R}$.
Then,

$$f(x+a) = f(x) + f'(x)a + \cdots + \frac{f^{(m)}(x)}{m!}a^m + \frac{f^{(m+1)}(c)}{(m+1)!}a^{m+1}$$

for some point $c \in (x, x+a)$.

- In the above theorem, if a is negative then the point $c \in (x+a, x)$
- The point c depends on x, a and m
- The proof is a repeated application of the Rolle's theorem
- Taylor's theorem also holds true in higher dimension

♣ For $f : \mathbb{R}^n \to \mathbb{R}$, its gradient vector

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

♣ A point $\tilde{x}$ is called a STATIONARY POINT of f if

$$\nabla f(\tilde{x}) = 0$$

♣ For $f : \mathbb{R}^n \to \mathbb{R}$, its Hessian is a $n \times n$ matrix with entries

$$\left(\nabla^2 f\right)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ *be twice continuously differentiable on* $\mathbb{R}^n$ *and let* $p \in \mathbb{R}^n$*. Then,*

$$f(x+p) = f(x) + \langle \nabla f(x + \ell p), p \rangle$$

for some $\ell \in (0,1)$ *and*

$$f(x+p) = f(x) + \langle \nabla f(x), p \rangle + \frac{1}{2} \left\langle \nabla^2 f(x+tp)p, p \right\rangle$$

for some $t \in (0,1)$*.*

- Above theorem is implied by the one dimensional Taylor's theorem
- For first assertion, f being once continuously differentiable suffices
- Idea is to work with $g : [0,1] \to \mathbb{R}$ defined as

$$g(s) := f(x+sp) \qquad \text{for } s \in [0,1].$$

Theorem

If x_* is a local minimizer of f and if f is continuously differentiable in an open neighbourhood of x_*, then

$$\nabla f(x_*) = 0.$$

- This isn't a sufficient condition: think of $f(x) = x^3$ at $x = 0$.
- Suppose not true, i.e. $\nabla f(x_*) \neq 0$
- Take $p := -\nabla f(x_*)$. Observe that

$$\langle \nabla f(x_*), p \rangle = - \|\nabla f(x_*)\|^2 < 0$$

- As ∇f is continuous at x_*, there exists a $T > 0$ such that

$$\langle \nabla f(x_* + sp), p \rangle < 0 \qquad \text{for all } s \in [0, T].$$

FIRST ORDER NECESSARY CONDITION (CONTD.)

- For any $t_* \in (0, T]$, Taylor's theorem says

$$f(x_* + t_* p) = f(x_*) + t_* \langle \nabla f(x_* + s_* p), p \rangle \qquad \text{for some } s_* \in (0, t_*)$$

- Recall that we also have

$$\langle \nabla f(x_* + sp), p \rangle < 0 \qquad \text{for all } s \in [0, T].$$

- Hence we deduce that

$$f(x_* + t_* p) < f(x_*) \qquad \text{for all } t_* \in (0, T].$$

- This contradicts the fact that x_* is a local minimizer of f

END OF LECTURE 1

THANK YOU FOR YOUR ATTENTION