OPTIMISATION

Optimisation : faire les choses de la meilleure façon possible

Exemples :

en navigation : trouver le chemin le plus court, ou le plus rapide, ou le plus économique pour aller d'un point A à un point B,

en production : trouver la forme d'une boîte de conserve de volume donné qui utilise la plus petite quantité de métal,

en planification : trouver la tournée la plus courte qui dessert un ensemble de villes donné,

en télécommunications : trouver où placer un nombre donné d'antennes relais pour couvrir un territoire donné,

etc...

Dans un problème d'optimisation, on trouve donc :

toujours :
- un objet que l'on cherche à déterminer : une trajectoire, un nombre (réel ou entier), une séquence (permutation), un vecteur, etc...
- un ou plusieurs critères que l'on souhaite minimiser ou maximiser : longueur, temps, coût, etc...

parfois :
- des contraintes : rester sur la route (GPS), un volume (boîte de conserve), revenir à son point de départ (tournée), etc...
- des paramètres aléatoires (embouteillages, météo, etc...)

Le but d'un problème d'optimisation est donc de trouver $\mathbf{x}$ (l'objet) appartenant à l'ensemble $S \subset \Omega$ qui minimise la fonction $f : \Omega \to \mathbb{R}$, ce que l'on peut écrire sous la forme :

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$
$$\text{sous la contrainte } \mathbf{x} \in S.$$

Ω peut être égal à $\mathbb{N}^n$ (vecteurs d'entiers), $\mathbb{R}^n$ (vecteur de réels), $\mathbb{L}p$ (espace de Lebesgue), etc...

Notes :

- Si $\Omega \subset \mathbb{R}^n$, on parle d'optimisation continue.
- Si Ω est fini ou dénombrable (par exemple : $\Omega = \mathbb{N}$), on parle d'optimisation discrète.
- Si Ω est un ensemble de fonctions, on parle de commande optimale.
- Si les données sont aléatoires, on parle d'optimisation stochastique.
- Si $f : \Omega \to \mathbb{R}^m$, on parle d'optimisation multicritère.

Nous ne traiterons ici que les problèmes d'**optimisation continue** ($\Omega \subset \mathbb{R}^n$), déterministes et monocritère ($m = 1$)

Définition

- Les composantes x_i $(i = 1, \cdots, n)$ de $\mathbf{x} \in \mathbb{R}^n$ sont appelées **variables de décision** du problème.
- La fonction f est appelée **fonction objectif** (ou parfois **fonction coût**, ou encore **fonction économique**).
- L'ensemble S est appelé **domaine réalisable** (ou **admissible**).

Le problème d'optimisation est parfois écrit sous la forme :

$$\min_{\mathbf{x} \in S} f(\mathbf{x}).$$

Ouverts, fermés, bornés

Définition : boule ouverte

Soit $\mathbf{x} \in \mathbb{R}^n$ et $r > 0$. On appelle **boule ouverte** de centre $\mathbf{x}$ et de rayon r l'ensemble

$$\mathcal{B}(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n \ tq \ \|\mathbf{y} - \mathbf{x}\| < r\}$$

Définition : ouvert, fermé

Un ensemble O est dit **ouvert** si

$$\forall \mathbf{x} \in O \ \exists r > 0 \ tq \ \mathcal{B}(\mathbf{x}, r) \subset O$$

Un ensemble F est dit **fermé** si son complémentaire $\overline{F}$ est un ouvert.

Exemples :

- $\mathbb{R}^n$ et $\varnothing$ sont des ouverts. Ce sont d'ailleurs aussi des fermés.
- Les boules ouvertes sont des ouverts.
- Les intervalles $]a, b[$ $(-\infty < a < b < +\infty)$ sont des ouverts.
- Les intervalles $]-\infty, a[$ et $]a, +\infty[$ $(-\infty < a < +\infty)$ sont des ouverts.
- Les intervalles $[a, b]$ $(-\infty < a < b < +\infty)$ sont des fermés.
- Les intervalles $]-\infty, a]$ et $[a, +\infty[$ $(-\infty < a < +\infty)$ sont des fermés.
- $A = \{(x, y) \ tq \ x^2 + y^2 < 1\}$ est un ouvert.
- $A \cup \{(0, 1)\}$ n'est ni ouvert, ni fermé.

Définition : borné

Un ensemble $F \subset E$ est dit **borné** si il existe une boule ouverte de E contenant F.

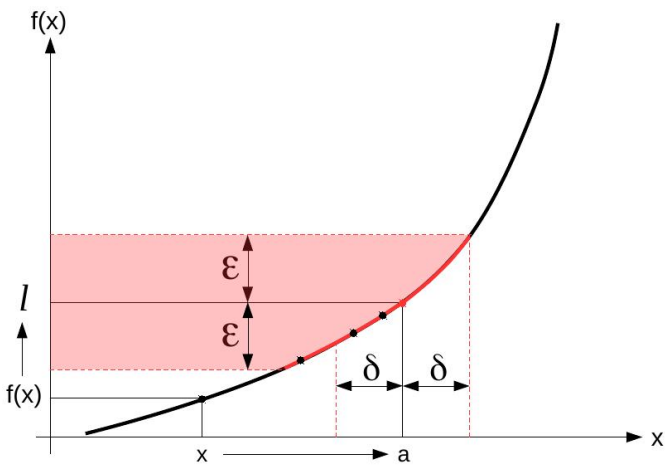
Exemples :

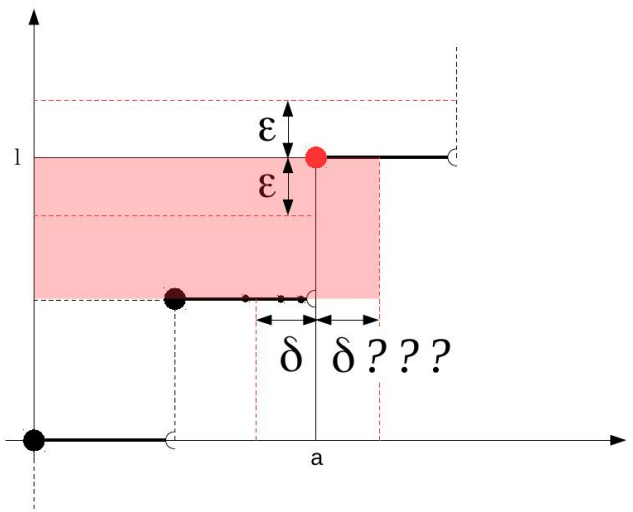
- $[0,1]$, $]-3,12]$ sont bornés : ils sont respectivement contenus par exemple dans les boules ouvertes $\mathcal{B}(0.5,2) =]-1.5,2.5[$ et $\mathcal{B}(0,13) =]-13,+13[$ par exemple.
- $]-\infty,0[$ n'est pas borné.
- $A = \{(x,y) \ tq \ x^2+y^2 < 1\}$ est un borné de $\mathbb{R}^2$: il est par exemple contenu dans la boule ouverte $\mathcal{B}((0,0),2) = \{(x,y) \ tq \ x^2+y^2 < 4\}$).
- $B = \{\frac{1}{n}\}_{n\in\mathbb{N}^*}$ est un ensemble borné de $\mathbb{R}$: il est par exemple contenu dans la boule ouverte $\mathcal{B}(1,1) =]0,2[$.

Limite

On dira que $f(x)$ tend vers l lorsque x tend vers a, et on notera $\lim_{x \to a} f(x) = l$ si on peut rendre $f(x)$ *aussi proche que l'on veut de* l à condition de choisir x *suffisamment proche de* a :

$$\forall \epsilon > 0,\ \exists \delta > 0 \ tq\ |x - a| \leq \delta \implies |f(x) - l| \leq \epsilon$$





définition : limite

	Limite finie, l	limite infinie $+\infty$ $(-\infty)$
En $a \in \mathbb{R}$	$\lim\limits_{x \to a} f(x) = l$ $\forall \epsilon > 0, \exists \delta > 0$ $\lvert x - a \rvert \leq \delta \Rightarrow \lvert f(x) - l \rvert \leq \epsilon$	$\lim\limits_{x \to a} f(x) = +\infty$ $\forall M > 0, \exists \delta > 0$ $\lvert x - a \rvert \leq \delta \Rightarrow f(x) \geq M$ $(\leq -M)$
En $+\infty$ $(-\infty)$	$\lim\limits_{x \to +\infty} f(x) = l$ $\forall \epsilon > 0, \exists K > 0$ $x \geq K \Rightarrow \lvert f(x) - l \rvert \leq \epsilon$ $(\leq -K)$	$\lim\limits_{x \to +\infty} f(x) = +\infty$ $\forall M > 0, \exists K > 0$ $x \geq K \Rightarrow f(x) \geq M$ $(\leq -K)$ $(\leq -M)$

Continuité

définition : continuité

Soient $f : \mathcal{D} \subset \mathbb{R} \to \mathbb{R}$ et $a \in \mathbb{R}$ un réel fixé. On dit que :

- f est continue en a ssi $\lim\limits_{\substack{x \to a \\ x \in D}} f(x) = f(a)$,
- f est continue sur $\mathcal{D}$ si $\forall a \in \mathcal{D}$, f est continue en a.

Remarque : si f définie en a et si $\lim\limits_{\substack{x \to a \\ x \in D}} f(x) = l$, alors nécessairement :

$$l = f(a).$$

En effet, par définition de la limite :

$$\forall \epsilon > 0,\ \exists \delta > 0\ tq\ |x - a| \leq \delta \implies |f(x) - l| \leq \epsilon.$$

Puisque f est définie en a, on peut prendre en particulier $x = a$. Dans ce cas, $|x - a| \leq \delta$ est vrai pour tout $\delta > 0$ et la proposition ci-dessus devient :

$$\forall \epsilon > 0,\ |f(x) - l| \leq \epsilon.$$

$|f(x) - l|$ est inférieur à toute quantité positive : il est donc nul et on a bien $f(x) = l$.

Dérivabilité

définition : dérivabilité

Soit I un intervalle ouvert de $\mathbb{R}$ et $f : I \to \mathbb{R}$. Alors :

- Pour $a \in I$ donné, on dit que f est *dérivable* lorsque la limite
$$\lim_{\substack{x \to a \\ x \in I - \{a\}}} \frac{f(x) - f(a)}{x - a}$$
existe et est finie. On note alors cette limite $f'(a)$.
- f est dite *dérivable sur* I si f est dérivable en tout $a \in I$.
- f est dite *continûment dérivable sur* I et notée C^1 si f est dérivable sur I et sa dérivée est continue sur I.

proposition : dérivabilité $\implies$ continuité

Soit I un intervalle ouvert de $\mathbb{R}$ et $f : I \to \mathbb{R}$.
Si f est dérivable en $a \in I$, alors elle est continue en a.

Limite, continuité

Les notions de limite et de continuité se généralisent en un point fixé

définition : limite

Soit une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ et $\mathbf{a} \in \mathbb{R}^n$. Nous disons que la limite de f lorsque $\mathbf{x}$ tend vers $\mathbf{a}$ est l pour dénoter la propriété suivante :

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = l \iff \forall \epsilon > 0, \exists \delta > 0 \ tq \ \|\mathbf{x} - \mathbf{a}\| \leq \delta \Rightarrow |f(\mathbf{x}) - l| \leq \epsilon$$

définition : continuité

Soit $D \subset \mathbb{R}^n$ un sous ensemble de $\mathbb{R}^n$, $f : D \to \mathbb{R}$ une fonction définie sur D et $a \in D$. Nous disons que

- f est continue en $\mathbf{a}$ si et seulement si $\lim\limits_{\substack{\mathbf{x}\in D \\ \mathbf{x}\to\mathbf{a}}} f(\mathbf{x}) = f(\mathbf{a})$
- f est continue sur D si pour tout $\mathbf{a} \in D$ f est continue en $\mathbf{a}$.

proposition : caractérisation de la continuité

Soit f une fonction de D dans $\mathbb{R}$. Soit $\mathbf{a} \in D$. Alors f est continue en $\mathbf{a}$ si et seulement si pour toute suite $(\mathbf{x}_m)_{m\in\mathbb{N}}$ qui tend vers $\mathbf{a}$, $(f(\mathbf{x}_m))_{m\in\mathbb{N}}$ tend vers $f(\mathbf{a})$.

Exemple : on considère sur $\mathbb{R}^2$ l'application définie par

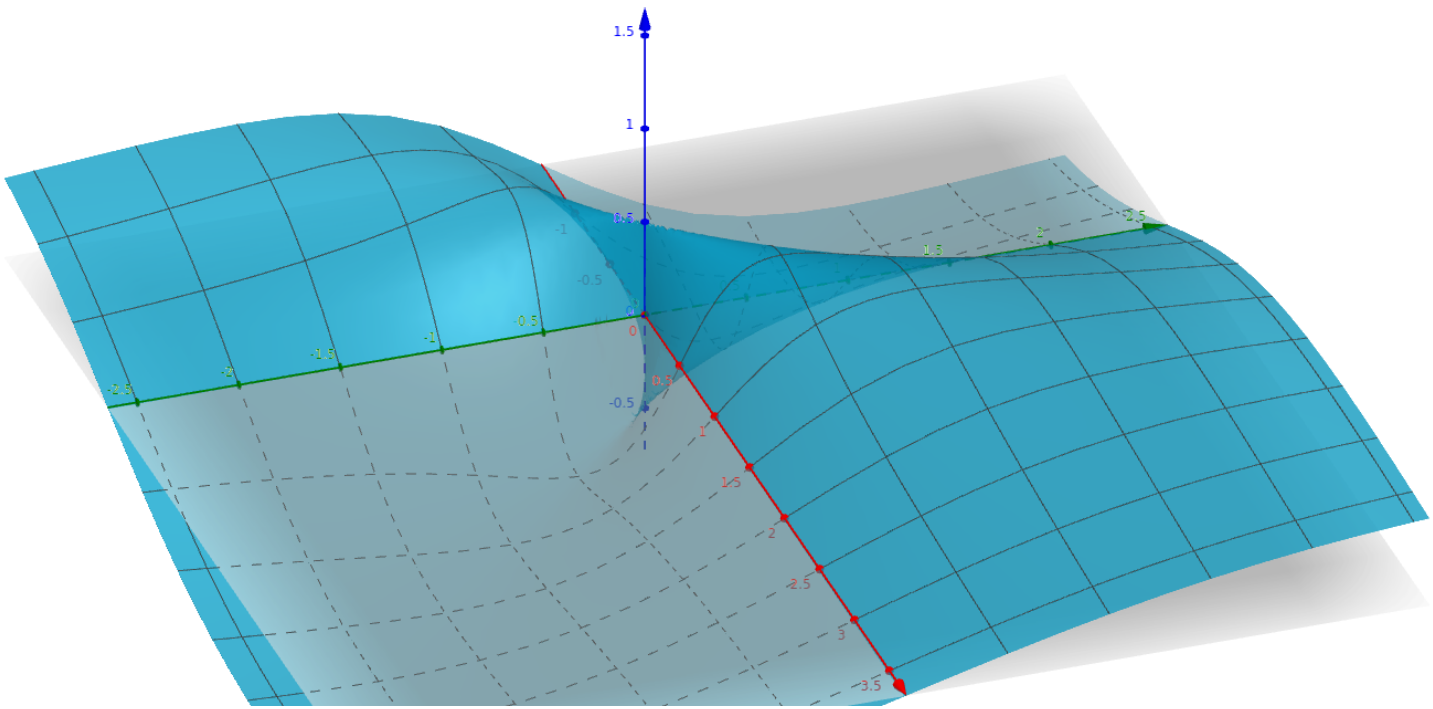
$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} \text{ si } (x,y) \neq (0,0), \\ 0 \text{ sinon.} \end{cases}$$

Sur $\mathbb{R}^2 \setminus (0,0)$, f est une fraction rationnelle dont le dénominateur ne s'annule pas et est donc continue. D'autre part :

- $f(x,0) = 0$ et tend donc vers 0 quand x tend vers 0,
- $f(0,y) = 0$ et tend donc vers 0 quand y tend vers 0

Ainsi, quand on se rapproche de 0 en suivant la direction Ox ou la direction Oy, f tend bien vers 0.

Pourtant, f n'est pas continue. En effet, considérons la suite $u_n = \left(\frac{1}{n}, \frac{1}{n}\right)$ pour $n \in \mathbb{N}^*$. Quand n tend vers $+\infty$, cette suite tend bien vers $(0,0)$ mais $f(u_n)) = \frac{1}{2}$ ne tend bien sûr pas vers $f(0,0) = 0$.



Calcul différentiel

Dans tout ce qui suit, on considère une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$

a) Gradient

définition : gradient

On appelle **gradient** de $f : \mathbb{R}^n \to \mathbb{R}$ au point $\mathbf{x}$ et on note $\nabla f(\mathbf{x})$ le vecteur des dérivées partielles de f :

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{x}) \\ \dfrac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

Propriétés :

- L'opérateur gradient est linéaire : $\forall f, g$ de classe C^1, $\forall (a,b) \in \mathbb{R}^2$,

$$\nabla\left(af(\mathbf{x}) + bg(\mathbf{x})\right) = a\nabla f(\mathbf{x}) + b\nabla g(\mathbf{x}),$$

- $\forall (\mathbf{u}, \mathbf{x}) \in \mathbb{R}^n \times \mathbb{R}^n$, $\nabla_{\mathbf{x}}\left({}^t\mathbf{x} \cdot \mathbf{u}\right) = \nabla_{\mathbf{x}}\left({}^t\mathbf{u} \cdot \mathbf{x}\right) = \mathbf{u}$,
- $\forall \mathbf{x} \in \mathbb{R}^n$ et $\mathbf{M}$ une matrice $n \times n$, $\nabla_{\mathbf{x}}\left({}^t\mathbf{x}\mathbf{M}\mathbf{x}\right) = \left(\mathbf{M} + {}^t\mathbf{M}\right)\mathbf{x}$.

définition : différentiabilité, C^1

Soit $f : \mathbb{R}^n \to \mathbb{R}$ telle que chacune des composantes de $\nabla f(\mathbf{x})$ existe et est continue en tout point $\mathbf{x}$ de $\mathbb{R}^n$. Alors f est différentiable en $\mathbf{x}$ et on dit que f est de classe C^1 sur $\mathbb{R}^n$.

b) hessienne

définition : hessienne

On appelle **hessienne** de $f : \mathbb{R}^n \to \mathbb{R}$ au point $\mathbf{x}$ et on note $\nabla^2 f(\mathbf{x})$ la matrice des dérivées partielles de ∇f :

$$\left(\nabla^2 f(\mathbf{x})\right)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \qquad \forall (i,j) \in \{1, \cdots, n\}^2.$$

On a donc :

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix}.$$

définition : fonction de classe C^2

Soit $f : \mathbb{R}^n \to \mathbb{R}$ telle que chacune des composantes de $\nabla^2 f(\mathbf{x})$ existe et est continue en tout point $\mathbf{x}$ de $\mathbb{R}^n$. Alors f est deux fois différentiable en $\mathbf{x}$ et on dit que f est de classe C^2 sur $\mathbb{R}^n$.

Théorème : Schwarz

Soit $f : \mathbb{R}^n \to \mathbb{R}$ de classe C^2. Alors $\forall \mathbf{x} \in \mathbb{R}^n$, $\nabla^2 f(\mathbf{x})$ est symétrique.

$\nabla^2 f(\mathbf{x})$ symétrique peut aussi s'écrire :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}),$$

ou encore :

$$\nabla^2 f(\mathbf{x}) = {}^t\nabla^2 f(\mathbf{x}).$$

c) jacobienne

définition : jacobienne

On appelle **jacobienne** de $f : \mathbb{R}^n \to \mathbb{R}^m$ au point $\mathbf{x}$ et on note $\mathbf{J}_f(\mathbf{x})$ la matrice de terme général :

$$\left(\mathbf{J}_f(\mathbf{x})\right)_{i,j} = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x})\right) \qquad \forall (i,j) \in \{1, \cdots, m\} \times \{1, \cdots, n\}.$$

On a donc :

$$\mathbf{J}_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

Remarques :

- la i^{e} ligne de $\mathbf{J}_f(\mathbf{x})$ est la transposée du gradient de f_i,
- la hessienne est la jacobienne du gradient : $\nabla^2 f(\mathbf{x}) = \mathbf{J}_{\nabla f}(\mathbf{x})$.

Développement de Taylor-Young

proposition : développement de Taylor-Young à l'ordre n

Soit $f : \mathbb{R} \to \mathbb{R}$ de classe C^n. Au voisinage du point x_0, on a :

$$\underbrace{f(x_0+h) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \cdots + \frac{h^n}{n!}f^{(n)}(x_0)}_{\text{partie régulière : } P_n(h)} + \underbrace{h^n\varepsilon(h)}_{\text{reste : } r(h)},$$

où $\lim\limits_{h\to 0} \varepsilon(h) = 0$. On écrit aussi :

$$f(x_0+h) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \cdots + \frac{h^n}{n!}f^{(n)}(x_0) + o(h^n)$$

Ce développement se généralise dans $\mathbb{R}^n$. En particulier, on a à l'ordre 2 :

$$\boxed{f(\mathbf{x_0}+\mathbf{h}) = f(\mathbf{x_0}) + {}^t\nabla f(\mathbf{x_0})\cdot\mathbf{h} + \frac{1}{2}\,{}^t\mathbf{h}\nabla^2 f(\mathbf{x_0})\mathbf{h} + o(\|\mathbf{h}\|^2).}$$

Remarques :

- La notation $o(h^n)$ (notation de Landau) représente une fonction de h qui est *négligeable devant* h^n. Plus précisément :

$$f \underset{a}{=} o(g) \iff f(x) = g(x)\epsilon(x) \text{ avec } \lim_{x \to a} \epsilon(x) = 0$$

  Par conséquent, $r(h) = o(h^n) \Leftrightarrow r(h) = h^n\epsilon(h)$ avec $\lim_{x \to 0} \epsilon(x) = 0$, et donc : $o(h^n) = h^n\epsilon(h)$ et $o(\|h\|^2) = \|h\|^2\epsilon(\|h\|)$
- La partie régulière du développement limité de f à l'ordre n en x_0 est un polynôme en x de degré n dans lequel $x_0,\ f(x_0),\ f'(x_0),\ f''(x_0), \cdots, f^{(n)}(x_0)$ sont des **constantes**.
- Une fonction ne peut admettre qu'un développement limité d'ordre n.
- La partie régulière du DL d'ordre n de $f + g$ est la somme des parties régulières des DL d'ordre n de f et de g.
- La partie régulière du DL d'ordre n de fg se compose des termes de degré au plus égal à n du produit des parties régulières des DL d'ordre n de f et de g.

Développements limités usuels en 0 (à savoir par coeur)

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots + x^n + o(x^n)$$

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n + o(x^n)$$

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + o(x^n)$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + o(x^n)$$

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} \cdots - \frac{x^n}{n} + o(x^n)$$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \cdots + (-1)^{n-1}\frac{x^n}{n} + o(x^n)$$

$$\sin x = x - \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + o\left(x^{2n+2}\right)$$

$$\cos x = 1 - \frac{x^2}{2!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + o\left(x^{2n+1}\right)$$

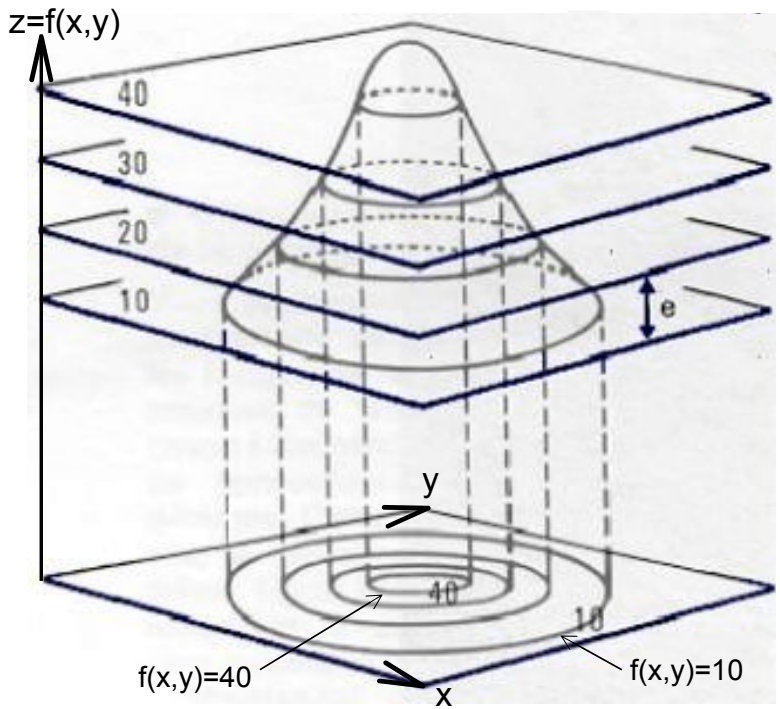
Courbes de niveau

définition : courbes de niveau

On appelle courbe de niveau d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ l'ensemble des points

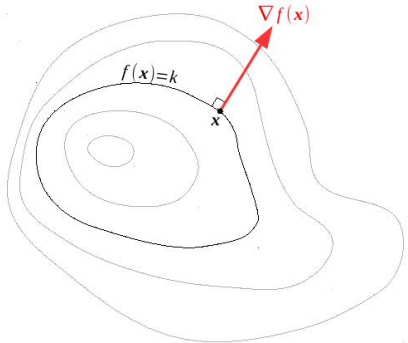
$$\{\mathbf{x} \in \mathbb{R}^n,\ f(\mathbf{x}) = k\} \qquad \text{(courbe de niveau } k\text{)}.$$

Exemple dans $\mathbb{R}^2$:



proposition :

$\forall \mathbf{x} \in \mathbb{R}^n$, le gradient de f en $\mathbf{x}$ est toujours orthogonal à la courbe de niveau passant par $\mathbf{x}$.



En effet, si f est C^1 au voisinage de $\mathbf{x}_0$, elle admet un DL à l'ordre 1 en ce point :

$$f(\mathbf{x}) = f(\mathbf{x}_0) + {}^t\nabla f(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0) + \|\mathbf{x}-\mathbf{x}_0\|\epsilon(\|\mathbf{x}-\mathbf{x}_0\|).$$

Or, sur une courbe de niveau, $f(\mathbf{x}) = cste = f(\mathbf{x}_0)$ donc :

$$\begin{aligned}
& {}^t\nabla f(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0) + \|\mathbf{x}-\mathbf{x}_0\|\epsilon(\|\mathbf{x}-\mathbf{x}_0\|) = 0 \\
\iff & {}^t\nabla f(\mathbf{x}_0)\frac{\mathbf{x}-\mathbf{x}_0}{\|\mathbf{x}-\mathbf{x}_0\|} + \epsilon(\|\mathbf{x}-\mathbf{x}_0\|) = 0 \text{ (si } \mathbf{x} \neq \mathbf{x}_0) \\
\implies & \lim_{\mathbf{x}\to\mathbf{x}_0} {}^t\nabla f(\mathbf{x}_0)\frac{\mathbf{x}-\mathbf{x}_0}{\|\mathbf{x}-\mathbf{x}_0\|} + \underbrace{\lim_{\mathbf{x}\to\mathbf{x}_0}\epsilon(\|\mathbf{x}-\mathbf{x}_0\|)}_{=0} = 0
\end{aligned}$$

Puisque $\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\mathbf{x}-\mathbf{x}_0}{\|\mathbf{x}-\mathbf{x}_0\|}$ est le vecteur tangent à la courbe de niveau en $\mathbf{x}_0$, celle-ci est donc bien normale au gradient en ce point.

Matrices (semi) définies-positives

définition : matrices (semi) définies-positives

Soit la forme quadratique $Q(\mathbf{z}) = {}^t\mathbf{z}\mathbf{A}\mathbf{z} \quad \forall \mathbf{z} \in \mathbb{R}^{\mathbf{n}}$ où $\mathbf{A}$ est une matrice symétrique réelle. On dit que la matrice A est :

- **définie-positive** si $\forall \mathbf{z} \neq \mathbf{0},\ Q(\mathbf{z}) > 0$,
- **semi définie-positive** si $\forall \mathbf{z} \neq \mathbf{0},\ Q(\mathbf{z}) \geq 0$,

Rappel : $\mathbf{A}$ est symétrique $\iff \mathbf{A} = {}^t\mathbf{A}$.

Exemple : $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ est définie-positive. En effet :

$$\forall (x, y) \in \mathbb{R}^2 \setminus \{(0,0)\},\ \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 2y^2 > 0$$

Caractérisation : la matrice symétrique réelle **A** est

définie-positive ssi tous ses principaux déterminants mineurs successifs sont strictement positifs (critère de Sylvester) :

$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \rightarrow \begin{cases} \det\begin{pmatrix} a_{11} \end{pmatrix} > 0, \\ \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \\ \det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} > 0. \end{cases}$$

Caractérisation équivalente : toutes les valeurs propres de **A** sont strictement positives. (N.B. : puisque **A** est symétrique réelle, ses valeurs propres sont nécessairement réelles.)

semi définie-positive ssi tous ses déterminants mineurs principaux sont positifs ou nuls (N.B. : déterminants formés en barrant $0, 1, \cdots, n$ lignes et colonnes se croisant sur la diagonale).

$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} \Rightarrow \det\left(a_{11}\right), \det\left(a_{22}\right), \det\left(a_{33}\right)$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \Rightarrow \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \det\begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix}$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \Rightarrow \det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Caractérisation équivalente : toutes les valeurs propres de $\mathbf{A}$ sont positives ou nulles.

définie-négative ssi $-\mathbf{A}$ (qui est aussi symétrique) est définie-positive.

Attention : cela ne signifie pas que tous ses principaux déterminants mineurs successifs sont strictement négatifs puisque $\det(-\mathbf{A}) = (-1)^n \times \det(\mathbf{A})$.

Caractérisation équivalente : toutes les valeurs propres de $\mathbf{A}$ sont strictement négatives.

semi définie-négative ssi $-\mathbf{A}$ est semi définie-positive.

Attention : cela ne signifie pas que tous ses déterminants mineurs principaux sont négatifs ou nuls.

Caractérisation équivalente : toutes les valeurs propres de $\mathbf{A}$ sont négatives ou nulles.

indéfinie ssi $\mathbf{A}$ n'est ni semi définie-positive, ni semi définie-négative, c'est-à-dire si

$$\exists(\mathbf{z_1}, \mathbf{z_2}) \text{ tq } {}^t\mathbf{z_1}\mathbf{A}\mathbf{z_1} < 0 \text{ et } {}^t\mathbf{z_2}\mathbf{A}\mathbf{z_2} > 0$$

Caractérisation équivalente : $\mathbf{A}$ possède une valeur propre strictement négative et une valeur propre strictement positive.

Remarques : Si $\mathbf{A}$ symétrique réelle est définie-positive, alors

- $\mathbf{A}$ est inversible et $\mathbf{A}^{-1}$ est aussi symétrique réelle définie-positive.
- $\mathbf{A}$ est semi définie-positive.

minimum, maximum, infimum, supremum

Définition : minorant, majorant

Soit E un sous-ensemble de $\mathbb{R}$. On dit que

- $m \in \mathbb{R} \cup \{-\infty, +\infty\}$ est un minorant de E ssi m est inférieur ou égal à tous les éléments de E.
- $M \in \mathbb{R} \cup \{-\infty, +\infty\}$ est un majorant de E ssi M est supérieur ou égal à tous les éléments de E.

$$m \in \mathbb{R} \cup \{-\infty, +\infty\} \text{ est un minorant de } E \iff \forall x \in E, m \leq x,$$
$$M \in \mathbb{R} \cup \{-\infty, +\infty\} \text{ est un majorant de } E \iff \forall x \in E, M \geq x$$

Si E admet un minorant (resp. : majorant) **fini**, alors E est dit minoré (resp. : majoré).

Définition : infimum, supremum

Soit E un sous-ensemble de $\mathbb{R}$. On appelle

- infimum de E le plus grand des minorants de E.
- supremum de E le plus petit des majorants de E.

$$\text{On les note : } \begin{cases} \inf(E) \in \mathbb{R} \cup \{-\infty, +\infty\} = \inf\limits_{x \in E}(x) \\ \sup(E) \in \mathbb{R} \cup \{-\infty, +\infty\} = \sup\limits_{x \in E}(x) \end{cases}$$

On remarquera que l'infimum et le supremum ne sont pas nécessairement finis. Lorsqu'ils le sont, E est minoré (resp. : majoré).

proposition

$$\text{Soit } E \subset \mathbb{R}. \text{ Alors : } \begin{cases} \inf(E) \in \mathbb{R} \iff E \text{ est minoré} \\ \sup(E) \in \mathbb{R} \iff E \text{ est majoré} \end{cases}$$

On remarquera aussi que l'infimum et le supremum de E n'appartiennent pas nécessairement à E. Lorsqu'ils y appartiennent, on parle de minimum et de maximum :

définition : minimum, maximum

Soit $E \subset \mathbb{R}$.

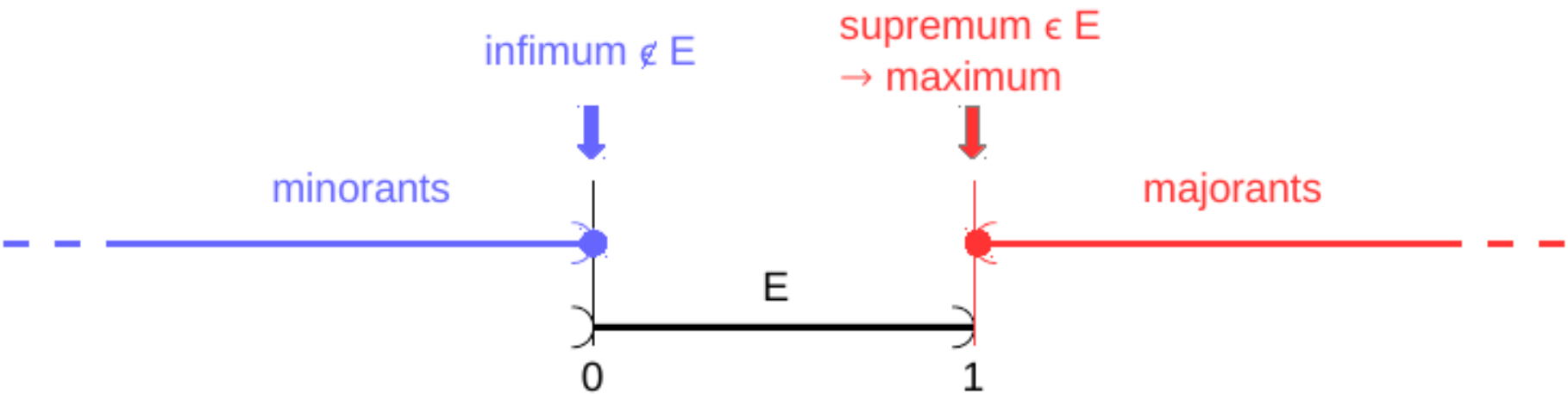
- $\inf(E)$ est appelé minimum si et seulement si $inf(E) \in E$.
- $\sup(E)$ est appelé maximum si et seulement si $\sup(E) \in E$.

Ils sont alors notés respectivement $\min(E)$ et $\max(E)$.

Exemple :

Considérons l'ensemble $E =]0, 1]$.

- L'ensemble des minorants de E est $[-\infty, 0]$ et l'ensemble de ses majorants est $[1, +\infty]$.
- $\inf(E) = 0$, $\sup(E) = 1$.
- Puisque $0 \notin E$, E n'a pas de minimum. En revanche, $1 \in E$ donc 1 est le maximum de E.



Définition du problème

On cherche à résoudre le problème suivant : $\boxed{\min_{\mathbf{x}\in S} f(x)}$

où $f : \mathbb{R}^n \to \mathbb{R}$ et $\mathbf{x} = {}^t(x_1, x_2, \cdots, x_n)$. Il s'agit donc de chercher le **minimum** (local ou global) de f sur S.

définition : minimum global

On dit que la fonction $f : S \subset \mathbb{R}^n \to \mathbb{R}$ atteint son **minimum global** au point $\mathbf{x}^*$ si et seulement si :

$$\mathbf{x}^* \in S \qquad \text{et} \qquad \forall \mathbf{x} \in S,\ f(\mathbf{x}^*) \leq f(\mathbf{x})$$

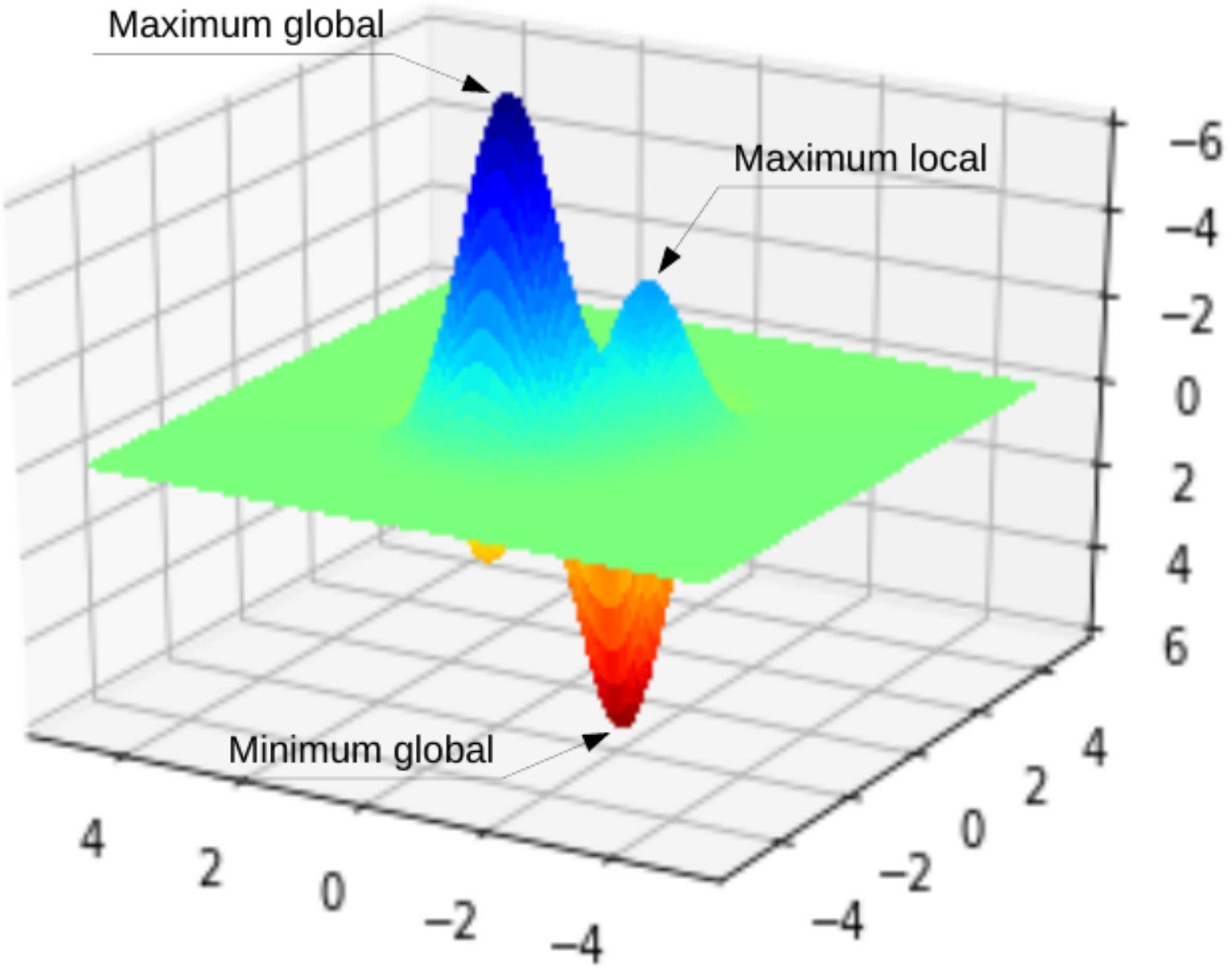
définition : minimum local

On dit que la fonction $f : S \subset \mathbb{R}^n \to \mathbb{R}$ atteint son **minimum local** au point $\mathbf{x}^*$ si et seulement si :

$$\mathbf{x}^* \in S \qquad \text{et} \qquad \exists \epsilon > 0 \ tq\ \forall \mathbf{x} \in S,\ \|\mathbf{x} - \mathbf{x}^*\| < \epsilon \Rightarrow f(\mathbf{x}^*) \leq f(\mathbf{x})$$

3 Optimum d'une fonction de plusieurs variables réelles

3.1 Quelques définitions : définition du problème



Remarques :

- L'ensemble des points $\mathbf{x} \in S$ tels que $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$ est l'intersection de S avec la boule ouverte de centre $\mathbf{x}^*$ et de rayon ϵ.
- Les définitions de **maximum global** et de **maximum local** s'obtiennent en inversant le signe des inégalités dans les définitions précédentes.
- $\mathbf{x}^*$ est un maximum local (resp. : global) de f ssi $\mathbf{x}^*$ est un minimum local (resp. : global) de $-f$. On a :

$$\max_{\mathbf{x} \in S} f(x) = -\min_{\mathbf{x} \in S} \left(-f(x)\right).$$

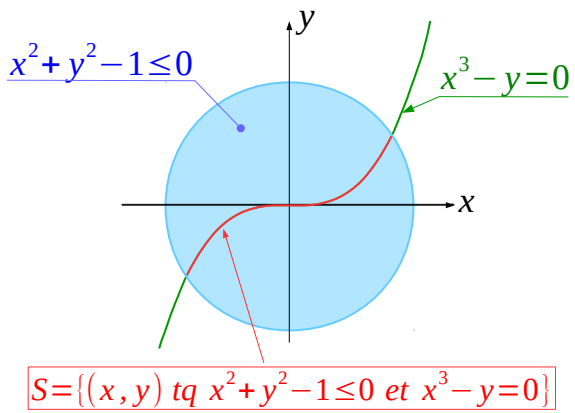
- Si $\mathbf{x}^*$ est un minimum global, alors c'est aussi un minimum local.

C.S. d'existence

On s'intéresse ici au cas où S est défini par des contraintes d'égalité ou d'inégalité : $S = \{\mathbf{x} \in \mathbb{R}^n \ tq \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \ et \ \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$

avec $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^p$ et $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^q$.

Exemple :



proposition :

Si $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^p$ et $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^q$ sont continues, alors

$S = \{\mathbf{x} \in \mathbb{R}^n \ tq \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \ et \ \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$ est un fermé de $\mathbb{R}^n$.

Suivant que S est borné ou non, on a alors les résultats suivants :

Cas où S est borné :

théorème : Weierstrass

Soit S un fermé borné non vide de $\mathbb{R}^n$ et $f : S \subset \mathbb{R}^n \to \mathbb{R}$ une application continue sur S.
Alors f est bornée et atteint ses bornes : f admet donc un minimum global sur S.

Par conséquent :

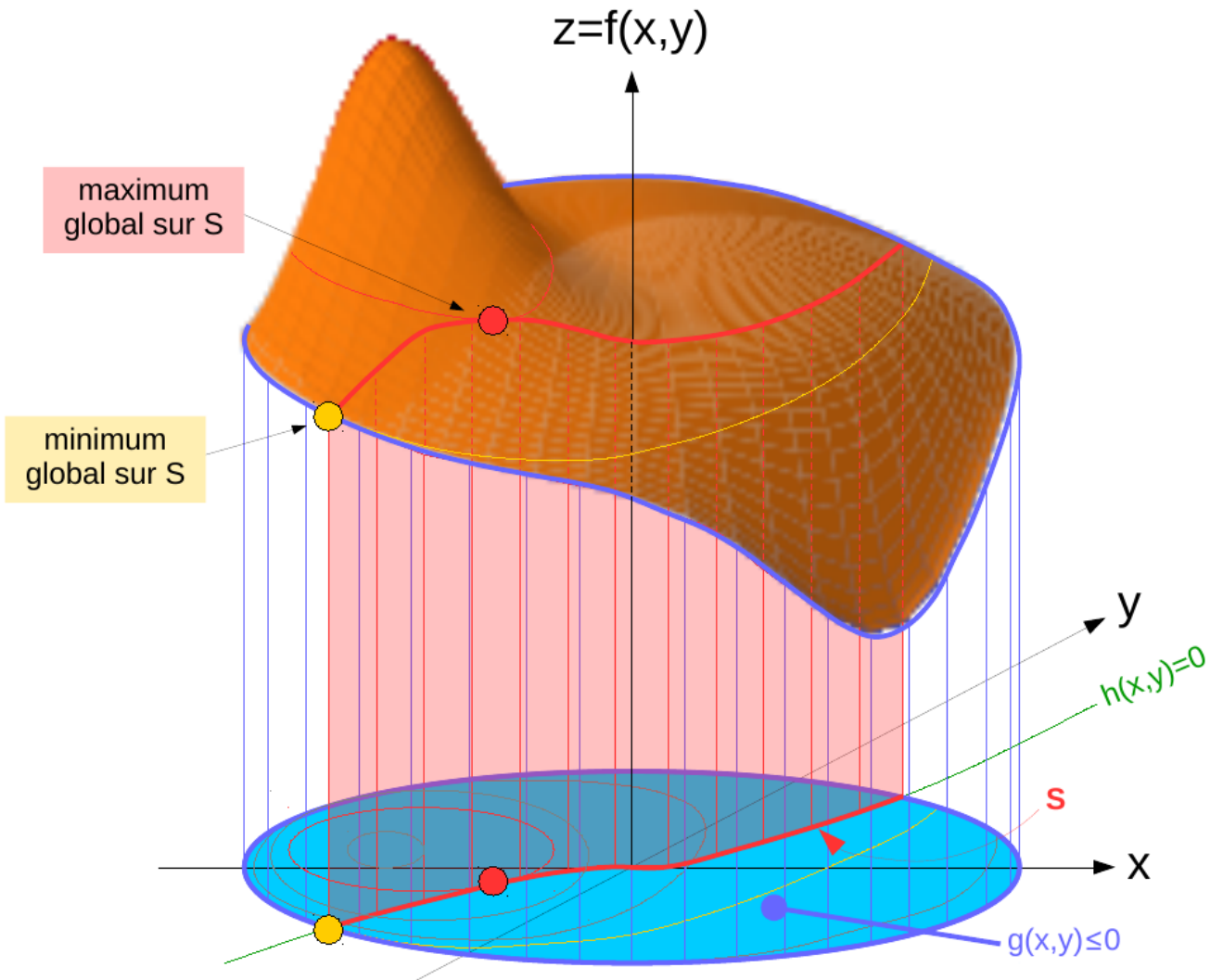
proposition :

Si f, $\mathbf{g}$ et $\mathbf{h}$ sont continues et si S est borné, alors

le problème $$\mathcal{P} : \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \\ \text{s.c. :} \begin{vmatrix} \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{vmatrix} \end{cases}$$

admet au moins une solution globale.

Exemple :



Cas où S est non borné (c'est le cas en particulier s'il n'y a pas de contraintes) :

définition : coercivité

Une application $f : S \subset \mathbb{R}^n \to \mathbb{R}$ est dite **coercive** (ou infinie à l'infini) si et seulement si :

$$\forall A \in \mathbb{R},\ \exists R > 0\ tq\ \forall \mathbf{x} \in S,\ \|\mathbf{x}\| \geq R \Rightarrow f(\mathbf{x}) \geq A$$

On note : $$\lim_{\substack{\|\mathbf{x}\| \to +\infty \\ \mathbf{x} \in S}} f(\mathbf{x}) = +\infty$$

Exemples de fonctions coercives :

- x^2 (il suffit de prendre $R \geq \sqrt{|A|}$),
- $\sqrt{x^2 + y^2}$ (avec par exemple $R \geq A$).

Exemples de fonctions non coercives :

- x (quand $x \to -\infty$),
- $x^2 - y^2$ (il suffit de prendre $x = 0$).

proposition : caractérisation de la coercivité

Si $f : S \subset \mathbb{R}^n \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}$ telles que $f(\mathbf{x}) \geq g(\|\mathbf{x}\|)$ avec $\lim_{t \to +\infty} g(t) = +\infty$, alors f est coercive.

Preuve : puisque $\lim_{t \to +\infty} g(t) = +\infty$, on a par définition :

$$\forall A \in \mathbb{R},\ \exists R > 0\ tq\ \forall t \in \mathbb{R}, t \geq R \Rightarrow g(t) \geq A.$$

Comme par ailleurs $f(\mathbf{x}) \geq g(\|\mathbf{x}\|)$, en posant $t = \|\mathbf{x}\|$ on a bien :

$$\forall A \in \mathbb{R},\ \exists R > 0\ tq\ \forall \mathbf{x} \in S, \|\mathbf{x}\| \geq R \Rightarrow f(\mathbf{x}) \geq g(\|\mathbf{x}\|) \geq A.$$

théorème (admis)

Soit S un fermé non vide de $\mathbb{R}^n$ et $f : S \subset \mathbb{R}^n \to \mathbb{R}$ une application continue et coercive sur S. Alors f admet un minimum global sur S.

Par conséquent :

Proposition

Si f, $\mathbf{g}$ et $\mathbf{h}$ sont continues et si f est coercive sur $S = \{\mathbf{x} \in \mathbb{R}^n \ tq \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \ et \ \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, alors le problème

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$
$$\text{s.c. : } \begin{vmatrix} \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{vmatrix}$$

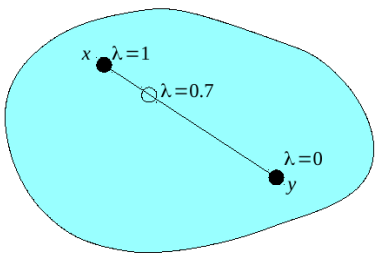
admet au moins une solution globale.

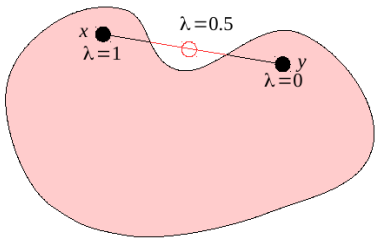
Convexité et optimisation

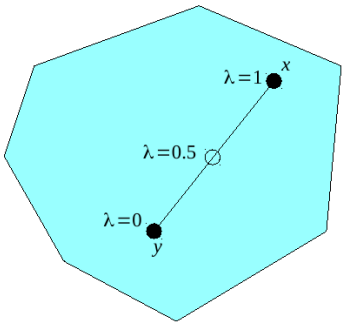
définition : ensemble convexe

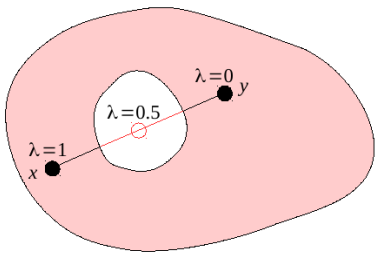
Soit $S \subset \mathbb{R}^n$. S est dit **convexe** si et seulement si :

$$\forall (\mathbf{x}, \mathbf{y}) \in S^2,\ \forall \lambda \in]0,1[,\ \lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in S.$$









CONVEXE **NON CONVEXE**

- $\mathbb{R}^n$ est convexe
- Les boules ouvertes et les boules fermées sont convexes

définition : fonction convexe

Une fonction $f : S \to \mathbb{R}$ est dite **convexe** si et seulement si :

$$\forall (\mathbf{x}, \mathbf{y}) \in S^2,\ \forall \lambda \in]0,1[,\ f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

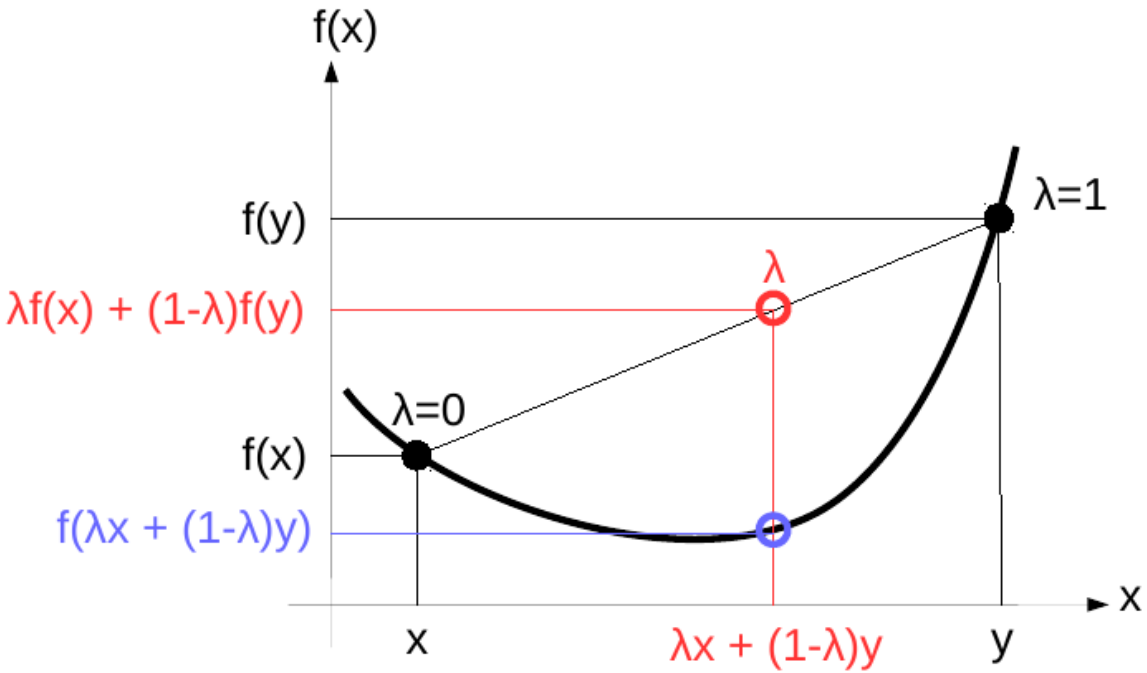
(**strictement convexe** si l'inégalité est stricte).

De la même façon, une fonction $f : S \to \mathbb{R}$ est dite **concave** si et seulement si :

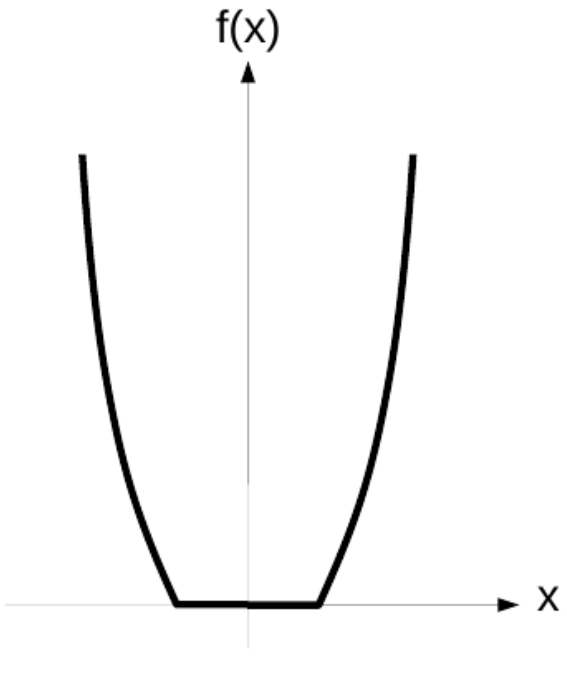
$$\forall (\mathbf{x}, \mathbf{y}) \in S^2,\ \forall \lambda \in]0,1[,\ f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

(**strictement concave** si l'inégalité est stricte).

Exemples : dans $\mathbb{R}$,



Fonction strictement convexe



Fonction convexe

Exemples de fonctions strictement convexes : x^2, e^x

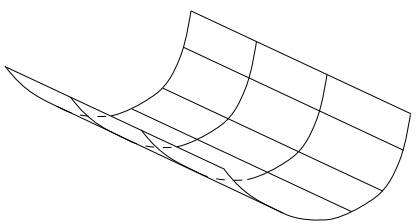
Exemples de fonctions simplement convexes : $|x|$, $\max(1, x^2)$

théorème

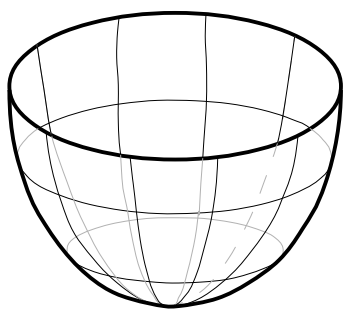
Soit $f : \mathbb{R}^n \to \mathbb{R}$ de classe C^2. Alors :

- $\nabla^2 f(\mathbf{x})$ est semi définie-positive $\forall \mathbf{x} \in \mathbb{R}^n \iff f$ est convexe,
- $\nabla^2 f(\mathbf{x})$ est définie-positive $\forall \mathbf{x} \in \mathbb{R}^n \Rightarrow f$ est strictement convexe.

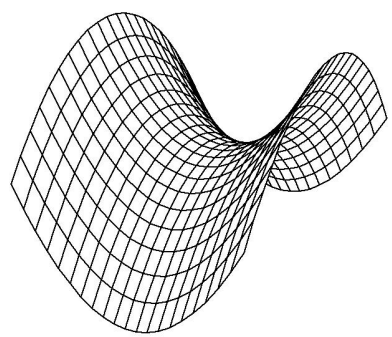
N.B. : dans le second cas, on a juste une implication. Par exemple $f(x) = x^4$ est strictement convexe et pourtant $f''(x) = 3x^2$ est nulle en $x = 0$ donc sa hessienne n'est pas définie-positive.



Semi définie-positive



Définie-positive



Indéfinie

théorème

Soient $S \subset \mathbb{R}^n$ un ensemble convexe et $f : S \to \mathbb{R}$.
Soit $\mathbf{x}^*$ un minimum local de f sur S.

1. Si f est convexe, alors $\mathbf{x}^*$ est un minimum global de f sur S,
2. Si f est strictement convexe, alors $\mathbf{x}^*$ est l'unique minimum global de f sur S.

Démonstration :

1. (par l'absurde) Soit $\mathbf{x}^* \in S$ un minimum local de f. Alors :

$$\exists \varepsilon > 0 \ tq \ \forall \mathbf{x} \in S \text{ vérifiant } \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon, \ f(\mathbf{x}) \geq f(\mathbf{x}^*).$$

Supposons que $\mathbf{x}^*$ ne soit pas un minimum global de f sur S :

$$\exists \mathbf{x}^+ \in S \ tq \ f(\mathbf{x}^+) < f(\mathbf{x}^*).$$

$\mathbf{x}^+ \notin \mathcal{B}(\mathbf{x}^*, \varepsilon)$ car $f(\mathbf{x}^+) < f(\mathbf{x}^*)$ donc $\|\mathbf{x}^+ - \mathbf{x}^*\| \geq \varepsilon$.

Considérons maintenant le point

$$\mathbf{x}_{\varepsilon/2} = \alpha\mathbf{x}^+ + (1-\alpha)\mathbf{x}^*$$

avec $\alpha = \dfrac{\varepsilon}{2\|\mathbf{x}^+ - \mathbf{x}^*\|} \in]0,1[$ (car $\|\mathbf{x}^+ - \mathbf{x}^*\| \geq \varepsilon$). On a alors :

$$\begin{aligned}
\|\mathbf{x}_{\varepsilon/2} - \mathbf{x}^*\| &= \|\alpha\mathbf{x}^+ + (1-\alpha)\mathbf{x}^* - \mathbf{x}^*\| \\
&= \alpha\|\mathbf{x}^+ - \mathbf{x}^*\| \\
&= \frac{\varepsilon}{2\|\mathbf{x}^+ - \mathbf{x}^*\|}\|\mathbf{x}^+ - \mathbf{x}^*\| \\
&= \frac{\varepsilon}{2} < \varepsilon
\end{aligned}$$

donc $\mathbf{x}_{\varepsilon/2} \in \mathcal{B}(\mathbf{x}^*, \varepsilon)$

Mais puisque f est convexe :

$$\begin{aligned} f(\mathbf{x}_{\varepsilon/2}) &\leq \alpha f(\mathbf{x}^+) + (1-\alpha) f(\mathbf{x}^*) \\ &< \alpha f(\mathbf{x}^*) + (1-\alpha) f(\mathbf{x}^*) = f(\mathbf{x}^*) \end{aligned}$$

donc $\mathbf{x}_{\varepsilon/2} \notin \mathcal{B}(\mathbf{x}^*, \varepsilon)$: contradiction. Par conséquent, $\mathbf{x}^*$ est un minimum global de f sur S.

2. (par l'absurde) Supposons qu'il existe deux minimum globaux de f sur S : $\mathbf{x_1}$ et $\mathbf{x_2}$. Alors $f(\mathbf{x_1}) = f(\mathbf{x_2}) = \min_{x \in S} f(\mathbf{x})$

   Puisque f est strictement convexe :

$$\begin{aligned} f\left(\frac{\mathbf{x_1} + \mathbf{x_2}}{2}\right) &< \frac{1}{2} f(\mathbf{x_1}) + \frac{1}{2} f(\mathbf{x_2}) \\ &< \frac{1}{2} \min_{x \in S} f(\mathbf{x}) + \frac{1}{2} \min_{x \in S} f(\mathbf{x}) \\ &< \min_{x \in S} f(\mathbf{x}) \end{aligned}$$

   ce qui est impossible.

corrolaire

Soit $S \subset \mathbb{R}^n$ un ensemble fermé non vide et convexe. Soit $f : S \to \mathbb{R}$ une fonction continue et strictement convexe. Si l'une des deux conditions suivantes est vérifiée :

- soit S est borné,
- soit f est coercive,

alors f admet un minimum global unique sur S.

théorème : conditions nécessaire d'extrémalité

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application différentiable. Alors :

$$\mathbf{x}^* \text{ est un minimum local de } f \Rightarrow \nabla f(\mathbf{x}^*) = \mathbf{0}.$$

Si de plus f est deux fois différentiable, alors :

$$\mathbf{x}^* \text{ est un minimum local de } f \Rightarrow \nabla^2 f(\mathbf{x}^*) \text{ est semi définie-positive.}$$

Démonstration : soit $\mathbf{x}^*$ un minimum local de f. Par définition, on a donc :

$$\exists \varepsilon > 0 \ tq \ \forall \mathbf{x} \in \mathcal{B}(\mathbf{x}^*, \varepsilon), \ f(\mathbf{x}^*) \leq f(\mathbf{x}).$$

De plus, $\forall \mathbf{u} \in \mathbb{R}^n, \ \exists \theta > 0 \ tq \ \mathbf{x}^* + \theta \mathbf{u} \in \mathcal{B}(\mathbf{x}^*, \varepsilon)$ (prendre $\theta < \frac{\varepsilon}{\|\mathbf{u}\|}$). Par conséquent :

$$\forall (\mathbf{u}, \theta) \in \mathbb{R}^n \times \mathbb{R}^{+*} \ tq \ \mathbf{x}^* + \theta \mathbf{u} \in \mathcal{B}(\mathbf{x}^*, \varepsilon), \ f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \theta \mathbf{u}).$$

Or f est différentiable en $\mathbf{x}^*$ donc :

$$f(\mathbf{x}^* + \theta\mathbf{u}) = f(\mathbf{x}^*) + \langle \nabla f(\mathbf{x}^*), \theta\mathbf{u} \rangle + o(\|\theta\mathbf{u}\|),$$

ou encore :

$$\lim_{\theta \to 0} \frac{f(\mathbf{x}^* + \theta\mathbf{u}) - f(\mathbf{x}^*)}{\theta} = \langle \nabla f(\mathbf{x}^*), \mathbf{u} \rangle$$

et comme $\dfrac{f(\mathbf{x}^* + \theta\mathbf{u}) - f(\mathbf{x}^*)}{\theta} \geq 0$, on en déduit que :

$$\langle \nabla f(\mathbf{x}^*), \mathbf{u} \rangle \geq 0.$$

Cette inégalité étant vraie pour tout $\mathbf{u}$ de $\mathbb{R}^n$, elle est vraie aussi pour $-\mathbf{u}$ donc $-\langle \nabla f(\mathbf{x}^*), \mathbf{u} \rangle \geq 0$. Il s'en suit que $\langle \nabla f(\mathbf{x}^*), \mathbf{u} \rangle = 0$ et donc que :

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

Montrons maintenant la seconde partie du théorème. Pour cela, on écrit le développement de Taylor-Young à l'ordre 2 de f en $\mathbf{x}^*$:

$$f(\mathbf{x}^* + \theta\mathbf{u}) = f(\mathbf{x}^*) + \theta\ {}^t\nabla f(\mathbf{x}^*)\mathbf{u} + \frac{1}{2}\theta^2\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u} + \|\theta\mathbf{u}\|^2\varepsilon(\|\theta\mathbf{u}\|)$$

Puisque $\mathbf{x}^*$ est un minimum local de f, $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et donc :

$$f(\mathbf{x}^* + \theta\mathbf{u}) - f(\mathbf{x}^*) = \frac{1}{2}\theta^2\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u} + \theta^2\|\mathbf{u}\|^2\varepsilon(\|\theta\mathbf{u}\|)$$

ou encore :

$$\lim_{\theta\to 0} \frac{f(\mathbf{x}^* + \theta\mathbf{u}) - f(\mathbf{x}^*)}{\theta^2} = \frac{1}{2}\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u},$$

et comme $\dfrac{f(\mathbf{x}^* + \theta\mathbf{u}) - f(\mathbf{x}^*)}{\theta^2} \geq 0$, on a donc :

$$\forall \mathbf{u} \in \mathbb{R}^n,\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u} \geq 0$$

- $\nabla f(\mathbf{x}^*) = \mathbf{0}$ est appelée **condition nécessaire d'optimalité du premier ordre**. Les points $\mathbf{x}^*$ vérifiant cette condition sont appelés **points stationnaires** ou **points critiques**.
- $\nabla^2 f(\mathbf{x}^*)$ semi définie-positive est appelée **condition nécessaire d'optimalité du deuxième ordre**.

Remarques :

1. Il ne s'agit là de CN d'optimalité QUE dans le cadre des hypothèses qui y sont attachées (f définie sur $\mathbb{R}^n$ et différentiable en $\mathbf{x}^*$) : $f(x) = \sqrt{x}$ est définie sur $\mathbb{R}^+$ et est minimale en 0 mais n'y est même pas dérivable.
2. Ce ne sont pas des conditions suffisantes : pour $f(x) = x^3$, on a $f'(0) = f''(0) = 0$ mais 0 n'est pas un extremum.

$\Rightarrow$ Les points stationnaires sont de bons candidats... mais ce ne sont que des candidats.

théorème : conditions suffisantes d'extremalité locale

Si $f : \mathbb{R}^n \to \mathbb{R}$ est de classe C^2 et si $\mathbf{x}^*$ vérifie :

$$\begin{cases} \nabla f(\mathbf{x}^*) = 0 \\ \nabla^2_f(\mathbf{x}^*) \text{ définie-positive (respectivement : définie-négative)} \end{cases}$$

alors $\mathbf{x}^*$ est un minimum local (resp. : maximum local) de f

Démonstration. On écrit le développement de Taylor-Young à l'ordre 2 en un point stationnaire $\mathbf{x}^*$:

$$f(\mathbf{x}^* + \theta\mathbf{u}) = f(\mathbf{x}^*) + {}^t\nabla f(\mathbf{x}^*) \cdot \theta\mathbf{u} + \frac{1}{2}\, {}^t(\theta\mathbf{u})\nabla^2 f(\mathbf{x}^*)\theta\mathbf{u} + \|\theta\mathbf{u}\|^2 \varepsilon\left(\|\theta\mathbf{u}\|^2\right)$$

avec $\|\mathbf{u}\| = 1$ et $\theta \geq 0$. Puisque $\mathbf{x}^*$ est un point stationnaire, $\nabla f(\mathbf{x}^*) = \mathbf{0}$ et :

$$\frac{f(\mathbf{x}^* + \theta\mathbf{u}) - f(\mathbf{x}^*)}{\theta^2} = \frac{1}{2}\, {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u} + \varepsilon\left(\theta^2\right)$$

Or, puisque $\lim_{\theta \to 0} \varepsilon\left(\theta^2\right) = 0$:

$$\forall \epsilon > 0,\ \exists \delta > 0\ tq\ \theta \leq \delta \Rightarrow \left|\varepsilon\left(\theta^2\right)\right| \leq \epsilon$$

Par exemple, avec $\epsilon = \frac{\frac{1}{2}\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u}}{2}$ (> 0 car $\nabla^2 f(\mathbf{x}^*)$ est DP et $\|\mathbf{u}\| \neq \mathbf{0}$) :

$$\exists \delta > 0\ tq\ \theta \leq \delta \Rightarrow \left|\varepsilon\left(\theta^2\right)\right| \leq \frac{\frac{1}{2}\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u}}{2}$$
$$\Rightarrow \frac{1}{2}\ {}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u} + \varepsilon\left(\theta^2\right) > 0.$$

Par conséquent, en notant $\mathbf{h} = \theta\mathbf{u}$:

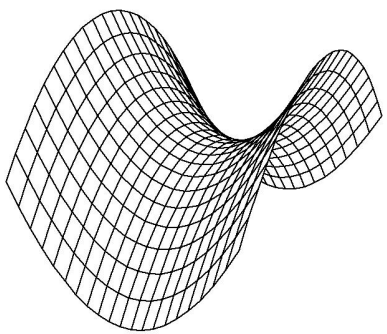
$$\forall \mathbf{h} \in \mathbb{R}^n,\ \exists \delta > 0\ tq\ \|\mathbf{h}\| \leq \delta \Rightarrow f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) > 0.$$

$\mathbf{x}^*$ est donc bien un minimum local.

Remarques :

- On obtient un résultat similaire en inversant le signe de l'inégalité si $\nabla^2 f(\mathbf{x})$ est définie-négative.
- Si $\nabla^2 f(\mathbf{x})$ n'est que *semi* définie-positive, alors $\frac{1}{2}\,{}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u}$ peut être nul. Et comme $\epsilon(\theta^2)$ peut être négatif, on ne peut plus trouver de valeur de θ, même très petite, telle que $\frac{1}{2}\,{}^t\mathbf{u}\nabla^2 f(\mathbf{x}^*)\mathbf{u} + \varepsilon\left(\theta^2\right) > 0$: dans ce cas, $\mathbf{x}^*$ n'est pas un minimum local. On ne peut donc rien conclure. En revanche, on notera la définition suivante :

Si $\nabla^2 f(\mathbf{x}^*)$ est indéfinie (c'est-à-dire si elle n'est ni SDP, ni SDN, ou encore si elle admet deux valeurs propres non nulles et de signe différent), alors $\mathbf{x}^*$ est appelé point selle (ou point d'inflexion en dimension 1).



Si dans le théorème précédent on remplace l'hypothèse $\nabla^2 f(\mathbf{x}^*)$ DP par l'hypothèse plus forte $\nabla^2 f(\mathbf{x})$ DP pour tout $\mathbf{x}$ (et non plus seulement en $\mathbf{x}^*$), alors :

- $\nabla^2 f$ est en particulier DP en $\mathbf{x}^*$ donc $\mathbf{x}^*$ est un minimum local,
- f est strictement convexe (cf théorème §3.3). Le minimum local $\mathbf{x}^*$ est donc aussi l'unique minimum global.

On en déduit le théorème suivant :

théorème : conditions suffisantes d'optimalité globale

Si $f : \mathbb{R}^n \to \mathbb{R}$ est de classe C^2 et si $\mathbf{x}^*$ vérifie :

$$\begin{cases} \nabla f(\mathbf{x}^*) = 0 \\ \nabla^2_f(\mathbf{x}) \text{ définie-positive (respectivement : définie-négative) } \forall \mathbf{x} \in \mathbb{R}^n \end{cases}$$

alors $\mathbf{x}^*$ est l'unique minimum global (resp. : maximum global) de f.

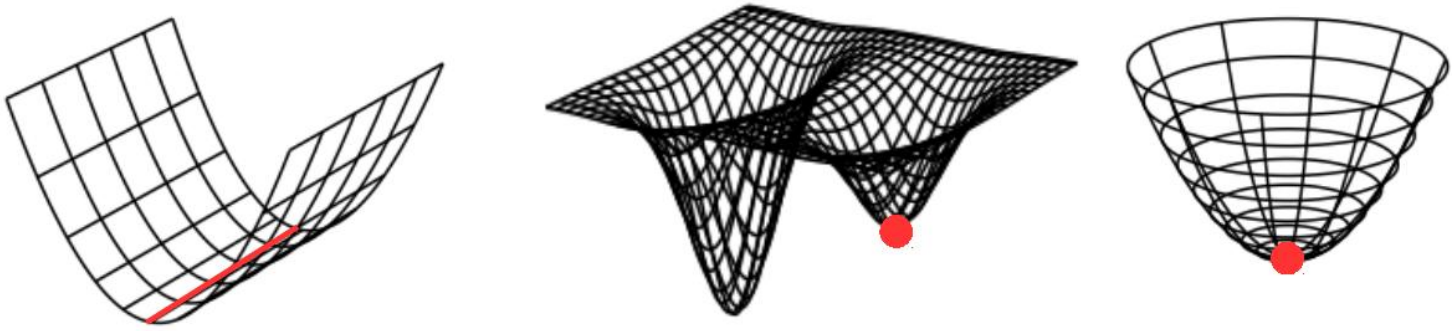
Si maintenant $\nabla^2 f(\mathbf{x})$ est seulement SDP pour tout $\mathbf{x}$, alors f est convexe et $\mathbf{x}^*$ est un minimum global de f, mais il n'est pas nécessairement unique :

théorème : conditions suffisantes d'optimalité globale (2)

Si $f : \mathbb{R}^n \to \mathbb{R}$ est de classe C^2 et si $\mathbf{x}^*$ vérifie :

$$\begin{cases} \nabla f(\mathbf{x}^*) = 0 \\ \nabla^2_f(\mathbf{x}) \text{ semi définie-positive (respectivement : définie-négative) } \forall \mathbf{x} \in \mathbb{R}^n \end{cases}$$

alors $\mathbf{x}^*$ est un minimum global (resp. : maximum global) de f.



SDP partout | DP localement | DP partout

EN RÉSUMÉ :

Soit $f(\mathbf{x})$ une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ deux fois différentiable.

Conditions nécessaires d'optimalité locale :

- Si $\mathbf{x}^*$ est un minimum local, alors :
  - $\nabla f(\mathbf{x}^*) = 0$ (CN d'ordre 1),
  - $\nabla^2 f(\mathbf{x}^*)$ est SDP (CN d'ordre 2).

Conditions suffisantes d'optimalité locale (CS d'ordre 2) :

- Si $\nabla f(\mathbf{x}^*) = 0$ et si $\nabla^2 f(\mathbf{x}^*)$ est :
  - DP, alors $\mathbf{x}^*$ est un minimum local de f,
  - non-définie : $\mathbf{x}^*$ est un point de selle (ou point col).

Conditions suffisantes d'optimalité globale (CS d'ordre 2) :

- Si $\nabla f(\mathbf{x}^*) = 0$ et si $\nabla^2 f(\mathbf{x})$, **pour tout x**, est :
  - SDP, alors f est convexe et $\mathbf{x}^*$ est un minimum global de f,
  - DP, alors f est strictement convexe et $\mathbf{x}^*$ est l'unique minimum global de f.

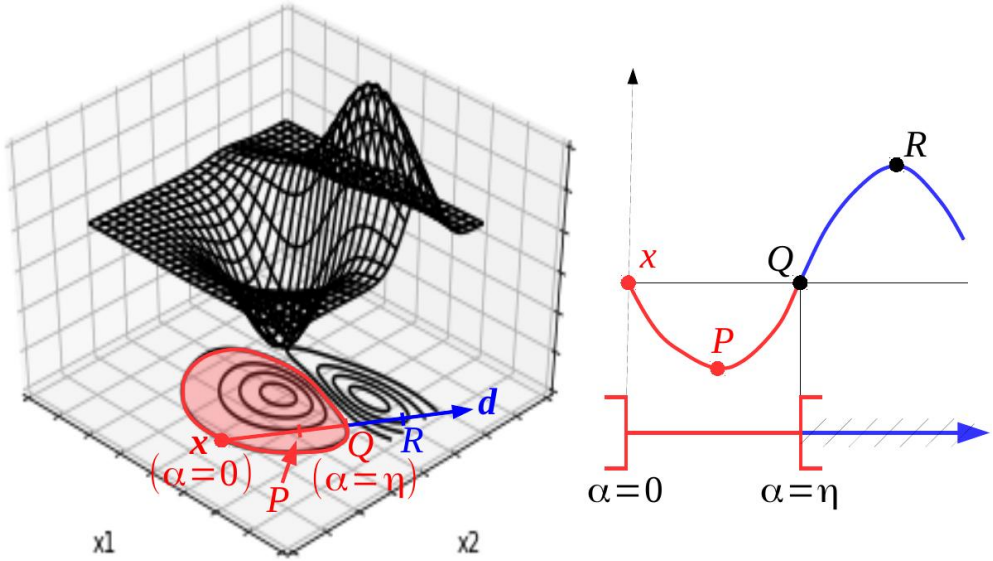
Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable et admettant au moins un minimum.

On cherche à résoudre : $\boxed{\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}).}$

définition : direction de descente

Une direction **d** est appelée **direction de descente** en un point **x** si :

$$\exists \eta > 0 \ tq \ \forall \alpha \in]0, \eta[, \ f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x}).$$



Remarque :

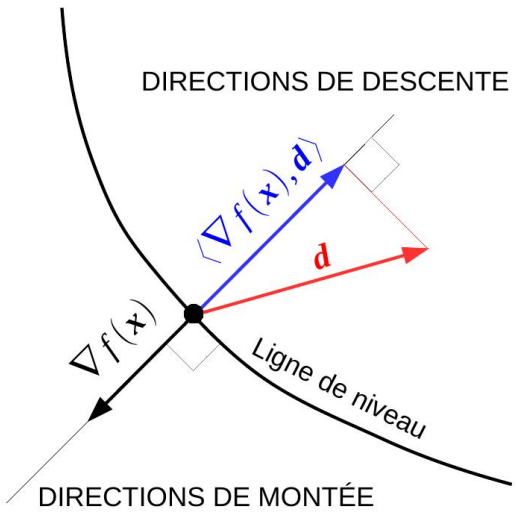
si f est différentiable, sa dérivée directionnelle en $\mathbf{x}$ dans la direction $\mathbf{d}$ s'écrit :

$$D_{\mathbf{d}} f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = {}^t \nabla f(\mathbf{x}) \cdot \mathbf{d}.$$

Or $\mathbf{d}$ est une direction de descente lorsque $D_{\mathbf{d}} f(\mathbf{x}) < 0$, donc :

proposition : caractérisation d'une direction de descente

$\mathbf{d}$ est une direction de descente en $\mathbf{x} \iff {}^t \nabla f(\mathbf{x}_k) \cdot \mathbf{d}_k < 0$



Principe des méthodes de descente :

A partir d'un point initial $\mathbf{x}_0$ choisi arbitrairement, on construit une suite d'itérés $(\mathbf{x}_k)_{k\in\mathbb{N}}$ définie par :

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

où $\mathbf{d}_k \in \mathbb{R}^n$ est une direction de descente de f en $\mathbf{x}_k$ et où $\alpha_k > 0$ est le pas effectué dans cette direction, que l'on choisit tel que $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k)$.

algorithme type

1. $k = 0$, $\mathbf{x}_k = \mathbf{x}_0$.
2. Tant que le critère d'arrêt n'est pas satisfait :
   1. trouver $\mathbf{d}_k$ telle que ${}^t\nabla f(\mathbf{x}_k) \cdot \mathbf{d}_k < 0$,
   2. choisir $\alpha_k > 0$ tel que $f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k)$,
   3. $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$,
   4. $k = k + 1$.
3. $\tilde{\mathbf{x}}^* = \mathbf{x}_k$, $\tilde{\mathbf{y}}^* = f(\mathbf{x}_k)$.

On définit plusieurs critères d'arrêt :

- $\|\nabla f(\mathbf{x}_k)\| \leq \varepsilon$: point stationnaire. Si f strictement convexe ($\nabla^2_f(\mathbf{x}_k)$ DP), c'est un minimum local.
- $|f(\mathbf{x}_k) - f(\mathbf{x}_{k-1})| \leq \varepsilon$ (ou $\leq \varepsilon|f(\mathbf{x}_k)|$) : f ne diminue quasiment plus entre deux itérations.
- $\|\mathbf{x}_k - \mathbf{x}_{k-1}\| \leq \varepsilon$ (ou $\leq \varepsilon\|\mathbf{x}_k\|$) : $\mathbf{x}$ ne varie quasiment plus entre deux itérations.
- $k > k_{max}$: on interrompt l'algorithme à un nombre d'itération maximum fixé à l'avance.
- $t > t_{max}$: on interrompt l'algorithme au bout d'un temps maximum fixé à l'avance.

Il reste à trouver $\mathbf{d}_k$ et α_k à chaque itération pour que $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$.

Méthode du gradient (ou de *plus grande pente*) :

Direction : on choisit la direction de plus grande descente

$$\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$$

Pas de déplacement : on choisit la valeur qui minimise $f(\mathbf{x}_{k+1})$.

$f(\mathbf{x}_{k+1}) = f\left(\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)\right)$ est une fonction de α_k que l'on peut noter $g_k(\alpha_k)$: il faut donc résoudre le problème d'optimisation unidimensionnelle :

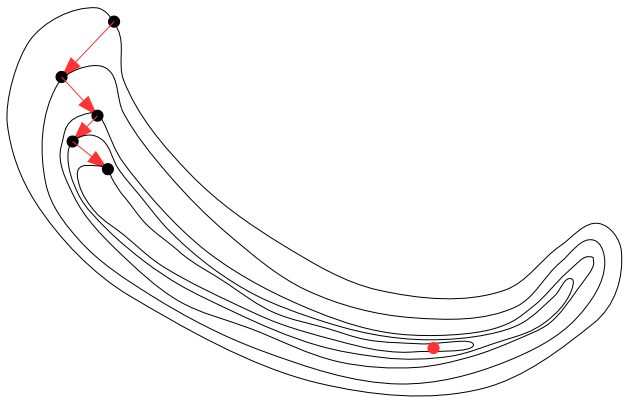
$$\min_{\alpha>0} g_k(\alpha)$$

Si g n'est pas connue explicitement, on se contente de trouver α_k tel que :

$$f\left(\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)\right) < f(\mathbf{x}_k) \Rightarrow \begin{cases} \text{règle de Wolfe} \\ \text{règle d'Armijo} \\ \alpha_k = \mathit{Cste} \end{cases}$$

Méthode du gradient à pas fixe

Dans la pratique, résoudre $\min_{\alpha>0} g_k(\alpha)$ est coûteux et parfois peu efficace car les directions successives sont **orthogonales**, d'où une progression en zig-zag.



On préfère alors choisir α constant :

- suffisamment petit pour assurer la convergence,
- suffisamment grand pour converger rapidement.

Comment trouver le bon compromis ?

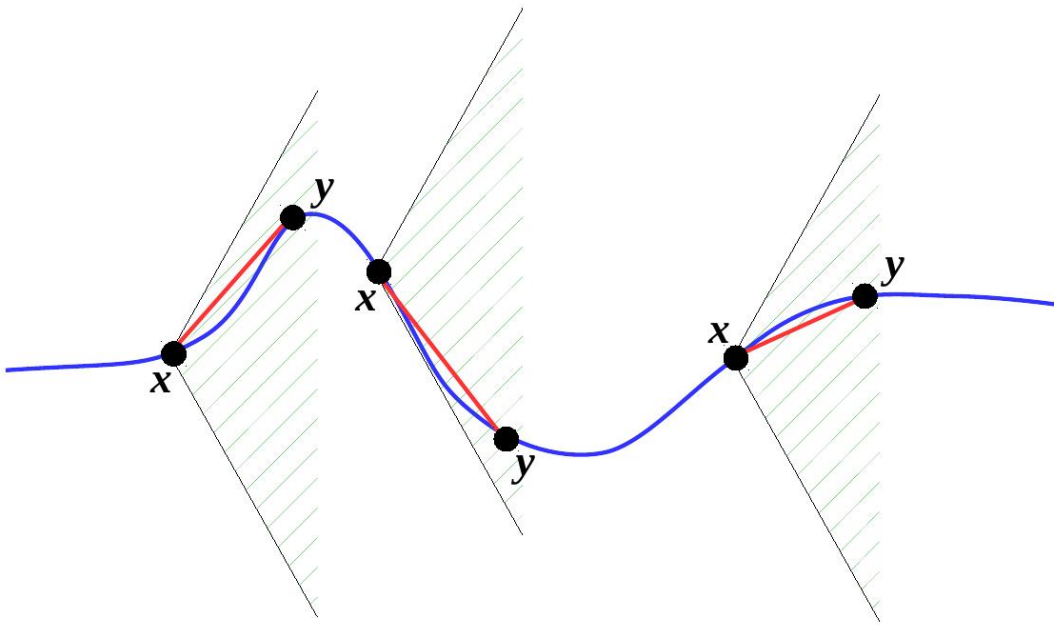
On va supposer que f est différentiable, admet un minimum et qu'elle est de **gradient lipschitz**.

définition : fonction lipschitz

La fonction $g : \mathbb{R}^n \to \mathbb{R}^m$ est dite **lipschitz** sur $\mathbb{R}^n$ ssi il existe $L \in \mathbb{R}$ tq :

$$\forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n,\ \|g(\mathbf{y}) - g(\mathbf{x})\| \leq L\|\mathbf{y} - \mathbf{x}\|$$

Tout segment reliant 2 points du graphe de g a une pente (absolue) inférieure à L.



En particulier : $g : \mathbb{R} \to \mathbb{R}$ lipschitz $\iff$ g' bornée.

Propriétés : si ∇f est lipschitz sur $\mathbb{R}$,

- $\forall \mathbf{x} \in \mathbb{R}^n,\ \left\|\nabla^2 f(\mathbf{x})\right\| \leq L$ (norme de la hessienne inférieure à L),
- $\forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n,\ f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2$.

On en déduit une condition suffisante de convergence de l'algorithme de gradient à pas fixe :

proposition : CS de convergence

Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable, admettant un minimum et telle que ∇f soit L-lipschitz.

1. Si $\alpha < \frac{2}{L}$, alors l'algorithme du gradient à pas fixe est un algorithme de descente,
2. Si $\alpha < \frac{2}{L}$, alors l'algorithme du gradient à pas fixe converge.

Démonstration :

1. Soit $(\mathbf{x}_k)_{k\in\mathbb{N}}$ une suite d'itérés générée par l'algorithme du gradient à pas fixe α à partir d'un point $\mathbf{x}_0$. On a donc $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha\nabla f(\mathbf{x}_k)$
   D'après la 2^{e}propriété ci-dessus :

$$\begin{aligned} f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) &\leq \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &\leq \langle \nabla f(\mathbf{x}_k), -\alpha\nabla f(\mathbf{x}_k) \rangle + \frac{L}{2}\|-\alpha\nabla f(\mathbf{x}_k)\|^2 \\ &\leq -\alpha\|\nabla f(\mathbf{x}_k)\|^2 + \frac{L}{2}\alpha^2\|\nabla f(\mathbf{x}_k)\|^2 \\ &\leq \alpha\left(\frac{L}{2}\alpha - 1\right)\|\nabla f(\mathbf{x}_k)\|^2 \end{aligned}$$

   Par conséquent, $\alpha\left(\frac{L}{2}\alpha - 1\right) < 0 \Rightarrow f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) < 0$, soit :

$$0 < \alpha < \frac{2}{L} \Rightarrow f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$$

   remarque : $f(\mathbf{x}_{k+1})$ mini si $\alpha\left(\frac{L}{2}\alpha - 1\right)$ mini, soit $\alpha^* = \frac{1}{L}$.

2. Supposons que $\alpha < \frac{2}{L}$. D'après le résultat précédent, $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ donc la suite $(f(x_k))_{k\in\mathbb{N}}$ est décroissante. De plus, puisque nous avons supposé que f admet un minimum m, cette suite est minorée. Par conséquent, elle converge.

En outre :

$$f(\mathbf{x}_0) - f(\mathbf{x}_1) \geq \alpha\left(1 - \frac{L}{2}\alpha\right)\|\nabla f(\mathbf{x}_0\|^2$$

$$f(\mathbf{x}_1) - f(\mathbf{x}_2) \geq \alpha\left(1 - \frac{L}{2}\alpha\right)\|\nabla f(\mathbf{x}_1)\|^2$$

$$\vdots$$

$$f(\mathbf{x}_{N-1}) - f(\mathbf{x}_N) \geq \alpha\left(1 - \frac{L}{2}\alpha\right)\|\nabla f(\mathbf{x}_{N-1})\|^2$$

$$f(\mathbf{x}_0) - f(\mathbf{x}_N) \geq \sum_{k=0}^{N-1} \alpha\left(1 - \frac{L}{2}\alpha\right)\|\nabla f(\mathbf{x}_k)\|^2 .$$

Or $f(x_N) \geq m$ donc $f(\mathbf{x}_0) - m \geq f(\mathbf{x}_0) - f(\mathbf{x}_N)$ et $\alpha\left(1 - \frac{L}{2}\alpha\right)$ est une constante $C > 0$. Par conséquent :

$$C \sum_{k=0}^{N-1} \|\nabla f(\mathbf{x}_k)\|^2 \leq f(\mathbf{x}_0) - f(\mathbf{x}_N) \leq f(\mathbf{x}_0) - M$$

$$\Rightarrow \sum_{k=0}^{N-1} \|\nabla f(\mathbf{x}_k)\|^2 \leq \underbrace{\frac{f(\mathbf{x}_0) - M}{C}}_{Cste}$$

Il s'agit d'une série à termes positifs majorée donc convergente. Par conséquent, la suite $\left(\|\nabla f(\mathbf{x}_k)\|^2\right)_{k\in\mathbb{N}}$ converge vers 0, de même que la suite $\left(\|\nabla f(\mathbf{x}_k)\|\right)_{k\in\mathbb{N}}$: la suite $(\mathbf{x}_k)_{k\in\mathbb{N}}$ converge donc vers un point stationnaire.

Méthode de Newton

La méthode de plus forte pente n'utilise que l'information de gradient (ordre 1). En utilisant des informations supplémentaires comme la hessienne (ordre 2), on peut construire des algorithmes plus efficaces.

On sait que les points stationnaires $\bar{\mathbf{x}}$ de f vérifient nécessairement

$$\nabla f(\bar{\mathbf{x}}) = \mathbf{0} \iff \begin{cases} \dfrac{\partial f}{\partial x_1}(\bar{x}_1, \cdots, \bar{x}_n) = 0 \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\bar{x}_1, \cdots, \bar{x}_n) = 0 \end{cases}$$

Il s'agit d'un système de n équations non linéaires à n inconnues. Pour le résoudre, on va procéder itérativement en **linéarisant localement** ∇f au voisinage du point courant $\mathbf{x}_k$, c'est à dire en remplaçant ∇f par son approximation à l'ordre 1 en $\mathbf{x}_k$.

$$\widetilde{\nabla f}_k(\mathbf{x}) = \nabla f(\mathbf{x}_k) + \underbrace{J_{\nabla f}(\mathbf{x}_k)}_{\nabla^2 f(\mathbf{x}_k)}(\mathbf{x} - \mathbf{x}_k).$$

A chaque itération, on cherche donc à résoudre $\widetilde{\nabla f}_k(\bar{\mathbf{x}}_k) = \mathbf{0}$, soit :

$$\nabla^2 f(\mathbf{x}_k)(\bar{\mathbf{x}}_k - \mathbf{x}_k) = -\nabla f(\mathbf{x}_k) \qquad \text{(système linéaire)}.$$

Puisque $\widetilde{\nabla f}_k(\mathbf{x})$ n'est qu'une approximation de ∇f au voisinage de $\mathbf{x}_k$, la solution $\bar{\mathbf{x}}_k$ du système précédent (si $\nabla^2 f(\mathbf{x}_k)$ est inversible, et en particulier si elle est DP) n'est qu'une approximation de $\bar{\mathbf{x}}$. On va donc plutôt considérer que ce système fournit une direction de descente $\mathbf{d}_k = \bar{\mathbf{x}}_k - \mathbf{x}_k$, et la suite des itérés est alors :

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \quad \text{avec} \quad \mathbf{d}_k = -\left(\nabla^2 f(\mathbf{x}_k)\right)^{-1} \nabla f(\mathbf{x}_k)$$

avantages : plus rapide que la méthode de plus forte pente.

inconvénients : calcul de la hessienne + résolution d'un système linéaire, convergence non assurée...

Remarque : on obtient le même résultat si l'on cherche les points stationnaires $\bar{\mathbf{x}}_k$ de l'approximation de f (et non plus de ∇f) à l'ordre 2.

$$\tilde{f}_k(\mathbf{x}) = f(\mathbf{x}_k) + {}^t\nabla f(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k) + \frac{1}{2}\,{}^t(\mathbf{x}-\mathbf{x}_k)\nabla^2 f(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k).$$

Ces points stationnaires vérifient $\nabla \tilde{f}_k(\bar{\mathbf{x}}_k) = 0$, soit :

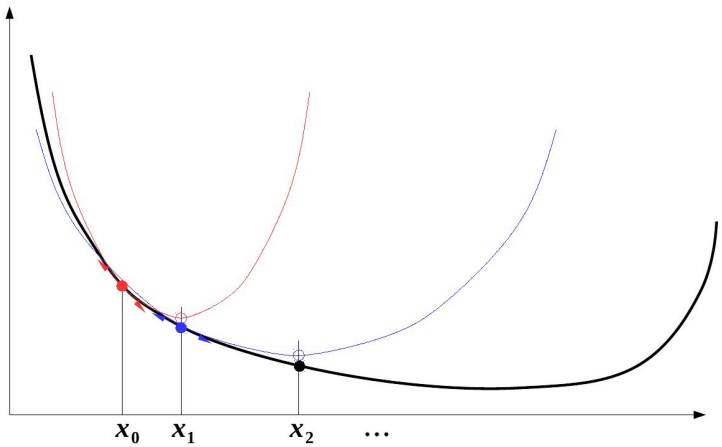
$$\mathbf{0} + \nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)(\bar{\mathbf{x}}_k - \mathbf{x}_k) = 0$$
$$\iff \nabla^2 f(\mathbf{x}_k)(\bar{\mathbf{x}}_k - \mathbf{x}_k) = -\nabla f(\mathbf{x}_k).$$

Cette formulation est appelée **problème quadratique tangent**.

Remarque : $\nabla^2 \tilde{f}_k(\mathbf{x}) = \nabla^2 f(\mathbf{x}_k)$.

$\nabla^2 \tilde{f}_k(\mathbf{x})$ ne dépend donc pas de x et

$$\left.\begin{matrix}\text{DP}\\ \nabla^2 f(\mathbf{x}_k)\ \text{ ou}\\ \text{DN}\end{matrix}\right| \Rightarrow \bar{\mathbf{x}}_k \text{ est extr. glob. unique.}$$



Méthodes quasi-Newton

Plutôt que de calculer la hessienne et résoudre le système linéaire à chaque itération, on va construire une suite de matrices $(\mathbf{H}_k)$ qui approxime $\left(\nabla^2 f(\mathbf{x}_k)\right)^{-1}$ à chaque itération et telle que $\lim_{k\to+\infty} \mathbf{H}_k = \left(\nabla^2 f(\mathbf{x}_k)\right)^{-1}$.

L'algorithme s'écrit alors :

$$\boxed{\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{H}_k \nabla f(\mathbf{x}_k)}$$

En général, $\mathbf{H}_{k+1} = \mathbf{H}_k + \mathbf{Hc}_k$, où $\mathbf{Hc}_k$ est un terme de correction qui permet à $\mathbf{H}_k$ de s'approcher progressivement de $\left(\nabla^2 f(\mathbf{x}_k)\right)^{-1}$ tout en conservant à chaque itération les caractéristiques de l'inverse de la hessienne au voisinage d'un minimum de f, à savoir qu'elle soit :

- symétrique,
- définie-positive.

Il existe plusieurs façons de construire la suite de matrices $\mathbf{H}_k$:

Davidon-Fletcher-Powel (DFP) :

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{\mathbf{H}_k \mathbf{y}_k \, {}^t\mathbf{y}_k \mathbf{H}_k}{{}^t\mathbf{y}_k \mathbf{H}_k \mathbf{y}_k} + \frac{\mathbf{s}_k \, {}^t\mathbf{s}_k}{{}^t\mathbf{y}_k \mathbf{s}_k},$$

Broyden-Fletcher-Goldfarb-Shanno (BFGS) :

$$\mathbf{H}_{k+1} = \left(\mathbf{I} - \frac{\mathbf{s}_k \, {}^t\mathbf{y}_k}{{}^t\mathbf{y}_k \mathbf{s}_k}\right) \mathbf{H}_k \left(\mathbf{I} - \frac{\mathbf{y}_k \, {}^t\mathbf{s}_k}{{}^t\mathbf{y}_k \mathbf{s}_k}\right) + \frac{\mathbf{s}_k \, {}^t\mathbf{s}_k}{{}^t\mathbf{y}_k \mathbf{s}_k}$$

$$\text{ou encore } \mathbf{H}_{k+1} = \mathbf{H}_k - \frac{\mathbf{s}_k \, {}^t\mathbf{y}_k \mathbf{H}_k + \mathbf{H}_k \mathbf{y}_k \, {}^t\mathbf{s}_k}{{}^t\mathbf{s}_k \mathbf{y}_k} + \left(1 + \frac{{}^t\mathbf{y}_k \mathbf{H}_k \mathbf{y}_k}{{}^t\mathbf{s}_k \mathbf{y}_k}\right) \frac{\mathbf{s}_k \, {}^t\mathbf{s}_k}{{}^t\mathbf{s}_k \mathbf{y}_k}.$$

avec $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$, $\quad \mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\quad \mathbf{H}_0 = \mathbf{I}$.

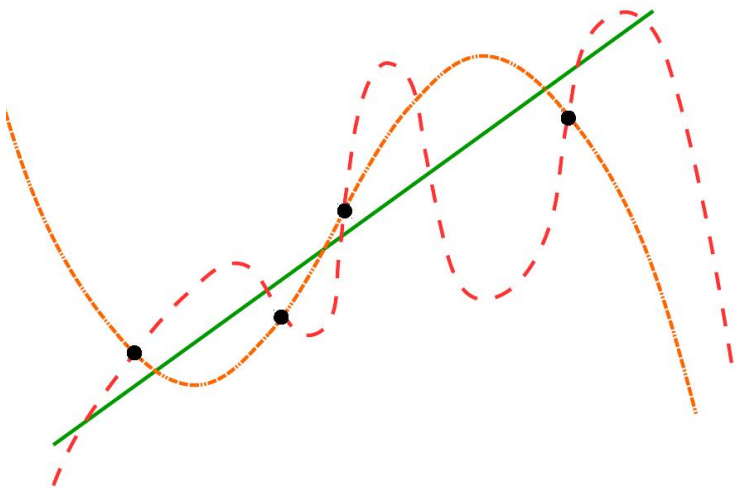
Avantages	Inconvénients
plus rapide que gradient calcul gradient uniquement	un peu moins rapide que Newton

Moindres carrés

On veut représenter un ensemble de points $(\mathbf{x}_i, y_i)_{i=1,\cdots,M}$ par une fonction paramétrique $\tilde{y} = f_{\mathbf{u}}(\mathbf{x})$, où $\mathbf{u} = {}^t(u_1, \cdots, u_N)$ est le vecteur des N paramètres de $f_{\mathbf{u}}$. En général, $M \gg N$.

par exemple : $N = 2$,

$$\mathbf{u} = {}^t(u_0, u_1)$$
$$\text{et } f_{\mathbf{u}}(x) = u_0 + u_1 x$$



On cherche $\mathbf{u}$ tel que $f_{\mathbf{u}}(\mathbf{x})$ passe *au mieux* entre les points $(\mathbf{x}_i, y_i)_{i=1,\cdots,M}$.

$$\boxed{\min_{\mathbf{u}\in\mathbb{R}^N} \sum_{i=1}^{M} (f_{\mathbf{u}}(\mathbf{x}_i) - y_i)^2} \qquad \text{(moindres carrés).}$$

Si $f_{\mathbf{u}}(\mathbf{x})$ dépend **linéairement** de $\mathbf{u}$, alors :

$$f_{\mathbf{u}}(x) = \sum_{k=1}^{N} \alpha_k(\mathbf{x}) u_k = {}^t\boldsymbol{\alpha}(\mathbf{x}) \cdot \mathbf{u}$$

où $\boldsymbol{\alpha}(\mathbf{x}) = {}^t(\alpha_1(\mathbf{x}), \cdots \alpha_N(\mathbf{x}))$. Par exemple, si $f_{\mathbf{u}}(x) = u_1 + u_2 x + u_3 x^2$, on a $\boldsymbol{\alpha}(x) = {}^t(1, x, x^2)$. Pour chacune des M valeurs de $\mathbf{x}_i$, l'écart entre la valeur prédite par $f_{\mathbf{u}}$ et la valeur attendue y_i s'écrit alors :

$$\begin{cases} f_{\mathbf{u}}(\mathbf{x}_1) - y_1 = {}^t\boldsymbol{\alpha}(\mathbf{x}_1) \cdot \mathbf{u} - y_1 \\ f_{\mathbf{u}}(\mathbf{x}_2) - y_2 = {}^t\boldsymbol{\alpha}(\mathbf{x}_2) \cdot \mathbf{u} - y_2 \\ \vdots \\ f_{\mathbf{u}}(\mathbf{x}_M) - y_M = {}^t\boldsymbol{\alpha}(\mathbf{x}_M) \cdot \mathbf{u} - y_M \end{cases}$$

ou encore :

$$\begin{pmatrix} f_{\mathbf{u}}(\mathbf{x}_1) - y_1 \\ \vdots \\ f_{\mathbf{u}}(\mathbf{x}_M) - y_M \end{pmatrix} = \begin{pmatrix} \alpha_1(\mathbf{x}_1) & \cdots & \alpha_N(\mathbf{x}_1) \\ \vdots & & \vdots \\ \alpha_1(\mathbf{x}_M) & \cdots & \alpha_N(\mathbf{x}_M) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix}$$
$$= \mathbf{Au} - \mathbf{y}.$$

donc $\sum_{i=1}^{M} (f_{\mathbf{u}}(\mathbf{x}_i) - y_i)^2 = {}^t(\mathbf{Au} - \mathbf{y}) \cdot (\mathbf{Au} - \mathbf{y}) = \|\mathbf{Au} - \mathbf{y}\|^2$ et le problème des moindres carrés s'écrit :

$$\boxed{\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{Au} - \mathbf{y}\|^2} \qquad \text{(MC)}$$

où $\mathbf{A} \in \mathbb{R}^{M \times N}$ et $\mathbf{y} \in \mathbb{R}^M$.

définition : équations normales

L'équation

$$ {}^t\mathbf{A}\mathbf{A}\mathbf{u} = {}^t\mathbf{A}\mathbf{y} \qquad \text{(EN)} $$

est appelé **système d'équations normales** associé au problème (MC).

théorème : EN $\iff$ MC

$\mathbf{u}^*$ est solution de (EN) si et seulement si $\mathbf{u}^*$ est solution de (MC).
De plus, si $\mathbf{A}$ est de rang plein, $\mathbf{u}^*$ est l'unique solution.

- $\mathbf{A}$ $(m \times n)$ avec $m \geq n$ est de rang plein si toutes ses colonnes sont linéairement indépendantes : $\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$,

Rappels :

- ${}^t\mathbf{x}\mathbf{y} = {}^t\mathbf{y}\mathbf{x}$,
- ${}^t(\mathbf{A}\mathbf{u} - \mathbf{y}) = {}^t(\mathbf{A}\mathbf{u}) - {}^t\mathbf{y}$,
- ${}^t(\mathbf{A}\mathbf{u}) = {}^t\mathbf{u}\,{}^t\mathbf{A}$.

Démonstration : Notons $g(\mathbf{u}) = \frac{1}{2}\|\mathbf{Au}-\mathbf{y}\|^2 = \frac{1}{2}\,{}^t(\mathbf{Au}-\mathbf{y})\cdot(\mathbf{Au}-\mathbf{y})$.
On a :

$$\begin{aligned}
{}^t(\mathbf{Au}-\mathbf{y})\cdot(\mathbf{Au}-\mathbf{y}) &= ({}^t\mathbf{u}\,{}^t\mathbf{A} - {}^t\mathbf{y})\cdot(\mathbf{Au}-\mathbf{y}),\\
&= {}^t\mathbf{u}\,{}^t\mathbf{AAu} - {}^t\mathbf{u}\,{}^t\mathbf{Ay} - {}^t\mathbf{yAu} + {}^t\mathbf{yy},\\
&= {}^t\mathbf{u}\,{}^t\mathbf{AAu} - {}^t\mathbf{u}\,{}^t\mathbf{Ay} - {}^t({}^t\mathbf{Ay})\mathbf{u} + {}^t\mathbf{yy},\\
&= {}^t\mathbf{u}\,{}^t\mathbf{AAu} - {}^t\mathbf{u}\,{}^t\mathbf{Ay} - {}^t\mathbf{u}\,{}^t\mathbf{Ay} + {}^t\mathbf{yy},\\
&= {}^t\mathbf{u}\,{}^t\mathbf{AAu} - 2\,{}^t\mathbf{u}\,{}^t\mathbf{Ay} + {}^t\mathbf{yy},
\end{aligned}$$

d'où :

$$\begin{aligned}
\nabla g(\mathbf{u}) &= \frac{1}{2}\nabla_{\mathbf{u}}({}^t\mathbf{u}\,{}^t\mathbf{AAu}) - \nabla_{\mathbf{u}}({}^t\mathbf{u}\,{}^t\mathbf{Ay}) + \frac{1}{2}\nabla_{\mathbf{u}}({}^t\mathbf{yy}),\\
&= {}^t\mathbf{AAu} - {}^t\mathbf{Ay} + \mathbf{0}\\
\text{et } \nabla^2 g(\mathbf{u}) &= {}^t\mathbf{AA} - \mathbf{0}
\end{aligned}$$

Par conséquent :

MC $\Rightarrow$ EN : si $\mathbf{u}^*$ est un minimum local de g, alors $\nabla g(\mathbf{u}) = \mathbf{0}$ et ${}^t\mathbf{A}\mathbf{A}\mathbf{u} = {}^t\mathbf{A}\mathbf{y}$,

EN $\Rightarrow$ MC : si ${}^t\mathbf{A}\mathbf{A}\mathbf{u} = {}^t\mathbf{A}\mathbf{y}$, alors $\nabla g(\mathbf{u}) = \mathbf{0}$ et $\mathbf{u}^*$ est un point critique de g.
De plus, $\forall \mathbf{z} \neq \mathbf{0}$, ${}^t\mathbf{z}\,({}^t\mathbf{A}\mathbf{A})\,\mathbf{z} = {}^t(\mathbf{A}\mathbf{z})\,\mathbf{A}\mathbf{z} = \|\mathbf{A}\mathbf{z}\|^2 \geq 0$ donc ${}^t\mathbf{A}\mathbf{A} = \nabla^2 g(\mathbf{u})$ est SDP quel que soit $\mathbf{u}$: g est donc convexe et $\mathbf{u}^*$ est un minimum global de g.

Si en outre $\mathbf{A}$ est de rang plein, alors $\mathbf{A}\mathbf{z} = \mathbf{0} \Rightarrow \mathbf{z} = \mathbf{0}$ et $\nabla^2 g(\mathbf{u})$ est DP quel que soit $\mathbf{u}$: g est donc strictement convexe et $\mathbf{u}^*$ est le minimum global unique de g.

Remarque : lorsque ${}^t\mathbf{A}\mathbf{A}$ est DP, elle est inversible et :

$$\boxed{\mathbf{u}^* = ({}^t\mathbf{A}\mathbf{A})^{-1}\,{}^t\mathbf{A}\mathbf{y}.}$$

Sommes finies

Dans le problème précédent, on minimisait un critère de moindres carrés pour lequel la fonction était linéaire en ses paramètres :

$$\min_{\mathbf{u}\in\mathbb{R}^N} \sum_{i=1}^{M} (f_{\mathbf{u}}(\mathbf{x}_i) - y_i)^2.$$

On va ici généraliser ce problème de minimisation d'une somme finie à une fonction quelconque ψ :

$$\boxed{\min_{\mathbf{u}\in\mathbb{R}^N} \sum_{i=1}^{M} \psi(\mathbf{u}, \mathbf{x}_i).}$$

Les méthodes de gradient (plus forte pente, Newton, quasi-Newton) nécessitent de calculer $\nabla_{\mathbf{u}} \sum_{i=1}^{M} \psi(\mathbf{u}, \mathbf{x}_i) = \sum_{i=1}^{M} \nabla_{\mathbf{u}} \psi(\mathbf{u}, \mathbf{x}_i)$ à chaque itération

$$\Rightarrow \text{ irréalisable si } M \text{ est très grand.}$$

On va donc *estimer* $\nabla_{\mathbf{u}} \sum_{i=1}^{M} \psi(\mathbf{u}, \mathbf{x}_i)$ à partir d'un échantillon réduit de $P \ll M$ valeurs de $\mathbf{x_i}$: c'est la méthode du **gradient stochastique**.

Algorithme du gradient stochastique

1. $k = 0$, $\mathbf{u_k} = \mathbf{u_0}$.
2. Tant que le critère d'arrêt n'est pas satisfait :
   - tirer aléatoirement P valeurs de i dans $\{1, \cdots, M\}$, soit $\{i_1, \cdots, i_P\}$,
   - calculer les P gradients $\nabla_{\mathbf{u}}\psi(\mathbf{u}_k, \mathbf{x}_{i_p})_{p\in\{i_1,\cdots,i_P\}}$ et $\mathbf{d}_k = \sum_{p=1}^{P} \nabla_{\mathbf{u}}\psi(\mathbf{u}_k, \mathbf{x}_{i_p})$,
   - $\mathbf{u}_{k+1} = \mathbf{u}_k - \alpha_k \mathbf{d}_k$,
   - $k = k + 1$.
3. $\widetilde{\mathbf{u}}^* = \mathbf{u}_k$.

Remarques :

1. $\sum_{p=1}^{P} \nabla_{\mathbf{u}} \psi(\mathbf{u}_k, \mathbf{x}_{i_p})$ n'est pas nécessairement une direction de descente pour le critère complet $\Rightarrow$ inutile de chercher α_k qui réduit sa valeur. On choisira une suite (α_k) prédéfinie (par exemple $\alpha_k = Cste$).

2. Le critère d'arrêt est généralement un nombre maxi d'itération ou un temps de calcul.

Convergence des algorithmes

définition : convergence

On dit qu'un algorithme converge vers une solution $\mathbf{x}^*$ lorsque la suite $(\mathbf{x}_k)$ converge vers un point $\mathbf{x}^*$, c'est-à-dire lorsque :

$$\lim_{k \to +\infty} \|\mathbf{x}_k - \mathbf{x}^*\| = 0.$$

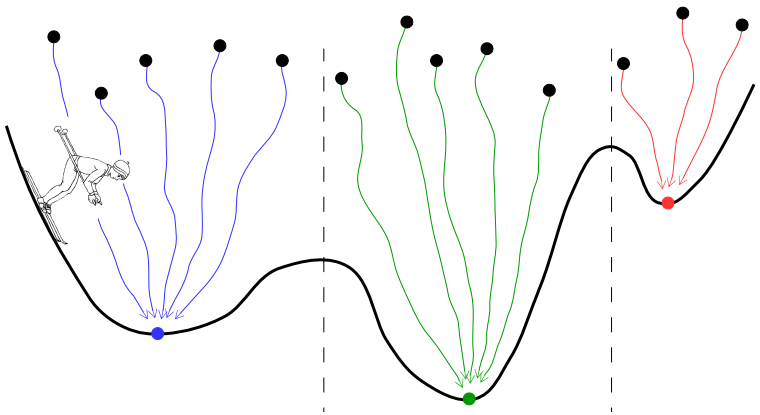
définition : convergence locale

Un algorithme converge **localement** si partant d'une solution initiale $\mathbf{x}_0$ voisine de $\mathbf{x}^*$ il converge vers $\mathbf{x}^*$.

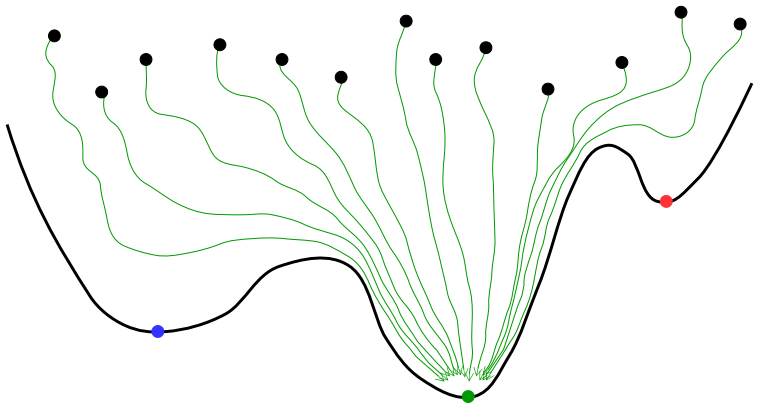
définition : convergence globale

Un algorithme converge **globalement** s'il converge vers $\mathbf{x}^*$ quelle que soit la solution initiale $\mathbf{x}_0$.

Exemple : la méthode de la plus forte pente va conduire à l'optimum situé dans la "vallée" à laquelle appartient la solution initiale.



Convergence locale



Convergence globale

définition : taux de convergence

Soit une suite $(\mathbf{x}_k) \in \mathbb{R}^n$ qui converge vers $\mathbf{x}^* \in \mathbb{R}^n$ avec $\forall k \in \mathbb{N},\ \mathbf{x}_k \neq \mathbf{x}^*$.
La convergence de $(\mathbf{x}_k) \in \mathbb{R}^n$ est

- **linéaire** s'il existe $\tau \in]0,1[$ tel que $\lim_{k\to+\infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = \tau$.
  Le n$^{\text{bre}}$ de chiffres significatifs exacts augmente d'une valeur asymptotiquement constante à chaque itération $\rightarrow$ plus grande pente,
- **superlinéaire** si $\lim_{k\to+\infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 0$.
  Le n$^{\text{bre}}$ de chiffres significatifs exacts augmente d'une valeur asymptotiquement infinie à chaque itération $\rightarrow$ quasi-Newton,
- d'**ordre p** si $\exists p > 1 \in \mathbb{N}, \exists \tau > 0 \in \mathbb{R}$ *tq* $\lim_{k\to+\infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|^p} = \tau$.
  Le n$^{\text{bre}}$ de chiffres significatifs exacts double asymptotiquement à chaque itération $\rightarrow$ Newton.

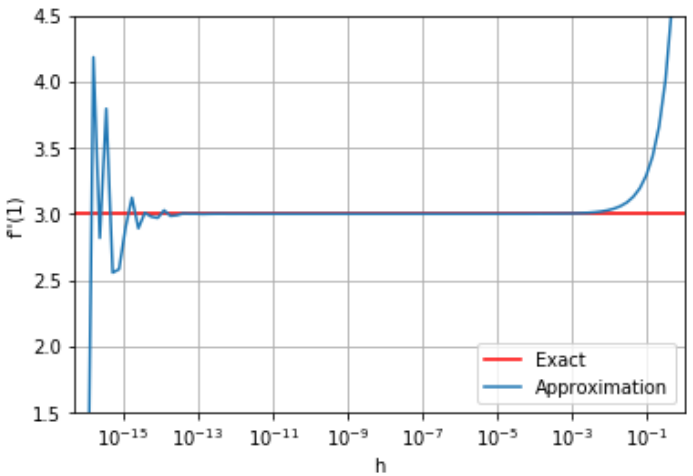
Approximation numérique du gradient

Dans la pratique, on ne connait pas l'expression mathématique de f (fonction "boîte noire") mais seulement son expression numérique $\tilde{f}(\mathbf{x})$, et on ne sait donc pas calculer ∇f. On estime alors ses composantes par :

$$\frac{\partial f}{\partial x_i} \simeq \Delta_i(\mathbf{x}, h) = \frac{\tilde{f}(\mathbf{x} + h\mathbf{e_i}) - \tilde{f}(\mathbf{x})}{h}$$

h doit être :

- suffisamment petit pour que $\dfrac{f(\mathbf{x} + h\mathbf{e_i}) - f(\mathbf{x})}{h}$ soit proche de $\dfrac{\partial f}{\partial x_i}(\mathbf{x})$,
- mais pas trop sinon $\tilde{f}(\mathbf{x} + h\mathbf{e_i})$ et $\tilde{f}(\mathbf{x})$ sont trop proches pour que leur différence puisse être représentée par l'ordinateur avec une précision suffisante.



Dérivée de $f(x) = x^3$ en $x = 1$

En notant :

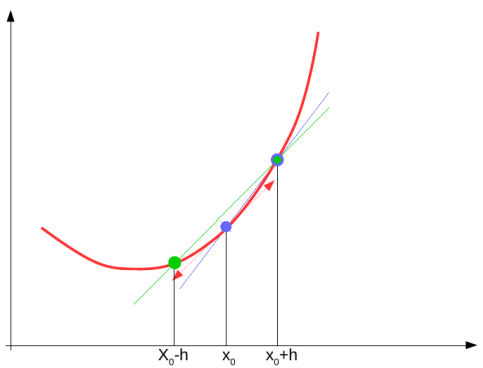
- $\varepsilon_A = \sup_{\mathbf{x}\in[\mathbf{x},\mathbf{x}+h\mathbf{e_i}]} \left|\tilde{f}(\mathbf{x}) - f(\mathbf{x})\right|$ l'erreur machine maxi sur $[\mathbf{x}, \mathbf{x}+h\mathbf{e_i}]$,
- $\varepsilon_B = \sup_{\mathbf{x}\in[\mathbf{x},\mathbf{x}+h\mathbf{e_i}]} |f''(\mathbf{x})|$ une borne de $\nabla^2 f$ sur $[\mathbf{x}, \mathbf{x}+h\mathbf{e_i}]$,

on montre que :

1. $\left|\frac{\partial f}{\partial x_i}(\mathbf{x}) - \Delta_i(x,h)\right| \leq \frac{2\varepsilon_A}{h} + \frac{\varepsilon_B h}{2}$ (borne de l'erreur d'approximation du gradient),
2. cette borne est mini pour $h^* = 2\sqrt{\frac{\varepsilon_A}{\varepsilon_B}}$.

On peut améliorer cette approximation :

$$\delta(x,h) = \frac{\tilde{f}(\mathbf{x}+h\mathbf{e_i}) - \tilde{f}(\mathbf{x}-h\mathbf{e_i})}{2h}$$



On cherche maintenant $\mathbf{x}^*$ qui minimise $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ tout en respectant des contraintes. Le problème s'écrit :

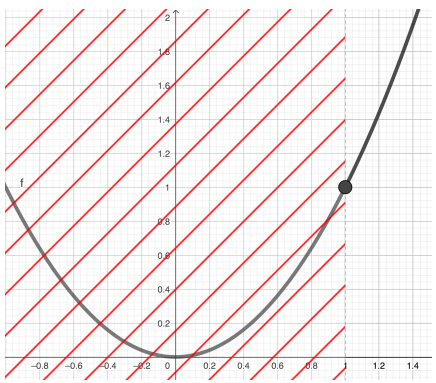
$$(\mathcal{P}) : \quad \min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) \quad \text{s.c.} \quad \begin{vmatrix} g_j(\mathbf{x}) \leq 0 \ \forall j \in \{1, \cdots, J\}, \\ h_k(\mathbf{x}) = 0 \ \forall k \in \{1, \cdots, K\} \end{vmatrix}$$

Attention :

la condition $\nabla f(\bar{\mathbf{x}}) = 0$ n'est maintenant plus nécessaire pour que $\bar{\mathbf{x}}$ soit un point stationnaire, et a fortiori un extremum.

Exemple :

$$\min_{x\in\mathbb{R}} x^2 \quad \text{s.c. } x \geq 1$$



définition : lagrangien et multiplicateurs de Lagrange

On appelle **lagrangien** d'un problème d'optimisation sous la forme standard $(\mathcal{P})$ la fonction $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ définie par :

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) + \sum_{k=1}^{K} \mu_k h_k(\mathbf{x})$$
$$= f(\mathbf{x}) + {}^t\boldsymbol{\lambda}\mathbf{g}(\mathbf{x}) + {}^t\boldsymbol{\mu}\mathbf{h}(\mathbf{x})$$

où $\boldsymbol{\lambda}$ et $\boldsymbol{\mu}$ sont appelés **multiplicateurs de Lagrange**.

Exemple :

$$\max_{\mathbf{x}\in\mathbb{R}^2} x_1^2 + x_2^2 \quad \text{s.c.} \begin{vmatrix} x_1 - x_2^2 \geq 1, \\ x_1 + x_2 \leq 2. \end{vmatrix} \longrightarrow \min_{\mathbf{x}\in\mathbb{R}^2} -(x_1^2 + x_2^2) \quad \text{s.c.} \begin{vmatrix} -x_1 + x_2^2 + 1 \leq 0, \\ x_1 + x_2 - 2 \leq 0. \end{vmatrix}$$

$$\Rightarrow \boxed{L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -(x_1^2 + x_2^2) + \lambda_1(-x_1 + x_2^2 + 1) + \lambda_2(x_1 + x_2 - 2).}$$

théorème : conditions nécessaires d'optimalité de KKT

Soit $I = \{j \ tq \ g_j(\bar{\mathbf{x}}) = 0\}$ (ensemble des contraintes actives).
Si $\forall (j,k) \in I \times \{1, \cdots, K\}$ les gradients $\nabla g_j(\bar{\mathbf{x}})$ et $\nabla h_k(\bar{\mathbf{x}})$ sont linéairement indépendants [a] et si $\bar{\mathbf{x}}$ est une solution de $(\mathcal{P})$, alors $\bar{\mathbf{x}}$ satisfait les conditions suivantes dites *conditions de Karush-Kuhn-Tucker (KKT)* :

$$\exists \boldsymbol{\lambda}, \boldsymbol{\mu} \ tq \ : \quad (\text{KKT}) \begin{cases} \nabla f(\bar{\mathbf{x}}) + \sum_{j=1}^{J} \lambda_j \nabla g_j(\bar{\mathbf{x}}) + \sum_{k=1}^{K} \mu_k \nabla h_k(\bar{\mathbf{x}}) = \mathbf{0}, \\ g_j(\bar{\mathbf{x}}) \leq 0 \ (j = 1, \cdots, J), \\ h_k(\bar{\mathbf{x}}) = 0 \ (k = 1, \cdots, K), \\ \lambda_j g_j(\bar{\mathbf{x}}) = 0 \ (j = 1, \cdots, J), \\ \lambda_j \geq 0. \end{cases}$$

a. C'est-à-dire si l'ensemble des gradients de toutes les contraintes actives ($\nabla g_j(\bar{\mathbf{x}})$ et $\nabla h_k(\bar{\mathbf{x}})$) forme une famille libre.

Interprétation des conditions nécessaires de KKT

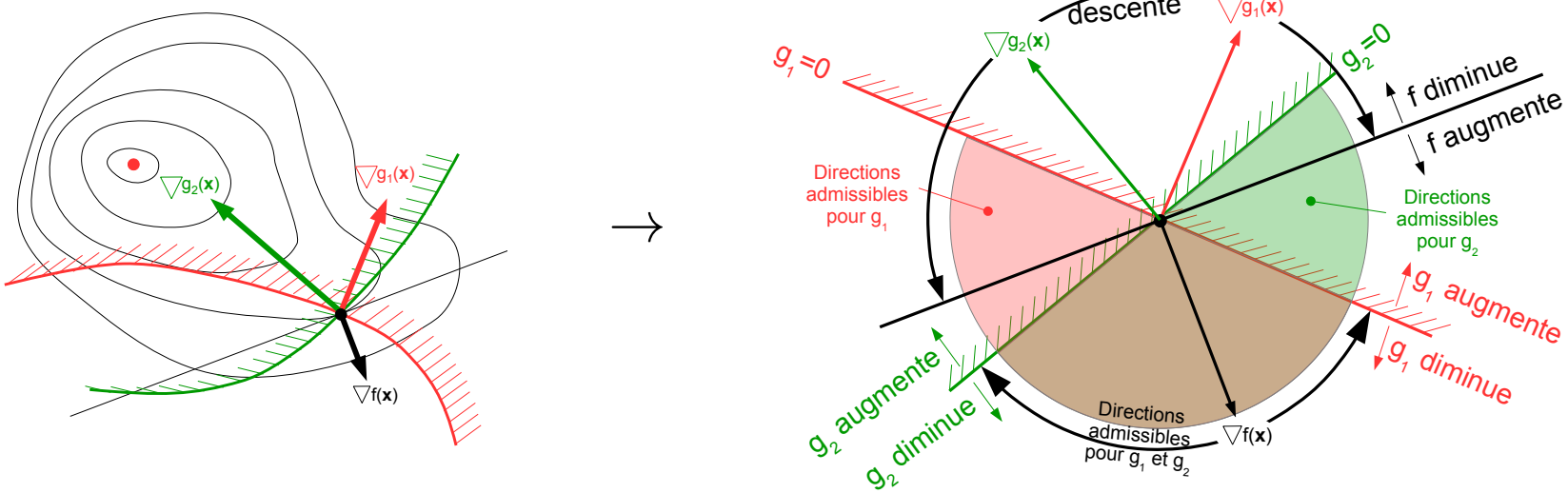
- L'indépendance des contraintes actives s'appelle la **contrainte de qualification**.
- 1re condition : le gradient du lagrangien doit être nul à l'optimum (condition nécessaire d'optimalité pour un problème non contraint).
- 2e et 3e conditions : contraintes satisfaites à l'optimum.
- 4e condition :
  - soit g_j est active à l'optimum ($g_j(\bar{\mathbf{x}}) = 0$) et dans ce cas $\lambda_j \neq 0$,
  - soit g_j n'est pas active ($g_j(\bar{\mathbf{x}}) \neq 0$) et dans ce cas $\lambda_j = 0$: la contrainte n'intervient donc pas dans le lagrangien.
- 5e condition : un accroissement de g_j (qui rendrait donc sa valeur positive et ferait que cette contrainte ne serait plus respectée) augmenterait la valeur de $\lambda_j g_j(\mathbf{x})$ qui peut être vu comme une pénalisation de la fonction objectif f dans le lagrangien.
- Pas de contrainte de signe sur μ_k : le dernier terme du lagrangien peut donc s'écrire indifféremment $+\sum_{k=1}^{K} \mu_k \nabla h_k(\bar{\mathbf{x}})$ ou $-\sum_{k=1}^{K} \mu_k \nabla h_k(\bar{\mathbf{x}})$

Un point $\bar{\mathbf{x}}$ qui satisfait les conditions de KKT est appelé **point de KKT**.

Les conditions de KKT traduisent le fait qu'à l'optimum, le gradient de f doit être une combinaison linéaire négative des gradients des contraintes actives

$$\nabla f(\bar{\mathbf{x}}) = -\sum_{j=1}^{J} \lambda_j \nabla g_j(\bar{\mathbf{x}}) - \sum_{k=1}^{K} \mu_k \nabla h_k(\bar{\mathbf{x}})$$

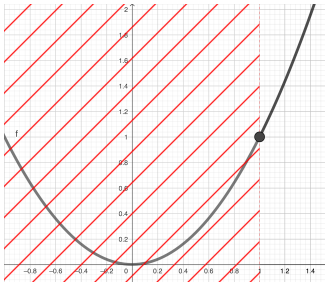
c'est-à-dire qu'il n'existe plus de direction de descente admissible :



Exemple :

$\min_{x\in\mathbb{R}} f(x) = x^2$

s.c. $x \geq 1 \Rightarrow g(x) = 1 - x \leq 0.$



f et g sont continues
f est coercive ($\lim_{|x|\to\infty} f(x) = +\infty$)
donc $\exists$ solution globale.

Son lagrangien est $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = x^2 + \lambda(1 - x)$ et KKT s'écrit :

$$(\text{KKT}) \begin{cases} 2x - \lambda = 0 & \text{(1a)} \\ 1 - x \leq 0 & \text{(1b)} \\ \lambda(1 - x) = 0 & \text{(1c)} \\ \lambda \geq 0. & \text{(1d)} \end{cases}$$

1. g n'est pas active $\Rightarrow g(x) = 1 - x < 0 \overset{(1c)}{\Longrightarrow} \lambda = 0 \overset{(1a)}{\Longrightarrow} x = 0$: contradiction avec l'hypothèse $g(x) = 1 - x < 0$, donc g est active.
2. g est active $\Rightarrow g(x) = 1 - x = 0 \Rightarrow x = 1 \overset{(1a)}{\Longrightarrow} \lambda = 2 \geq 0$ (1d) : les conditions de KKT sont vérifiées.

Par conséquent, $x^* = 1$: c'est bien la solution du problème.

Remarquons que puisqu'il n'y a qu'une seule contrainte, on vérifie sa qualification par le fait que $\nabla g(\mathbf{x}^*) = g'(1) = 1 \neq 0$.

Algorithmes d'optimisation sous contraintes : méthode SQP

On suppose que l'on connaît les contraintes actives. Les autres contraintes ($g_i(\mathbf{x}) < 0$) n'interviennent pas et le problème s'écrit donc :

$$(\mathcal{P}) \; : \; \begin{array}{l} \min\limits_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.c. } \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{array}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ et $h : \mathbb{R}^n \to \mathbb{R}^m$ sont de classe C^2. Le lagrangien du problème s'écrit :

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \mu_1 h_1(\mathbf{x}) + \cdots + \mu_m h_m(\mathbf{x}) = f(\mathbf{x}) + {}^t\boldsymbol{\mu}\mathbf{h}(\mathbf{x})$$

et les conditions de KKT s'écrivent $\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{0} \\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{cases}$

Il s'agit d'un système non linéaire de $n+m$ équations à $n+m$ inconnues : c'est un problème de recherche de zéro que l'on écrit $\mathbf{F}(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{0}$ où

$$\begin{aligned} \mathbf{F} : \mathbb{R}^{n+m} &\longrightarrow \mathbb{R}^{n+m} \\ (\mathbf{x}, \boldsymbol{\mu}) &\longmapsto \mathbf{F}(\mathbf{x}, \boldsymbol{\mu}) = \begin{bmatrix} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) \\ \mathbf{h}(\mathbf{x}) \end{bmatrix} \end{aligned}$$

Pour résoudre ce système, on va procéder itérativement en linéarisant F localement au voisinage du point courant $(\mathbf{x}_k, \boldsymbol{\mu}_k)$, c'est-à-dire en l'assimilant à la partie régulière de son développement limité à l'ordre 1 :

$$\widetilde{\mathbf{F}}_k(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{F}(\mathbf{x}_k, \boldsymbol{\mu}_k) + \mathbf{J}_F(\mathbf{x}_k, \boldsymbol{\mu}_k) \begin{pmatrix} \mathbf{x} - \mathbf{x}_k \\ \boldsymbol{\mu} - \boldsymbol{\mu}_k \end{pmatrix}$$

L'itéré $(\mathbf{x}_{k+1}, \boldsymbol{\mu}_{k+1})$ est alors la solution de $\widetilde{\mathbf{F}}_k(\mathbf{x}_{k+1}, \boldsymbol{\mu}_{k+1}) = 0$ à l'itération k, qui est un problème linéaire :

$$\mathbf{J}_F(\mathbf{x}_k, \boldsymbol{\mu}_k) \begin{pmatrix} \mathbf{x}_{k+1} - \mathbf{x}_k \\ \boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k \end{pmatrix} = -\mathbf{F}(\mathbf{x}_k, \boldsymbol{\mu}_k).$$

Explicitons la matrice $\mathbf{J}_F(\mathbf{x}_k, \boldsymbol{\mu}_k)$ et le vecteur $\mathbf{F}(\mathbf{x}_k, \boldsymbol{\mu}_k)$.

$$\mathbf{J}_F(\mathbf{x}_k, \boldsymbol{\mu}_k) = \begin{pmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial L}{\partial x_1}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial L}{\partial x_1}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) & \frac{\partial}{\partial \mu_1}\left(\frac{\partial L}{\partial x_1}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) & \cdots & \frac{\partial}{\partial \mu_m}\left(\frac{\partial L}{\partial x_1}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial L}{\partial x_n}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) & \cdots & \frac{\partial}{\partial x_n}\left(\frac{\partial L}{\partial x_n}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) & \frac{\partial}{\partial \mu_1}\left(\frac{\partial L}{\partial x_n}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) & \cdots & \frac{\partial}{\partial \mu_m}\left(\frac{\partial L}{\partial x_n}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) \\ \frac{\partial h_1}{\partial x_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}_k) & \frac{\partial h_1}{\partial \mu_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_1}{\partial \mu_m}(\mathbf{x}_k) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}_k) & \frac{\partial h_m}{\partial \mu_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_m}{\partial \mu_m}(\mathbf{x}_k) \end{pmatrix}$$

Mais $\frac{\partial L}{\partial x_i}(\mathbf{x}_k, \boldsymbol{\mu}_k) = \frac{\partial f}{\partial x_i}(\mathbf{x}_k) + \boldsymbol{\mu}_1 \frac{\partial h_1}{\partial x_i}(\mathbf{x}_k) + \cdots + \boldsymbol{\mu}_m \frac{\partial h_m}{\partial x_i}(\mathbf{x}_k)$, donc :

$$\frac{\partial}{\partial \mu_j}\left(\frac{\partial L}{\partial x_i}\right)(\mathbf{x}_k, \boldsymbol{\mu}_k) = \frac{\partial h_j}{\partial x_i}(\mathbf{x}_k)$$

De plus, $\frac{\partial h_i}{\partial \mu_j}(\mathbf{x}_k) = 0$ (h ne dépend pas de $\boldsymbol{\mu}$). On a donc :

$$\mathbf{J}_F(\mathbf{x}_k, \boldsymbol{\mu}_k) = \left(\begin{array}{ccc|ccc} \frac{\partial^2 L}{\partial x_1^2}(\mathbf{x}_k, \boldsymbol{\mu}_k) & \cdots & \frac{\partial^2 L}{\partial x_n \partial x_1}(\mathbf{x}_k, \boldsymbol{\mu}_k) & \frac{\partial h_1}{\partial x_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_m}{\partial x_1}(\mathbf{x}_k) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial^2 L}{\partial x_1 \partial x_n}(\mathbf{x}_k, \boldsymbol{\mu}_k) & \cdots & \frac{\partial^2 L}{\partial x_n^2}(\mathbf{x}_k, \boldsymbol{\mu}_k) & \frac{\partial h_1}{\partial x_n}(\mathbf{x}_k) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}_k) \\ \hline \frac{\partial h_1}{\partial x_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}_k) & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1}(\mathbf{x}_k) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}_k) & 0 & \cdots & 0 \end{array}\right)$$

ou encore : $\mathbf{J}_F(\mathbf{x}_k, \boldsymbol{\mu}_k) = \begin{bmatrix} \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k) & {}^t\mathbf{J}_h(\mathbf{x}_k) \\ \mathbf{J}_h(\mathbf{x}_k) & \mathbf{0} \end{bmatrix}$

Par ailleurs, $\mathbf{F}(\mathbf{x}, \boldsymbol{\mu}) = \begin{bmatrix} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) \\ \mathbf{h}(\mathbf{x}) \end{bmatrix}$, avec :

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) + \mu_1 \frac{\partial h_1}{\partial x_1}(\mathbf{x}) + \cdots + \mu_m \frac{\partial h_m}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) + \mu_1 \frac{\partial h_1}{\partial x_n}(\mathbf{x}) + \cdots + \mu_m \frac{\partial h_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots, \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix} + \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial h_m}{\partial x_1}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial h_1}{\partial x_n}(\mathbf{x}) & \cdots & \frac{\partial h_m}{\partial x_n}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$$

$$= \nabla f(\mathbf{x}) + {}^t\mathbf{J}_h(\mathbf{x})\boldsymbol{\mu}$$

Finalement, à chaque itération, $\mathbf{x}_{k+1}$ et $\boldsymbol{\mu}_{k+1}$ sont solution du système **linéaire** suivant :

$$\begin{bmatrix} \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k) & {}^t\mathbf{J}_h(\mathbf{x}_k) \\ \mathbf{J}_h(\mathbf{x}_k) & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{x}_{k+1} - \mathbf{x}_k \\ \boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k \end{pmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}_k) - {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\mu}_k \\ -\mathbf{h}(\mathbf{x}_k) \end{bmatrix}$$

Remarque : Ces équations sont les équations de KKT du sous-problème suivant, appelé **problème quadratique tangent** :

$$\begin{cases} \min\limits_{\mathbf{d_x}} {}^t\nabla f(\mathbf{x}_k)\mathbf{d_x} + \frac{1}{2}\,{}^t\mathbf{d_x}\nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d_x} \\ \text{s.c. : } \mathbf{h}(\mathbf{x}_k) + \mathbf{J}_h(\mathbf{x}_k)\mathbf{d_x} = \mathbf{0} \text{ (approximation linéaire de } \mathbf{h}). \end{cases} \Bigg| \begin{array}{l} \text{(critère hybride, avec } \nabla f \text{ dans} \\ \text{la partie linéaire et } \nabla^2 L \text{ dans} \\ \text{la partie quadratique)} \end{array}$$

où $\mathbf{d_x} = \mathbf{x} - \mathbf{x}_k$.

En effet, le lagrangien de ce sous-problème s'écrit :

$$\mathcal{L}(\mathbf{d_x}, \boldsymbol{\lambda}) = {}^t\nabla f(\mathbf{x}_k)\mathbf{d_x} + \frac{1}{2}\,{}^t\mathbf{d_x}\nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d_x} + {}^t(\mathbf{h}(\mathbf{x}_k) + \mathbf{J}_h(\mathbf{x}_k)\mathbf{d_x})\boldsymbol{\lambda}$$

Son gradient est donc :

$$\nabla_{\mathbf{d_x}}\mathcal{L}(\mathbf{d_x}, \boldsymbol{\lambda}) = \nabla f(\mathbf{x}_k) + \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d_x} + {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\lambda}$$

Les équations de KKT (conditions nécessaires d'optimalité) du problème quadratique tangent sont donc :

$$\begin{cases} \nabla_{\mathbf{d_x}}\mathcal{L}(\mathbf{d_x}, \boldsymbol{\lambda}) = \mathbf{0}, \\ \mathbf{h}(\mathbf{x}_k) + \mathbf{J}_h(\mathbf{x}_k)\mathbf{d_x} = \mathbf{0}. \end{cases} \Leftrightarrow \begin{cases} \nabla f(\mathbf{x}_k) + \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d_x} + {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\lambda} = \mathbf{0}, \\ \mathbf{h}(\mathbf{x}_k) + \mathbf{J}_h(\mathbf{x}_k)\mathbf{d_x} = \mathbf{0}. \end{cases}$$

On introduit $- {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\mu}_k$ des deux côtés du signe égale dans la première équation :

$$\begin{cases} \nabla f(\mathbf{x}_k) + \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d}_\mathbf{x} + {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\lambda} - {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\mu}_k = - {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\mu}_k, \\ \mathbf{h}(\mathbf{x}_k) + \mathbf{J}_h(\mathbf{x}_k)\mathbf{d}_\mathbf{x} = \mathbf{0}. \end{cases}$$

$$\Leftrightarrow \begin{cases} \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d}_\mathbf{x} + {}^t\mathbf{J}_h(\mathbf{x}_k)(\boldsymbol{\lambda} - \boldsymbol{\mu}_k) = -\nabla f(\mathbf{x}_k) - {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\mu}_k, \\ \mathbf{J}_h(\mathbf{x}_k)\mathbf{d}_\mathbf{x} = -\mathbf{h}(\mathbf{x}_k). \end{cases}$$

soit encore, en posant $\boldsymbol{\lambda} = \boldsymbol{\mu}_{k+1}$:

$$\begin{bmatrix} \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k) & {}^t\mathbf{J}_h(\mathbf{x}_k) \\ \mathbf{J}_h(\mathbf{x}_k) & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{x}_{k+1} - \mathbf{x}_k \\ \boldsymbol{\mu}_{k+1} - \boldsymbol{\mu}_k \end{pmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}_k) - {}^t\mathbf{J}_h(\mathbf{x}_k)\boldsymbol{\mu}_k, \\ -\mathbf{h}(\mathbf{x}_k) \end{bmatrix}.$$

qui est bien la solution du problème initial.

Algorithme SQP (programmation quadratique séquentielle)

1. $k = 0$, $\mathbf{x}_k = \mathbf{x}_0$, $\boldsymbol{\mu}_k = \boldsymbol{\mu}_0$
2. Tant que critère d'arrêt non satisfait :
   - Évaluer $\mathbf{h}(\mathbf{x}_k)$, $\nabla f(\mathbf{x}_k)$, $\mathbf{J}_h(\mathbf{x}_k)$ et $\nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)$.
   - Résoudre le problème d'optimisation quadratique suivant :

$$\begin{cases} \min\limits_{\mathbf{d_x}} {}^t\nabla f(\mathbf{x}_k)\mathbf{d_x} + \dfrac{1}{2}\,{}^t\mathbf{d_x}\nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \boldsymbol{\mu}_k)\mathbf{d_x}, \\ \text{s.c. : } \mathbf{h}(\mathbf{x}_k) + \mathbf{J}_h(\mathbf{x}_k)\mathbf{d_x} = \mathbf{0}, \end{cases}$$

   - $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d_x}, \quad \boldsymbol{\mu}_{k+1} = \boldsymbol{\lambda}, \quad k := k + 1.$

Remarque : Lorsqu'il existe des contraintes inégalités $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, on se ramène au problème précédent en choisissant un ensemble de contraintes actives $\mathcal{A}$ qui sont considérées comme des contraintes égalité, les autres étant tout simplement ignorées. Le choix des contraintes inégalités à mettre dans $\mathcal{A}$ peut être guidé par la **méthode des contraintes actives** :

Algorithme des contraintes actives

1. Choisir une contrainte active i : $\mathcal{A} = \{i\}$.
2. Résoudre le problème d'optimisation sous contraintes égalités (SQP)
3. Si l'une des contraintes inégalité $j \notin \mathcal{A}$ est violée, alors $\mathcal{A} = \mathcal{A} \bigcup \{j\}$ et revenir au point 2
4. Si l'un des multiplicateurs de Lagrange λ_i est strictement négatif, alors $\mathcal{A} = \mathcal{A} - \{i\}$ et revenir au point 2
5. Sinon, on a trouvé la solution.

Méthodes de pénalités

Principe : remplacer le problème $(\mathcal{P})$ d'optimisation sous contraintes par une suite de problèmes $(\mathcal{P}_R)$ non contraints :

$$(\mathcal{P}) \ : \ \begin{cases} \min\limits_{\mathbf{x}} \ f(\mathbf{x}) \\ \text{s.c.} \ : \begin{vmatrix} \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ \mathbf{h}(\mathbf{x}) = \mathbf{0}. \end{vmatrix} \end{cases} \quad \rightsquigarrow \quad (\mathcal{P}_R) \ : \ \min_{\mathbf{x}} \ f(\mathbf{x}) + \Omega(R, \mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}))$$

où $\begin{vmatrix} R : \text{paramètre mis à jour à chaque itération,} \\ \Omega : \text{fonction de pénalisation.} \end{vmatrix}$

Pénalité parabolique : $\Omega(R, h(\mathbf{x})) = R\,(h(\mathbf{x}))^2$.

Pénalité logarithmique : $\Omega(R, g(\mathbf{x})) = -R \log(-g(\mathbf{x}))$.

Pénalité inverse : $\Omega(R, g(\mathbf{x})) = \dfrac{R}{g(\mathbf{x})}$.

Méthode des points intérieurs

Dans la méthode SQP, on de ramène à une succession de problèmes d'optimisation sous contraintes d'**égalité** en faisant des hypothèses sur les contraintes actives.

Avec la méthode des points intérieurs, on ne fait plus d'hypothèse sur les contraintes actives mais on transforme les contraintes inégalités en contraintes égalités par l'introduction de variables supplémentaires :

$$(\mathcal{P}) \ : \ \begin{cases} \min\limits_{\mathbf{x}} \ f(\mathbf{x}) \\ \text{s.c.} \ : \begin{vmatrix} \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ \mathbf{h}(\mathbf{x}) = \mathbf{0}. \end{vmatrix} \end{cases} \quad \rightsquigarrow \quad (\mathcal{P}_{\mathbf{s}}) \ : \ \begin{cases} \min\limits_{\mathbf{x},\mathbf{s}} \ f(\mathbf{x}) - \tau \sum\limits_{j=1}^{J} \log s_j \\ \text{s.c.} \ : \begin{vmatrix} \mathbf{g}(\mathbf{x}) + \mathbf{s} = \mathbf{0}, \\ \mathbf{h}(\mathbf{x}) = \mathbf{0}. \end{vmatrix} \end{cases}$$

$\mathbf{s}$ est appelé *vecteur des variables d'écart*.

$\tau > 0$ est un *paramètre de pénalisation*.

- Les s_j sont nécessairement > 0 donc $\mathbf{g}(\mathbf{x}) < \mathbf{0} \longrightarrow$points intérieurs.
- $\tau > 0$ décroit à chaque itération $\longrightarrow \tau \sum \log s_j$ a de moins en moins d'influence.

Le lagrangien de $(\mathcal{P}_\mathbf{s})$ est :

$$L(\mathbf{x}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \tau) = f(\mathbf{x}) - \tau \sum_{j=1}^{J} \log s_j + \sum_{j=1}^{J} \lambda_j (g_j(\mathbf{x}) + s_j) + \sum_{k=1}^{K} \mu_k h_k(\mathbf{x})$$

d'où :

$$\nabla_{\mathbf{x},\mathbf{s}} L(\mathbf{x}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \tau) = \begin{pmatrix} \frac{\partial f}{\partial x_1} + \lambda_1 \frac{\partial g_1}{\partial x_1} + \cdots + \lambda_J \frac{\partial g_J}{\partial x_1} + \mu_1 \frac{\partial h_1}{\partial x_1} + \cdots + \mu_K \frac{\partial h_K}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} + \lambda_1 \frac{\partial g_1}{\partial x_N} + \cdots + \lambda_J \frac{\partial g_J}{\partial x_N} + \mu_1 \frac{\partial h_1}{\partial x_N} + \cdots + \mu_K \frac{\partial h_K}{\partial x_N} \\ -\frac{\tau}{s_1} + \lambda_1 \\ \vdots \\ -\frac{\tau}{s_J} + \lambda_J \end{pmatrix}$$

En imposant $\lambda_j - \frac{\tau}{s_j} = 0$ et en tenant compte du fait que $\mathbf{s}$ doit être positif, les équations de KKT de $(\mathcal{P}_\mathbf{s})$ s'écrivent donc :

$$(KKT_\mathbf{s}) \ : \quad \begin{cases} \nabla f(\bar{\mathbf{x}}) + {}^t\boldsymbol{\lambda}\nabla \mathbf{g}(\bar{\mathbf{x}}) + {}^t\boldsymbol{\mu}\nabla \mathbf{h}(\bar{\mathbf{x}}) = \mathbf{0}, \\ \mathbf{g}(\bar{\mathbf{x}}) + \mathbf{s} = 0, \\ \mathbf{h}(\bar{\mathbf{x}}) = 0, \\ \lambda_j s_j = \tau \ \ \forall j \in \{1, \cdots, J\}, \\ \mathbf{s} > \mathbf{0} \\ \boldsymbol{\lambda} \geq \mathbf{0}. \end{cases}$$

Il s'agit des équations de KKT du problème initial $(\mathcal{P})$ dans lesquelles on a introduit une perturbation $\mathbf{s}$.

Les méthodes d'optimisation sous contraintes vues précédemment consistent à déterminer une solution $\mathbf{x}^*$ en introduisant des multiplicateurs associés $\boldsymbol{\lambda}^*$. Ces méthodes sont appelées **méthodes primales** et les multiplicateurs $\boldsymbol{\lambda}^*$ n'en sont qu'un sous-produit.

Avec les **méthodes duales**, l'effort est porté sur la recherche de $\boldsymbol{\lambda}^*$ et c'est cette fois la solution $\mathbf{x}^*$ qui en est un sous-produit. Ces méthodes consistent à transformer le problème primal $(\mathcal{P})$, dont la solution est $\mathbf{x}^*$, en un problème dual $(\mathcal{D})$ dont la solution est $\boldsymbol{\lambda}^*$.
On cherche donc à résoudre le problème d'optimisation sous contraintes suivant :

$$(\mathcal{P}) \ : \quad \begin{cases} \min\limits_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) \\ \text{s.c.} \ \begin{vmatrix} g_j(\mathbf{x}) \leq 0 \ \forall j \in \{1, \cdots, J\}, \\ h_k(\mathbf{x}) = 0 \ \forall k \in \{1, \cdots, K\} \end{vmatrix} \end{cases}$$

Le lagrangien de ce problème est :

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + {}^t\boldsymbol{\lambda}\mathbf{g}(\mathbf{x}) + {}^t\boldsymbol{\mu}\mathbf{h}(\mathbf{x})$$

définition : fonction duale

On appelle **fonction duale** (ou **fonction duale lagrangienne**) la fonction :

$$\begin{array}{rccl} \phi & : & \mathbb{R}^{J+K} & \longrightarrow & \mathbb{R} \cup \{-\infty, +\infty\} \\ & & \boldsymbol{\lambda}, \boldsymbol{\mu} & \longrightarrow & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf\limits_{\mathbf{x} \in \mathbb{R}^N} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \end{array}$$

Rappel : $\inf\limits_{\mathbf{x} \in \mathbb{R}^N} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R} \cup \{-\infty, +\infty\}$.

définition : problème dual

Soit $X_\phi = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^{J+K} \ tq \ \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}$.
On appelle **problème dual** de $\mathcal{P}$ le problème d'optimisation suivant :

$$(\mathcal{D}) \ : \quad \begin{cases} \max\limits_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^{J+K}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.c. :} \begin{vmatrix} \lambda_i \geq 0 \ \ \forall i \in \{1, \cdots, J\}, \\ (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in X_\phi \end{vmatrix} \end{cases}$$

$(\mathcal{P})$ est appelé **problème primal** de $(\mathcal{D})$

théorème : dualité faible

Soit $\mathbf{x}^*$ une solution du problème primal $(\mathcal{P})$ et soit $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ une solution du problème dual $(\mathcal{D})$. Alors :

$$\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq f(\mathbf{x}^*)$$

La solution du problème dual nous donne ainsi une indication (borne inférieure) sur le problème primal.

Dans le cas d'un problème convexe linéaire, c'est-à-dire si f et g_j sont convexes et h_k est linéaire, la **dualité forte** s'écrit :

$$\boxed{\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*).}$$

et il est alors équivalent de résoudre $(\mathcal{P})$ ou $(\mathcal{D})$.

Exemple :

$$(\mathcal{P}) \ : \quad \boxed{\begin{array}{l} \min\limits_{\mathbf{x}\in\mathbb{R}^2} \ x_1^2 + x_2^2 \\ \text{s.c.} \quad x_1 \geq 1 \end{array}}$$

Son lagrangien $\mathcal{P}$ s'écrit :

$$L(\mathbf{x}, \lambda) = x_1^2 + x_2^2 + \lambda(1 - x_1)$$

1. Résolution du problème primal

$$\text{KKT} : \begin{cases} 2x_1 - \lambda = 0 \\ 2x_2 = 0 \implies x_2 = 0 \\ x1 \geq 1 \\ \lambda(1 - x_1) = 0 \\ \lambda \geq 0 \end{cases}$$

- $\lambda = 0 \implies x_1 = 0 < 1$: impossible,
- $\lambda > 0 \implies x_1 = 1$ et $\lambda = 2 > 0$,

donc $\lambda^* = 2$, $\mathbf{x}^* = (1, 0)$ et $f(\mathbf{x}^*) = 1$.

2 Résolution du problème dual

La fonction duale de $(\mathcal{P})$ est $\phi(\lambda) = \inf_{\mathbf{x}\in\mathbb{R}^2} L(\mathbf{x},\lambda) = \inf_{\mathbf{x}\in\mathbb{R}^2} \left[x_1^2 + x_2^2 + \lambda(1-x_1)\right]$

$$L(\mathbf{x},\lambda)\text{minimum} \implies \nabla_{\mathbf{x}} L(\mathbf{x},\lambda) = \mathbf{0} \implies \begin{cases} 2x_1^* - \lambda = 0 \\ 2x_2^* = 0. \end{cases} \implies \begin{cases} x_1^* = \dfrac{\lambda}{2} \\ x_2^* = 0. \end{cases}$$

N.B. : il s'agit bien d'un minimum car $\nabla^2_{\mathbf{x}} L(\mathbf{x}^*,\lambda) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ est DP.

Par conséquent : $\phi(\lambda) = \frac{\lambda^2}{4} + 0 + \lambda - \frac{\lambda^2}{2} = \lambda - \frac{\lambda^2}{4}$.

En outre, $X_\phi = \{\boldsymbol{\lambda} \in \mathbb{R} \ tq \ \lambda - \frac{\lambda^2}{4} > -\infty\} = \mathbb{R}$. Le problème dual s'écrit donc :

$$\mathcal{D} \ : \quad \boxed{\begin{array}{l} \max\limits_{\lambda\in\mathbb{R}} \ \lambda - \dfrac{\lambda^2}{4} \\ \text{s.c.} \quad \lambda \geq 0 \end{array}}$$

$$\lambda - \frac{\lambda^2}{4} \text{ maxi } \Rightarrow \frac{d}{d\lambda}(\lambda - \frac{\lambda^2}{4}) = 0$$
$$\Rightarrow 1 - \frac{\lambda}{2} = 0$$
$$\Rightarrow \lambda^* = 2 \text{ donc } x_1^* = 1 \text{ et } x_2^* = 0.$$

Finalement, $\phi(\lambda^*) = 1 = f(x_1^*, x_2^*)$. Il n'y a pas de saut de dualité.

Cas particulier d'un problème primal linéaire :

$$(\mathcal{P}) \; : \; \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} \; {}^t\mathbf{c}\mathbf{x} \\ \text{s.c.} \quad \begin{vmatrix} \mathbf{A}\mathbf{x} = \mathbf{b} \quad (\mathbf{A} : \text{ matrice } m \times n) \\ \mathbf{x} \geq \mathbf{0} \; (\iff \; -x_i \leq 0 \; \forall i = 1, \cdots, n) \end{vmatrix} \end{cases}$$

Son lagrangien s'écrit $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \; {}^t\mathbf{c}\mathbf{x} - {}^t\boldsymbol{\lambda}\mathbf{x} + {}^t\boldsymbol{\mu}(\mathbf{b} - \mathbf{A}\mathbf{x})$

et sa fonction duale est : $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf\limits_{\mathbf{x}\in\mathbb{R}^n} \; \left[{}^t\mathbf{c}\mathbf{x} - {}^t\boldsymbol{\lambda}\mathbf{x} + {}^t\boldsymbol{\mu}(\mathbf{b} - \mathbf{A}\mathbf{x})\right]$

$$= \inf_{\mathbf{x}\in\mathbb{R}^n} \; \left[{}^t\boldsymbol{\mu}\mathbf{b} - \; {}^t\left({}^t\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\lambda} - \mathbf{c}\right)\mathbf{x}\right]$$

$$= \begin{cases} {}^t\boldsymbol{\mu}\mathbf{b} & \text{si } \; {}^t\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\lambda} - \boldsymbol{c} = \mathbf{0}, \\ -\infty & \text{sinon.} \end{cases}$$

Pour que $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ soit bornée, il faut donc que ${}^t\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\lambda} - \boldsymbol{c} = \mathbf{0}$.

Le problème dual s'écrit alors :

$$\begin{cases} \max\limits_{\boldsymbol{\mu},\boldsymbol{\lambda}} \ {}^t\boldsymbol{\mu}\mathbf{b} \\ \text{s.c.} \ \begin{vmatrix} \boldsymbol{\lambda} \geq \mathbf{0} \\ {}^t\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\lambda} - \boldsymbol{c} = \mathbf{0}. \end{vmatrix} \end{cases} \iff \begin{cases} \max\limits_{\boldsymbol{\mu}\in\mathbb{R}^m} \ {}^t\mathbf{b}\boldsymbol{\mu} \\ \text{s.c.} \ \begin{vmatrix} \boldsymbol{\lambda} \geq \mathbf{0} \\ \boldsymbol{\lambda} = \boldsymbol{c} - {}^t\mathbf{A}\boldsymbol{\mu}. \end{vmatrix} \end{cases} \iff \boxed{\begin{cases} \max\limits_{\boldsymbol{\mu}\in\mathbb{R}^m} \ {}^t\mathbf{b}\boldsymbol{\mu} \\ \text{s.c.} \ {}^t\mathbf{A}\boldsymbol{\mu} \leq \boldsymbol{c}. \end{cases}}$$

Considérons maintenant le problème primal suivant : $\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} \ -{}^t\mathbf{b}\mathbf{x} \\ \text{s.c.} \ \mathbf{A}\mathbf{x} \leq \boldsymbol{c}. \end{cases}$

Son lagrangien s'écrit $L(\mathbf{x},\boldsymbol{\lambda}) = -{}^t\mathbf{b}\mathbf{x} + {}^t\boldsymbol{\lambda}(\mathbf{A}\mathbf{x} - \boldsymbol{c})$

et sa fonction duale est

$$\begin{aligned} \phi(\boldsymbol{\lambda}) &= \inf_{\mathbf{x}\in\mathbb{R}^n} \left[-{}^t\mathbf{b}\mathbf{x} + {}^t\boldsymbol{\lambda}(\mathbf{A}\mathbf{x} - \boldsymbol{c}) \right] \\ &= \inf_{\mathbf{x}\in\mathbb{R}^n} \left[{}^t({}^t\mathbf{A}\boldsymbol{\lambda} - \mathbf{b})\mathbf{x} - {}^t\boldsymbol{\lambda}\boldsymbol{c} \right] \\ &= \begin{cases} -{}^t\boldsymbol{\lambda}\mathbf{c} & \text{si } {}^t\mathbf{A}\boldsymbol{\lambda} - \boldsymbol{b} = \mathbf{0}, \\ -\infty & \text{sinon.} \end{cases} \end{aligned}$$

Pour que $\phi(\boldsymbol{\lambda})$ soit bornée, il faut donc que ${}^t\mathbf{A}\boldsymbol{\lambda} - \boldsymbol{b} = \mathbf{0}$.

Le problème dual s'écrit alors : $\begin{cases} \max\limits_{\boldsymbol{\lambda}\in\mathbb{R}^m} & -{}^t\boldsymbol{\lambda}\mathbf{c} \\ \text{s.c.} & : \begin{vmatrix} {}^t\mathbf{A}\boldsymbol{\lambda} - \boldsymbol{b} = \mathbf{0} \\ \boldsymbol{\lambda} \geq \mathbf{0} \end{vmatrix} \end{cases} \iff \boxed{\begin{cases} \min\limits_{\boldsymbol{\lambda}\in\mathbb{R}^m} & {}^t\mathbf{c}\boldsymbol{\lambda} \\ \text{s.c.} & : \begin{vmatrix} {}^t\mathbf{A}\boldsymbol{\lambda} = \boldsymbol{b} \\ \boldsymbol{\lambda} \geq \mathbf{0} \end{vmatrix} \end{cases}}$

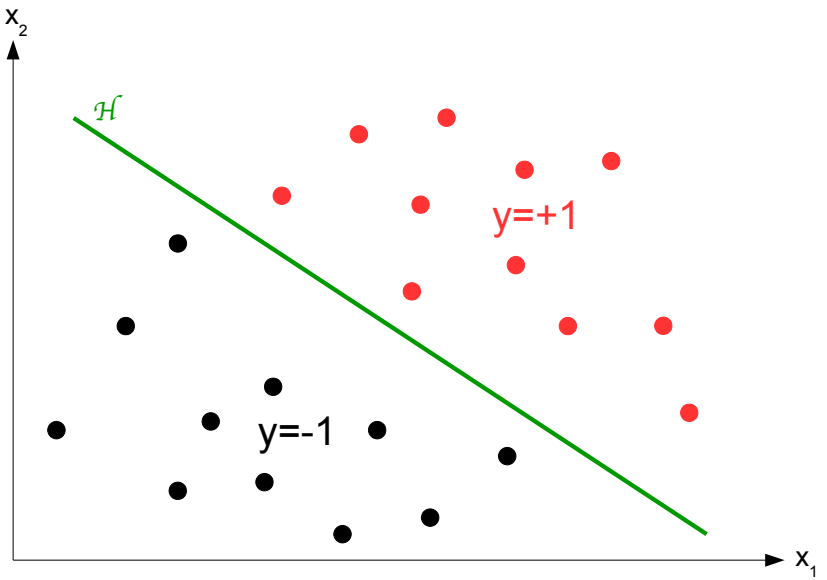
On retrouve le problème original : **le dual du dual (appelé bidual) est le primal**.

PRIMAL		DUAL
$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & {}^t\mathbf{c}\mathbf{x} \\ \text{s.c.} & \begin{vmatrix} \mathbf{A}\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{vmatrix} \end{cases}$	$\longleftrightarrow$	$\begin{cases} \max\limits_{\mathbf{y}\in\mathbb{R}^m} & {}^t\mathbf{b}\mathbf{y} \\ \text{s.c.} & {}^t\mathbf{A}\mathbf{y} - \boldsymbol{c} \leq \mathbf{0}. \end{cases}$
$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & {}^t\mathbf{c}\mathbf{x} \\ \text{s.c.} & \mathbf{A}\mathbf{x} + \boldsymbol{b} \leq \mathbf{0}. \end{cases}$	$\longleftrightarrow$	$\begin{cases} \max\limits_{\mathbf{y}\in\mathbb{R}^m} & {}^t\mathbf{b}\mathbf{y} \\ \text{s.c.} & : \begin{vmatrix} {}^t\mathbf{A}\mathbf{y} + \boldsymbol{c} = \mathbf{0} \\ \mathbf{y} \geq \mathbf{0} \end{vmatrix} \end{cases}$

Les séparateurs à vaste marge

Soit un ensemble de points $\mathbf{x}_{i_{(i=1,\cdots,M)}}$ de $\mathbb{R}^n$ à chacun desquels on associe une valeur $y_i = \pm 1$ qui permet de les partager en deux classes :

$$\begin{cases} \mathcal{C}_{-1} = \{\mathbf{x}_i \ tq \ y_i = -1\} \\ \qquad \text{et} \\ \mathcal{C}_{+1} = \{\mathbf{x}_i \ tq \ y_i = +1\} \end{cases}$$



On suppose ces deux classes séparées linéairement par un hyperplan :

$$\mathcal{H} \ : \ {}^t\mathbf{u}\mathbf{x} + u_0 = 0$$

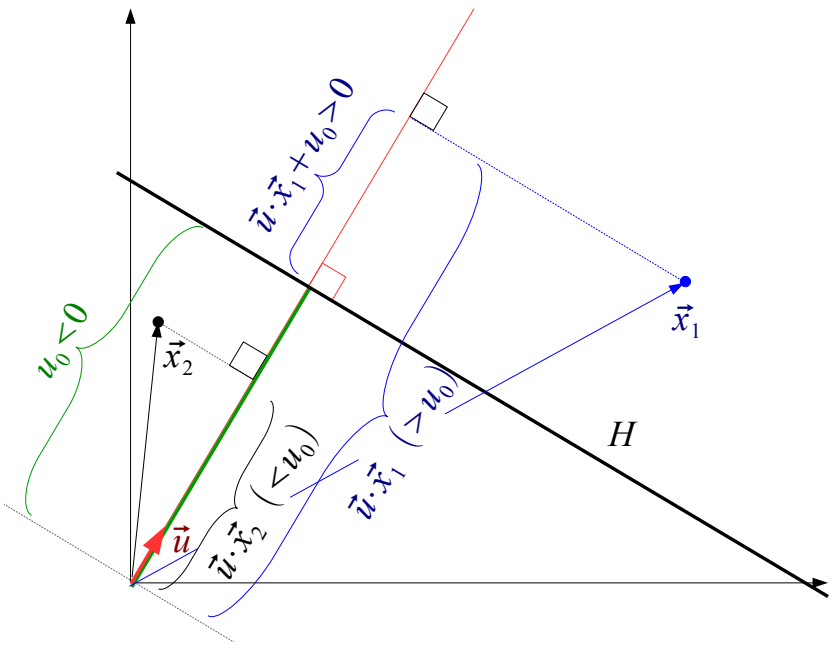
où $\mathbf{u}$ est un vecteur normal à cet hyperplan.

Rappels :

- $\frac{1}{\|\mathbf{u}\|}\,{}^t\mathbf{uv}$ est le module signé de la projection de $\mathbf{v}$ sur $\mathbf{u}$.
- La distance signée d'un point $\mathbf{x}$ à l'hyperplan $\mathcal{H}$ est :

$$d(\mathbf{x}_i, \mathcal{H}) = \frac{1}{\|\mathbf{u}\|}\left({}^t\mathbf{ux} + u_0\right).$$

Elle est positive quand $\mathbf{x}$ est du côté de $\mathcal{H}$ pointé par $\mathbf{u}$.



On choisit d'attribuer la classe $\mathcal{C}_{+1}$ aux points $\mathbf{x}_i$ tels que $d(\mathbf{x}_i, \mathcal{H}) \geq 0$ et $\mathcal{C}_{-1}$ aux autres points :

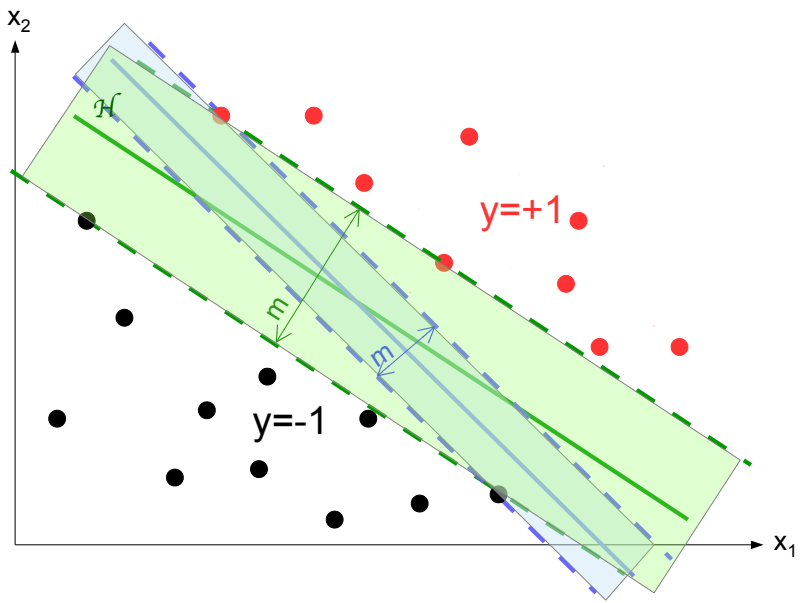
$$\left(\frac{1}{\|\mathbf{u}\|}\,{}^t\mathbf{ux}_i + \frac{u_0}{\|\mathbf{u}\|}\right) y_i \geq 0, \quad i = 1, \cdots, M.$$

Puisque $y_i = \pm 1$, $\left(\frac{1}{\|\mathbf{u}\|}\,{}^t\mathbf{u}\mathbf{x}_i + \frac{u_0}{\|\mathbf{u}\|}\right) y_i$ est toujours positif et représente la distance du point $\mathbf{x}_i$ à l'hyperplan $\mathcal{H}$. La distance m du point le plus proche de $\mathcal{H}$ est donc :

$$m = \min_{i\in\{1,\cdots,M\}} d(\mathbf{x}_i, \mathcal{H}) = \min_{i\in\{1,\cdots,M\}} \left(\frac{1}{\|\mathbf{u}\|}\,{}^t\mathbf{u}\mathbf{x}_i + \frac{u_0}{\|\mathbf{u}\|}\right) y_i.$$

Parmi tous les hyperplans séparant les deux classes, on choisit celui qui maximise la marge m :

$$\begin{cases} \max\limits_{(\mathbf{u},u_0,m)\in\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}} m \\ s.c. \quad \left(\frac{1}{\|\mathbf{u}\|}\,{}^t\mathbf{u}\mathbf{x}_i + \frac{u_0}{\|\mathbf{u}\|}\right) y_i \geq m, \\ \qquad\qquad\qquad\qquad i = 1, \cdots, M. \end{cases}$$



Remarque : le problème est mal posé car si $(\mathbf{u}, u_0)$ est solution, alors $\forall \lambda > 0,\ (\lambda\mathbf{u}, \lambda u_0)$ est aussi solution. On pose alors :

$$\mathbf{w} = \frac{\mathbf{u}}{m\|\mathbf{u}\|} \qquad \text{et} \qquad w_0 = \frac{u_0}{m\|\mathbf{u}\|}.$$

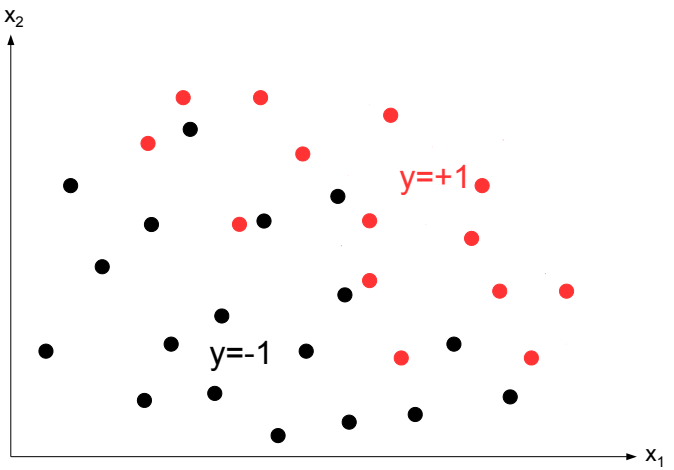
Ainsi, $\|\mathbf{w}\| = \left\| \frac{\mathbf{u}}{m\|\mathbf{u}\|} \right\| = \frac{1}{m}$ et maximiser m est donc équivalent à minimiser $\|\mathbf{w}\|$, ou encore à minimiser $\frac{1}{2}\|\mathbf{w}\|^2$.
Par ailleurs, $\left(\frac{{}^t\mathbf{u}}{\|\mathbf{u}\|}\mathbf{x}_i + \frac{u_0}{\|\mathbf{u}\|} \right) y_i \geq m \iff ({}^t\mathbf{w}\mathbf{x}_i + w_0)\, y_i \geq 1$ et le problème devient :

$$\boxed{\begin{cases} \min\limits_{(\mathbf{w}, w_0) \in \mathbb{R}^n \times \mathbb{R}} & \frac{1}{2}\|\mathbf{w}\|^2 \\ s.c. \ : & ({}^t\mathbf{w}\mathbf{x}_i + w_0)\, y_i \geq 1, \quad i = 1, \cdots, M. \end{cases}}$$

C'est un problème d'optimisation à $n+1$ inconnues et M contraintes.

Problème primal **non linéairement séparble**

En pratique, les données à classer ne sont pas toujours linéairement séparables : dans ce cas, il n'existe pas d'hyperplan de $\mathbb{R}^n$ séparant les deux classes et le problème précédent n'a alors pas de solution.



On va alors pénaliser le critère proportionnellement au dépassement des points mal classés :

$$\min_{(\mathbf{w},w_0,\boldsymbol{\xi})\in\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^M} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M}\xi_i$$

$$s.c. \ : \begin{cases} \left({}^t\mathbf{w}\mathbf{x}_i + w_0\right) y_i \geq 1-\xi_i, & i=1,\cdots,M, \\ \xi_i \geq 0, & i=1,\cdots,M. \end{cases}$$

C'est un problème d'optimisation à $n+1+M$ inconnues et $2M$ contraintes.

Problème **dual** non linéairement séparble

Lagrangien du problème primal non linéaire :

$$L(\mathbf{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2}\,{}^t\mathbf{w}\mathbf{w} + C\sum_{i=1}^{M}\xi_i - \sum_{i=1}^{M}\lambda_i\left[({}^t\mathbf{w}\mathbf{x}_i + w_0)y_i - 1 + \xi_i\right] - \sum_{i=1}^{M}\mu_i\xi_i.$$

Fonction duale :

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{w}, w_0, \boldsymbol{\xi}} L(\mathbf{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

En notant $(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*)$ la solution de ce problème, $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ s'écrit donc :

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \frac{1}{2}\,{}^t\mathbf{w}^*\mathbf{w}^* - \sum_{i=1}^{M}\lambda_i\,{}^t\mathbf{w}^*\mathbf{x}_i y_i - w_0^*\sum_{i=1}^{M}\lambda_i y_i + \sum_{i=1}^{M}\lambda_i + \sum_{i=1}^{M}(C - \lambda_i - \mu_i)\xi_i^*. \end{aligned}$$

$(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*)$ point stationnaire de L donc $\nabla_{\mathbf{w},w_0,\boldsymbol{\xi}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}) = \mathbf{0}$

$$\Rightarrow \begin{cases} \nabla_{\mathbf{w}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{w}^* - \sum_{i=1}^{M} \lambda_i \mathbf{x}_i y_i = \mathbf{0} \implies {}^t\mathbf{w}^*\mathbf{w}^* = \sum_{i=1}^{M} \lambda_i \, {}^t\mathbf{w}^*\mathbf{x}_i y_i, \\ \nabla_{w_0} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -\sum_{i=1}^{M} \lambda_i y_i = 0 \iff {}^t\mathbf{y}\boldsymbol{\lambda} = 0, \\ \nabla_{\boldsymbol{\xi}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0} \iff C - \lambda_i - \mu_i = 0 \; \forall i \in \{1, \cdots, M\}, \text{ avec } \begin{cases} \mu_i \geq 0, \\ \lambda_i \geq 0, \end{cases} \\ \qquad\qquad\qquad\qquad \implies 0 \leq \lambda_i \leq C \;\; \forall i \in \{1, \cdots, M\}, \end{cases}$$

donc

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} {}^t\mathbf{w}^*\mathbf{w}^* - \sum_{i=1}^{M} \lambda_i \, {}^t\mathbf{w}^*\mathbf{x}_i y_i - w_0^* \sum_{i=1}^{M} \lambda_i y_i + \sum_{i=1}^{M} \lambda_i + \sum_{i=1}^{M} (C - \lambda_i - \mu_i)\xi_i^*,$$

$$= \frac{1}{2} {}^t\mathbf{w}^*\mathbf{w}^* - {}^t\mathbf{w}^*\mathbf{w} - 0 + \sum_{i=1}^{M} \lambda_i + 0,$$

$$= -\frac{1}{2} {}^t\mathbf{w}^*\mathbf{w}^* + \sum_{i=1}^{M} \lambda_i.$$

Puisque $\mathbf{w}^* = \sum_{i=1}^{M} \lambda_i \mathbf{x}_i y_i$, on a ${}^t\mathbf{w}^*\mathbf{w}^* = \sum_{i=1}^{M}\sum_{j=1}^{M} \lambda_i\lambda_j \, {}^t\mathbf{x}_i\mathbf{x}_j y_i y_j = {}^t\boldsymbol{\lambda}\mathbf{A}\boldsymbol{\lambda}$, où $\mathbf{A}$ est la matrice $M \times M$ de terme général $a_{ij} = {}^t\mathbf{x}_i\mathbf{x}_j y_i y_j$.

En notant $\mathbf{e} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, on a donc $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = g(\boldsymbol{\lambda}) = -\frac{1}{2}\, {}^t\boldsymbol{\lambda}\mathbf{A}\boldsymbol{\lambda} + {}^t\mathbf{e}\boldsymbol{\lambda}$.

Remarques :

- la fonction duale g ne dépend plus de $\boldsymbol{\mu}$,
- les points $\mathbf{x}_i$ n'interviennent plus que par les produits scalaires ${}^t\mathbf{x}_i\mathbf{x}_j$.

Finalement, le problème dual s'écrit :

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^M} \ \frac{1}{2}\, {}^t\boldsymbol{\lambda}\mathbf{A}\boldsymbol{\lambda} - {}^t\mathbf{e}\boldsymbol{\lambda}$$
$$\text{s.c.} \begin{vmatrix} {}^t\mathbf{y}\boldsymbol{\lambda} = \mathbf{0}, \\ 0 \le \lambda_i \le C, \quad i = 1, \cdots, M. \end{vmatrix}$$

Fonction de décision :

$$\widetilde{y} = \text{signe}({}^t\mathbf{w}^*\mathbf{x} + w_0^*)$$

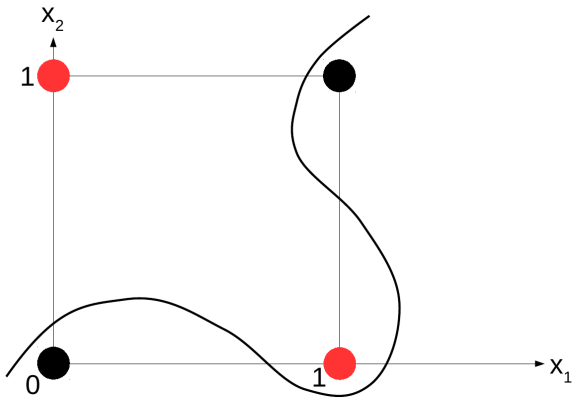
avec $\mathbf{w}^* = \sum_{i=1}^{M} \lambda_i \mathbf{x}_i y_i$

et $w_0 = \frac{1}{y_m} - {}^t\mathbf{w}^*\mathbf{x}_m$

C'est un problème d'optimisation à M inconnues et $M+1$ contraintes.

Le problème du XOR :

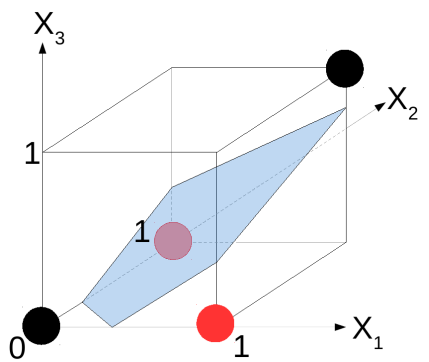
Lorsque les données ne sont clairement pas linéairement séparables, un hyperplan séparateur ne suffit plus. C'est le cas du OU exclusif.



En plongeant le problème dans un espace de dimension supérieure, il peut exister un hyperplan séparateur.

$$\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \xrightarrow{\mathbf{\Phi}} \mathbf{X} = (X_1, X_2, X_3) \in \mathbb{R}^3$$

$$avec \begin{cases} X_1 = x_1 \\ X_2 = x_2 \\ X_3 = x_1 x_2 \end{cases}$$



En notant $\mathbf{\Phi}$ la fonction qui permet de passer de l'espace initial de dimension n à ce nouvel espace de dimension $n' > n$, on peut réécrire notre problème dual non linéaire de la façon suivante :

$$\begin{cases} \min\limits_{\boldsymbol{\lambda}\in\mathbb{R}^M} \min \dfrac{1}{2}\displaystyle\sum_{i=1}^{M}\sum_{j=1}^{M} \lambda_i\lambda_j y_i y_j \ {}^t\mathbf{\Phi}(\mathbf{x_i})\mathbf{\Phi}(\mathbf{x_j}) - {}^t\mathbf{e}\boldsymbol{\lambda} \\ \text{s.c.} \left| \begin{array}{l} {}^t\mathbf{y}\boldsymbol{\lambda} = \mathbf{0}, \\ 0 \leq \lambda_i \leq C, \quad i = 1, \cdots, M. \end{array} \right. \end{cases}$$

Mais n' très grand (voire infini) $\Rightarrow$ calcul de ${}^t\mathbf{\Phi}(\mathbf{x_i})\mathbf{\Phi}(\mathbf{x_j})$ impossible.
$\Rightarrow$ **fonctions noyaux** : réalisent facilement ce produit scalaire en grande dimension.
Par exemple :

$$\begin{aligned} K \ : \ \mathbb{R}^2 \times \mathbb{R}^2 &\longrightarrow \ \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) \ &\longrightarrow K(\mathbf{x}, \mathbf{y}) = ({}^t\mathbf{xy})^2 \end{aligned}$$

réalise le produit scalaire pour la fonction :

$$\begin{aligned} \mathbf{\Phi} \ : \ \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ \mathbf{x} = {}^t(x_1, x_2) &\longrightarrow \mathbf{\Phi(x)} = {}^t(x_1^2, x_2^2, \sqrt{2}x_1x_2). \end{aligned}$$

En effet : $\mathbf{\Phi(x)} \cdot \mathbf{\Phi(y)} = x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2 = (x_1y_1 + x_2y_2)^2 = ({}^t\mathbf{xy})^2$.

Exemples de fonctions noyau :

- linéaire : $K(\mathbf{x},\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$,
- polynomial : $K(\mathbf{x},\mathbf{y}) = (c+\mathbf{x}\cdot\mathbf{y})^d$,
- gaussien : $K(\mathbf{x},\mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}$,
- laplacien : $K(\mathbf{x},\mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|}{\sigma}}$

Le problème devient alors :

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^M} \min \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j) - {}^t\mathbf{e}\boldsymbol{\lambda}$$
$$\text{s.c.} \left| \begin{array}{l} {}^t\mathbf{y}\boldsymbol{\lambda} = \mathbf{0}, \\ 0 \le \lambda_i \le C, \quad i = 1,\cdots,M. \end{array} \right.$$

Fonction de décision :

$$\widetilde{y}(\mathbf{x}) = \text{signe}\left(\sum_{i=1}^{M}\lambda_i y_i K(\mathbf{x}_i,\mathbf{x}) + w_0\right)$$

avec $w_0 = y_m - \sum_{i=1}^{M}\lambda_i y_i K(\mathbf{x}_i,\mathbf{x}_m)$

et (x_m, y_m) : point de la marge.