Outline

Numerical methods for optimisation

Marie Postel
Laboratoire Jacques-Louis Lions
Sorbonne Université

AIMS Master 2023-2024

postel.pages.math.cnrs.fr/aims

Practical issues

- Final grade : weighted sum of following grades
  - Python and math team assignments (at least 3)
  - Final written exam (2 hours 16/02)
- There will be at least 3 Python hands on sessions in place of regular classes
- Each hands on session will be followed by an evening session to complete the program before handing it in

Course objective

- Introduction to numerical methods of Optimization
- Improve programming skills
- Implementation and test of algorithms

Why Python 3.1

- Ideal for building algorithm prototypes
- Flexible interactive graphics
- Widely used in business and all scientific sectors
- Performance worse than in a compiled language of high level (C++, Fortran)

Course map

- Introduction
  - Introduction of Optimization
  - Differential calculus revisions
  - Convexity revisions
  - Numerical approximation of derivatives
- Un-constrained Continuous Optimization
  - Optimality conditions
  - Nonlinear equations (Fixed point, Newton and Quasi-Newton)
  - Descent/Gradient algorithms
- Constrained Continuous Optimization
  - Duality
  - Optimality conditions with equality constraints
  - SQP algorithm
  - Optimality conditions with inequality constraints
  - Uzawa algorithm

Different categories of optimization

- Discrete optimization : variables in a discrete set
  - Combinatorial <-> linear programming
  - "NP-complete" (nondeterministic polynomial-time complete)
  - Logistics, Economy (Traveling salesman, Knapsack, etc.)
  - Heuristic methods : Hill climbing, Simulated annealing, Ant colony, etc.
- Continuous optimization : variables within a range of values
  - Infinite dimensions : calculus of variations, shape optimization, control theory
  - Finite dimension : includes the discretization of above problems

Definition of a minimum

Def : Let $f : V \to \mathbb{R}$ with V normed vector space.
$x^\star \in D_a \subset V$ **achieves**

- a local minimum on D_a if there exists $\varepsilon > 0$ such that

$$f(x^\star) \leq f(x) \quad \text{for all} \quad x \in D_a \quad \text{t.q. } \|x - x^\star\| \leq \varepsilon.$$

- a strict local minimum if there exists $\varepsilon > 0$ such that

$$f(x^\star) < f(x) \quad \text{for all} \quad x \in D_a \quad \text{s. t. } x \neq x^\star \text{ and } \|x - x^\star\| \leq \varepsilon.$$

Definition of a minimum

Def : $x^\star \in D_a$ **achieves**

- a global minimum on D_a if

$$f(x^\star) \leq f(x) \quad \text{for all} \quad x \in D_a.$$

- a strict global minimum if

$$f(x^\star) < f(x) \quad \text{for all} \quad x \in D_a \quad \text{s. t. } x \neq x^\star.$$

It is sometimes said that $x^\star$ **is a minimum** of $f(x)$, but this is a misnomer. The exact term, if $x^\star$ **realizes** a minimum of f, is that it is a **minimizer** of f, denoted

$$x^\star = \underset{x \in D_a}{\operatorname{argmin}} f(x)$$

Definition of a maximum

To find the maximum of f we search the minimum of $-f$.

General Optimization problem

exemple $f : \mathbb{R}^n \to \mathbb{R}$ $f(x) = \|x\|^2 = \sum_{i=1}^{n} x_i^2$

Definition : Let $F : V \to \mathbb{R}$ with V normed vector space. F is coercive iff $\lim_{\|x\| \to +\infty} F(x) = +\infty$.

Property : If F is continuous, F has a minimum on every compact set $\subset E$

Property : A function $F(x)$ from a finite dimensional space V into $\mathbb{R}$ which is continuous and coercive admits at least one minimum

Optimization applied to differential problems : Calculus of Variations

Let $V_0 = \{u \in C^2([0,1]),\ u(0) = u(1) = 0\}$ and $g : \mathbb{R}^3 \to \mathbb{R}$, $g \in \mathcal{C}^1$.

$$\mathcal{J}(u) = \int_0^1 g(x, u(x), u'(x))dx, \quad u \in V_0.$$

$$D\mathcal{J}(u)(v) = \langle -\frac{d}{dx}\frac{\partial g}{\partial u'}(x,u,u') + \frac{\partial g}{\partial u}(x,u,u'), v\rangle_{L^2([0,1])}.$$

Euler-Lagrange Theorem: An extremum of $\mathcal{J}$ satisfies

$$-\frac{d}{dx}\frac{\partial g}{\partial u'}(x,u,u') + \frac{\partial g}{\partial u}(x,u,u') = 0$$

Infinite dimension example

Let $V_0 = \{u \in C^2([0,1]),\ u(0) = u(1) = 0\}$, $f \in C^1([0,1])$ and $g : \mathbb{R}^3 \to \mathbb{R}$, $g \in \mathcal{C}^1$.

$= \int_0^1 g(x, u(x), u'(x))\,dx$

$$\mathcal{J}(u) = \int_0^1 \frac{1}{2}u'(x)^2 + \frac{1}{2}u(x)^2 - f(x)u(x)dx, \quad u \in V_0.$$

$$g(x,u,u') = \frac{1}{2}u'^2 + \frac{1}{2}u^2 - f(x)u.$$

$$\begin{aligned} D\mathcal{J}(u)(v) &= \langle -\frac{d}{dx}\frac{\partial g}{\partial u'}(x,u,u') + \frac{\partial g}{\partial u}(x,u,u'), v\rangle \\ &= \int_0^1 (-u'' + u - f)vdx. \end{aligned}$$

$$\mathcal{J}(\bar{u}) = \min \mathcal{J}(u) \text{ iff } \begin{cases} -u'' + u = f \\ u(0) = u(1) = 0 \end{cases}$$

Boundary value problem

Canonical Continuous Optimization problem on $\mathbb{R}^n$

Find the extrema of a function $f(x)$ defined on $\mathbb{R}^n$ (or part of $\mathbb{R}^n$ in the case of a optimization with constraints).

Find

$$\inf_{x \in \mathbb{R}^n} f(x),$$

under constraints

equality constraints

$$\begin{aligned} C^E(x) &= 0, \\ C^I(x) &\preceq 0 \quad (\Leftrightarrow C^I_i(x) \leq 0, \ i = 1, \ldots, p) \end{aligned}$$

inequality constraints

with

$$\begin{aligned} f &: \mathbb{R}^n \longrightarrow \mathbb{R}, \\ C^I &: \mathbb{R}^n \longrightarrow \mathbb{R}^p, \\ C^E &: \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad f, C^I, C^E, \text{smooth} \end{aligned}$$

Admissible domain

$$D_a = \{x \in \mathbb{R}^n, \ C^E(x) = 0, \ C^I(x) \preceq 0\}$$

Example 1: linear programming

Lagrange multiplier

$f : \mathbb{R}^n \to \mathbb{R}$ linear

$f(x) = c^T x = \langle c, x \rangle = \sum_{i=1}^{n} c_i x_i$

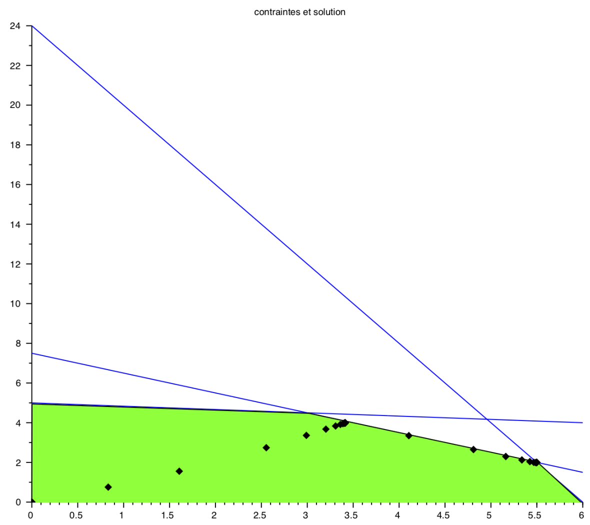
Generic problem

$$(P) \left\{ \begin{array}{ccccl} \text{Minimize} & c^T x & & & c \text{ and } x \in \mathbb{R}^n, \\ \text{under constraints} & Ax \preceq b & & \text{with} & b \in \mathbb{R}^m, \\ & x \succeq 0 & & & A \in \mathcal{M}_{m\times n}(\mathbb{R}) \end{array} \right.$$

Example and admissible domain

$$A = \begin{pmatrix} 1 & 6 \\ 2 & 2 \\ 4 & 1 \end{pmatrix}, \; b = \begin{pmatrix} 30 \\ 15 \\ 24 \end{pmatrix}$$

$$c = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$



Rewrite example 1 in canonical form

$$\inf_{x\in\mathbb{R}^n} f(x) \text{ under constraints } \begin{cases} C^E(x) = 0, \\ C^I(x) \preceq 0 \end{cases}$$

$Ax \le b$, $x \geq 0$

- f : $\mathbb{R}^n \to \mathbb{R}$ $f(x) = c^T x$
- C^E :
- C^I : $\mathbb{R}^n \longrightarrow \mathbb{R}^{m+n}$
- D_a, def, nature

$$C^I(x) = \begin{pmatrix} Ax - b \\ -x \end{pmatrix} \begin{matrix} m \le \\ n \le \end{matrix}$$

Solve with the Python toolbox linprog

u

```
scipy.optimize.linprog (c, A_ub=None,
B_ub=None, A_eq=None, B_eq=None, bounds=None,
method='interior-point', callback=None,
options=None, x0=None)
```

The problem must be written in the form expected by the program

$$\begin{aligned} \min_x \quad & c^T x \\ \text{such that} \quad & A_{ub} x \preceq B_{ub} \\ & A_{eq} x = B_{eq} \\ & \ell \preceq x \preceq u \end{aligned}$$

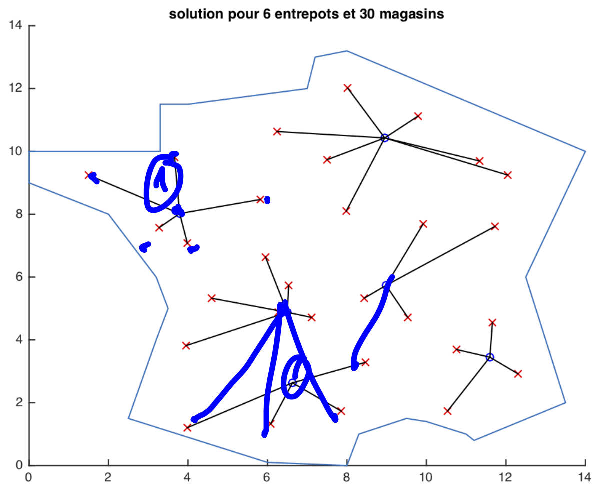
lower upper

Solve with the Python toolbox linprog

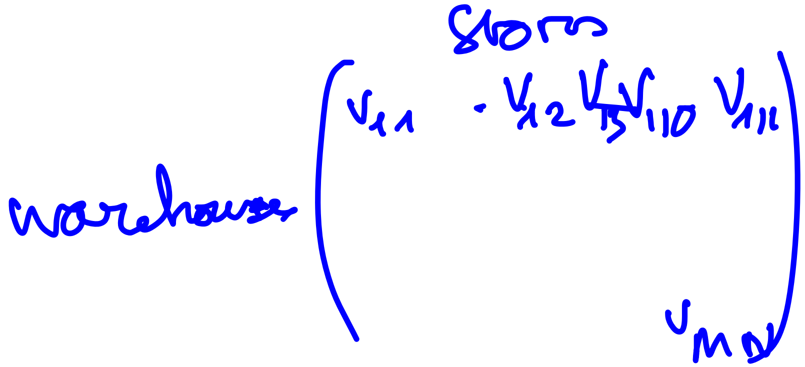
```
c = [-2, -1]
Aub = [[1, 6], [2, 2], [4,1]]
Bub = [30,15,24]
lu = (0., None)
bounds=2*[lu]
res = scipy.optimize.linprog(c, A_ub=Aub,
b_ub=Bub, bounds=bounds)
```

Practical example

- A company stores a commodity in M warehouses.
- Each warehouse i ($i = 1, \ldots, M$) has a quantity q_i of goods in stock.
- The company has a network of N stores.
- Each store j ($j = 1, \ldots, N$) ordered a quantity r_j of goods.
- The problem is to minimize the cost of delivering goods to stores.



Mathematical modelling



Let us denote

- $v_{i,j}$ the quantity of merchandise shipped from warehouse i to store j
- $Q = \sum_{i=1}^{M} q_i$ the total quantity of goods available in the warehouses
- $R = \sum_{j=1}^{N} r_j$ the total quantity of goods ordered by the stores, assuming $Q \geq R$
- $D_{i,j}$ the cost of unit transport from the warehouse i to the store j, directly proportional to the distance between the store and the warehouse.

Rewriting as a linear programming problem

The problem (whose unknowns are the $v_{i,j}$) is therefore to minimize

$$f(v) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} D_{i,j} v_{i,j}$$

linear

with respect to v, under the constraints

(i) $v_{i,j} \geq 0$ we do not return goods from a store to a warehouse

(ii) $\sum_{j=0}^{N-1} v_{i,j} \leq q_i$ a warehouse cannot supply more than its stock

(iii) $\sum_{i=0}^{M-1} v_{i,j} = r_j$ each store must receive the requested quantity

Solve with the Python toolbox linprog

```
scipy.optimize.linprog (c, A_ub=None,
B_ub=None, A_eq=None, B_eq=None, bounds=None,
method='interior-point', callback=None,
options=None, x0=None)
```

The problem must be written in the form expected by the program

$$\begin{aligned} \min_x \quad & c^T x \\ \text{such that} \quad & A_{ub} x \preceq B_{ub} \\ & A_{eq} x = B_{eq} \\ & \ell \preceq x \preceq u \end{aligned}$$

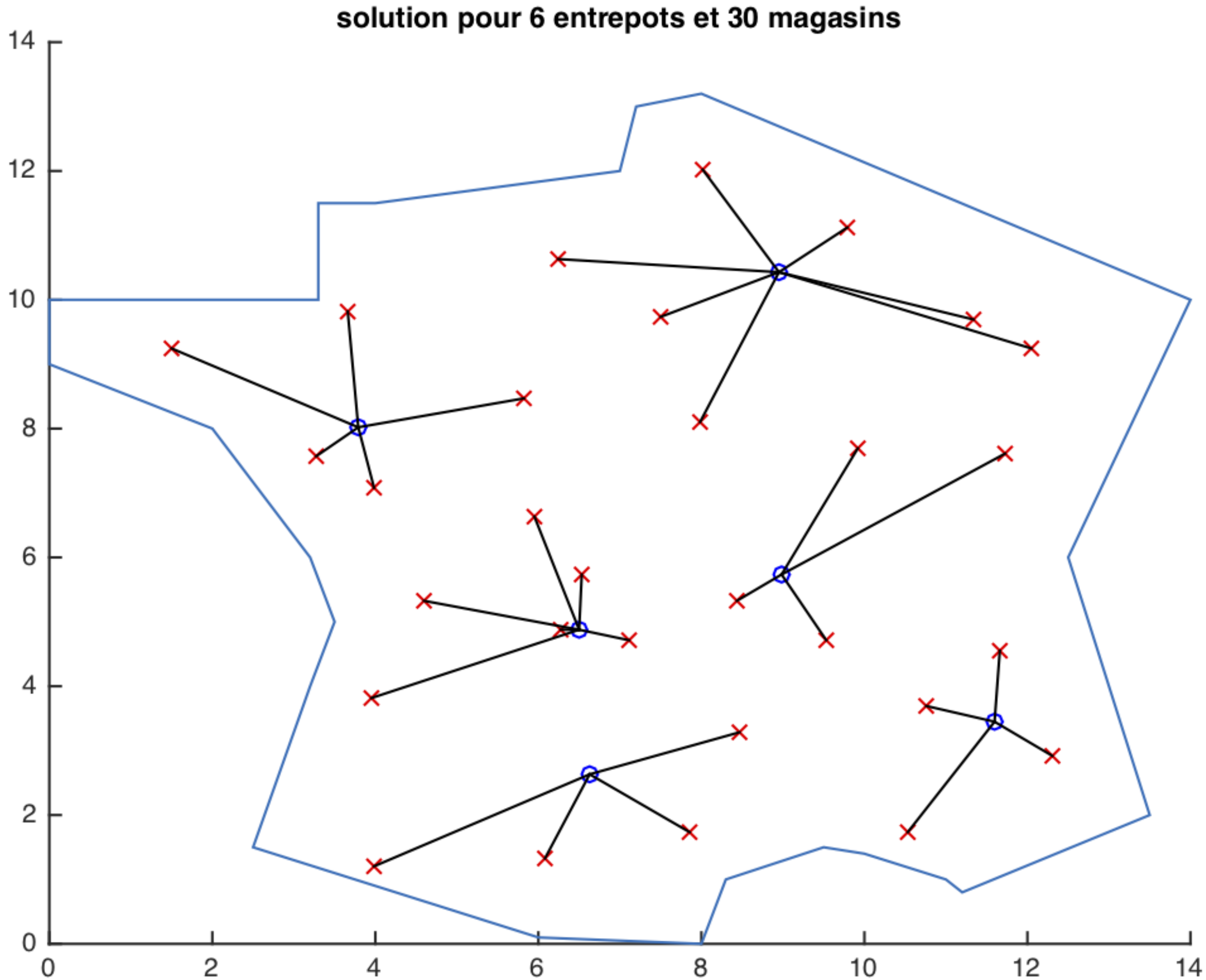
`c,x, B_ub, b_eq` are 1-D arrays or lists, `A_ub, A_eq` are 2-D arrays or lists. `bounds` is a list of 2 1-D lists

Solve with the Python toolbox linprog

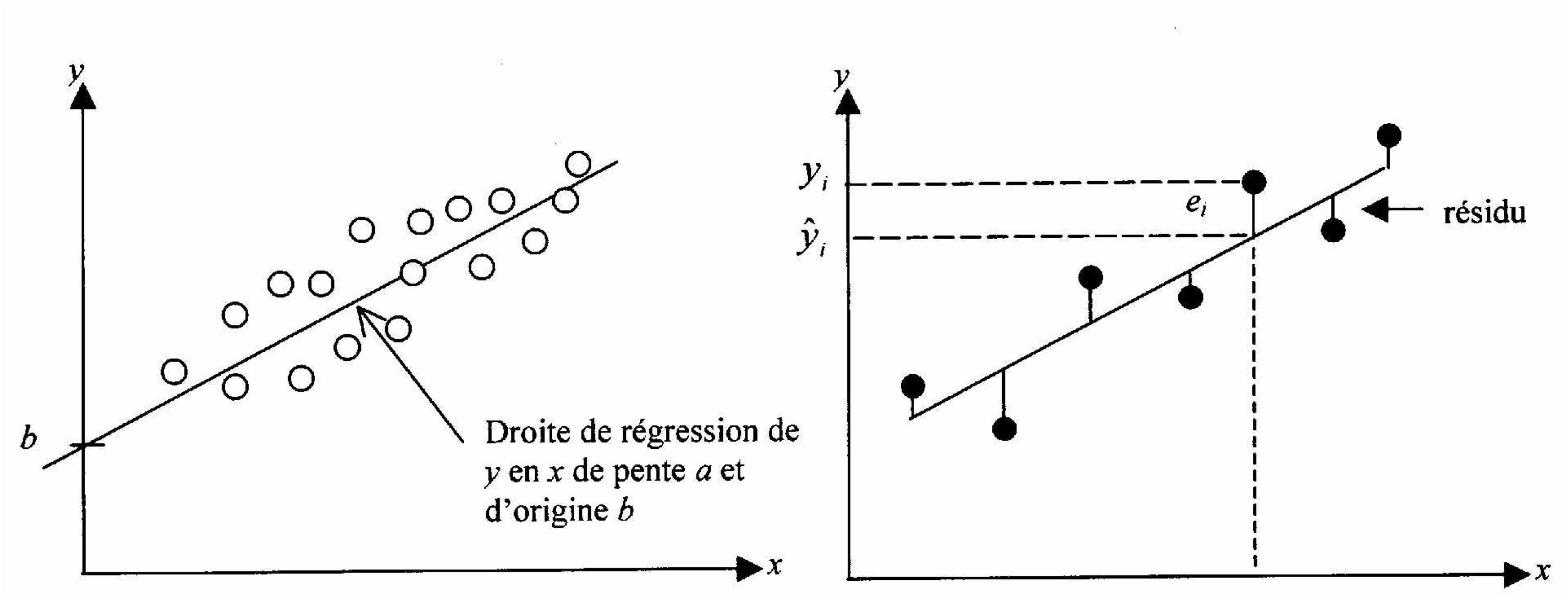
We must therefore define Python structures

- $x \in \mathbb{R}^{NM\times 1}$ contains the solution matrix v unrolled in columns
- $c \in \mathbb{R}^{NM\times 1}$ contains the matrix D unrolled in columns
- $Aub \in \mathcal{M}_{M,MN}(\mathbb{R})$ contains 1s in the right places so that $(Aub\ x)_i = \sum_{j=0}^{N-1} v_{i,j}$ for $i = 0, \ldots, M-1$
- $Bub \in \mathbb{R}^{M}$ contains q (the warehouse stocks)
- $Aeq \in \mathcal{M}_{N,MN}(\mathbb{R})$ contains 1s in the right places so that $(Aeq\ x)_j = \sum_{i=0}^{M-1} v_{i,j}$ for $j = 0, \ldots, N-1$
- $Beq \in \mathbb{R}^{N}$ contains r (the store orders)
- $\ell \in \mathbb{R}^{NM\times 1} = 0$

Solution Matlab du problème d'optimisation linéaire



Example 2 : Least Squares

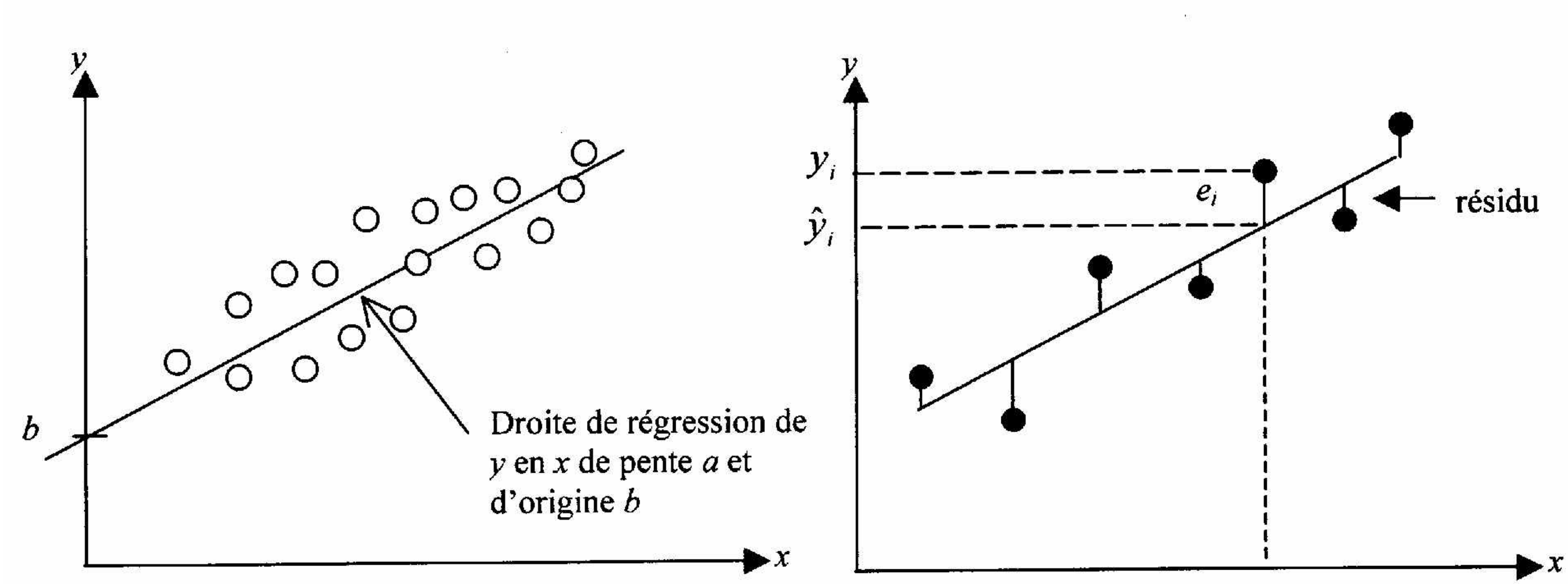


Data $x_i, y_i, i = 1, \ldots, n$

Linear model $y = ax + b$

Minimize $\sum_{i=1}^{n}(ax_i + b - y_i)^2$ with respect to $(a, b) \in \mathbb{R}^2$

Example 2 : Least Squares - vector formulation



Data $x_i, y_i, i = 1, \ldots, n$
Linear model $y = ax + b$

$\sum (ax_i + b.1 - y_i)^2 = f(a,b)$

$$X = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad P = \begin{pmatrix} a \\ b \end{pmatrix}$$

Minimize $\| Y - XP \|_2^2$ with respect to $P \in \mathbb{R}^2$

Rewrite Least Square problem in canonical form

$$\inf_{x \in \mathbb{R}^n} f(x) \text{ under constraints } \begin{cases} C^E(x) = 0, \\ C^I(x) \preceq 0 \end{cases}$$

X and Y are parameters

- f : $\mathbb{R}^2 \rightarrow \mathbb{R}$
- C^E
- C^I
- D_a, def, nature

$f(P) = \| XP - Y \|^2$

$D_a = \mathbb{R}^2$

Example 3 : Non differentiable convex Optimization

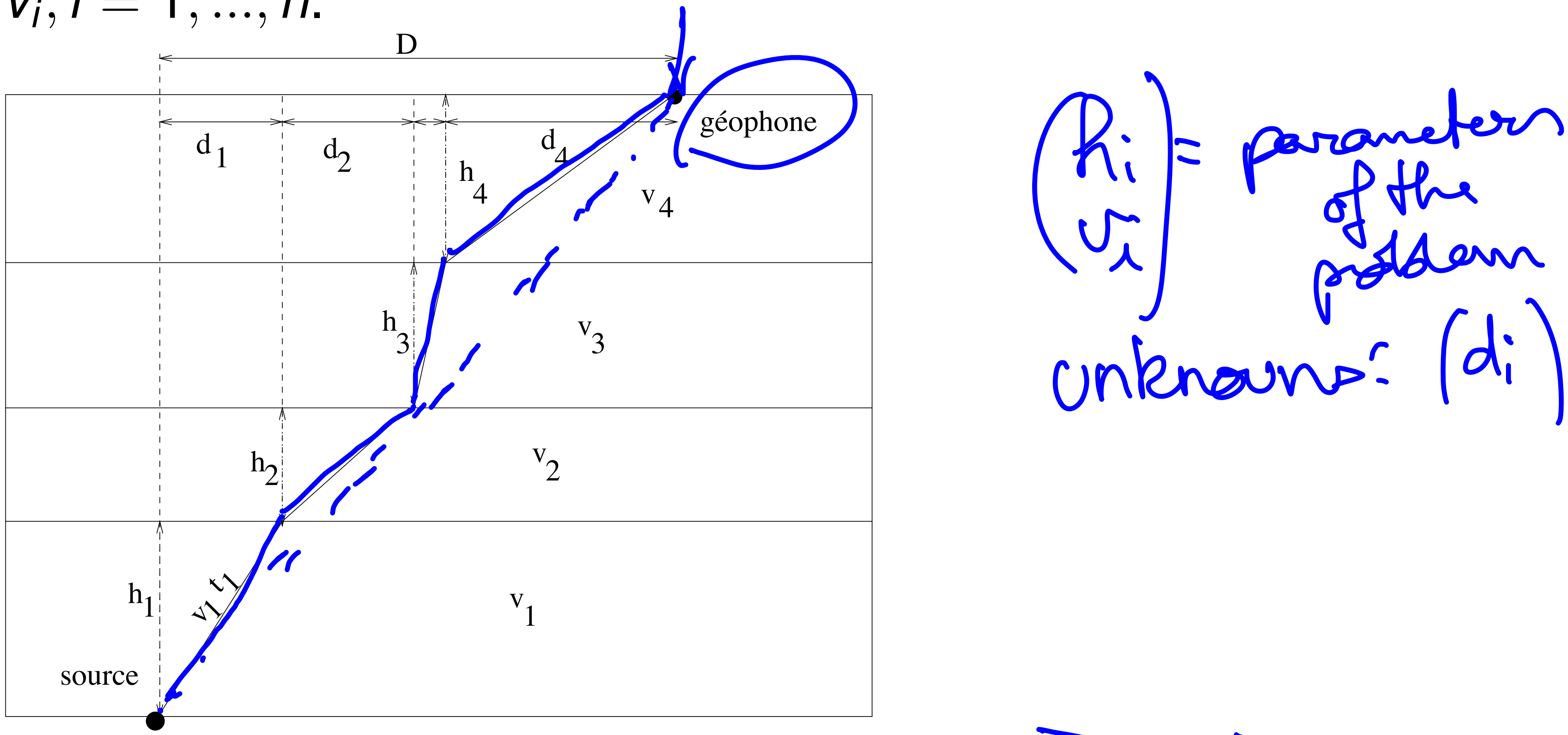
Parsimonious Least Squares Lasso (least absolute shrinkage and selection operator)

- Sociological models (e.g. explanation of academic success as a function of social, family, medical factors, etc.)
- Data $Y = (y_i)_{i=1,\dots,n}$, $X = (x_{i,j})_{i=1,\dots,n, j=1,\dots,p}$
- Linear model $\tilde{Y} = XP$, with $P \in \mathbb{R}^p$, using as few factors as possible
- Minimize $\|Y - XP\|_2^2 + \alpha \|P\|_1$

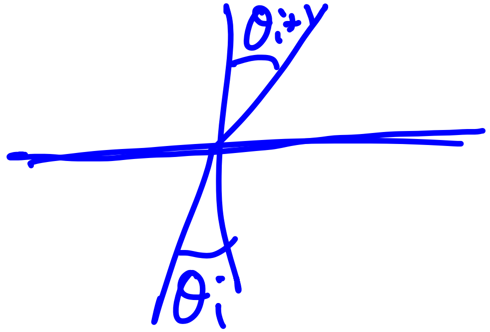
$\|P\|_1 = \sum |P_i|$

Example 5: Wave propagation in a stratified medium by "ray tracing"

n parallel layers of thickness $h_i, i = 1, ..., n$.
In each layer the speed of propagation is constant and equals $v_i, i = 1, ..., n$.



Optimization problem



Path followed by the seismic wave from the source to a geophone on the surface, at distance D from the vertical of the epicenter.

Descartes' law

$$\frac{\sin(\theta_i)}{v_i} = constant$$

Find the minimum travel time $\sum_i t_i$ under constraints

$D = \sum_{i=1}^{n} d_i$.

$d_i^2 + h_i^2 = v_i^2 t_i^2$

Canonical form of the optimization problem

$$\inf_{x\in\mathbb{R}^n} f(x) \text{ under constraints } \left\{ \begin{array}{l} C^E(x) = 0, \\ C^I(x) \preceq 0 \end{array} \right.$$

- f
- C^E
- C^I
- D_a, def, nature

Canonical form of the optimization problem

Choice of unknowns: $(t_i)_{i=1,\dots,n}$ or $(d_i)_{i=1,\dots,n}$

- If $X = (t_i)_{i=1,\dots,n}$
$$f(X) = \sum_{i=1}^{n} x_i \text{ and } C^E(x) = \sum_{i=1}^{n} \sqrt{v_i^2 {x_i}^2 - h_i^2} - D$$
- If $X = (d_i)_{i=1,\dots,n}$
$$f(X) = \sum_{i=1}^{n} \frac{\sqrt{{x_i}^2 + h_i^2}}{v_i} \text{ and } C^E(x) = \sum_{i=1}^{n} x_i - D$$

Equality or inequality constraints

$$C^E(x) = \sum_{i=1}^{n} x_i - D = 0 \quad \Leftrightarrow \quad \begin{cases} C_1^I(x) = \sum_{i=1}^{n} x_i - D \leq 0 \\ C_2^I(x) = D - \sum_{i=1}^{n} x_i \leq 0 \end{cases}$$

Special case of absolute values

$$\inf_{x\in\mathbb{R}^n} f(x) \text{ under constraints } |g(x)| \preceq b$$
$$\text{with } g:\mathbb{R}^n \to \mathbb{R}^d \text{ and } b \in \mathbb{R}^d_+$$

$|g_i(x)| \le b_i$, $i = 1,\dots,d$

Define C^I

$\cdot\ |g_i(x)| \le b_i \iff -b_i \le g_i(x) \le b_i$

$\iff \begin{cases} g_i(x) - b_i \le 0 \\ \text{and} \\ -b_i - g_i(x) \le 0 \end{cases}$

$$C^I : \mathbb{R}^n \longrightarrow \mathbb{R}^{2d}$$
$$x \longmapsto C^I(x) = \begin{pmatrix} g(x) - b \\ \\ -b - g(x) \end{pmatrix}$$

$g(x) - b$ ← bloc of d components

$-b - g(x)$ ← bloc of d components

Example 6: Epidemy model

give 6 values for P
solve the ODE problem
for [0, 2.5]

SIRC model

- $S(t)$, proportion of *susceptibles* persons
- $I(t)$, proportion of *infected* persons
- $R(t)$, proportion of *recovered* persons
- $C(t)$, proportion of *cross immuned* persons

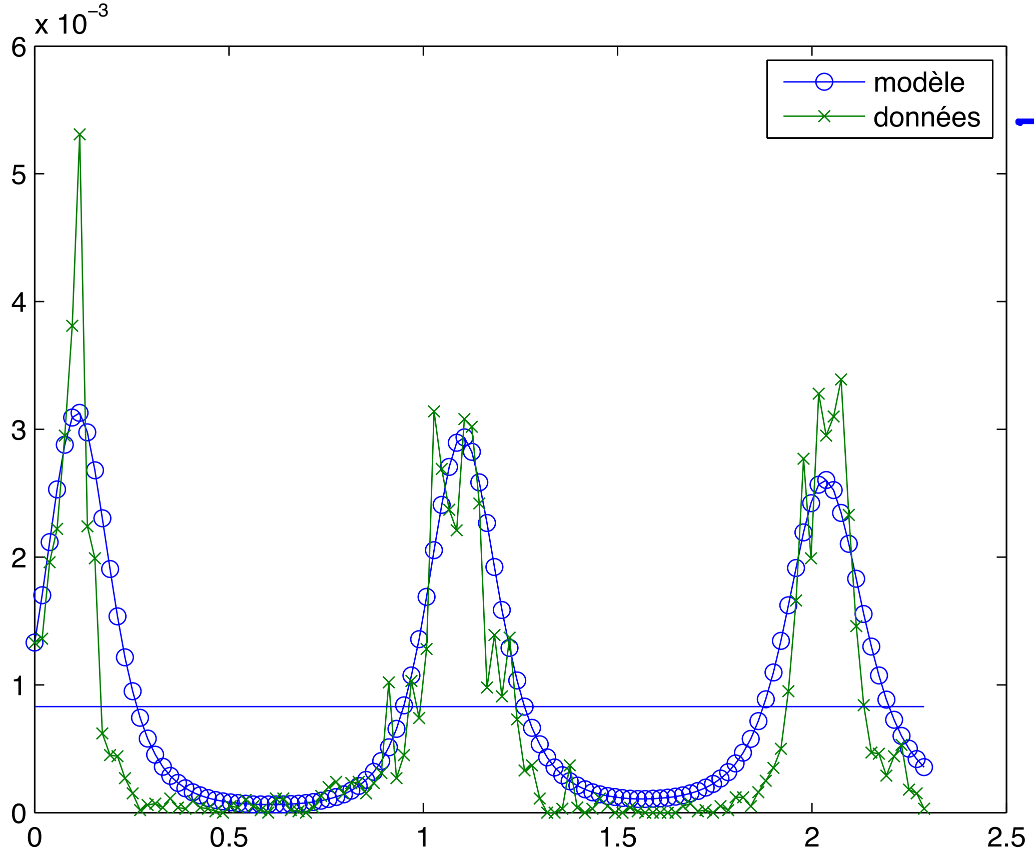
$$\frac{dS(t)}{dt} = \begin{cases} \dot{S}(t) & = & \mu(1-S) - \beta SI + \gamma C, \\ \dot{I}(t) & = & \beta SI + \sigma\beta CI - (\mu+\alpha)I, \\ \dot{R}(t) & = & (1-\sigma)\beta CI + \alpha I - (\mu+\delta)R, \\ \dot{C}(t) & = & \delta R - \beta CI - (\mu+\gamma)C, \end{cases} \tag{1}$$

Parameters $P = (\mu, \alpha, \beta, \gamma, \delta, \sigma)$ $(M = 6)$

Adequation of model with data

Data $(\tilde{I}_j)_{j=1,\dots,d}$ to be compared with values predicted by the model $(I(t_j))_{j=1,\dots,d}$ green points

Proportion of flu in Paris region between Jan 2007 and April 2009 (source : "Reseau Sentinelle")



→ nb of people who have the flu
~ I(t)

Rewrite Epidemic problem in canonical form

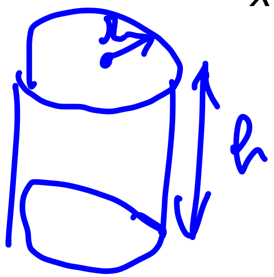
$$\inf_{x \in \mathbb{R}^n} f(x) \text{ under constraints } \begin{cases} C^E(x) = 0, \\ C^I(x) \preceq 0 \end{cases}$$

- f
- C^E
- C^I
- D_a, def, nature

Exercise

A cylindrical container should hold $20\pi m^3$. The price of the material constituting the bottom and the cover is 10 euros $/m^2$, that of the material constituting the sides is 8 euros $/m^2$. Write the optimisation problem to find the dimensions (radius r and height h) of the most economical container.

$$\inf_{x \in \mathbb{R}^n} f(x) \text{ under constraints } \begin{cases} C^E(x) = 0, \\ C^I(x) \preceq 0 \end{cases}$$



$S_c = \pi r^2$ → top and bottom

$S_{side} = 2\pi r h$

Constraint : volume $= 20\pi m^3 = \pi r^2 h$

- $f(r, h) = 10 \times S_{b+t} + 8 \times S_{side} = 20\pi r^2 + 16\pi r h$
- $C^E(r, h) = r^2 h - 20$
- $C^I : \mathbb{R}^2 \to \mathbb{R}^2$ $\quad r, h \geq 0 \quad C^I(r, h) = \begin{pmatrix} -r \\ -h \end{pmatrix}$
- D_a, def, nature

Outline

1st order differentiability

$o(\|h\|)$ little o of $\|h\|$

$O(\|h\|)$

a function $f(h)$ is a $o(\|h\|)$ if $\frac{f(h)}{\|h\|} \xrightarrow[h \to 0]{} 0$

a $O(\|h\|)$ if $\exists C > 0$ s.t. $f(h) \leq C\|h\|$ for h small enough

Definition : Let E and F be two normed vector spaces. Let f be an application of E in F. We say that f is differentiable in the sense of Fréchet at x if there exists a continuous linear map L from E into F such that for all $h \in E$

$$f(x+h) = f(x) + L(h) + o(\|h\|),$$

and we note $Df(x) = Df_x = L$, the differential of f at the point x.

$$f(x+h) = f(x) + Df(x)(h) + o(\|h\|)$$

$$Df_x(h)$$

Directional derivatives

Definition : We say that f is differentiable in the sense of Gâteaux at x if for all $h \in E$, the function $g(t) = f(x + th)$ is differentiable. We denote by $Df(x)$ the differential map of f in x which applies to $h \in E$

$$Df(x)h = \frac{df(x + th)}{dt}_{|t=0}.$$

and $Df(x)h$ is the directional derivative at x according to the vector h.

Property : If a function is differentiable (in the sense of Fréchet) then its differential in the sense of Gâteaux exists (the converse not always being true).

Partial derivatives

$\mathbb{R}^n$ $e_i = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ i^{th} row

E finite dimension
$(e_i)_{i=1,\dots,n}$ basis of E
$x = \sum_{i=1}^{n} x_i e_i$
Partial derivative

$$\frac{\partial f(x)}{\partial x_i} = Df(x)e_i$$

Examples

1. $f : \mathbb{R} \to \mathbb{R}$, differentiable on $\mathbb{R}$
2. $L : E \to F$, linear with E and F nvs
3. $A : E \to F$, affine: $A(x) = L(x) + b$ with linear L and $b \in F$
4. $f(X) = \|X\|_2^2$, with $X \in \mathbb{R}^n$
5. $f : \mathbb{R}^2 \to \mathbb{R} : (x_1, x_2) \mapsto \begin{cases} 0 & \text{if } (x_1, x_2) = (0, 0) \\ \frac{x_1 x_2}{x_1^2 + x_2^2} & \text{else} \end{cases}$
6. $f : \mathbb{R}^2 \to \mathbb{R} : (x_1, x_2) \mapsto \begin{cases} 0 & \text{if } (x_1, x_2) = (0, 0) \\ \frac{(x_2^2 - x_1)^2}{x_1^2 + x_2^4} & \text{else} \end{cases}$
7. $f : \mathbb{R}^2 \to \mathbb{R} : (x_1, x_2) \mapsto \begin{cases} 0 & \text{if } (x_1, x_2) = (0, 0) \\ \frac{x_1^2 x_2}{x_1^2 + x_2^2} & \text{else} \end{cases}$

$f : \mathbb{R} \to \mathbb{R}$, differentiable on $\mathbb{R}$

link between $df(x)$ and $f'(x)$

$$f(x+h) = f(x) + h f'(x) + o(|h|)$$

$$f(x+h) = f(x) + Df(x)(h) + o(|h|)$$

$Df(x) : \mathbb{R} \to \mathbb{R}$ linear function

$$h \mapsto Df(x)(h) = f'(x) h$$