

# Optimization Methods.

## Seminar 1. Introduction.

Alexandr Katrutsa

Moscow Institute of Physics and Technology  
Department of Control and Applied Mathematics

September 4, 2018

# Survey for students

- Name
- Base chair
- Final project in Computer Science course
- Familiarity with  $\text{\LaTeX}$
- Expectations of this course
- Favorite course for 2 study-years and why
- Future directions: academia, industry, etc.

# Short syllabus — general

Fall term (theory):

- Convex analysis
- Duality theory
- Optimality conditions

Spring term (practice):

- Methods for unconstrained optimization problems first and second orders
- Methods for constrained optimization problems
- Linear programming: simplex method and others
- Optimal methods
- ...

# Fall term syllabus — detailed

- Lecture and seminar — once a week
- Two problem sets
- Final test in the end of term  
(and midterm test in the middle of the term)
- Oral exam in the end of term. Grading policy — average the following grades:
  - grade for the term
  - grade for the exam
- Minitests at the beginning of every seminar
- Homework after every seminar —  $\text{\LaTeX}$

# Expected outcome of this course

- Formalization of the problem: how to select an element from a set?
- Justification of the decision correctness
- Wide range of applications:
  - machine learning: classification, clustering, regression
  - molecular modeling
  - risk analysis
  - portfolio optimization
  - optimal control
  - signal processing
  - parameters estimation in statistic
  - many-many others<sup>1</sup>

---

<sup>1</sup><http://www.cvxpy.org/en/latest/examples/index.html>

# Prerequisites

- Linear algebra
- Calculus
- Programming: Python (NumPy, SciPy, CVXPY) or MATLAB
- Selected topics of numerical analyses

Main steps of usage optimization methods in solving real-world problems:

1. Define objective function
2. Define feasible set
3. Optimization problem statement and its analysis
4. Selection of the best algorithm for stated problem
5. Algorithm implementation and verification its correctness

# Problem statement

$$\begin{aligned} & \min_{\mathbf{x} \in X} f_0(\mathbf{x}) \\ & \text{s.t. } f_i(\mathbf{x}) = 0, \quad i = 1, \dots, p \\ & \quad f_j(\mathbf{x}) \leq 0, \quad j = p + 1, \dots, m, \end{aligned}$$

- $\mathbf{x} \in \mathbb{R}^n$  — target variable
- $f_0(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$  — objective function
- $f_j(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$  — constraints functions

Example: selection items to invest money and defining how much money invest in every item:

- $\mathbf{x}$  — amount of money to invest in every item
- $f_0$  — total risk or revenue deviation
- $f_j$  — budget constraints, min/max invest in item, minimal accepted revenue...



# How to solve?

General case:

- NP-hard
- randomized algorithms: time vs stability

BUT some families of problems can be solved very fast!

- Linear programming
- Least squares problem
- Lowrank approximation
- Convex optimization

# History

- 1940s — linear programming
- 1950s — quadratic programming
- 1960s — geometric programming
- 1990s — polynomial interior point methods for nonlinear convex optimization problems

# Modern challenges

- Huge-scale optimization problems ( $\sim 10^8 - 10^{12}$ )
- Distributed optimization
- Fast first and higher order methods
- Stochastic optimization methods: scalability vs. accuracy
- Non-convex structured optimization problems
- Applications of convex optimization

# Linear programming

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^T \mathbf{x} \\ \text{s.t. } \mathbf{a}_i^T \mathbf{x} \leq c_i, \quad i = 1, \dots, m \end{aligned}$$

- no analytical solution
- effective algorithms exist
- developed technology
- simplex method is included in the TOP-10 algorithms of XX century<sup>2</sup>

---

<sup>2</sup><https://archive.siam.org/pdf/news/637.pdf>

# Least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|_2^2,$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^m$

- analytical solution:  $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$
- effective algorithms exist
- developed technology
- statistical interpretation

# Lowrank approximation

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \quad & \|\mathbf{A} - \mathbf{X}\|_F \\ \text{s.t.} \quad & \text{rank}(\mathbf{X}) \leq k \end{aligned}$$

## Theorem (Eckart–Young, 1993)

Let  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$  be a singular value decomposition of matrix  $\mathbf{A}$ , where  $\mathbf{U} = [\mathbf{U}_k, \mathbf{U}_{r-k}] \in \mathbb{R}^{m \times r}$ ,  $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_k, \dots, \sigma_r)$ ,  $\mathbf{V} = [\mathbf{V}_k, \mathbf{V}_{r-k}] \in \mathbb{R}^{n \times r}$  and  $r = \text{rank}(\mathbf{A})$ . Then the solution of the problem (14) can be written as:

$$\mathbf{X} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^\top,$$

where  $\hat{\mathbf{U}} \in \mathbb{R}^{m \times k}$ ,  $\hat{\mathbf{\Sigma}} = \text{diag}(\sigma_1, \dots, \sigma_k)$ ,  $\hat{\mathbf{V}} \in \mathbb{R}^{n \times k}$ .

Algorithms to compute SVD are stable and fast simultaneously

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} f_0(\mathbf{x}) \\ \text{s.t. } f_i(\mathbf{x}) \leq b_i, \quad i = 1, \dots, m \end{aligned} \tag{1}$$

- $f_0, f_i$  — convex functions:

$$f(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2),$$

where  $\alpha, \beta \geq 0$  and  $\alpha + \beta = 1$ .

- no analytical solution
- effective algorithms exist
- recognition convex optimization problem is often difficult
- techniques to reformulate a problem to the form (1) exist

# What problem is easier?

Search independent alphabet  
of the maximum cardinality

$$\begin{aligned} \max_{\mathbf{x} \in \mathbb{R}^n} \quad & \sum_{i=1}^n x_i \\ \text{s.t.} \quad & x_i^2 - x_i = 0 \quad i = 1, \dots, n \\ & x_i x_j = 0 \quad \forall (i, j) \in \Gamma, \end{aligned}$$

where  $\Gamma$  is a set of pairs

Truss design

$$\begin{aligned} \min \quad & -2 \sum_{i=1}^k \sum_{j=1}^m c_{ij} x_{ij} + x_{00} \\ \text{s.t.} \quad & \sum_{i=1}^k x_i = 1 \\ & \lambda_{\min}(\mathbf{A}) \geq 0, \end{aligned}$$

where  $\mathbf{A} =$

$$\begin{bmatrix} x_1 & \dots & \dots & \sum_{j=1}^m b_{pj} x_{1j} \\ \vdots & \ddots & & \vdots \\ & & x_k & \sum_{j=1}^m b_{pj} x_{kj} \\ \sum_{j=1}^m b_{pj} x_{1j} & \dots & \sum_{j=1}^m b_{pj} x_{kj} & x_{00} \end{bmatrix}$$



# Why convexity is so important?

R. Tyrrell Rockafellar (1935 —)

The great watershed in optimization is not between linearity and non-linearity, but convexity and non-convexity.

- Local optimum is global
- Necessary optimality condition is sufficient

Questions:

- Is any convex optimization problem solved efficiently?
- Could non-convex optimization problems be solved efficiently?

- Administrative issues
- Objective of the Optimization Methods course
- General optimization problem statement
- Classical optimization problems