

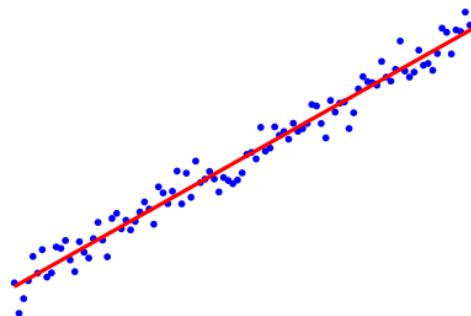
# Introductory Course on Non-smooth Optimisation

## Lecture 00 - Introduction

# **Outline**

## **1 Introduction**

## Example: least square estimation



Given cluster of points  $(h_i, v_i)_{i=1,\dots,m}$ , find a line  $v = ah + b$  such that it minimises

$$\sum_{i=1}^m \|ah_i + b - v_i\|^2.$$

Let

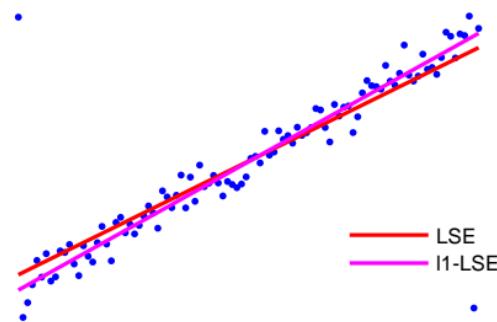
$$A = \begin{bmatrix} \vdots & \vdots \\ h_1 & 1 \\ \vdots & \vdots \end{bmatrix}, \quad x = \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} \vdots \\ v_i \\ \vdots \end{pmatrix},$$

then the above problem is equivalent to

$$\min_{x \in \mathbb{R}^2} \|Ax - w\|^2.$$

Closed form solution if  $A^T A$  has full rank:  $x^* = (A^T A)^{-1} A^T w$ .

## Example: least square with outliers



LSE is not robust to outliers. To deal with outliers which are sparse, consider

$$\min_{x \in \mathbb{R}^2} \|Ax - w\|_1.$$

Challenge: non-smooth, and no closed form solution.

Solvers: ADMM (alternating direction methods of multipliers)...

## Example: sparse logistic regression

Given two clusters of points  $(z_i, y_i) \in \mathbb{R}^n \times \{\pm 1\}$ ,  $i = 1, \dots, m$ , find a separation hyperplane via

$$\min_{(b,x) \in \mathbb{R} \times \mathbb{R}^n} \mu \|x\|_1 + \frac{1}{m} \sum_{i=1}^m f(\langle x, z_i \rangle + b, y_i),$$

where  $f(u_i, y_i) = \log(1 + e^{-u_i y_i})$ .

Requirement:  $x$  is sparse, that is  $x$  has few non-zero elements.

## Example: image deblur

$x_{\text{ob}}$	$w$	recovered $x$
-----------------	-----	---------------

How to blur an image

$$w = H * x_{\text{ob}} + \omega,$$

where  $H \in \mathbb{R}^{m \times n}$  is blur kernel,  $\omega \in \mathbb{R}^m$  is additive noise.

How to deblur? Sharp edges are the most important part of images.

$$\min_{x \in \mathbb{R}^{m \times n}} \mu \|\nabla x\|_1 + \frac{1}{2} \|H * x - w\|^2.$$

$\|\nabla x\|$  promotes sharp edges: out of the solutions of the LSE  $\frac{1}{2} \|H * x - w\|^2$ , finding the one  $x$  which has “proper” sharp edges...

## Example: matrix decomposition

$w$

$x_I$

$x_s$

Forward mixture model,

$$w = x_{\text{ob},I} + x_{\text{ob},s} + \omega,$$

$x_{\text{ob},I} \in \mathbb{R}^{m \times n}$  is low-rank,  $x_{\text{ob},s} \in \mathbb{R}^{m \times n}$  is sparse and  $\omega \in \mathbb{R}^{m \times n}$  is noise.

How to decompose  $w$  into low-rank part plus sparse part?

$$\min_{x_I, x_s \in \mathbb{R}^{m \times n}} \mu_1 \|x_s\|_1 + \mu_2 \|x_I\|_* + \frac{1}{2} \|x_s + x_I - w\|_F^2.$$

$\|x_I\|_* = \sum_i \sigma_i$ , where  $(\sigma_i)_{i=1,\dots,\text{rank}(x_I)}$  are the singular values of  $x_I$ .

## Example: linear inverse problems

$$w = H x_{\text{ob}} \odot \omega$$

**Forward model:**

$$w = (Hx_{\text{ob}}) \odot \omega.$$

**Goal:** recover  $x_{\text{ob}}$

**Challenge:** ill-posed

**Hope:** prior knowledge of  $x_{\text{ob}}$

- Regularisation: promoting low-complexity structure to the solution...
- Examples:

Sparsity  $\ell_1$ -norm,  $\ell_{1,2}$ -norm,  $\ell_p$ -norm,  $\ell_0$  pseudo-norm

Analysis sparsity total variation, wavelet, dictionary...

Low-rank nuclear norm, rank function

Constraints simplex, non-negativity...

Nerual networks CNN...

# Optimisation problem

Least square

$$\min_{x \in \mathbb{R}^n} \|Ax - w\|^2.$$

Least square with outliers

$$\min_{x \in \mathbb{R}^n} \|Ax - w\|_1.$$

Sparse logistic regression

$$\min_{x \in \mathbb{R}^n} \mu\|x\|_1 + \frac{1}{m} \sum_{i=1}^m f(\langle x, z_i \rangle + b, y_i).$$

Image deblur

$$\min_{x \in \mathbb{R}^{m \times n}} \mu\|\nabla x\|_1 + \frac{1}{2}\|H * x - w\|^2.$$

Principal component pursuit

$$\min_{x_l, x_s \in \mathbb{R}^{m \times n}} \mu_1\|x_s\|_1 + \mu_2\|x_l\|_* + \frac{1}{2}\|w - x_l - x_s\|^2.$$

## Challenges

Optimisation problems

$$\min_{x \in \mathbb{R}^n} F(x) + R(x),$$

subject to  $x \in \Omega$ .

Eg  $F = \frac{1}{2}\|Ax - w\|^2$ ,  $R = \|x\|_1$  and  $\Omega = \{x | f_i(x) \leq b_i, i = 1, \dots, m\}$ .

- No closed form solution: iterative scheme
- Non-smooth: difficult to evaluate
- Non-linear: linearisation or approximation
- **Non-convex: no global minimiser guarantee**
- Composite: need proper numerical schemes
- High dimension: computational demanding
- Others: hardware limitation e.g. mobile devices

## Goals and contents

### Goals

- recognise/formulate problems as optimisation problems
- characterise the property of solution
- familiar with and able to apply first-order methods

### Contents

- convex analysis and set-valued analysis
- first-order methods
- examples and applications

**NB:** rigorous mathematical proofs will not be the focus of this course...

## First-order methods

### Brief history

- Origins from numerical PDE dates back to 1950s
- Received attention since 1970s
- Tremendous development since new century...

# First-order methods

## Brief history

- Origins from numerical PDE dates back to 1950s
- Received attention since 1970s
- Tremendous development since new century...

## Applications

- Before 1990, mostly in operation research (e.g. linear programming)
- Since 1990, becomes ubiquitous in signal/image processing, inverse problems, data science, statistics, machine learning, game theory...

## First-order methods

First-order methods (FoM)...

$F$  Gradient descent, Heavy-ball

$R$  Proximal Point Algorithm (PPA), inertial PPA

$F + R$  Forward–Backward splitting (FB), inertial FB, Nesterov/FISTA

$F = \frac{1}{m} \sum_i f_i$ : stochastic gradient methods

$R_1 + R_2$  Douglas–Rachford splitting

$F + R(\mathcal{W}\cdot)$  Class of Primal–Dual splitting

Alternating Direction Method of Multipliers (ADMM)

$F + \sum_{i=1}^r R_i$  Three-operator splitting ( $r = 2$ )

Forward–Douglas–Rachford ( $r = 2, R_2 = \iota_{\mathcal{V}}(\cdot)$ )

Generalized Forward–Backward splitting ( $r \geq 2$ )

— ...

# Schedule

## Schedule

- Convex optimisation: 12 lectures
- Non-convex optimisation: 2 lectures
- Stochastic optimisation: 2 lectures

## Projects

- Convex optimisation: about 4
- Non-convex optimisation: 1
- Stochastic optimisation: 1

Only 1 or 2 projects will be mandatory, and they will be done in groups.

Programming language: MATLAB, Python.

## References

### References

- S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004.
- R. T. Rockafellar. Convex analysis. Princeton university press, 2015.
- A. Beck. First-order methods in optimization. Vol. 25. SIAM, 2017.
- H. H. Bauschke and P. L. Combettes. Convex analysis and monotone operator theory in Hilbert spaces. Vol. 408. New York: Springer, 2011.
- B. Polyak. Introduction to optimization. Optimization Software, 1987
- Y. Nesterov. Introductory lectures on convex optimization: A basic course. Vol. 87. Springer Science & Business Media, 2013.