Nonlinear Optimization I

Ben Grimmer

Office Hours : Mon 4-5pm
Wed 9-10am

TAs: Office Hours TBD

Roadmap

1. Syllabus
2. Example Optimization
3. Overview of Course Topics
4. Calculus/Geometry Review

1. Syllabus / Grading

This class will have four components:

$\boxed{\text{Homeworks}}$ ~5 total (~2 weeks)

Mostly proof-based

Some programming

HW1 posted (due 9/14)

$\boxed{\text{Midterm}}$ Takehome, Oct 15-19

$\boxed{\text{Final}}$ Takehome, TBD

$\boxed{\text{Participation}}$ Optionally, engaging lecture, OH, questions.

A Rubric is given by weighting these four components:

H = Student's Homework weight
M = " " Midterm weight
F = " " Final " "
P = " " Participation " "

Scanned with CamScanner

Given your scores in each component

C_H = Homework Score

C_M = Midterm Score

C_F = Final Score

C_P = Participation Score,

I will maximize your grade over all "reasonable*" rubrics:

2. Example Optimization

$$
\begin{cases}
\max \; (H \cdot C_H + M C_M + F C_F + P \cdot C_P)/100 \\
\text{s.t.} \quad P = 100 - H - M - F \\
\qquad H + M + F \le 100 \\
\qquad H, M \ge 15 \\
\qquad F \ge M \\
\qquad 50 \le M + F \le 80 \\
\qquad H + M + F \ge 90 \\
\qquad (H, M, F) \in \mathbb{R}^3
\end{cases}
$$

Free 3D printed "feasible regions" in my office.

3. Overview Course Topics

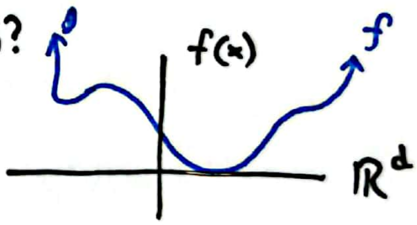
We interested in problems of form

$$\begin{cases} \min\limits_{x \in \mathbb{R}^d} f(x) \, . & f : \mathbb{R}^d \to \mathbb{R} \\ \text{s.t. } x \in S \, . & S \subseteq \mathbb{R}^d \end{cases}$$

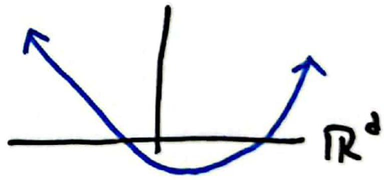
Primarily, $S = \mathbb{R}^d$ (unconstrained optimization)

Half of the class

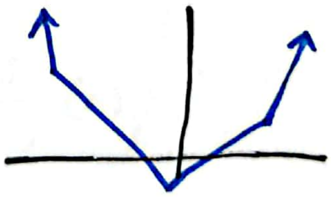
- Optimality Conditions (What makes a good solution?)
- First-Order Method (given $x \mapsto \nabla f(x)$).
  - What if f is smooth?



  - What if f is convex?



  - What if f is nonsmooth?



  - What if f is stochastic?
    $f(x) = \mathbb{E}_{\xi}[f(x, \xi)]$

Scanned with CamScanner

2nd Half of class

- Second-Order Method (given $x \mapsto (\nabla f(x), \nabla^2 f(x))$)
  - Newton's Method
  - Quasi-Newton Methods
  - Trust-Region Methods

  More expensive per step, but much faster

- Conjugate Gradient Methods, Linear Programming, Zeroth-Order Methods.

  Time Permitting

4. Calculus Review

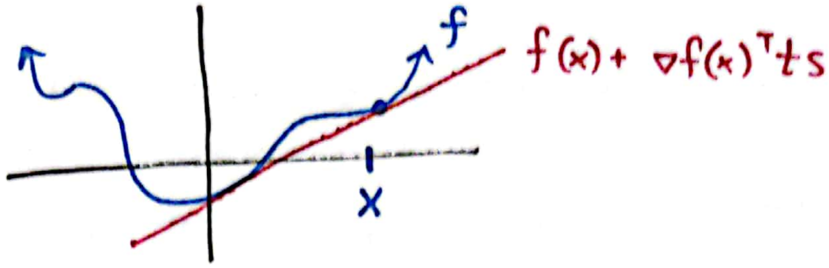
Consider function $f: \mathbb{R}^d \to \mathbb{R}$.

The gradient of f at $x \in \mathbb{R}^d$ is given by

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{pmatrix}.$$

Equivalently, the gradient is the unique vector s.t.

$$\forall s \in \mathbb{R}^d \quad \lim_{t \to 0} \frac{f(x+ts) - (f(x) + \nabla f(x)^T ts)}{t} = 0$$

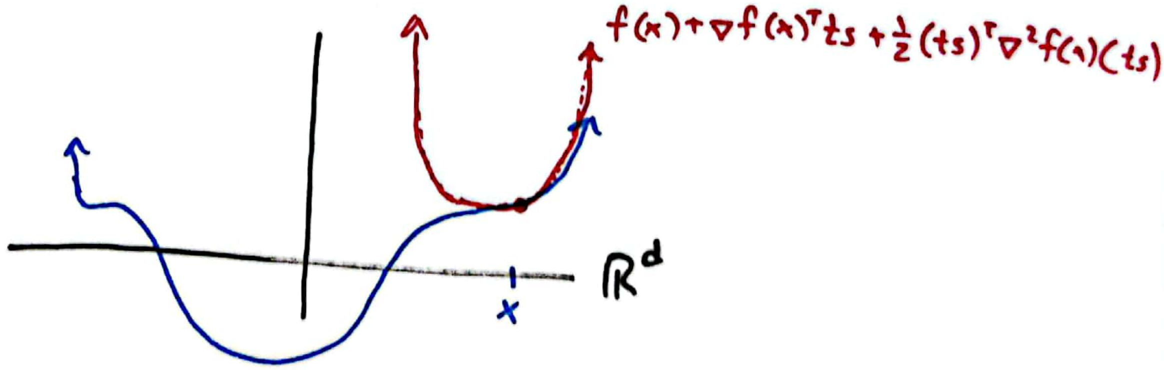


The Hessian of f at $x \in \mathbb{R}^d$ is given by

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \cdots & \frac{\partial^2}{\partial x_d x_1} f(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_1 x_d} f(x) & \cdots & \frac{\partial^2}{\partial x_d \partial x_d} f(x) \end{pmatrix}$$

Equivalently, the Hessian is the unique linear operator $\forall s \in \mathbb{R}^d$

$$\lim_{t \to 0} \frac{f(x+ts) - \left(f(x) + \nabla f(x)^T ts + \frac{1}{2}(ts)^T \nabla^2 f(x)\,(ts)\right)}{t^2} = 0.$$



Theorem (1D Restrictions of Multivariate Func)

Let $f : \mathbb{R}^d \to \mathbb{R}$ and consider $x, s \in \mathbb{R}^d$

Define $\phi(t) = f(x+ts)$. $\phi : \mathbb{R} \to \mathbb{R}$

If f is differentiable

$$\phi'(t) = \nabla f(x+ts)^T s$$

If f is twice differentiable

$$\phi''(t) = s^T \nabla^2 f(x+ts)\, s .$$