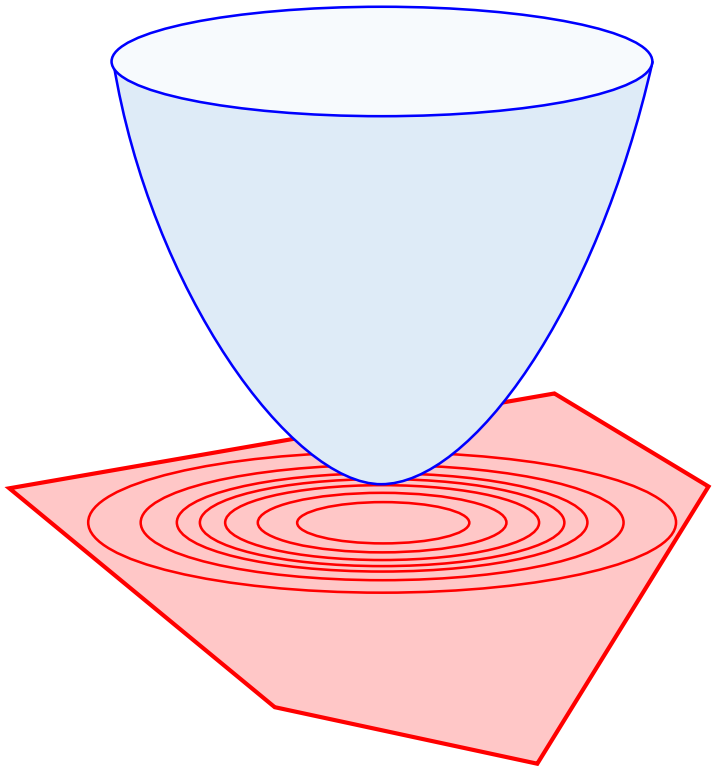
CS774: Optimization Techniques



Purushottam Kar

IIT KANPUR

Course Details

- **Name**: CS774(A)/CS698E: Optimization Techniques
- **Nickname**: OPT
- **Instructor**: Purushottam "Puru" Kar (purushot@cse.iitk.ac.in)
- **Teaching Assistant(s)**: TBD
- **Lectures**: Mondays, Thursdays 1200-1330 hrs, KD 101
- **Office hours**:

  Puru: Fridays 1600-1700 hrs (by appointment)

  TA: TBD
- **Website**: http://tinyurl.com/opt16-17a
- **Internal**: http://tinyurl.com/opt16-17ai

Auditors

- Please send a mail to Puru confirming your decision
- This is important for your email to be added to the announcement list
- Feel free to participate in all aspects of the course
  - Attend lectures
  - Assist creditors in scribing notes
  - Submit assignments – will be graded*
  - Appear for examinations – will be graded*
  - Participate in project groups

Grading Scheme

- 25%: Assignments
  - Paper-pen or TeXed
  - Programming-based
- 25%: End-semester examination
- 15%: Mid-semester examination
- 20%: Term Project
- 15%: Scribing lecture notes for **one** lecture OR **one** paper review
  - Typeset in LaTeX

Obtaining (near) publishable results in the project may get you a straight A!

Scribing Duties and Paper Review

- Please indicate preference on online poll
- Allotment will be done randomly keeping preferences in mind
- Can swap duties (scribe ←→ scribe) (scribe ←→review)
  - Need to inform Puru beforehand!
- Use the prescribed style file
  - Available on internal website
  - Do not edit style file – ask Puru in case of doubts
  - Sample scribes also present

To Dos/Donts for Scribes and Paper Reviews

- Take pride in your scribed notes – they will be put up publicly
- Benefit you and your classmates in preparation and reference
- Well explained, details worked out
- Well referenced, proper citations, acknowledgements
- Properly formatted definitions, theorems, lemmata
- Illustrations when necessary

- Paper review should also present an overview of the state of the art
- Possible starting point for project work for you or your friends
- Be critical, but also fair to the authors of the paper you review

Project

- Objective of the course
  - Lectures act as enablers – introducing basics, tools
  - Project investigation is where thorough instruction takes place
- Some project ideas to be put up on internal website
  - Expect list of suggested ideas by weekend/early next week
  - Discuss with friends, Puru for more ideas
- Project topic needs to be related to the course
- Project has to be substantial
  - Reading projects should survey an entire (sub)area of optimization
  - Implementation projects should try to contribute to a tool or a package
  - Research projects should try to push the boundaries of an area

Project

- Form groups of 2 (1 or 3 allowed as special case)
  - Auditors can join project groups but wont be counted
  - Make groups known to Puru by sending a mail copying all group members
- Project proposals (written) due before class on 18th August, 2016
  - Please do not wait till 18th August to discuss topics with friends, Puru
- Mid-term presentations: 29th September, 2016
- Final presentations + report: 10th November, 2016
- Breakup
  - Mid-term presentation: 5%
  - Final presentation: 10%
  - Final report: 5%

Reference Material

- No textbook for the course
- Reference list up on website
- Locally cached copies for some to be put up on internal website

- [**BRT**] D. Bertsekas, Nonlinear programming, 1999.
- [**BUB**] S. Bubeck, Convex Optimization: Algorithms and Complexity, 2015.
- [**BVB**] S. Boyd and L. Vandenberghe, Convex Optimization, 2003.
- [**HTW**] T. Hastie, R. Tibshirani and M. J. Wainwright, Statistical Learning with Sparsity: the Lasso and Generalizations, 2015.
- [**HZN**] E. Hazan. Introduction to Online Convex Optimization, 2015.
- [**NST**] Y. Nesterov, Introductory lectures on convex optimization, 2003.
- [**SNW**] S. Sra, S. Nowozin, and S. Wright (eds). Optimization for Machine Learning, 2011.

Use of Unfair Means

- The following are prohibited – severe penalties
  - Copying answers in theory assignments
  - Copying code in programming assignments
  - Passing off known results as one's own
  - Manipulating experimental results

- The following are prohibited – credit deductions
  - Using material in scribes/reviews (figures, text) without acknowledging
  - Using help from auditors in projects/scribes without acknowledging

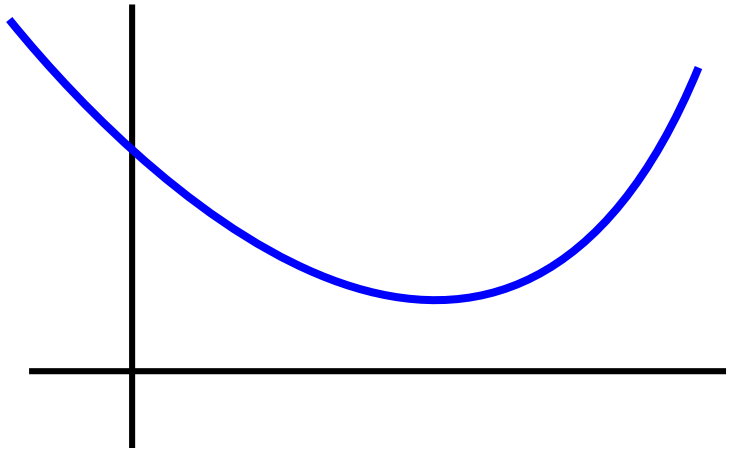
What is Optimization?

A Cartoon View of Optimization

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

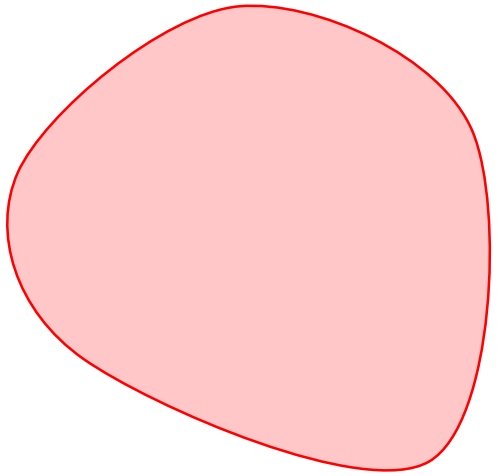
$f : \mathbb{R}^d \rightarrow \mathbb{R}$

Some function

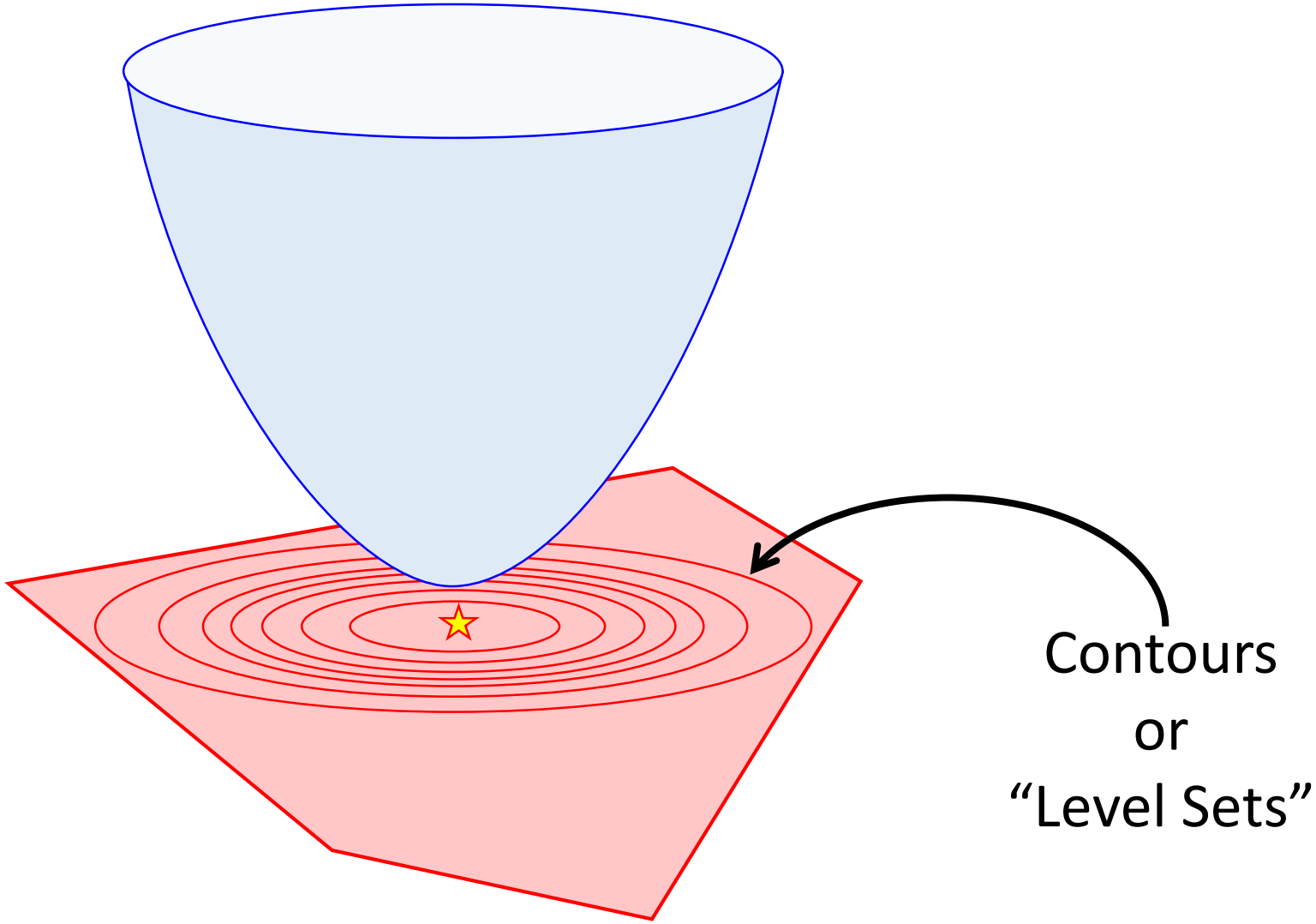


$\mathcal{C} \subseteq \mathbb{R}^d$

Some set



A Cartoon View of Optimization

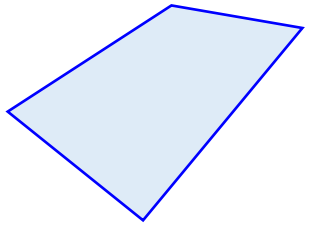


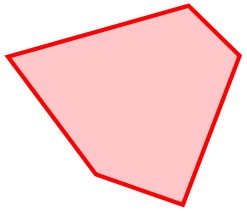
Examples

Linear Programming

$$\min_{\mathbf{x}\in\mathbb{R}^d} \mathbf{a}^\top \mathbf{x}$$

$$s.t.\ \mathbf{b}_i^\top \mathbf{x} \leq c_i$$

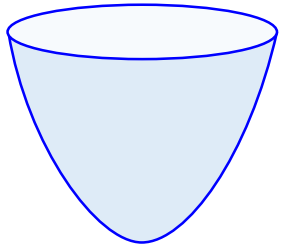
f 

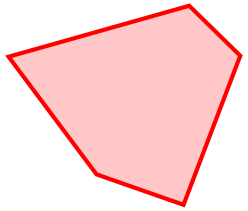
$\mathcal{C}$ 

Quadratic Programming

$$\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{a}^\top \mathbf{x}$$

$$s.t.\ \mathbf{b}_i^\top \mathbf{x} \leq c_i$$

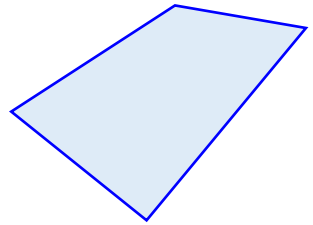
f 

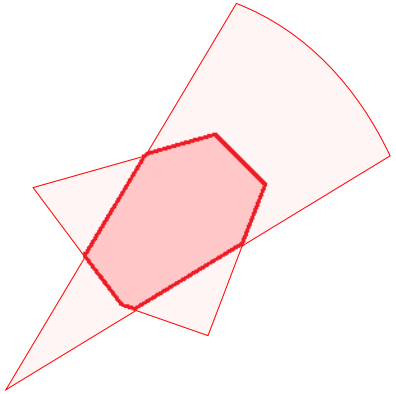
$\mathcal{C}$ 

Semidefinite Programming

$$\min_{\mathbf{X}\succeq\mathbf{0}} \mathbf{A}^\top \mathbf{X}$$

$$s.t.\ \mathbf{B}_i^\top \mathbf{X} \leq c_i$$

f 

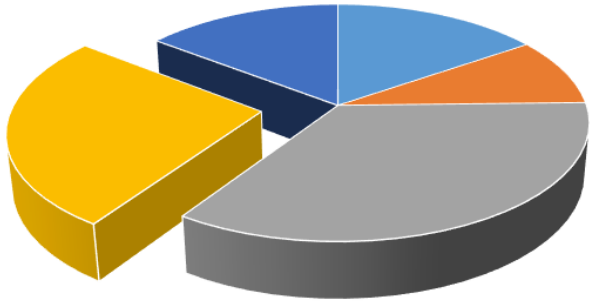
$\mathcal{C}$ 

Techniques

- Projected (Sub)gradient Methods
  - Stochastic, mini-batch variants
  - Primal, dual, primal-dual approaches
  - Coordinate update techniques
- Interior Point Methods
  - Barrier methods
  - Annealing methods
- Other Methods
  - Cutting plane methods
  - Accelerated routines
  - Proximal methods
  - Distributed optimization
  - Derivative-free optimization

Why Study Optimization?

Resource Allocation

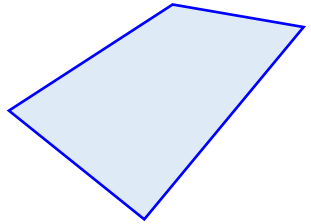


$$\max_{j_1, j_2, \ldots, j_T} \sum_{t=1}^{T} v(i_t, j_t)$$

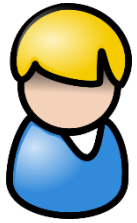
$$s.t. \sum_{t=1}^{T} \mathbb{I}\{j_t = j\} \leq G_j \ \forall j$$

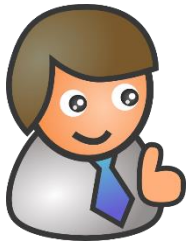
Advertisers

f

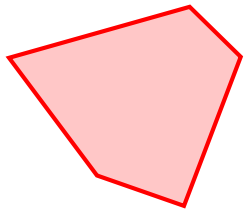


Users

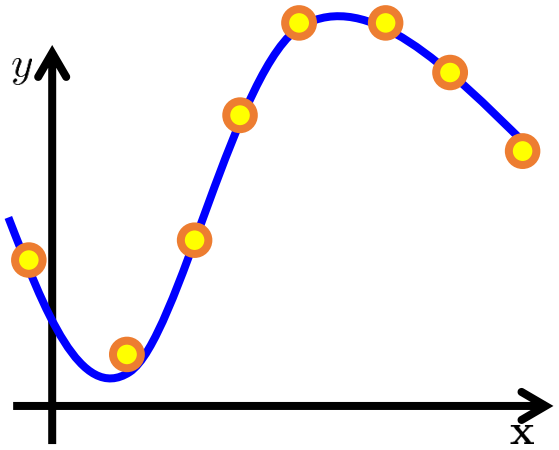


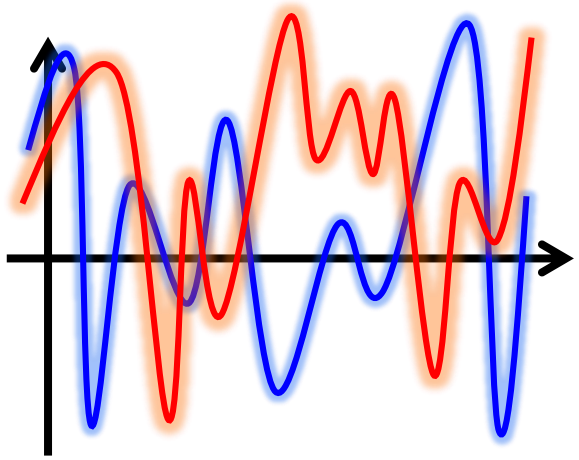


C



Finance and Economics



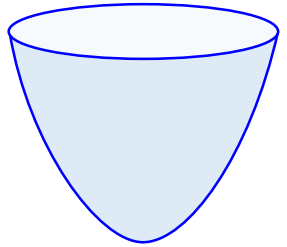


$$\min_{\mathbf{w}\in\mathbb{R}^p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \mathbf{w})^2$$

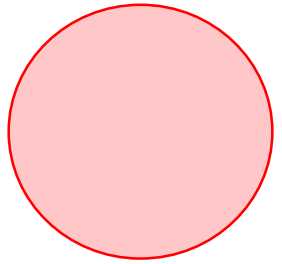
$$s.t.\ \|\mathbf{w}\|_2 \leq R$$

- Risk analysis in investment profiles
- Consumption assessment
- Demand and Supply prediction
- Trend analysis in time series

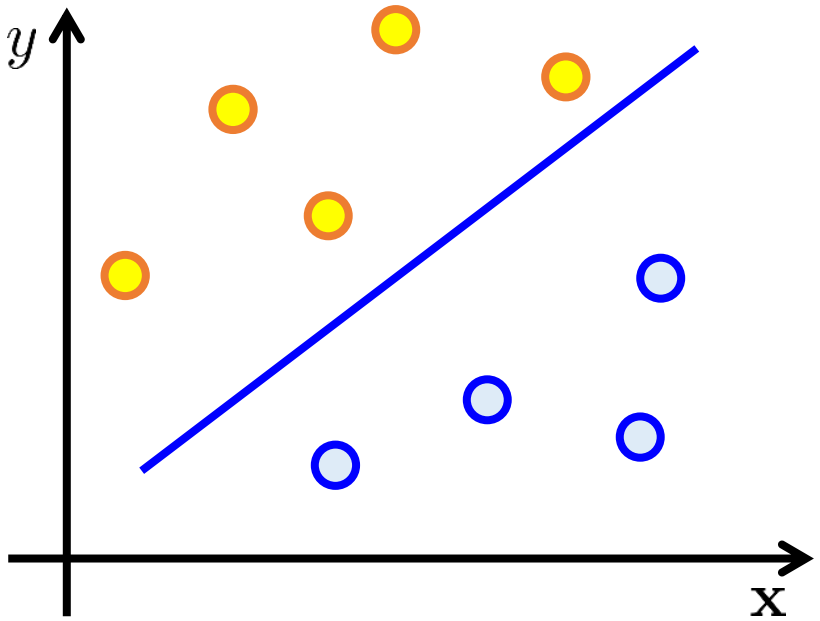
f



$\mathcal{C}$

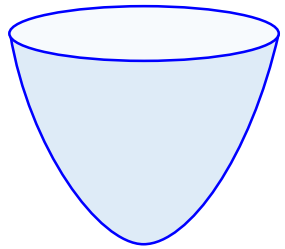


Fraud Detection



$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \ell(y_i, f(\mathbf{x}_i))$$

f



Users

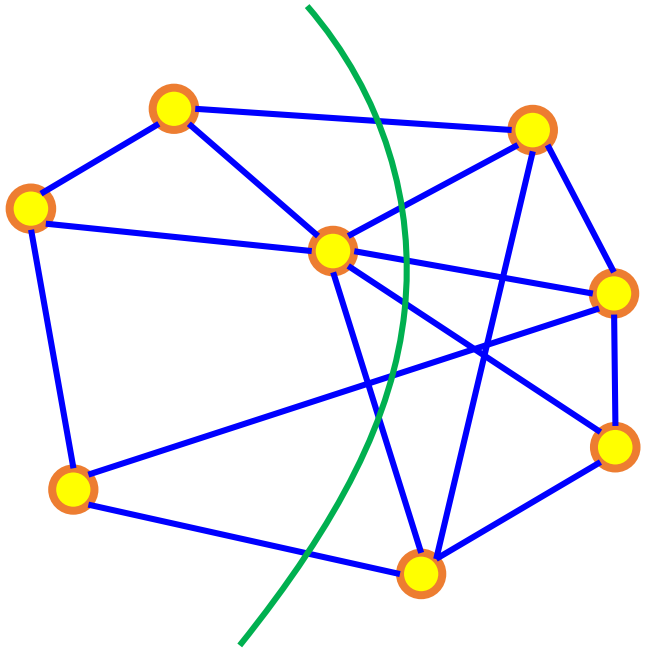
c

$*$

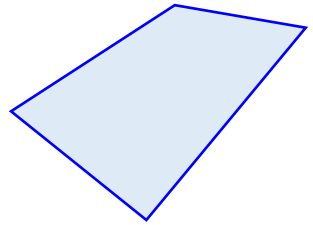
Image Segmentation

Input

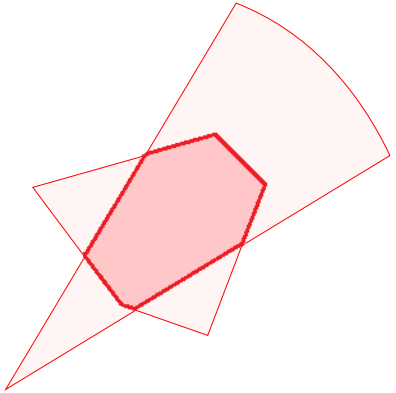
Output



f

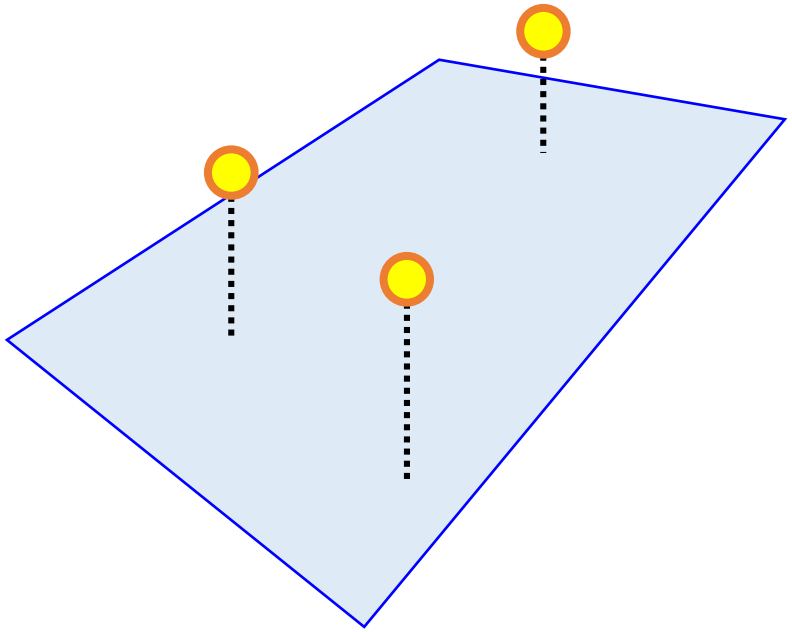


$\mathcal{C}$



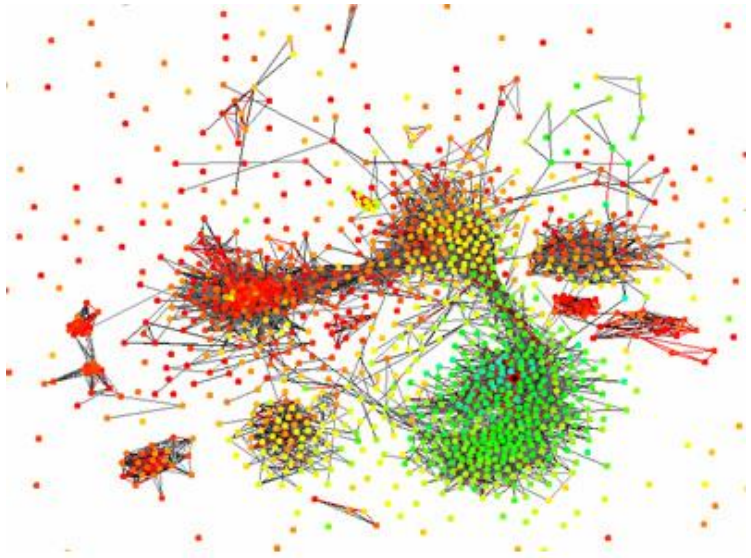
$$\min_{S \subset V} \frac{\text{cut}(S, \bar{S})}{\min\{|S|, |\bar{S}|\}}$$

Dimensionality Reduction



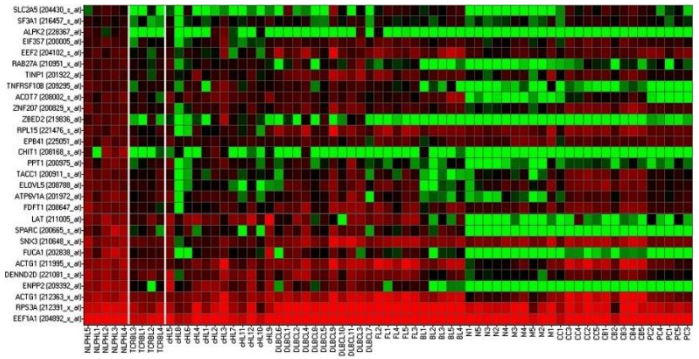
$$\min_{\mathbf{z}_1,\ldots,\mathbf{z}_n \in \mathbb{R}^d} \sum_{i=1}^{n} \ell(\mathbf{z}_i, \mathbf{x}_i)$$

where $\mathbf{x}_i \in \mathbb{R}^D$



visualcomplexity.com, infognition.com, opencv.org

Gene Expression Analysis

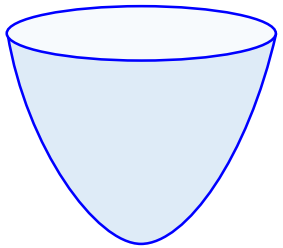


DNA micro-array gene expression data

$$\min_{\mathbf{w}\in\mathbb{R}^p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \mathbf{w})^2$$

$$s.t.\ \|\mathbf{w}\|_0 \le s \ll p$$

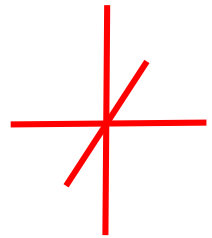
... n subjects

f 

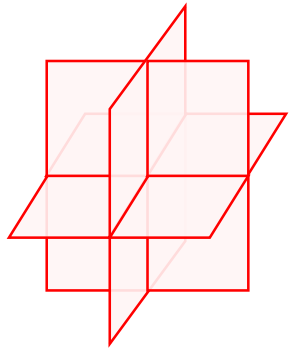
$$_i = (\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$$

Linear model: $y_i \approx \mathbf{x}_i^\top \mathbf{w}^*$

Challenge: $\mathbf{w}^*$ is sparse!

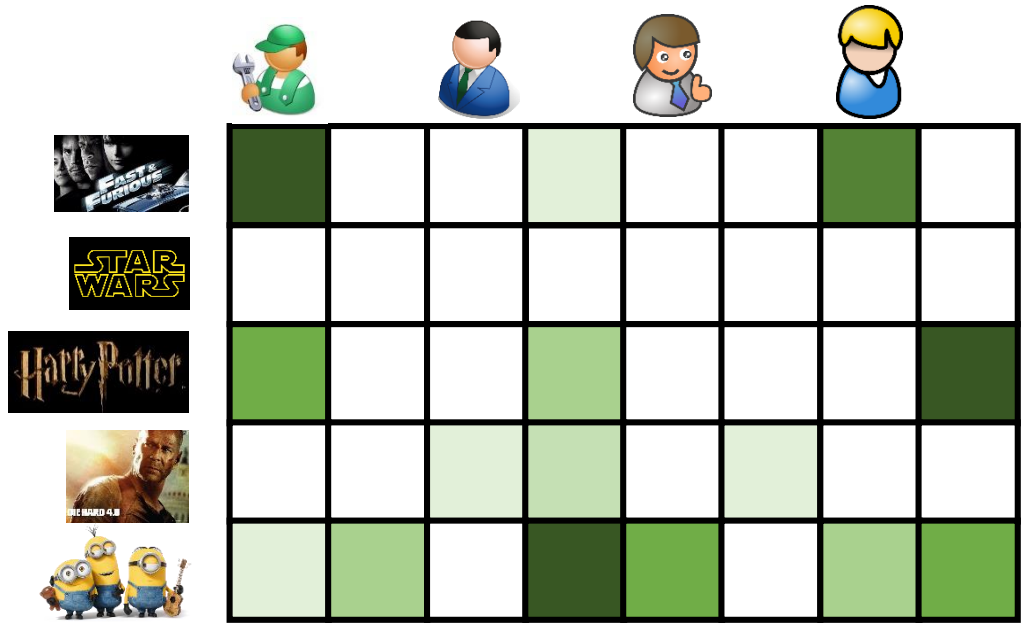
$\mathcal{C}$ 

$p = 3, s = 1$



$p = 3, s = 2$

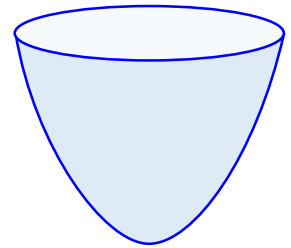
Recommender Systems



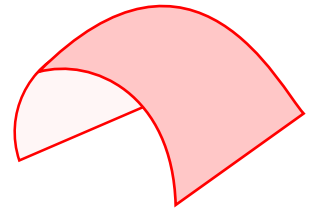
$$\min_{L \in \mathbb{R}^{m \times n}} \|X_\Omega - L_\Omega\|_F^2$$

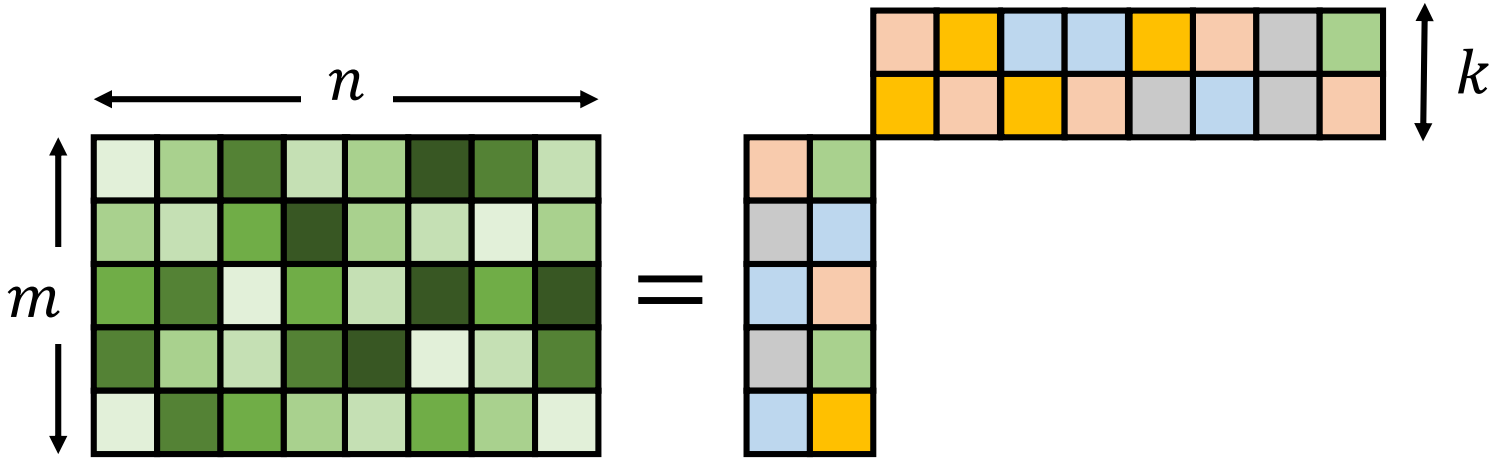
$$s.t.\ \mathrm{rank}(L) \leq k$$

f



$\mathcal{C}$

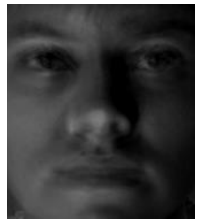


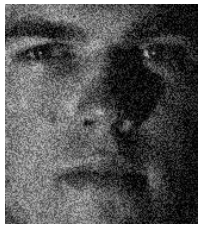


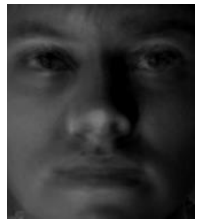
clker.com, clipartpanda.com, nj-clucker.com, hotelgardenmoshir.com

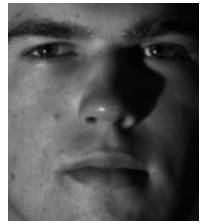
Image Reconstruction and Robust Face Recognition

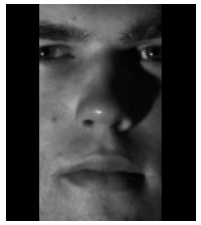

$=$
0.05

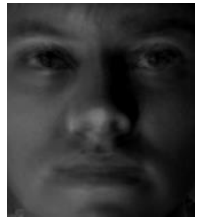
$+$
0.05

$+$
0.90



$\approx$
0.01

$+$
0.07

$+$
0.92

$=$



$\approx$
0.15

$+$
0.20

$+$
0.65

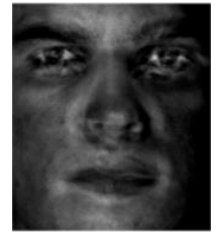
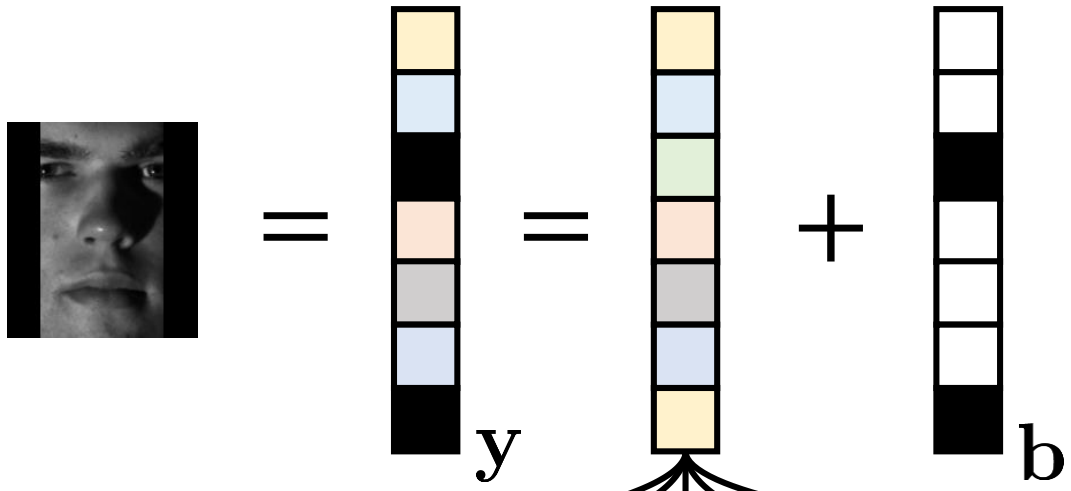
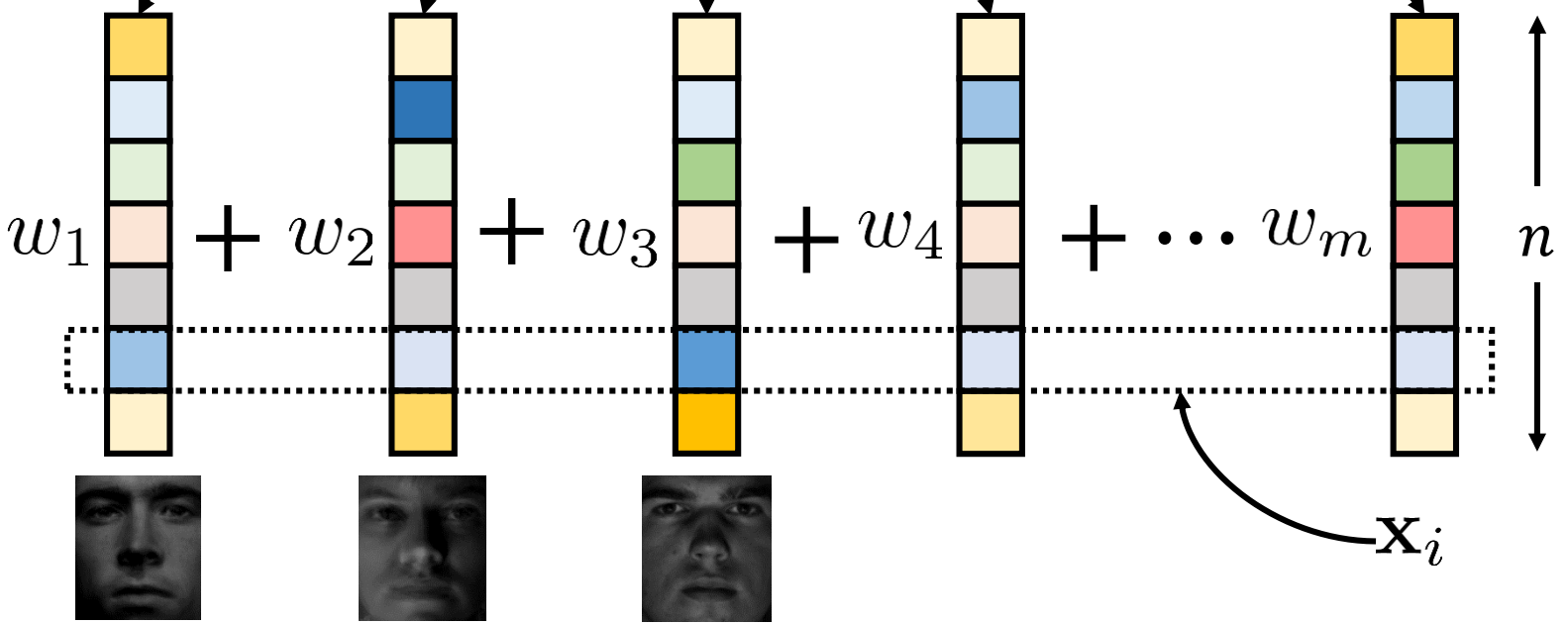
$=$


Image Denoising and Robust Face Recognition

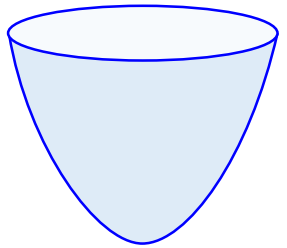


$$\min_{\mathbf{w}\in\mathbb{R}^p} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \mathbf{w} - b_i)^2$$

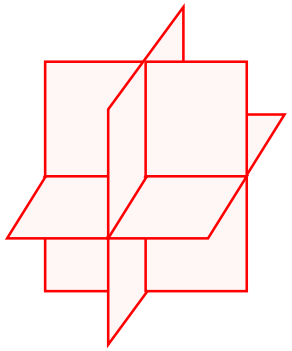
$$s.t.\ \|\mathbf{b}\|_0 \le k \ll n$$



f



$\mathcal{C}$



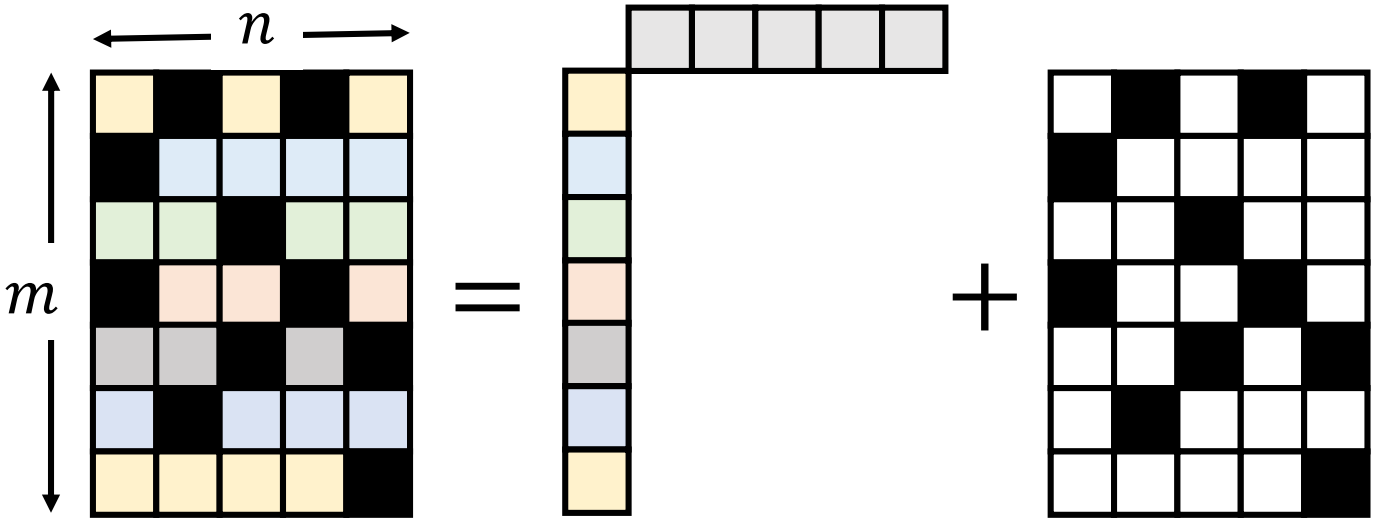
Large Scale Surveillance

- Foreground-background separation

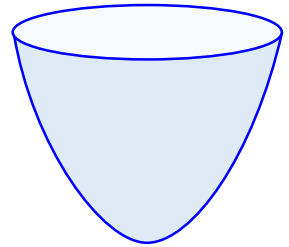
$$=\quad =\quad +$$

$$\min_{L \in \mathcal{M}_k^{m,n}} \|X - (L+S)\|_F^2$$

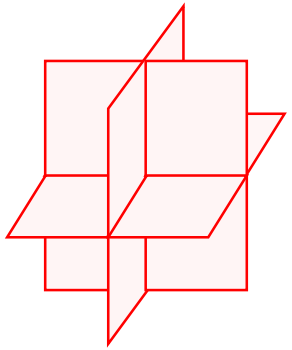
$$s.t.\ \|S\|_0 \leq s \ll mn$$



f



$\mathcal{C}$



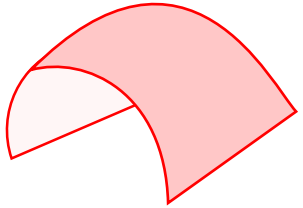
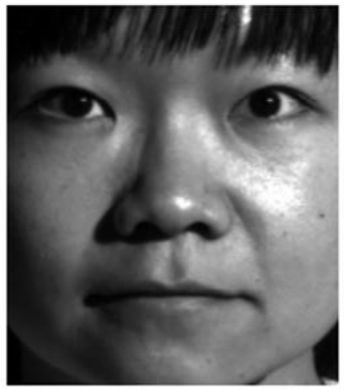


Image Reconstruction

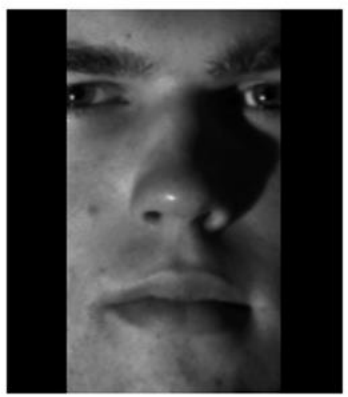


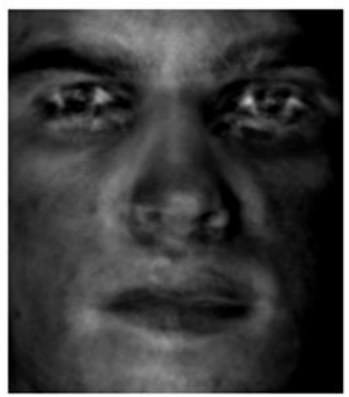
Original

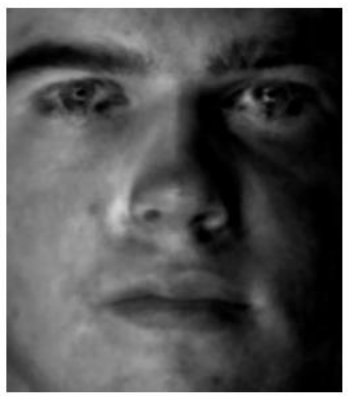
Input

OLS

TORRENT







[Bhatia *et al* 2015]

Foreground-background Separation

Convex Relaxation. Runtime: 1700 sec

Alt-Proj. Runtime: 70 sec

[Netrapalli *et al* 2014]

What we will do in this course

- Optimization Basics: a homage to the classical topics
- First Order and Second Order Methods
- Online and Stochastic Optimization Methods
- Non-convex Optimization Methods
- Based on interest and on demand

  Accelerated methods, Bayesian methods, Coordinate methods, Cutting plane methods, Interior point methods, Optimization methods for deep learning, Parallel and distributed methods, Robust optimization, Stochastic mini-batch methods, Submodular optimization, Variance reduced stochastic methods, Zeroth order methods
- Budget: 25 lectures – 1 (this one) – 2 (mid/end term project presentations)

Up Next

- Brief Introduction to Convex Analysis, Linear Algebra, and Probability Theory
- Basics of Optimization – Optimality, Convergence Rate, Duality