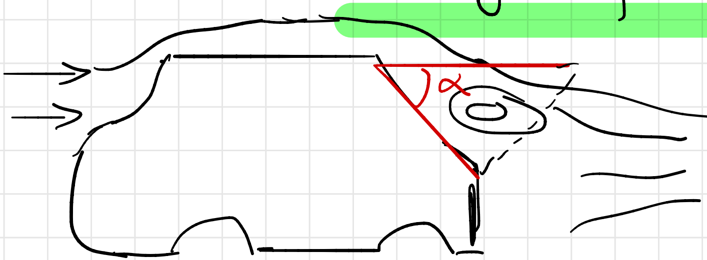
M1 Optimisation numérique

1) Introduction et rappels

1.1) Quelques problèmes d'optimisation

Exemple 1 : optimisation de forme aérodynamique



* Coefficient de traînée aérodynamique : C_x

* on cherche à minimiser C_x (pour réduire la consommation)

* calcul de C_x effectué à l'aide d'une simulation de l'écoulement de l'air autour du véhicule.

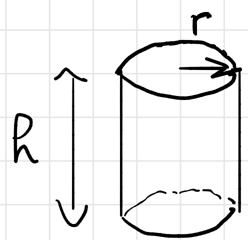
* Le problème d'optimisation consiste à minimiser C_x par rapport à la forme du véhicule (par exemple α angle de lunette)

$$\underbrace{\alpha}_{\text{variable(s)}} \longmapsto \underbrace{C_x(\alpha)}_{\text{fonction à minimiser}}$$

* On rajoute des contraintes aux variables de minimisation

$$\alpha \in [\alpha_{min}, \alpha_{max}]$$

* **Exemple 2** : optimisation de formes de pièces automobiles (capot, ...) vis à vis de critères mécaniques (résistance, ...)
* **Exemple 3** : problème de la canette



On cherche à minimiser la surface de la canette à volume fixé

* Variables : h et r (en cm)
* Fonction : $f(r,h) = 2\pi r^2 + 2\pi r h$
* Contraintes : $\pi r^2 h \geq V_0$ ($330\,cm^3$)

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

→ Exemple utilisé fréquemment ici (solution exacte : voir TD 1).

Les autres exemples d'illustration de ce cas seront des fonctions analytiques :

→ fonctions polynomiales (Rosenbrock)

→ fonctions régulières (Rastrigin, ...)

(voir TP)

1.2) Définitions

1.2.1) Problème d'optimisation

Un tel problème s'exprime :

$$\text{Min } f(x) \text{ où } x \in \Omega \qquad (PO)$$

$x \mapsto f(x)$ fonction de $\Omega \subset \mathbb{R}^n \to \mathbb{R}$ et

$\Omega = \{x \in \mathbb{R}^n,\ C_E(x) = 0$ et $C_I(x) \leq 0\}$ où

$$\begin{cases} C_E : \mathbb{R}^n \to \mathbb{R}^p \text{ (p contraintes égalité)} \\ C_I : \mathbb{R}^n \to \mathbb{R}^q \text{ (q contraintes inégalité)} \end{cases}$$

* $n \geq 1$: nombre de variables.
* f est une fonction quelconque (pas forcément régulière).

Questions :

→ Existence, unicité d'un minimum sur Ω ou sur une partie de Ω ?
(global) (local)

→ Construction d'une approximation de ce minimum.

→ Justification mathématique de la convergence de cette approximation

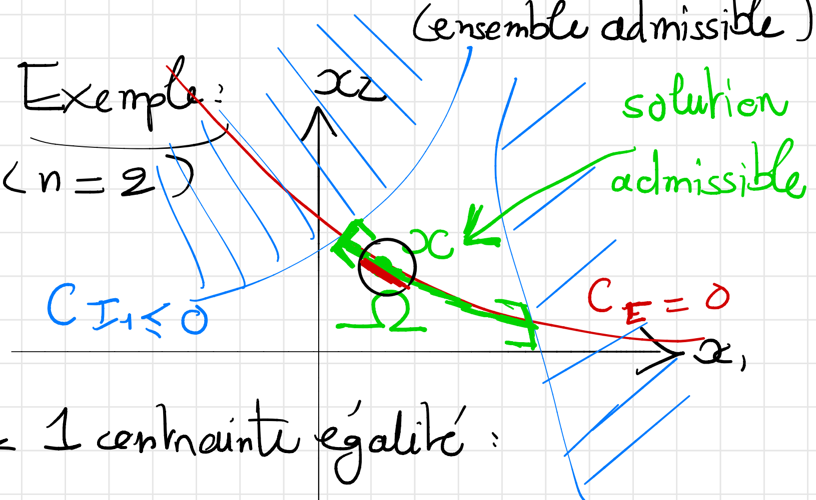
(remarque : pour maximiser une fonction g, on minimise $(-g)$)

1.2.2 Solution d'un problème (PO)

* Def : $x \in \mathbb{R}^n$ est une solution admissible de (PO) si $x \in \Omega$.

(ensemble admissible)

Exemple : ($n=2$)



* 1 contrainte égalité :

$$C_E(x_1, x_2) = 0$$

* 2 contraintes inégalité :

$$\begin{cases} C_{I_1}(x_1, x_2) \leq 0 \\ C_{I_2}(x_1, x_2) \leq 0 \end{cases}$$

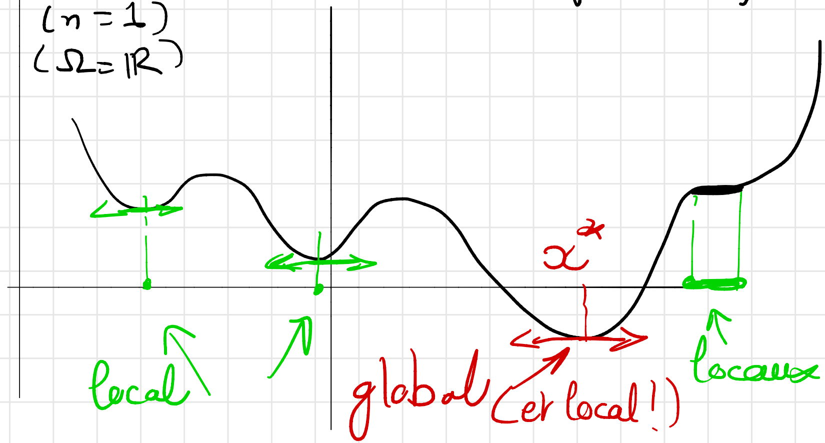
* $x^* \in \Omega$ est une solution globale de (PO) (ou minimum global) si

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

* $x^* \in \Omega$ est une solution locale de (PO) si il existe $\varepsilon > 0$ tq

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

($n=1$)
($\Omega = \mathbb{R}$)



1.2.3) Différentiabilité

* $f : \mathbb{R}^n \to \mathbb{R}$ est différentiable en $x \in \mathbb{R}^n$, s'il existe $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ tq

$$\forall h \in \mathbb{R}^n, \; f(x+h) = f(x) + \underbrace{Df(x)(h)}_{\text{partie linéaire de } f} + o(\|h\|)$$

On a $Df(x)(h) = \langle \underbrace{\nabla f(x)}_{\text{gradient de } f \text{ en } x}, h \rangle$

* f est C^1 si elle est différentiable partout $x \in \mathbb{R}^n$ et si $Df : x \mapsto Df(x)$ est C^0.

* Dérivée partielle :

$$\underset{(1 \leq i \leq n)}{\frac{\partial f}{\partial x_i}(x)} = \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon h_i) - f(x)}{\varepsilon}$$

où $h_i = (0, \dots, \underset{\uparrow}{1}, 0, \dots, 0)$

* f est C^1 si et seulement si elle admet des dérivées partielles et que celles-ci sont continues. Dans ce cas

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

$\star$ f est 2 fois différentiable en $x \in \mathbb{R}^n$ s'il existe $Hf(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ tq $\forall h \in \mathbb{R}^n$

$$f(x+h) = \underbrace{f(x) + \overbrace{\langle \nabla f(x), h \rangle}^{{}^t h . \nabla f(x)} + \frac{1}{2} \overbrace{\langle h, Hf(x).h \rangle}^{{}^t h \, Hf(x) \, h}}_{\text{partie quadratique}} + o(\|h\|^2)$$

$\star$ On appelle $Hf(x)$ la Hessienne de f en x.

Si f est C^2, on a

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

$(\in \mathcal{S}_n(\mathbb{R}))$

1.2.4. Représentation graphique de f en lignes de niveau

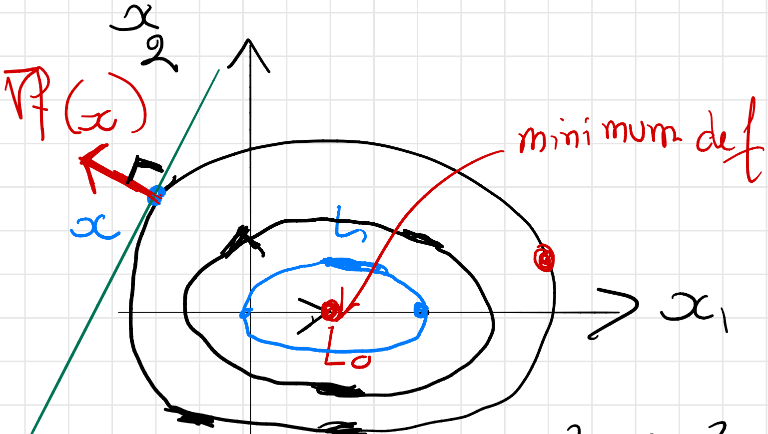
Déf. soit $f : \mathbb{R}^n \to \mathbb{R}$

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

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

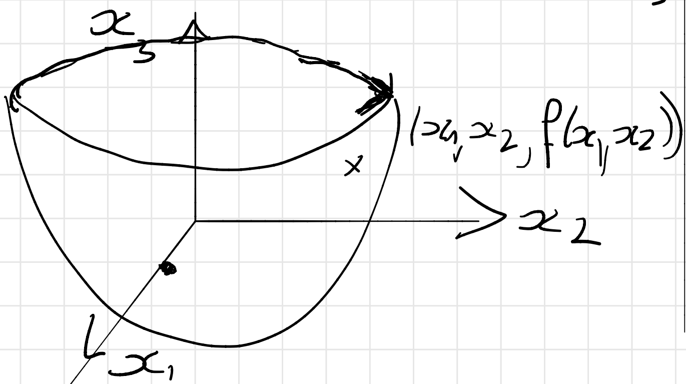
Cas particulier : $n = 2$

$$f : \mathbb{R}^2 \to \mathbb{R}$$



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

$$L_1 = \{(x_1,x_2) \,/\, (x_1-1)^2 + 2x_2^2 = 1\}$$



• Lien entre gradient et ligne de niveau :

En tout point $x \in \mathbb{R}^n$, $\nabla f(x)$ est orthogonal à la ligne de niveau en x et dirigé vers les valeurs supérieures par f.

En effet si $x+h \in L_{f(x)}$

$$\cancel{f(x+h)} = \cancel{f(x)} + \langle \nabla f(x), h \rangle + o(\|h\|)$$

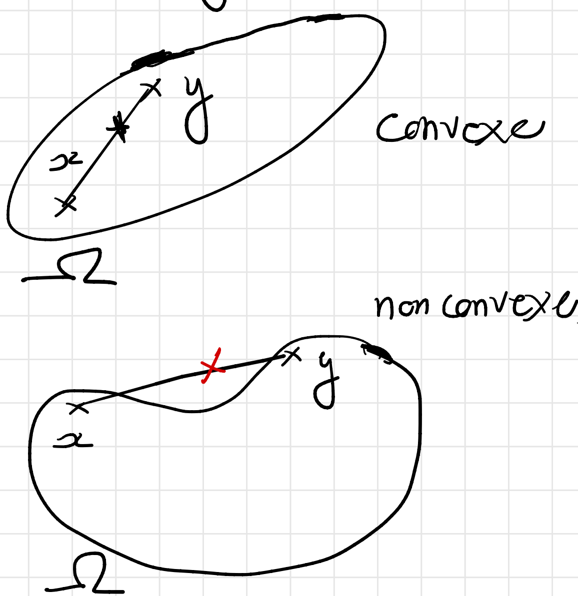
Par passage à la limite

$$\left\langle \nabla f(x), \frac{h}{\|h\|} \right\rangle \to 0 \text{ lorsque } h \to 0$$

1.2.5) Convexité

Def : $\Omega \subset \mathbb{R}^n$ est convexe ssi

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



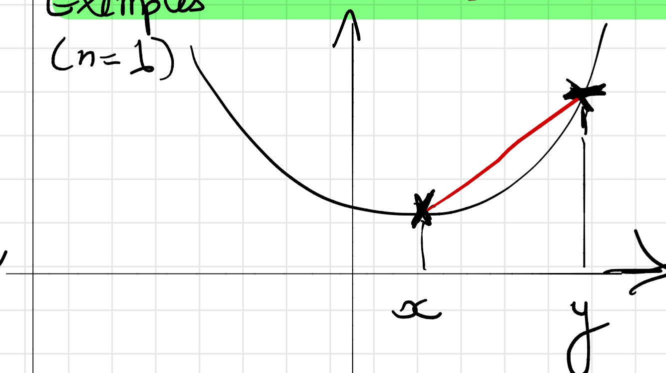
Def : Soit Ω convexe.

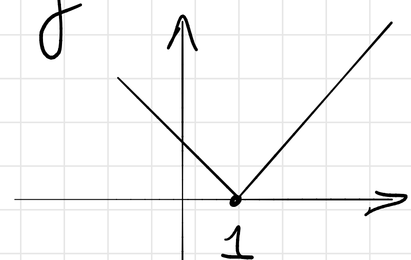
$f : \Omega \to \mathbb{R}$ est convexe (resp. strictement) ssi

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

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

Exemples $(n=1)$





* Si f est différentiable,

f convexe $\iff \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

(f au dessus de son hyperplan tangent)

* Si f est 2 fois différentiable,

f convexe $\iff \forall x \in \mathbb{R}^n, \forall h \in \mathbb{R}^n$

$$^t h \, Hf(x) h \geq 0$$

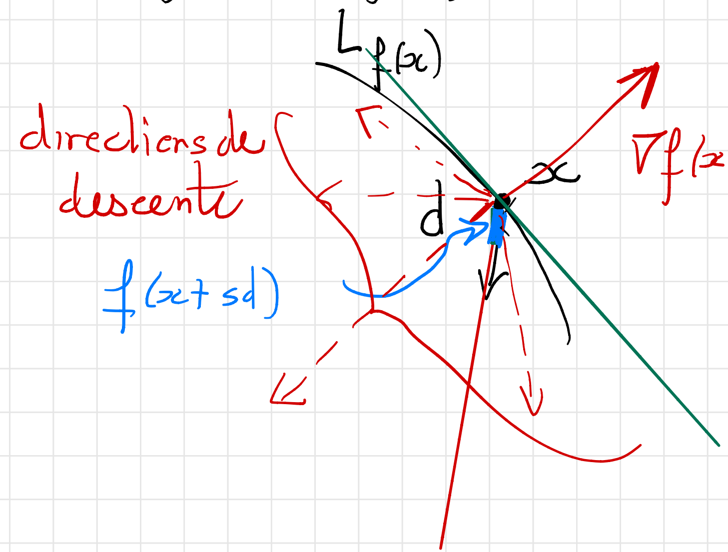
(matrice Hessienne positive)

(inégalités strictes par la stricte convexité).

1.2.6.) Direction de descente

Def : $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$. On dit que $d \in \mathbb{R}^n \setminus \{0\}$ est une direction de descente de f en x si

$$\exists \varepsilon > 0, \forall s \in]0, \varepsilon], f(x + sd) < f(x)$$



Propriété : si f est C^1, et d est une direction de descente de f en x, alors $\langle d, \nabla f(x) \rangle \leq 0$

(réciproque partielle avec inégalité stricte)

En effet,

$$f(x+sd) = f(x) + s\langle \nabla f(x), d \rangle + o(s)$$

ou

$$\langle \nabla f(x), d \rangle = \underbrace{\frac{f(x+sd) - f(x)}{s}}_{<0} + o(1)$$

(réciproque : TD)

et par passage à la limite ($s \to 0$)

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

* **Remarque** :

$x \in \mathbb{R}^n$ et $\nabla f(x) \neq 0$, alors $d = -\nabla f(x)$ est une direction de descente de f en x.

1.3) Conditions d'optimalité

Ici, on cherche à déterminer des conditions (nécessaires et/ou suffisantes) d'existence, voire d'unicité d'un minimum local d'une fonction

$$f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$$

(cas $\Omega = \mathbb{R}^n$: optimisation sans contraintes).

1.3.1) Cas d'un problème sans contrainte

Théorème (CN1 : condition nécessaire d'ordre 1). Soit f $C^1 : \mathbb{R}^n \to \mathbb{R}$ et $x \in \mathbb{R}^n$ un minimum local de f. Alors $\nabla f(x) = 0$ (pt critique)

En effet si $h \in \mathbb{R}^n$ et $\varepsilon > 0$

$$f(x+\varepsilon h) = f(x) + \varepsilon \langle \nabla f(x), h \rangle + o(\varepsilon)$$

D'où

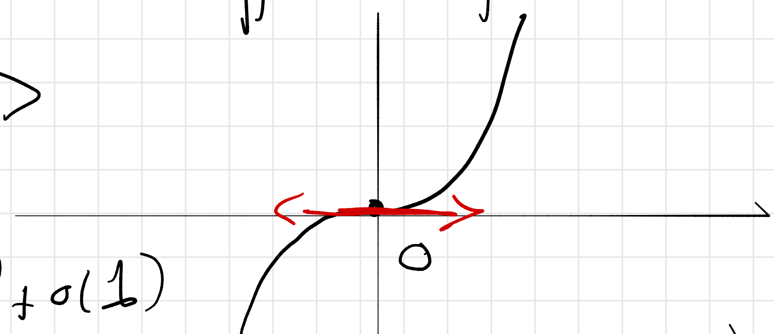
$$\langle \nabla f(x), h \rangle = \underbrace{\frac{f(x+\varepsilon h) - f(x)}{\varepsilon}}_{\geq 0} + o(1)$$

et $\langle \nabla f(x), h \rangle \geq 0$ pour tout $h \in \mathbb{R}^n$. En utilisant $h' = -h$, on trouve que

$$\langle \nabla f(x), h \rangle = 0 \quad \forall h \in \mathbb{R}^n$$

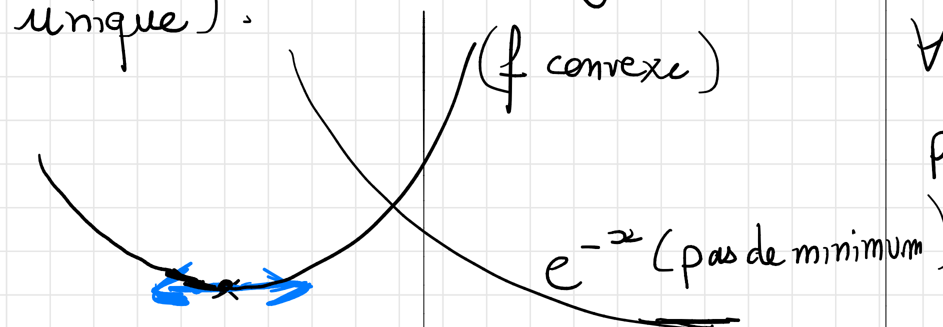
Ceci implique que $\nabla f(x) = 0$

* **Remarque** : il ne s'agit pas d'une condition suffisante : $f(x) = x^3$ $(n=1)$



(ou $f(x,y) = xy$ en $(0,0)$)

* A noter qu'il existe une réciproque partielle dans le cas des fonctions convexes : si f est convexe et $x \in \mathbb{R}^n$ est tel que* $\nabla f(x) = 0$, alors x est un minimum* global de f. (si f est strictement convexe, ce minimum est global et est unique).



Il existe une condition nécessaire d'ordre si f est C^2 :

Théorème (CN2) : soit f $C^2 : \mathbb{R}^n \to \mathbb{R}$

Soit x^* un minimum local de f.

Alors $\nabla f(x^*) = 0$ et $Hf(x)$ est une matrice symétrique positive :

$$\forall d \in \mathbb{R}^n \quad \langle d, Hf(x).d \rangle \geq 0$$

preuve : Taylor :

$$f(x^* + td) = f(x^*) + \cancel{\langle d, \nabla f(x^*) \rangle} + \frac{1}{2} t^2 \langle d, Hf(x).d \rangle + o(t^2)$$

$$t \to 0 : \Longrightarrow \langle d, Hf(x) d \rangle \geq 0$$

Il existe enfin une condition suffisante d'ordre 2 :

Théorème (CS2) : Soit f C^2, $\mathbb{R}^n \to \mathbb{R}$. Soit $x^* \in \mathbb{R}^n$. Si $\nabla f(x^*) = 0$ et $Hf(x^*)$ est définie positive, alors x^* est un minimum local de f.

Preuve :

$$f(x^* + t\underbrace{d}_{\neq 0}) = f(x^*) + \frac{1}{2} t^2 \overbrace{\langle d, Hf(b) d \rangle}^{>0} + o(t^2)$$

$\Rightarrow$ si t suffisament petit, non nul,

$$f(x^* + td) > f(x^*)$$

D'un point de vue pratique, pour rechercher un (ou les) minimum de f,

* on cherche les points critiques
* en ces points critiques, on recherche la nature de la Hessienne.

↗ si $\gg 0$: c'est un minimum

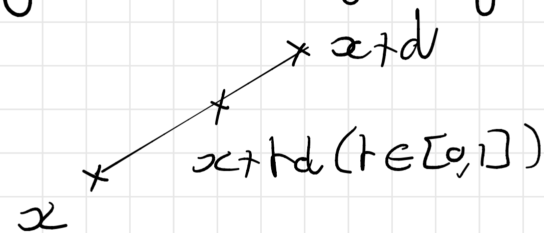
→ si non ≥ 0 : ce n'est pas un minimum

↘ sinon : on étudie localement f.

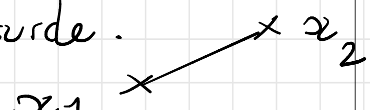
On peut disposer d'un outil supplémentaire pour conclure, à savoir la convexité éventuelle de f :

Théorème : soit f $C^2 : \mathbb{R}^n \to \mathbb{R}$ et convexe. Si $x^* \in \mathbb{R}^n$ est tel que $\nabla f(x^*) = 0$, alors x^* est un minimum global de f. Si de plus f est strictement convexe, alors x^* est unique.

preuve : + Taylor ordre 2 de type Taylor intégrale.

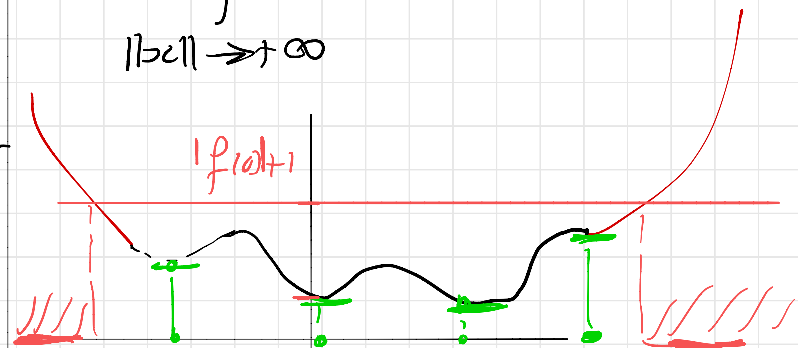


* Si f strictement convexe, on raisonne par l'absurde.



Un autre outil éventuel est la coercivité d'une fonction :

Déf : $f : \mathbb{R}^n \to \mathbb{R}$ est coercive si
$$\lim_{\|x\| \to +\infty} f(x) = +\infty$$



Proposition : si $f : \mathbb{R}^n \to \mathbb{R}$ est continue et coercive, alors elle admet au moins un minimum global.

(extension : $f : \Omega \to \mathbb{R}$, Ω fermé) preuve : soit $M = |f(0)| + 1$

il existe A tq si $\|x\| \geq A$, alors $f(x) \geq M$.

compact

Sur $B(0, A)$, f est continue, elle admet donc un minimum global, noté x^*. Il s'agit d'un minimum global sur $\mathbb{R}^n$: en effet,

$$\begin{cases} * \text{ si } x \in B(0,A), & f(x) \geq f(x^*) \\ * \text{ si } x \notin B(0,A), & f(x) \geq |f(0)| + 1 \geq f(x^*) + 1 \end{cases}$$

* Pour conclure à la présence d'un minimum dans le cas d'un point critique, on pourra par exemple essayer de montrer que :

$\left(\begin{array}{l} * \ f \text{ est coercive} \\ * \ f \text{ possède un unique pt critique} \end{array} \right.$

Dans ce cas, x^* point critique est un minimum global de f.

(voir exemples en TD)

Quelques "contre-exemples" :

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

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 \\ -3x_2^2 \end{pmatrix} = 0$$

et f possède un unique point critique : $(x_1,x_2)=(0,0)$.

En ce point $Hf(0,0)=\begin{pmatrix}2&0\\0&0\end{pmatrix}$

$Hf(0,0)\geq 0$

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

$(0,0)$

$f(0,t)=-t^3>f(0,0)$

et $(0,0)$ n'est pas un minimum local de f.

(f n'est ni convexe, ni coercive)

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

$(0,0)$ est l'unique point critique

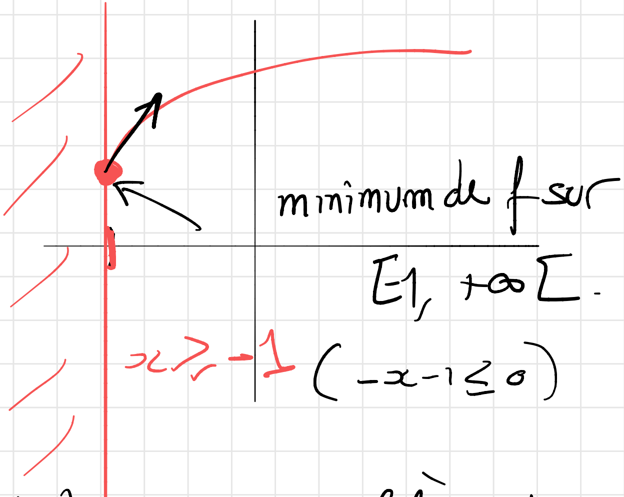
Dans les 2 cas $Hf(0,0)=0$ (matrice nulle)

Dans le 1^{er} cas, f est minimale en $(0,0)$

Dans le 2^e cas, f est maximale en $(0,0)$

1.3.2.) Cas d'un problème avec contraintes

Dans ce cas, il existe une nouvelle CN1 (appelée condition KKT). En aucun cas, on ne peut écrire que le gradient de f est nul en un point de minimum :



On s'intéresse au problème de minimisation de $f : \Omega \to \mathbb{R}$ où

$$\Omega = \left\{ x \in \mathbb{R}^n \,/\, \begin{array}{l} c_I(x) \leq 0 \\ c_E(x) = 0 \end{array} \right\}$$

nombre q

nombre p

Déf : on définit par le problème précédent le Lagrangien associé :

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \langle \lambda, c_E(x) \rangle + \langle \mu, c_I(x) \rangle$$

où $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^q$.

$$\sum_{i=1}^{p} \lambda_i c_{E_i}(x)$$

La CN1 s'écrit à l'aide du Lagrangien (relation KKT)

Karush Kuhn Tucker

Théorème : Soit $f : \Omega \to \mathbb{R}$ et x^* un minimum de f sur Ω. Alors

On suppose que les contraintes sont linéairement indépendantes en x^* (condition de qualification (*)). Alors : il existe $\lambda^* \in \mathbb{R}^p$ et $\mu^* \in \mathbb{R}^q$ (appelés multiplicateurs de Lagrange) tels que :

* $\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$
* $c_E(x^*) = 0$
* $c_I(x^*) \leq 0$
* $\mu^* \geq 0$ $\quad (\mu_i^* \geq 0 \ \forall i \in \{1, \dots, q\})$
* $\mu^* \odot c_I(x^*) = 0$ $\quad$ (si $c_{I_i}(x^*) < 0$ alors $\mu_i^* = 0$)

(CN1)

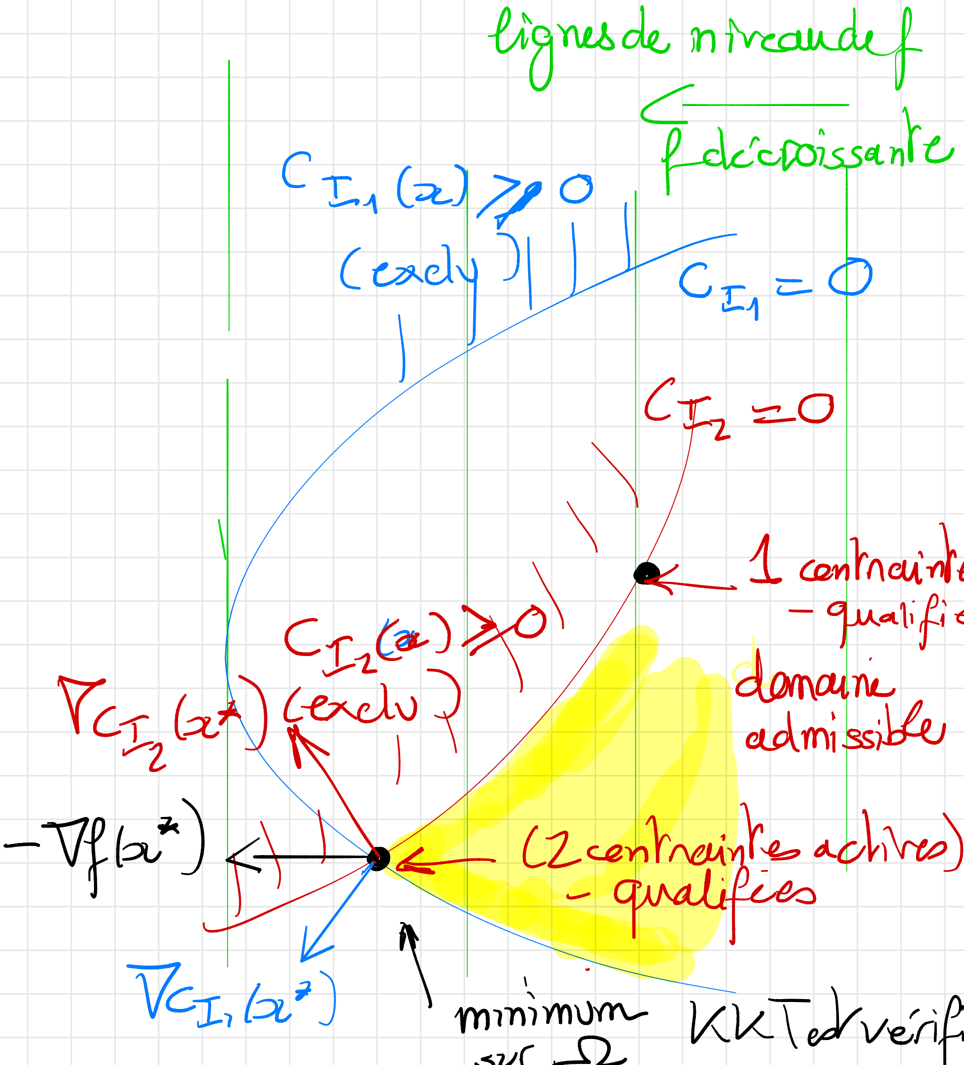
(*) $\{\nabla c_I(x^*), \nabla c_E(x^*)\}$ famille libre par les contraintes actives

Interprétation graphique de la condition

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$$

dans le cas où on a 2 contraintes inégalités et $n = 2$.

$$= \nabla_x f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla c_{I_i}(x^*) + \sum_{i=1}^{q} \underbrace{\mu_i}_{\geq 0} \nabla c_{E_i}(x^*) = 0$$



KKT et vérifié

KKT : en x^*, il existe $\mu_1^* \geq 0$ et $\mu_2^* \geq 0$ tq

$$\nabla f(x^*) + \mu_1^* \nabla c_1(x^*) + \mu_2^* \nabla c_2(x^*) = 0$$