CIMPA Course 0 :

Introduction to (continuous) optimization

Part 1 : theoretical background (monday, 14.30 → 16:00 local time) (−6h in France)

Includes { main definitions and vocabulary
optimality conditions

Part 2 : numerical backgroung (tuesday, 14.00 → 15.30)

Includes { main descent algorithms
examples of numerical implementation

PART 1 : Theoretical background

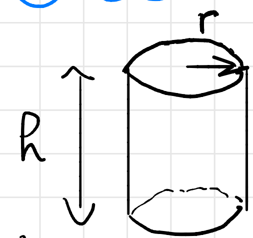
1.1. Some examples of optimization problems

Example 1 : Minimize the function

$$f : \begin{pmatrix} \mathbb{R}^2 \longrightarrow \mathbb{R} \\ (x,y) \longmapsto 100(y-x^2)^2 + (x-1)^2 \end{pmatrix}$$

(Rosenbrock function)

* Example 2 : The can problem



* Objective : minimize the can surface with a fixed volume
* Variables : h and r (in cm)
* Function : $f(r,h) = 2\pi r^2 + 2\pi r h$
* Constraints : $\pi r^2 h \geq V_0$

and $\begin{cases} r \geq r_0 > 0 \\ h \geq h_0 > 0 \end{cases}$

1.2. Definitions

1.2.1. Optimization problem

$$\min_{x \in \Omega} f(x) \quad (P_0) \text{ where :}$$

(number of variables)

$$f : \begin{pmatrix} \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R} \\ x \longmapsto f(x) \end{pmatrix} \quad \text{and}$$

$$\Omega = \left\{ x \in \mathbb{R}^n, \ c_E(x) = 0 \text{ and } c_I(x) \leq 0 \right\}$$

(f C^0, C^1, or C^2)

$$\begin{cases} c_E : \mathbb{R}^n \longrightarrow \mathbb{R}^p & \text{(equality constraints)} \\ c_I : \mathbb{R}^n \longrightarrow \mathbb{R}^q & \text{(inequality constraints)} \end{cases}$$

Some questions :

→ Existence (uniqueness) of solutions of (P_0) (local or global)

→ Approximation of solutions of (P_0) ⊕ convergence proof.

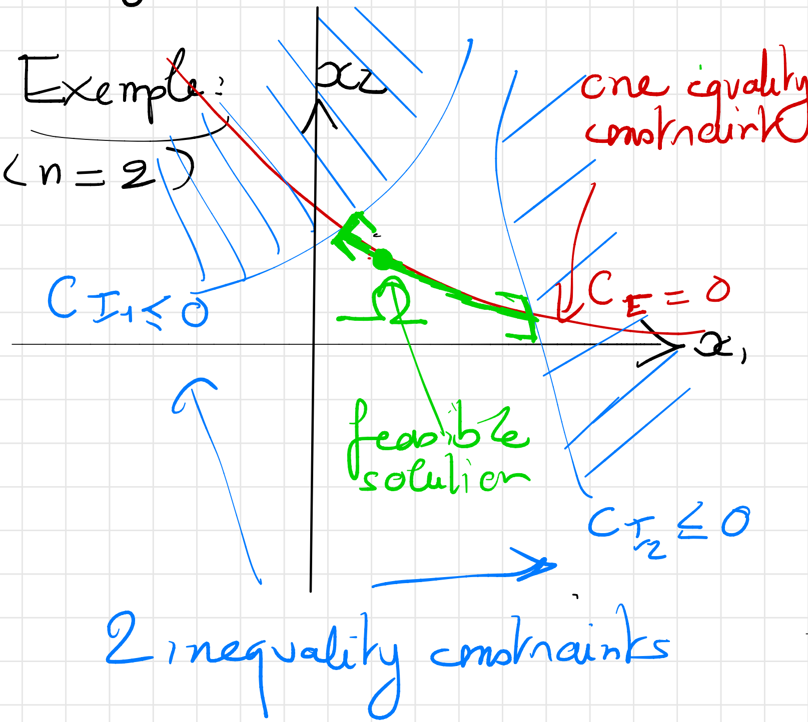
→ Robustness of the solution

→ $\left(\max_{x \in \Omega} f(x)\right) \equiv \min_{x \in \Omega}(-f(x))$

1.2.2 Feasible set and solutions

* Def : $x \in \mathbb{R}^n$ is a feasible solution of problem (P_0) if $x \in \Omega$.

Exemple : $(n = 2)$

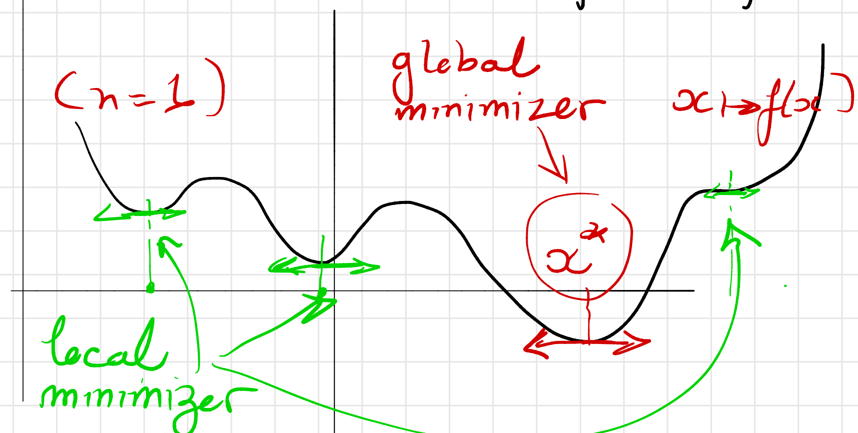


* $x^* \in \Omega$ is a global minimizer of f on Ω if :

$$\forall x \in \Omega, \quad f(x^*) \leq f(x)$$

* $x^* \in \Omega$ is a local minimizer of f on Ω if there exists $\varepsilon > 0$ such that :

$$\forall x \in \Omega \cap B(x^*, \varepsilon), \quad f(x^*) \leq f(x)$$



1.2.3 Differentiability

* $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $x \in \mathbb{R}^n$ if there exists a linear form $D(x) : \mathbb{R}^n \to \mathbb{R}$ such that:

$$\forall h \in \mathbb{R}^n,\ f(x+h) = \underbrace{f(x) + Df(x)(h)}_{\text{linear approximation of } f \text{ at } x} + o(\|h\|)$$

* Gradient: by definition:

$$Df(x)(h) = \langle \underbrace{\nabla f(x)}_{\text{gradient of } f\ (\in \mathbb{R}^n)}, h \rangle$$

* Partial derivatives:

$$\frac{\partial f}{\partial x_i}(x) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon h_i) - f(x)}{\varepsilon}$$

where $h_i = (0, \dots, 1, 0, \dots 0)$

* f is C^1 if and only if f has continuous partial derivatives. In this case:

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

5

* f is twice differentiable at $x \in \mathbb{R}^n$ if there exists $Hf(x) : \mathbb{R}^n \to \mathbb{R}^n$, linear, such that $\forall h \in \mathbb{R}^n$,

$$f(x+h) = \underbrace{f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2}\langle h, Hf(x).h \rangle}_{\text{quadratic part of } f \text{ at } x} + o(\|h\|^2)$$

* $Hf(x)$ is the Hessian of f at x.

When f is C^2, then

$$Hf(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \ddots & \\ \vdots & & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} \quad (\in S_n(\mathbb{R}))$$

(Hessian matrix of f at x)

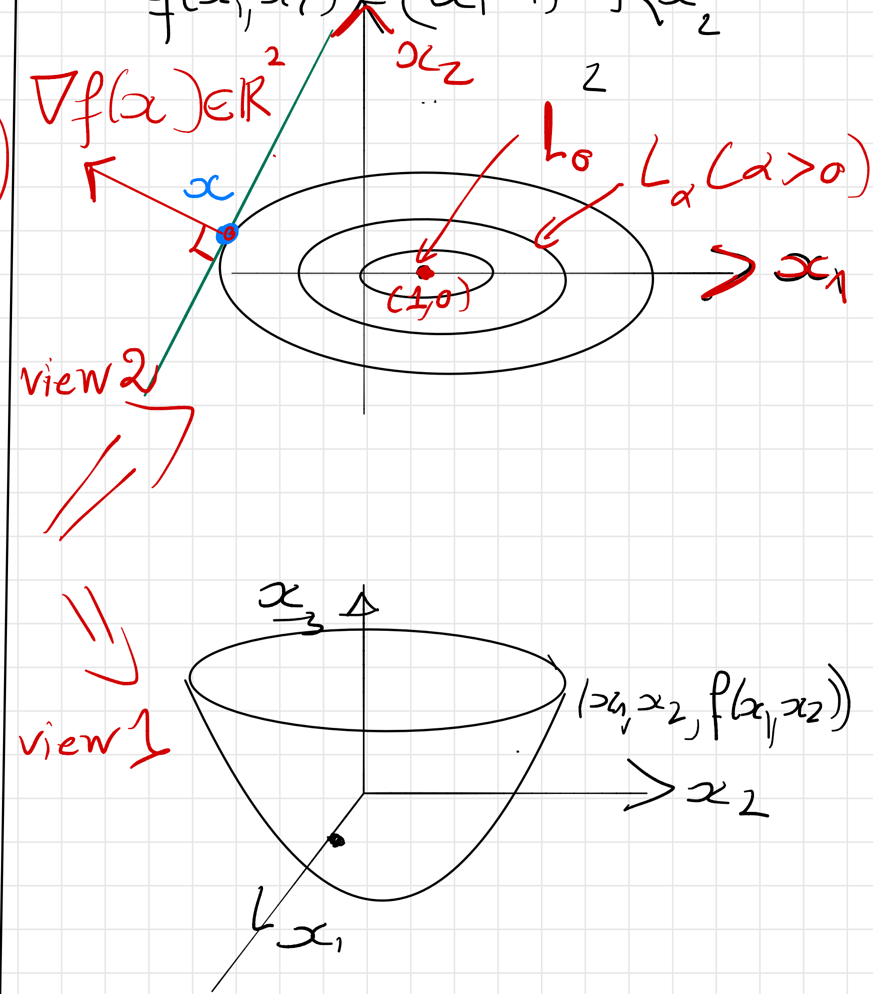
1.2.4. Level lines

$$L_\ell = \{x \in \mathbb{R}^n \,/\, f(x) = \ell\} \quad (\subset \mathbb{R}^n)$$

$(\ell \in \mathbb{R})$

* Particular case : $n = 2$

$$f(x_1, x_2) = (x_1 - 1)^2 + 2x_2^2$$



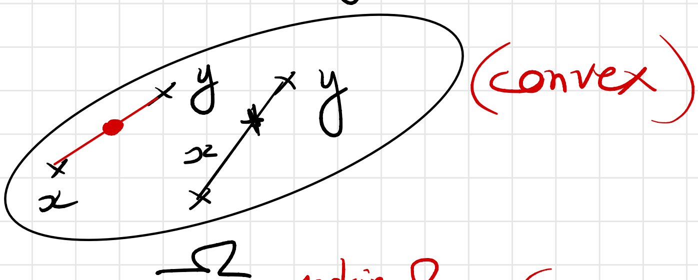
Remark : at $x \in \mathbb{R}^n$, if f is C^1, $\nabla f(x)$ is orthogonal at x, to the level line $L_{f(x)}$.

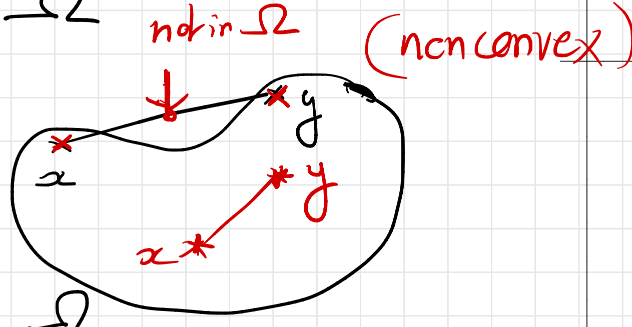
(exercice)

1.2.5. Convexity

Def: $\Omega \subset \mathbb{R}^n$ is a convex set if:

$$\forall (x,y)\in\Omega^2,\ \forall \lambda\in[0,1],\quad \lambda x+(1-\lambda)y\in\Omega$$



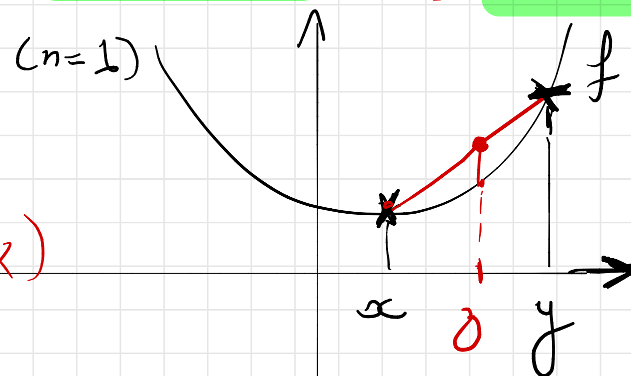


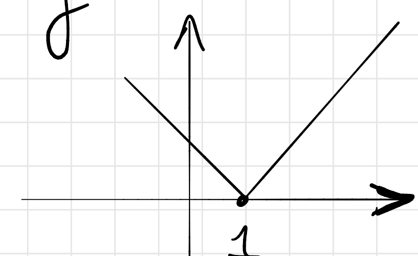
Def. Let Ω a convex set.

$f:\Omega\to\mathbb{R}$ is convex (resp. strictly) if

$\forall (x,y)\in\Omega,\ \forall\lambda\in[0,1]$, $(\forall\lambda\in]0,1[)$

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y) \quad (<)$$





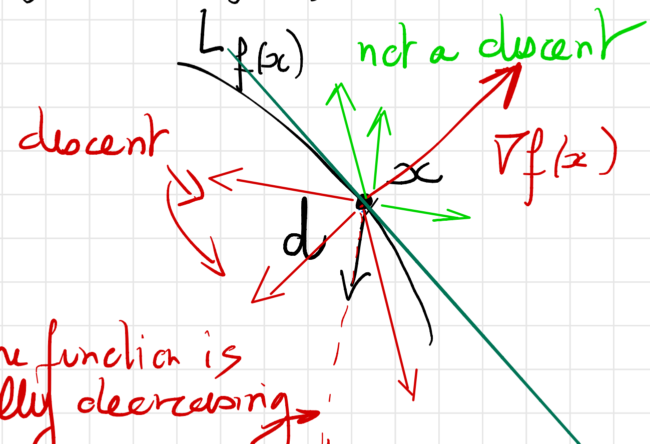
* If f is differentiable,
f is convex $\iff \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n$
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

* If f is twice differentiable,
f convex $\iff \forall x \in \mathbb{R}^n, \forall h \in \mathbb{R}^n$
$$\langle x, H(f)(x) \rangle \geq 0$$

(Hessien of f is positive, semi-definite matrix)

1.2.6 Descent direction

Def: $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$
$d \in \mathbb{R}^n \setminus \{0\}$ is a descent direction of f at x if:
$$\exists \varepsilon > 0, \forall s \in]0, \varepsilon], f(x + sd) < f(x)$$



The function is locally decreasing at x, in direction d

Proposition: if f is C^1 and d is a descent direction of f at $x \in \mathbb{R}^n$, then

$$\langle d, \nabla f(x) \rangle \leq 0$$

Proof: Taylor approximation of order 1 at x (and at direction d)

$$f(x+sd) = f(x) + \langle \nabla f(x), sd \rangle + \underset{(s \to 0)}{o(s)}$$

$$\Rightarrow \underbrace{\frac{f(x+sd) - f(x)}{s}}_{<0 \text{ when } s>0 \text{ is small}} = \langle \nabla f(x), d \rangle + \underset{s \to 0}{o(1)}$$

by passing to the limit when $s \to 0^+$, $\langle \nabla f(x), d \rangle \leq 0$

* Remark: if f is C^1, $x \in \mathbb{R}^n$ and $\nabla f(x) \neq 0$, then $d = -\nabla f(x)$ is a descent direction of f at x.

$$\left(\langle d, \nabla f(x) \rangle = \underbrace{-\|\nabla f(x)\|^2}_{<0} \right)$$

1.3 Optimality conditions

1.3.1. The unconstrained case

Theorem 1: Let $f \in C^1(\mathbb{R}^n, \mathbb{R})$ and x^* a local minimizer of f.
Then:

$$\boxed{\nabla f(x^*) = 0}$$

Proof: Taylor approximation of order 1:

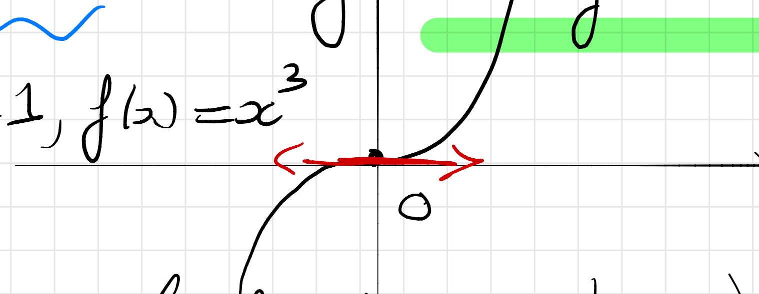
$$f(x^* + \underbrace{\varepsilon}_{>0} \underbrace{h}_{\in \mathbb{R}^n}) = f(x) + \langle \nabla f(x^*), \varepsilon h \rangle + \underset{\varepsilon \to 0}{o(\varepsilon)}$$

$$\Rightarrow \langle \nabla f(x^*), h \rangle = \underbrace{\frac{f(x^* + \varepsilon h) - f(x^*)}{\varepsilon}}_{\geq 0 \text{ when } \varepsilon \text{ is small}} + \underset{\varepsilon \to 0}{o(1)}$$

By passing to the limit $\varepsilon \to 0^+$, $\langle \nabla f(x^*), h \rangle \geq 0$. By repeating with $(-h)$ instead of h, $\langle \nabla f(x^*), -h \rangle \geq 0 \Rightarrow$ $\langle \nabla f(x^*), h \rangle = 0 \ \forall h \in \mathbb{R}^n$ and $\nabla f(x^*) = 0$ //

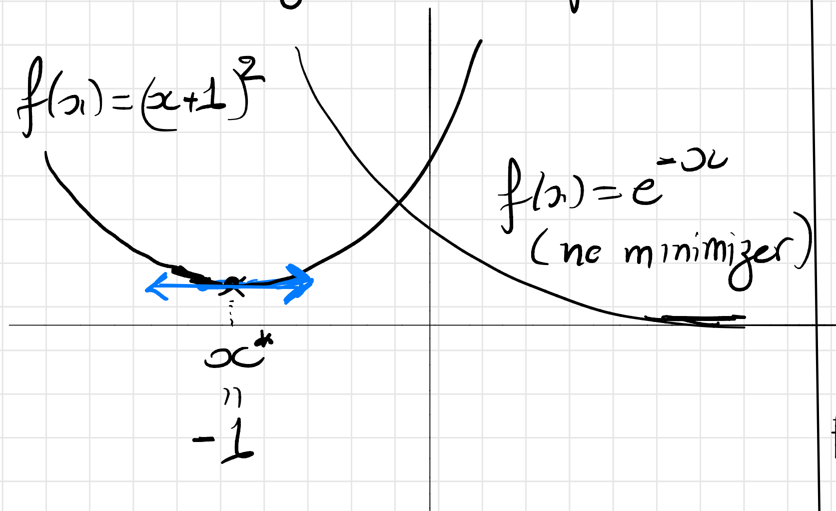
Remark: it is only a necessary condition:

$n = 1$, $f(x) = x^3$



(other example: $f(x,y) = xy$ at $(0,0)$)

The necessary condition becomes sufficient in the case of a convex function.



Theorem 2: Let f C^2: $\mathbb{R}^n \to \mathbb{R}$ and x^* a local minimizer of f.

Then: $\nabla f(x^*) = 0$ and

$H(f)(x)$ is positive semi-definite:

$$\forall h \in \mathbb{R}^n,\ \langle \underbrace{H(f)(x)}_{\text{Hessian matrix}}.h, h \rangle \geq 0$$

Proof: Taylor, order 2:

$$f(x^* + \underset{>0}{t}\underset{\in \mathbb{R}^n}{d}) = f(x^*) + \cancel{\langle \nabla f(x^*), td \rangle} + \frac{1}{2} t^2 \underbrace{\langle d, Hf(x^*).d \rangle}_{\geq 0} + o(t^2) \quad t \to 0$$

Remark: If f is C^2 and convex, the necessary condition becomes sufficient.

Theorem 3: Let f $C^2 : \mathbb{R}^n \to \mathbb{R}$ and $x^* \in \mathbb{R}^n$ such that:

→ $\nabla f(x^*) = 0$

→ $H(f)(x^*)$ is positive definite (*)

Then, x^* is a local minimizer of f. (sufficient condition)

Proof: (exercise) Let $d \in (\mathbb{R}^n)^*$.

$$f(x^*+td) = f(x^*) + t\langle \cancel{\nabla f(x^*)}, d\rangle + \frac{t^2}{2}\underbrace{\langle d, H(f)(x^*)d\rangle}_{>0} + o(t^2)_{t\to 0}$$

$$\Rightarrow \frac{f(x^*+td) - f(x^*)}{t^2} > 0 \quad (t \text{ small})$$

$$f(x^*+td) > f(x^*) \quad (t \text{ small})$$

(*) $\forall h \in \mathbb{R}^n\setminus\{0\}, \langle H(f)(x^*).h, h\rangle > 0$

Practically, to seek local minima of f $C^2 : \mathbb{R}^n \to \mathbb{R}$:

① → first, search for critical points. ($\nabla f(x) = 0$)

② → then, for any critical point:

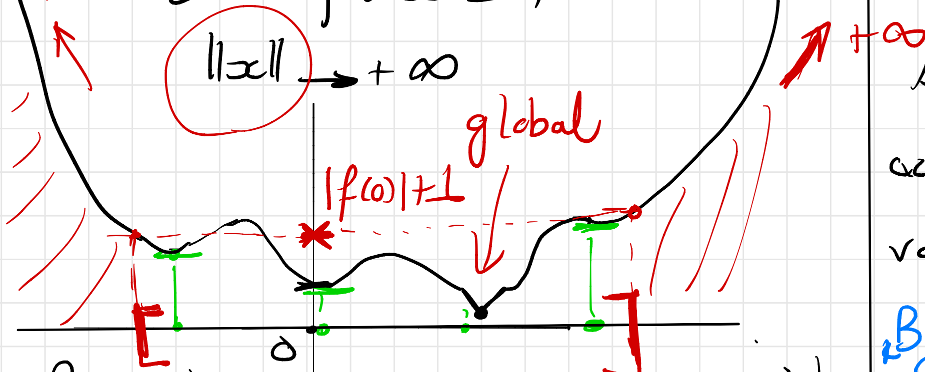
- if the Hessian is $\gg 0$ (positive definite): OK
- if the Hessian is not ≥ 0: N.OK
- else: local study of f.

* if f is convex, point ① is enough

* if f is only C^0, coercivity can help ...

Déf : $f : \mathbb{R}^n \to \mathbb{R}$ is coercive if

$$\lim_{\|x\| \to +\infty} f(x) = +\infty$$



Proposition : if f is continuous and coercive, then it has at least a global minima on $\mathbb{R}^n$.

Proof : (exercice)

Indication : restrict to a compact set (where a continuous function achieves its minimal and maximal values).

- By coercivity, $\exists M > 0$ such that $\forall x \in \mathbb{R}^n$, $\|x\| \geq M \Rightarrow f(x) \geq |f(0)| + 1$.
- On $B(0, M)$, f is continuous and achieves (at least) a local minimizer at x^*.
- x^* is a global minimizer on $\mathbb{R}^n$:

$\to \forall x \in \mathbb{R}^n \setminus B(0, M)$,

$$f(x^*) \leq f(0) \leq |f(0)| + 1 \leq f(x)$$

Some examples (or counter examples)

* $f(x,y) = 100(y-x^2)^2 + (x-1)^2$

(exercise: critical points? local? global?)

* critical points:

$\nabla f(x,y) = 0 \iff \begin{cases} -400x(y-x^2) + 2(x-1) = 0 \\ 200(y-x^2) = 0 \end{cases}$

$\iff \begin{cases} x = 1 \\ y = 1 \end{cases}$. Only one critical point

* At $(x,y) = (1,1)$

$Hf(x,y) = \begin{pmatrix} 802 & 400 \\ -400 & 200 \end{pmatrix} \gg 0$

$\Rightarrow (1,1)$ is a (local) minima

* Also, $f(x,y) \geq 0$ and $= 0$ iff $x = y = 1$

* $f(x_1, x_2) = x_1^2 - x_2^3$

* critical points:

* $\nabla f(x_1, x_2) = 0 \iff \begin{cases} 2x_1 = 0 \\ 3x_2^2 = 0 \end{cases} \iff x_1 = x_2 = 0$

* At $(0,0)$: $Hf(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \geq 0$

* It is not a minima as:

$f(0,+t) = -t^3 < 0 = f(0,0)$

$f(t,0) = t^2 > 0 = f(0,0)$

* $f(x_1, x_2) = x_1^4 + x_2^4$

* critical points

$\nabla f(x_1, x_2) = 0 \Rightarrow \begin{cases} 4x_1^3 = 0 \\ 4x_2^3 = 0 \end{cases} \iff x_1 = x_2 = 0$

* At $(0,0)$: $Hf(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \geq 0$

* It is a global minima as $f(x_1, x_2) \geq 0 = f(0,0)$

1.3.2 The constrained case

$f : \Omega \to \mathbb{R}$ where :

$$\Omega = \{ x \in \mathbb{R}^n \,/\, c_I(x) \leq 0,\ c_E(x) = 0 \}$$

q conditions

p conditions

(qualification constraint)

Definition : the Lagrangian of f is defined as :

$$\mathcal{L}(\underbrace{x, \lambda, \mu}) = f(x) + \underbrace{\langle \lambda, c_E(x) \rangle}_{\sum_{i=1}^{p} \lambda_i c_{E_i}(x)} + \langle \mu, c_I(x) \rangle$$

$\lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^q$

for any $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^q$

Theorem 4 : Let $f : \Omega \to \mathbb{R}$ and $x^* \in \Omega$. Assume that $\{ \nabla c_E(x^*), \nabla c_I(x^*) \}$ is a set of independant vectors of $\mathbb{R}^n$ and that x^* is a local minimizer of f on Ω. Then, there exists $\lambda^* \in \mathbb{R}^p$ and $\mu^* \in \mathbb{R}^q$ such that :

$$\begin{cases} * \ \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \left(= \nabla_x f(x) + \langle \lambda, \nabla c_E \rangle + \langle \mu, \nabla c_I \rangle \right) \\ * \ c_E(x^*) = 0 \\ * \ c_I(x^*) \leq 0 \\ * \ \mu^* \geq 0 \quad (\mu_i^* \geq 0 \ \forall i \in \{1, \dots, q\}) \\ * \ \mu^* \odot c_I(x^*) = 0 \ (\text{if } c_I(x^*) < 0, \text{ then } \mu^* = 0) \end{cases}$$

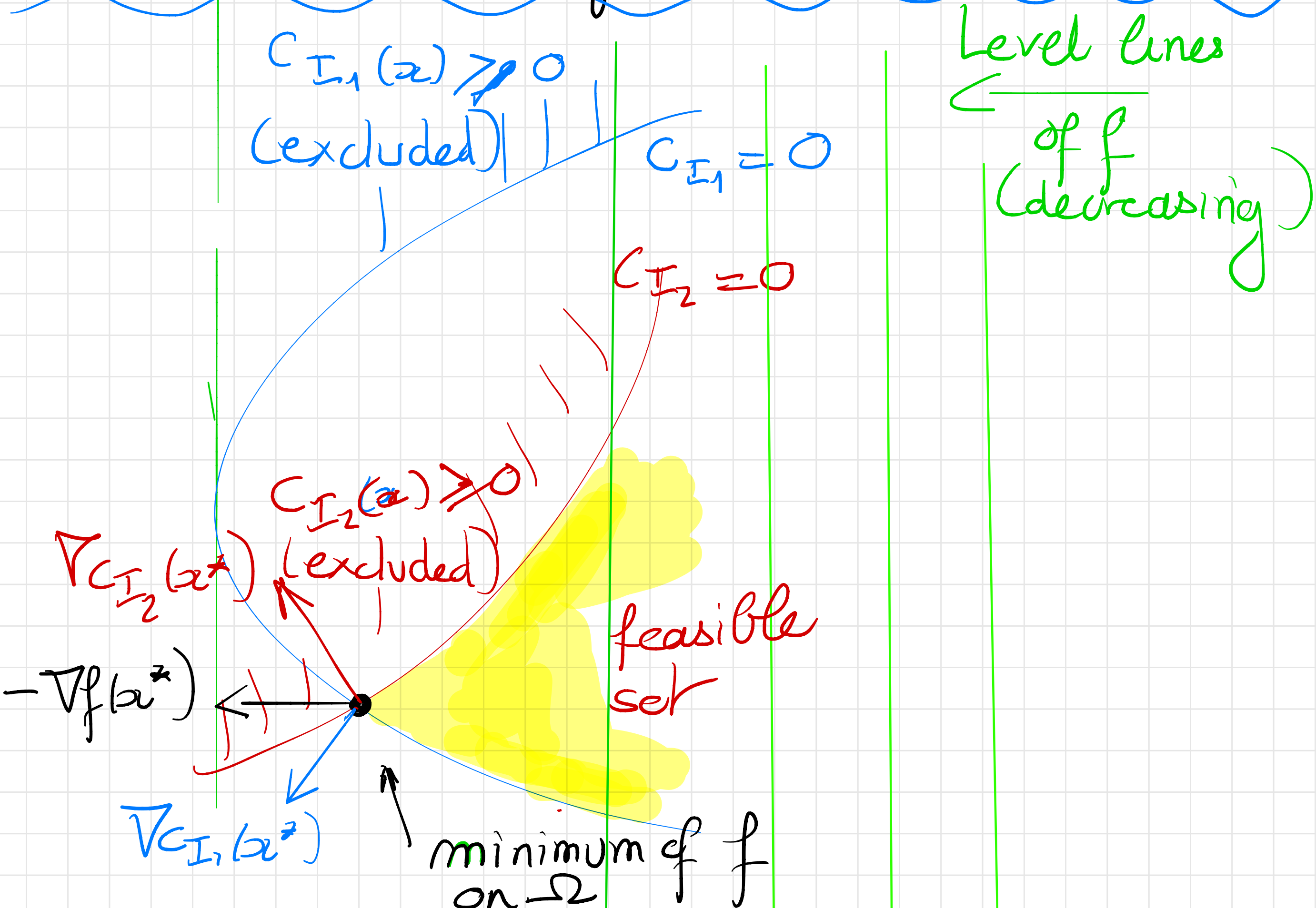
feasibility

complementary slackness

The system is called the KKT necessary condition of order 1.

λ^* and μ^* are called the Lagrange multipliers.

Graphical interpretation of the KKT condition ($n=2, p=0, q=2$)



Some exercices : Write and solve the KKT system

→ Minimize the can function

$f(r,h) = 2\pi r^2 + 2\pi r h$

on $\Omega = \{(r,h) \in (\mathbb{R}_+)^2 \ / \ \pi r^2 h \geq V_0\}$

* $L(r,h,\lambda) = 2\pi r^2 + 2\pi r h + \lambda(V_0 - \pi r^2 h)$

KKT:

$$\begin{cases} 4\pi r + 2\pi h - 2\pi \lambda r h = 0 \\ 2\pi r - \lambda \pi r^2 = 0 \\ \lambda \geq 0, \ (\pi r^2 h - V_0)\lambda = 0 \\ r > 0 \quad h > 0 \end{cases}$$

Eq 2 $\rightsquigarrow$ $\lambda r = 2$. If $\lambda > 0$,

$\pi r^2 h = V_0$ and $4\pi r = 2\pi h$ (Eq. 1) $\Rightarrow$

$4\pi r^3 = 2V_0$, $r = \sqrt[3]{\frac{V_0}{2\pi}}$, $h = \frac{V_0}{\pi r^2}$; $\lambda = 2/r$

(one solution of KKT)

Moreover, f is coercive:

$\lim\limits_{\substack{\|(r,h)\| \to \infty \\ (r,h) \in \Omega}} f(r,h) = +\infty$

As Ω is closed subset of $\mathbb{R}^n$, and f is coercive, it has a global minima on Ω. As this minima satisfies the KKT system, it is the previous solution.

* Also, write $h(r) = \frac{V_0}{\pi r^2}$ and study $r \to f(r, h(r))$

→ Minimize
$$g : \begin{pmatrix} \Omega_i \to \mathbb{R} \\ (x,y) \mapsto x^2+y^2+xy \end{pmatrix}$$

with $\Omega_1 = \{(x,y) \in \mathbb{R}^2 / x+y=3\}$

and $\Omega_2 = \{(x,y) \in \mathbb{R}^2 / x^2+y^2 \leq 4 \text{ and } x+y \geq 2\}$

* KKT on Ω_1 :

$\mathcal{L}(x,y,\lambda) = x^2+y^2+xy+\lambda(x+y-3) \Rightarrow$

$$\begin{cases} 2x+y+\lambda=0 \\ 2y+x+\lambda=0 \\ x+y=3 \end{cases} \iff \begin{cases} 9+2\lambda=0 \\ 2x+y=9/2 \\ x+y=3 \end{cases}$$

$$\iff \begin{cases} \lambda = -2/9 \\ x = 3/2 \\ y = 3/2 \end{cases} \quad \left(\tfrac{3}{2}, \tfrac{3}{2}, -\tfrac{2}{9}\right) \text{ unique solution}$$

* f is convex : $Hf(x,y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \gg 0$

on Ω convex $\Rightarrow \left(\tfrac{3}{2}, \tfrac{3}{2}\right)$ is a global minima

* KKT on Ω_2 :

$\mathcal{L}(x,y,\mu_1,\mu_2) = x^2+y^2+xy+\mu_1(x^2+y^2-4)+\mu_2(2-x-y)$

$$\begin{cases} 2x+y+2\mu_1 x - \mu_2 = 0 \\ 2y+x+2\mu_1 y - \mu_2 = 0 \\ \mu_1, \mu_2 \geq 0 ;\ x^2+y^2 \leq 4 \text{ and } x+y \geq 2 \\ \mu_1(x^2+y^2-4)=0 \text{ and } \mu_2(x+y-2)=0 \end{cases}$$

* Case 1 : $\mu_1 = 0, \mu_2 > 0$: KKT becomes :

$$\begin{cases} 2x+y-\mu_2=0 \\ 2y+x-\mu_2=0 \\ x+y=2 \end{cases} \Rightarrow \begin{cases} \mu_2 = 3 \\ 2x+y=3 \\ 2y+x=3 \end{cases} \Rightarrow \begin{cases} \mu_2=3 \\ y=1 \\ x=1 \end{cases}$$

* Case 2 : $\mu_1 > 0, \mu_2 = 0$: KKT becomes

$$\begin{cases} 2x+y+2\mu_1 x=0 \\ 2y+x+2\mu_2 y=0 \\ x^2+y^2=4 \end{cases} \Rightarrow \begin{cases} y^2-x^2=0 \\ x^2+y^2=4 \end{cases}$$

If $x=y(=\pm\sqrt{2})$, then $\mu_1=-\frac{3}{2}$: impossible

If $x=-y(=\pm\sqrt{2})$, then $\mu_1=-\frac{1}{2}$: impossible

Case 3: $\mu_1=\mu_2=0$, then $x=y=0$: impossible

Case 4: $\mu_1>0$ and $\mu_2>0$, then

$$\begin{cases} x+y=2 \\ x^2+y^2=4 \end{cases} \Rightarrow \begin{cases} x=2 \\ y=0 \end{cases} \text{or} \begin{cases} x=0 \\ y=2 \end{cases}.$$

In the first case: $\begin{cases} 4+4\mu_1-\mu_2=0 \\ 2-\mu_2=0 \end{cases} \Rightarrow$

$\mu_2=2$, $\mu_1=-\frac{1}{2}$: impossible.

(same in the second case).

As, Ω_2 is compact, there is at least a minimizer of f in Ω_2. It is necessarily the point $(x,y)=(1,1)$.