Optimisation

Dr. Yacouba Simporé
simplesaint@gmail.com

25 décembre 2020

Table des matières

1 Calcul différentiel et Analyse convexe **4**

1.1 Rappels 4

1.1.1 Norme sur $\mathbb{R}^n$ 4

1.1.2 Différentiel 4

1.1.3 Dérivée selon un vecteur 4

1.1.4 Gradient 5

1.1.5 Développement limité d'ordre 1 et 2 6

1.1.6 Formules 6

1.2 Analyse convexe 7

1.2.1 Généralités 7

1.2.2 Caractérisation de la forte convexité 9

1.2.3 Test de convexité – concavité 10

2 Optimisation sans contrainte **13**

2.1 Résultats d'existence 13

2.2 Algorithmes du gradient 14

2.2.1 Algorithme du gradient 14

2.2.2 Algorithme du gradient à pas constant 15

2.2.3 Algorithme du gradient à pas optimal 16

2.2.4 Algorithme du gradient conjugué 17

2.2.5 Algorithme de Newton 18

2.2.6 Théorèmes généraux de minimalité 19

3 Optimisation avec contraintes **20**

3.1 Résultats d'existence 20

3.2 Lagrangien du problème 21

3.2.1 Points selles du Lagrangien 22

3.3 Algorithme 23
3.3.1 Algorithme d'USAWA 24

Chapitre 1

Calcul différentiel et Analyse convexe

1.1 Rappels

1.1.1 Norme sur $\mathbb{R}^n$

Propriétés d'une norme

Soit N une norme sur $\mathbb{R}^n$, on a :

- $N(x) = 0 \iff x = 0_{\mathbb{R}^n}$;
- $\forall x, y \in \mathbb{R}^n,\ N(x+y) \leq N(x) + N(y)$;
- $\forall x \in \mathbb{R}^n,\ \forall \lambda \in \mathbb{R},\ N(\lambda x) = |\lambda| N(x)$;
- Inégalité de Cauchy-Schwarz (norme euclidienne)

$$\forall x, y \in \mathbb{R}^n, \quad |\langle x\,,\, y\rangle| \leq \|x\|\|y\|.$$

En rappel, $\langle x\,,\, y\rangle$ désigne le produit scalaire des vecteurs x et y, et $\|x\| = \sqrt{\langle x\,,\, x\rangle}$ la norme du vecteur x.

1.1.2 Différentiel

Soit U un ouvert de $\mathbb{R}^n$ et $x_0 \in U$. On dit que l'application $f : U \to \mathbb{R}^p$ est différentiable en x_0 ou admet un développement limité d'ordre 1 en x_0 s'il existe $df(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ continue telle que

$$f(x_0 + h) = f(x_0) + df(x_0)h + o(\|h\|).$$

1.1.3 Dérivée selon un vecteur

Soit h un vecteur fixé, il existe un réel $S > 0$ tel que

$$\forall t \in [-S, S],\ x_0 + th \in U \ (x_0 \in U,\ U \text{ un ouvert de } \mathbb{R}^n).$$

On pose, pour $x_0 \in U$ et $h \in \mathbb{R}^n$ comme ci-dessus,

$$\varphi_h(t) = f(x_0 + th).$$

Définition 1.1

On dit que f *est dérivable en* x_0 *selon le vecteur* h *lorsque* φ_h *est dérivable à l'origine. On note alors*

$$D_h f(x_0) = \varphi'_h(0).$$

En clair, on a

$$D_h f(x_0) = \varphi'_h(0) = \lim_{t\to 0} \frac{f(x_0 + th) - f(x_0)}{t}.$$

On note que la dérivée ainsi définie ci-dessus est un vecteur, dit ***vecteur dérivé de*** f ***au point*** x_0 ***selon le vecteur*** h.

Définition 1.2

Soit $e = (e_1, e_2, \dots, e_n)$ *la base canonique de* $\mathbb{R}^n$ *et* $j \in [\![1, n]\!]$.
On dit que f *admet une* ***dérivée partielle en*** x_0 ***selon la*** j***-ième variable*** *ou que* f *admet une* ***dérivée partielle au point*** x_0 ***suivant*** x_j *lorsque* f *admet une dérivée partielle au point* x_0 *selon le vecteur* e_j.
Cette dérivée est alors notée

$$\frac{\partial f}{\partial x_j}(x_0) = D_{e_j} f(x_0)$$

et est appelée la j*-ième dérivée partielle de* f *au point* x_0.
On note aussi parfois

$$\frac{\partial f}{\partial x_j}(x_0) = \lim_{t\to 0} \frac{f(x_0 + te_j) - f(x_0)}{t}.$$

Théorème 1.1

Si f *est différentiable en* x_0*, elle admet une dérivée selon tout vecteur* h.
En outre $df(x_0) \cdot h = D_h f(x_0)$.

On a alors, pour $e = (e_1, e_2, \dots, e_n)$ la base canonique de $\mathbb{R}^n$, les égalités qui suivent

$$df(x_0) \cdot h = D_h f(x_0) = \sum_{j=1}^{n} h_j D_{e_j} f(x_0), \qquad h = \sum_{j=1}^{n} h_j e_j,$$
$$= \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(x_0).$$

1.1.4 Gradient

Soit U un ouvert de $\mathbb{R}^n$, $x_0 \in U$ et $f : U \to \mathbb{R}$ une fonction différentiable en x_0.

Définition 1.3

Le gradient de f *en* x_0 *est le vecteur* $\overrightarrow{\text{grad}}(f)(x_0)$ *tel que, pour tout* $h \in \mathbb{R}^n$,

$$df(x_0) \cdot h = \left\langle \overrightarrow{\text{grad}}(f)(x_0)\,,\, h \right\rangle = \langle \nabla f(x_0)\,,\, h \rangle\,.$$

Le vecteur $\overrightarrow{\text{grad}}(f)(x_0)$ *est appelé* ***gradient de*** f ***au point*** x_0 *; le symbole* ∇ *est dit* ***opérateur nabla****, et* ∇f *est lu « **nabla f** ».*

Définition 1.4 (*Point critique*)

Soit $f : U \to \mathbb{R}$. *Un point* $x_0 \in U$ *tel que*

$$f \text{ est différentiable en } x_0 \quad et \quad df(x_0) = 0$$

est dit ***point critique de la fonction*** f.
L'équation des points critiques est donnée par : $\nabla f(x_0) = 0$.

1.1.5 Développement limité d'ordre 1 et 2

Soit f différentiable en x_0, $h \in \mathbb{R}^n$ fixé et $t \to 0$. Pour l'ordre 2, supposons f deux fois différentiable en x_0.

1. $f(x_0 + th) - f(x_0) = t \langle \nabla f(x_0) , h \rangle + o(t)$;
2. $f(x_0 + th) - f(x_0) = t df(x_0)h + \dfrac{t^2}{2} d^2 f(x_0)[h, h] + o(t^2)$;
3. $f(x_0 + th) - f(x_0) = t \langle \nabla f(x_0) , h \rangle + \dfrac{t^2}{2} \left\langle \nabla^2 f(x_0)h , h \right\rangle + o(t^2)$.

1.1.6 Formules

Formule de Taylor avec reste intégral

Supposons que f est de classe $\mathcal{C}^2$ dans un ouvert U de $\mathbb{R}^n$, à valeurs réelles.

Si le segment $[x_0, x_0 + h]$ est contenu dans U, alors on a :

$$f(x_0 + h) - f(x_0) = \langle \nabla f(x_0) , h \rangle + \int_0^1 (1 - t) \left\langle \nabla^2 f(x_0 + th)h , h \right\rangle dt.$$

Formule de Taylor avec reste de Lagrange

On suppose que f est deux fois différentiable sur un ouvert U de $\mathbb{R}^n$, que le segment $[x_0, x_0 + h]$ est contenu dans U et que, il existe une constante $C > 0$ telle que, pour tout $t \in [0, 1]$,

$$\left| \left\langle \nabla^2 f(x_0 + th) \cdot h , h \right\rangle \right| \leq C \|h\|^2.$$

Alors

$$|f(x_0 + h) - f(x_0) - \langle \nabla f(x_0) , h \rangle| \leq \frac{C}{2} \|h\|^2.$$

La différentielle d'ordre 2 de f en x_0 est :

$$d^2 f(x_0)[h, h] = \left\langle \nabla^2 f(x_0) \cdot h , h \right\rangle.$$

1.2 Analyse convexe

1.2.1 Généralités

Définition 1.5

*Un ensemble $U \subset \mathbb{R}^n$ est dit **convexe** si pour tout $x, y \in U$, le segment d'extrémités x et y est contenu dans U.*

Définition 1.6

1. *On dit que f est **convexe** sur U si*

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x), \quad \forall t \in [0,1],\ \forall x, y \in U.$$

2. *On dit que f est **strictement convexe** sur U si*

$$f(ty + (1-t)x) < tf(y) + (1-t)f(x), \quad \forall t \in]0,1[,\ \forall x, y \in U,\ x \neq y.$$

3. *On dit que f est **fortement convexe** sur U s'il existe $\alpha > 0$ tel que*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{1}{2}\alpha t(1-t)\|x-y\|^2, \quad \forall x, y \in U,\ \forall t \in [0,1].$$

4. *On dit que f est **concave (resp. strictement concave)** si $(-f)$ est **convexe (resp. strictement convexe)**.*

Proposition 1.1. *Soit $U \subset \mathbb{R}^n$ un ouvert, $V \subset U$ un ensemble convexe, et $f : U \longrightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ sur U.*

1. *Alors, les énoncés suivants sont équivalents*
   - *(a) f est convexe sur V ;*
   - *(b) $f(y) \geq f(x) + \langle \nabla f(x)\,,\, y - x \rangle\,, \quad \forall x, y \in V$;*
   - *(c) ∇f est **monotone sur** V, i.e.*

$$\langle \nabla f(y) - \nabla f(x)\,,\, y - x \rangle \geq 0, \quad \forall x, y \in V.$$

2. *Si de plus, f est de classe $\mathcal{C}^2$ sur V, alors f est convexe si et seulement si*

$$\left\langle \nabla^2 f(x) \cdot (y-x)\,,\, y - x \right\rangle \geq 0, \quad \forall x, y \in V.$$

Démonstration :

1. (a) $1 \implies 2$

On suppose donc que f est convexe sur V ; on a

$$\begin{aligned}
f(ty + (1-t)x) &\leq tf(y) + (1-t)f(x), \quad \forall x, y \in V,\ \forall t \in [0,1], \\
\implies f(x + t(y-x)) &\leq tf(y) + f(x) - tf(x), \\
\implies \frac{f(x + t(y-x)) - f(x)}{t} &\leq f(y) - f(x), \quad t \neq 0, \\
\implies \lim_{t \to 0} \frac{f(x + t(y-x)) - f(x)}{t} &\leq f(y) - f(x), \\
\implies \langle \nabla f(x)\,,\, y - x \rangle &\leq f(y) - f(x)
\end{aligned}$$

(b) $2 \implies 3$

On a

$$f(y) \geq f(x) + \langle \nabla f(x)\,,\, y - x\rangle \tag{1.1}$$

et

$$-f(y) \geq -f(x) + \langle \nabla f(y)\,,\, x - y\rangle \tag{1.2}$$

d'où

$$\begin{aligned}(1.1) \text{ et } (1.2) &\implies 0 \geq \langle \nabla f(x)\,,\, y - x\rangle - \langle \nabla f(y)\,,\, y - x\rangle \\ &\implies 0 \geq \langle \nabla f(x) - \nabla f(y)\,,\, y - x\rangle \\ &\implies \langle \nabla f(x) - \nabla f(y)\,,\, y - x\rangle \leq 0 \iff \langle \nabla f(y) - \nabla f(x)\,,\, y - x\rangle \geq 0.\end{aligned}$$

(c) $3 \implies 1$

Soit $k : [0,1] \to \mathbb{R}$ tel que $k(t) = f(ty + (1-t)x)$, on a $k'(t) = \langle \nabla f(ty + (1-t)x)\,,\, y - x\rangle$, donc

$$\begin{aligned}k'(u) = \langle \nabla f(uy + (1-u)x)\,,\, y - x\rangle \quad \text{et} \quad k'(v) &= \langle \nabla f(vy + (1-v)x)\,,\, y - x\rangle \\ \implies k'(u) - k'(v) &= \langle \nabla f(uy + (1-u)x) - \nabla f(vy + (1-v)x)\,,\, y - x\rangle\,.\end{aligned}$$

La monotonie de ∇f rend k' croissante. Donc k est convexe et

$$k(t) \leq tk(1) + (1-t)k(0)$$

c'est-à-dire

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x).$$

2. Supposons f, de classe $\mathcal{C}^2$, convexe. Il vient donc

$$\begin{aligned}&\langle \nabla f(x + th) - \nabla f(x)\,,\, th\rangle \geq 0 \\ \implies &\frac{1}{t^2}\langle \nabla f(x + th) - \nabla f(x)\,,\, th\rangle \geq 0, \quad t \neq 0, \\ \implies &\left\langle \frac{\nabla f(x + th) - \nabla f(x)}{t}\,,\, h\right\rangle \geq 0 \\ \implies &\lim_{t\to 0}\left\langle \frac{\nabla f(x + th) - \nabla f(x)}{t}\,,\, h\right\rangle \geq 0 \implies \left\langle \nabla^2 f(x)\cdot h\,,\, h\right\rangle \geq 0.\end{aligned}$$

En posant $h = y - x \in \mathbb{R}^n$, on obtient

$$\left\langle \nabla^2 f(x)\cdot(y - x)\,,\, y - x\right\rangle \geq 0.$$

Réciproquement, supposons que $\left\langle \nabla^2 f(x)\cdot(y - x)\,,\, y - x\right\rangle \geq 0$; posons

$$A = \langle \nabla f(y) - \nabla f(x)\,,\, y - x\rangle \quad \text{et} \quad h = y - x;$$

il vient donc que

$$A = \langle \nabla f(x + h) - \nabla f(x)\,,\, h\rangle\,.$$

Mais comme f est de classe $\mathcal{C}^2$, on a

$$\nabla f(x + h) - \nabla f(x) = \int_0^1 \nabla^2 f(x + th)h\,dt,$$

donc

$$A = \int_0^1 \left\langle \nabla^2 f(x + th)h\,,\, h\right\rangle dt \geq 0.$$

Par conséquent,

$$\langle \nabla f(y) - \nabla f(x)\,,\, y - x\rangle \geq 0,$$

et donc f est convexe. ■

Propriété 1.1 (Caractérisation de la convexité stricte). *Soit U un ouvert de $\mathbb{R}^n$, $V \subset U$ un convexe, et $f : U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Alors, les trois propositions suivantes sont équivalentes :*

1. *f est strictement convexe ;*
2. $f(y) > f(x) + \langle \nabla f(x) , y - x \rangle , \quad \forall x, y \in V, x \neq y$ *;*
3. *∇f est strictement monotone sur V, c'est-à-dire*

$$\langle \nabla f(x) - \nabla f(y) , x - y \rangle > 0, \quad \forall x, y \in V, x \neq y.$$

Si de plus, f est de classe $\mathcal{C}^2$ sur U, alors f est strictement convexe dès que

$$\left\langle \nabla^2 f(x)(y - x) , y - x \right\rangle > 0, \quad \forall x, y \in V,\ x \neq y.$$

1.2.2 Caractérisation de la forte convexité

Soit U un ouvert de $\mathbb{R}^n$, $V \subset U$ un convexe, et $f : U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$.

Alors, les trois propositions suivantes sont équivalentes.

1. f est fortement convexe.
2. $\exists \beta > 0,\ \forall x, y \in V :\ f(y) \geq f(x) + \langle \nabla f(x) , y - x \rangle + \beta \|y - x\|^2$.
3. Il existe $r > 0$ tel que : $\langle \nabla f(y) - \nabla f(x) , y - x \rangle \geq r \|y - x\|^2,\ \forall x, y \in V$.

Si de plus, f est de classe $\mathcal{C}^2$, alors f est fortement convexe si et seulement si il existe $\beta > 0$ tel que

$$\left\langle \nabla^2 f(x)(y - x) , y - x \right\rangle \geq \beta \|y - x\|^2, \quad \forall x, y \in V.$$

Définition 1.7

1. *On appelle **fonction elliptique** une fonction de classe $\mathcal{C}^1$ fortement convexe.*
2. *Soit $U \subset \mathbb{R}^n$ non borné, $f : U \to \mathbb{R}$.*
   *On dit que f est **coercive** sur U si*

$$\lim_{\substack{x \in U \\ \|x\| \to +\infty}} f(x) = +\infty.$$

   Cette définition s'écrit aussi $f\left(x^k\right) \to +\infty$, pour toute suite $\left(x^k\right)_{k \in \mathbb{N}} \subset U$ telle que $\left\|x^k\right\| \to +\infty$.

Proposition 1.2. *Soit $U \subset \mathbb{R}^n$ un ouvert, $V \subset U$ un ensemble convexe non borné et f une fonction de classe $\mathcal{C}^1$ et fortement convexe sur V. Alors f est coercive sur V.*

Démonstration : f étant fortement convexe sur V, on a

$$\exists \beta > 0 :\ \forall x, y \in V,\ f(y) \geq f(x) + \langle \nabla f(x) , y - x \rangle + \beta \|y - x\|^2.$$

Soit $x = a$ fixé. Il vient

$$\begin{aligned} & f(y) \geq f(a) + \langle \nabla f(a) , y - a \rangle + \beta \|y - a\|^2 \\ \Longrightarrow\ & f(y) \geq f(a) - |\langle \nabla f(a) , y - a \rangle| + \beta \|y - a\|^2, \end{aligned}$$

car $\langle \nabla f(a) , y - a \rangle \geq -|\langle \nabla f(a) , y - a \rangle|$.

Or

$$|\langle \nabla f(a)\,,\, y-a\rangle| \leq \|\nabla f(a)\|\|y-a\|,$$

d'où

$$\begin{aligned} f(y) &\geq f(a) - \|\nabla f(a)\|\|y-a\| + \beta\|y-a\|^2 \\ \Longrightarrow\ f(y) &\geq f(a) + \beta\|y-a\|^2 - \|\nabla f(a)\|\|y-a\|, \end{aligned}$$

d'où, si $\|y\| \to +\infty$, il s'ensuit $f(y) \to +\infty$: et donc f est coercive. ■

Propriété 1.2 (Fonctions convexes)**.** *Soit $U \subset \mathbb{R}^n$ un convexe, $p \in \mathbb{N}^*$, $f_1, \ldots, f_p$ des fonctions convexes définies de U à valeurs dans $\mathbb{R}$, $\alpha_1, \ldots, \alpha_p$ des constantes strictement positives. Posons*

$$f = \sum_{i=1}^{p} \alpha_i f_i.$$

Alors f est convexe.

Autrement dit, toute combinaison linéaire (à coefficients strictement positifs) de fonctions convexes est encore une fonction convexe.

De plus :

- *si au moins une des fonctions f_i est strictement convexe, alors f est strictement convexe ;*
- *si au moins une des fonctions f_i est fortement convexe, alors f est fortement convexe.*

1.2.3 Test de convexité – concavité

Activité

Énoncé

1. Calculer $J(X+th)$, sachant que $J(X) = \langle AX\,,\, X\rangle + \langle b\,,\, X\rangle + C$.
2. Calculer $dJ(X)\cdot h$, puis $d^2J(X)$.

Résolution

1. On a

$$\begin{aligned} J(X+th) &= \langle A(X+th)\,,\, X+th\rangle + \langle b\,,\, X+th\rangle + C \\ &= \langle AX\,,\, X\rangle + t\,\langle AX\,,\, h\rangle + t\,\langle Ah\,,\, X\rangle + t^2\,\langle Ah\,,\, h\rangle + \langle b\,,\, X\rangle + t\,\langle b\,,\, h\rangle + C. \end{aligned}$$

2. On en déduit :

$$\begin{aligned} &\Longrightarrow\ J(X+th) - J(X) = t\,(\langle AX\,,\, h\rangle + \langle Ah\,,\, X\rangle + \langle b\,,\, h\rangle) + t^2\,\langle Ah\,,\, h\rangle \\ &\Longrightarrow\ \frac{1}{t}\,(J(X+th) - J(X)) = \langle AX\,,\, h\rangle + \langle Ah\,,\, X\rangle + \langle b\,,\, h\rangle + t\,\langle Ah\,,\, h\rangle\,, \end{aligned}$$

d'où

$$dJ(X)\cdot h = \langle AX\,,\, h\rangle + \langle Ah\,,\, X\rangle + \langle b\,,\, h\rangle\,.$$

Il s'ensuit que

$$\begin{aligned} dJ(X+th)\cdot h &= \langle A(X+th)\,,\, h\rangle + \langle Ah\,,\, X+th\rangle + \langle b\,,\, h\rangle \\ &= \langle AX\,,\, h\rangle + t\,\langle Ah\,,\, h\rangle + \langle Ah\,,\, X\rangle + t\,\langle Ah\,,\, h\rangle + \langle b\,,\, h\rangle \\ \Longrightarrow\ dJ(X+th)\cdot h - dJ(X)\cdot h &= 2t\,\langle Ah\,,\, h\rangle\,, \end{aligned}$$

d'où

$$\begin{aligned} d^2J(X)[h,h] &= \lim_{t\to 0}\left[\frac{1}{t}\left(dJ(X+th)\cdot h - dJ(X)\cdot h\right)\right] = \lim_{t\to 0}\left[2\left\langle Ah\,,\,h\right\rangle\right] \\ \Longrightarrow\ d^2J(X)[h,h] &= 2\left\langle Ah\,,\,h\right\rangle. \end{aligned}$$

La matrice hessienne de J est $A + A^t$, et dans le cas où la matrice A est symétrique, on a

$$\nabla J(X) = 2AX + b.$$

Fonction d'une seule variable 2 fois dérivable

Soit f une fonction d'une seule variable deux fois dérivable. Alors f est

- convexe (resp. concave) sur un intervalle I si et seulement si :

$$\frac{d^2f(x)}{dx^2} \geq 0 \quad \left(\text{resp. } \frac{d^2f(x)}{dx^2} \leq 0\right);$$

- strictement convexe (resp. concave) si :

$$\frac{d^2f(x)}{dx^2} > 0 \quad \left(\text{resp. } \frac{d^2f(x)}{dx^2} < 0\right).$$

Fonction de deux variables de classe $\mathcal{C}^2$

Soit f une fonction de deux variables, de classe $\mathcal{C}^2$; on a les critères suivants (suffisants dans les deux cas stricts) :

	Convexe	Strictement convexe	Concave	Strictement concave
$\frac{\partial^2 f}{\partial x^2}$	positive	strictement positive	négative	strictement négative
$\frac{\partial^2 f}{\partial y^2}$	positive	strictement positive	négative	strictement négative
$\frac{\partial^2 f}{\partial x^2} \times \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$	positive	strictement positive	positive	strictement positive

Généralisation

Soit f une fonction de deux variables au moins deux fois dérivables, de matrice hessienne H.

Théorème 1.2

1. *Soit* $f : U \to \mathbb{R}$, U *un ouvert convexe.*
   f *est convexe sur* U *si et seulement si sa matrice hessienne est semi-définie positive pour tout* $x \in U$.
2. f *est concave si et seulement si sa matrice hessienne est semi-définie négative pour tout* $x \in U$.

Théorème 1.3

1. f *est convexe sur* U *si et seulement si les valeurs propres de sa matrice hessienne sont toutes positives ou nulles.*
2. f *est concave sur* U *si et seulement si sa matrice hessienne a toutes ses valeurs propres négatives ou nulles.*

Théorème 1.4

1. f *est convexe sur* U *si et seulement si* $\mathrm{Tr}(H) \geq 0$ *sur* U *et* $\det(H) \geq 0$ *sur* U.
2. f *est concave sur* U *si et seulement si* $\mathrm{Tr}(H) \leq 0$ *sur* U *et* $\det(H) \geq 0$ *sur* U.

Chapitre 2

Optimisation sans contrainte

2.1 Résultats d'existence

Définition 2.1 (*Suites minimisantes*)

Soit $(x_k)_{k\in\mathbb{N}}$ *une suite d'éléments de* $\mathbb{R}^n$, J *la fonctionnelle définie de* $\mathbb{R}^n$ *dans* $\mathbb{R}$, *et* $\mathcal{P}$ *le problème*

$$\begin{cases} \text{Trouver } x^* \in \mathbb{R}^n : \\ \\ J(x^*) = \min_{x\in\mathbb{R}^n} J(x). \end{cases}$$

Si $\lim_{k\to+\infty} J(x_k) = \inf_{x\in\mathbb{R}^n} J(x)$, *alors on dit que* (x_k) *est une suite minimisante de la fonctionnelle* J.

Théorème 2.1

Soit f *une fonction continue et coercive sur* $\mathbb{R}^n$, *alors il existe une solution au problème suivant :*

$$(\mathcal{P}) : \min_{x\in\mathbb{R}^n} f(x).$$

Démonstration : Une suite minimisante (x_k) est bornée par coercivité, car $(f(x_k))$ est majorée. Elle admet donc une sous-suite $x_{k_j} \to x^*$. Par continuité, $f(x^*) = \lim_j f(x_{k_j}) = \inf_{x\in\mathbb{R}^n} f(x)$. ■

Théorème 2.2

Si la fonction f *est convexe, continue et coercive sur* $\mathbb{R}^n$, *alors l'ensemble des solutions du problème de minimisation* $(\mathcal{P})$ *est convexe.*
En outre, si f *est strictement convexe, alors le minimum est unique.*

Démonstration : Comme f est continue et coercive, alors le problème de minimisation admet au moins une solution.
Soit x et y deux solutions du problème :

$$f(x) = f(y) = \min_{x\in\mathbb{R}^n} f(x).$$

Si f est convexe, alors

$$\begin{aligned} & f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \\ \Longrightarrow\ & f(\lambda x + (1-\lambda)y) \leq \min_{x \in \mathbb{R}^n} f(x), \end{aligned}$$

donc $\lambda x + (1-\lambda)y$ est aussi solution du problème et donc l'ensemble des solutions est bien convexe.
Si f est strictement convexe, avec x et y deux solutions distinctes, alors on a :

$$\begin{aligned} & f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \\ \Longrightarrow\ & f(\lambda x + (1-\lambda)y) < \min_{x \in \mathbb{R}^n} f(x), \end{aligned}$$

ce qui constitue une contradiction.
Ainsi, si f est strictement convexe, alors la solution du problème est unique. ■

Théorème 2.3

Soit f une fonction convexe, coercive, différentiable. Alors il y a équivalence entre les problèmes

$$(\mathcal{P}_1) : \min_{x \in \mathbb{R}^n} f(x) \qquad \textit{et} \qquad (\mathcal{P}_2) : \nabla f(x) = 0.$$

2.2 Algorithmes du gradient

On se propose de donner des méthodes numériques qui sont spécifiques à l'optimisation. Ces méthodes consistent à construire la suite approximante $(x_k)_{k \in \mathbb{N}}$ par récurrence. C'est-à-dire :

- on se donne un point initial $x_0 \in \mathbb{R}^n$;
- on définit $x_{k+1} = x_k + \rho_k d_k$, avec
  - $d_k \in \mathbb{R}^n$, des vecteurs appelés ***direction de descente*** ;
  - $\rho_k > 0$, des scalaires appelés ***facteurs de descente***.

On définit ainsi une suite $(x_k)_{k \in \mathbb{N}}$ telle que

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné}, \\ x_{k+1} = x_k + \rho_k d_k, \\ J(x_{k+1}) \leq J(x_k)\,. \end{cases}$$

2.2.1 Algorithme du gradient

Par développement en séries de Taylor, on établit :

$$J(x_{k+1}) = J(x_k + \rho_k d_k) = J(x_k) + \rho_k \langle \nabla J(x_k)\ ,\ d_k \rangle + \rho_k \|d_k\| \varepsilon(\rho_k d_k)\,.$$

D'où, comme on veut avoir

$$J(x_{k+1}) \leq J(x_k)\,, \qquad \forall k \in \mathbb{N},$$

on tire qu'on peut poser

$$d_k = -\nabla J(x_k)\,.$$

En effet, il vient alors

$$J(x_{k+1}) = J(x_k) - \rho_k \|\nabla J(x_k)\|^2 + \rho_k \|\nabla J(x_k)\| \varepsilon(-\rho_k \nabla J(x_k))\,,$$

et, si le gradient est non nul, pour un petit pas positif,

$$J(x_{k+1}) \leq J(x_k)\,.$$

2.2.2 Algorithme du gradient à pas constant

Théorème 2.4

Considérons une fonctionnelle convexe continue satisfaisant les hypothèses suivantes

- $\langle \nabla f(x) - \nabla f(y)\,,\, x - y\rangle \geq \alpha \|x-y\|^2, \qquad \alpha > 0,\ \forall x, y \in \mathbb{R}^n$;
- $\exists M > 0 :\ \|\nabla f(x) - \nabla f(y)\| \leq M\|x-y\|, \qquad \forall x, y \in \mathbb{R}^n.$

Si $0 < \rho < \frac{2\alpha}{M^2}$, *alors l'algorithme suivant converge.*

- *Initialisation :* $x_0 \in \mathbb{R}^n$ *donné avec* $\nabla f(x_0) \neq 0$
- *Itération :*
  - *pour* $k \geq 0$,
  - *si* $\nabla f(x_k) \neq 0$, *alors*
$$x_{k+1} = x_k - \rho \nabla f(x_k)\,,$$
  - *sinon, arrêt.*

Démonstration : Montrons que si $0 < \rho < \frac{2\alpha}{M^2}$, alors la suite donnée par $(x_k)_{k\in\mathbb{N}}$: $x_{k+1} = x_k - \rho\nabla f(x_k)$ converge vers la solution du problème de minimisation

$$(\mathcal{P}) :\ \min_{x\in\mathbb{R}^n} f(x).$$

Soit x^* solution du problème $(\mathcal{P})$; on a :

$$\begin{aligned}
x_{k+1} - x^* &= x_k - \rho\nabla f(x_k) - x^* - (-\rho\nabla f(x^*)) \\
&= x_k - x^* - \rho\left(\nabla f(x_k) - \nabla f(x^*)\right) \\
\Longrightarrow\ \|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \rho\left(\nabla f(x_k) - \nabla f(x^*)\right)\|^2 \\
&\leq \|x_k - x^*\|^2 - 2\rho\,\langle x_k - x^*\,,\, \nabla f(x_k) - \nabla f(x^*)\rangle + \rho^2\|\nabla f(x_k) - \nabla f(x^*)\|^2 \\
&\leq \|x_k - x^*\|^2 + \rho^2 M^2\|x_k - x^*\|^2 - 2\rho\alpha\|x_k - x^*\|^2 \\
\|x_{k+1} - x^*\|^2 &\leq \left(1 - 2\rho\alpha + \rho^2M^2\right)\|x_k - x^*\|^2.
\end{aligned}$$

Posons $g(\rho) = 1 - 2\rho\alpha + \rho^2M^2$ et étudions g sur $\left]0, \frac{2\alpha}{M^2}\right[$. On a :

$$g'(\rho) = -2\alpha + 2\rho M^2 = 0 \iff \rho = \frac{\alpha}{M^2}.$$

Ce qui permet de définir le tableau de variation suivant pour g :

ρ	0		$\frac{\alpha}{M^2}$		$\frac{2\alpha}{M^2}$
$g'(\rho)$		-	0	+	
$g(\rho)$	1	↘	$1 - \frac{\alpha^2}{M^2}$	↗	1

Ainsi, pour tout $\rho \in \left]0, \frac{2\alpha}{M^2}\right[$, on a : $g(\rho) < 1$, et donc

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 \leq g(\rho)\|x_k - x^*\|^2 \\
\Longrightarrow\ &\|x_k - x^*\|^2 \leq (g(\rho))^k\|x_0 - x^*\|^2.
\end{aligned}$$

Il en découle

$$\lim_{k\to+\infty} \|x_k - x^*\|^2 = 0 \qquad \text{d'où} \qquad x_k \underset{k\to+\infty}{\longrightarrow} x^*.$$

■

2.2.3 Algorithme du gradient à pas optimal

Algorithme général Soit f une fonctionnelle différentiable sur $\mathbb{R}^n$.

- Initialisation : $x_0 \in \mathbb{R}^n$ donné avec $\nabla f(x_0) \neq 0$;
- Itération :
  - pour $k \geq 0$,
  - si $\nabla f(x_k) \neq 0$, alors
    - détermination de $\rho_k^* \in \mathbb{R}$: $f(x_k - \rho_k^* \nabla f(x_k)) = \min_{\rho_k \in \mathbb{R}} f(x_k - \rho_k \nabla f(x_k))$;
    - définition de x_{k+1} par
    $$x_{k+1} = x_k - \rho_k^* \nabla f(x_k) ;$$
  - sinon, arrêt.

Théorème 2.5

Soit f une fonctionnelle elliptique, alors la méthode du gradient à pas optimal est bien défini et converge vers la solution.

Cas linéaire Soit A une matrice carrée symétrique définie positive et J la fonctionnelle définie sur $\mathbb{R}^n$ par :

$$J(X) = \frac{1}{2} \langle AX , X \rangle - \langle b , X \rangle + C,$$

avec b un vecteur de $\mathbb{R}^n$ et $C \in \mathbb{R}$. On considère le problème

$$(\mathcal{P}) : \min_{x \in \mathbb{R}^n} J(X).$$

Théorème 2.6

Le problème $(\mathcal{P})$ admet une unique solution $X^* \in \mathbb{R}^n$ et X^* est l'unique solution de l'équation $AX = b$.

Théorème 2.7

Soit f une fonctionnelle elliptique alors la méthode du gradient à pas optimal est bien défini et converge.

Cas linéaire Soit A une matrice carrée symétrique définie positive, et J une fonctionnelle définie sur $\mathbb{R}^n$ par

$$J(X) = \frac{1}{2} \langle AX , X \rangle - \langle b , X \rangle + C$$

b étant un vecteur de $\mathbb{R}^n$ et $C \in \mathbb{R}$. On considère le problème

$$(\mathcal{P}) : \min_{X \in \mathbb{R}^n} J(X).$$

Théorème 2.8

Le problème $(\mathcal{P})$ admet une solution unique $X^* \in \mathbb{R}^n$ et X^* est l'unique solution de l'équation $AX = b$.

Démonstration : On a

$$\begin{aligned} \nabla J(X) &= \frac{1}{2}\left(A^t + A\right)X - b \\ \Longrightarrow\ \nabla^2 J(X) &= A. \end{aligned}$$

Il en découle que J est strictement convexe, et il ne reste plus donc qu'à vérifier que J est coercive. ■

Théorème 2.9

Soit A une matrice symétrique définie positive, $b \in \mathbb{R}^n$ et $C \in \mathbb{R}$, la matrice A étant symétrique définie positive. Alors le pas optimal à l'itération d'ordre $k \geq 0$ est donné par

$$\rho_k = \frac{\langle \nabla J(X_k)\ ,\ \nabla J(X_k)\rangle}{\langle A\nabla J(X_k)\ ,\ \nabla J(X_k)\rangle}.$$

Démonstration : . ■

2.2.4 Algorithme du gradient conjugué

Cas linéaire

Théorème 2.10

Soit A une matrice symétrique définie positive. On considère la fonctionnelle J définie par

$$J(X) = \frac{1}{2}\langle AX\,,\,X\rangle - \langle b\,,\,X\rangle + C.$$

Alors, l'algorithme suivant converge vers la solution de

$$(\mathcal{P}) : \min_{X\in\mathbb{R}^n} J(X).$$

- Initialisation
  - $X_0 \in \mathbb{R}^n$ donné, avec $\nabla J(X_0) \neq 0$;
  - $r_0 = b - AX_0$, $\rho_0 = r_0$;
- Itération
  - $\alpha_k = \dfrac{\langle r_k\,,\,r_k\rangle}{\langle A\rho_k\,,\,\rho_k\rangle}$
  - $X_{k+1} = X_k + \rho_k\alpha_k$;
  - $r_{k+1} = r_k - \alpha_k A\rho_k$; arrêt si $r_{k+1} = 0$;
  - $B_k = \dfrac{\langle r_{k+1}\,,\,r_{k+1}\rangle}{\langle r_k\,,\,r_k\rangle}$;
  - $\rho_{k+1} = r_{k+1} + B_k\rho_k$.

De plus, l'algorithme converge en au plus n itérations en arithmétique exacte.

Cas général

Soit J elliptique à gradient lipschitzien. L'algorithme suivant

- Initialisation : $X_0 \in \mathbb{R}^n$, $d_0 = -\nabla J(X_0) \neq 0$;
- Itération
  - $X_{k+1} = X_k + \rho_k d_k$, où $\rho_k \in \mathbb{R}_+$ est l'unique solution qui minimise $J(X_k + \rho_k d_k)$;
  - $B_k = \dfrac{\|\nabla J(X_{k+1})\|^2}{\|\nabla J(X_k)\|^2}$;
  - $d_{k+1} = -\nabla J(X_{k+1}) + B_k d_k$;
  - arrêter si $\|\nabla J(X_{k+1})\|^2 = 0$; sinon poursuivre ;

converge vers la solution du problème $(\mathcal{P})$: $\min_{X \in \mathbb{R}^n} J(X)$.

2.2.5 Algorithme de Newton

Soit J une fonction deux fois continûment différentiable, $J \in \mathcal{C}^2(\mathbb{R}^n)$, et soit $\nabla^2 J$ sa matrice hessienne. On peut construire une approximation de ∇J au voisinage de X en posant :

$$\nabla J(Y) \simeq \nabla J(X) + \nabla^2 J(X) \cdot (Y - X).$$

Pour J coercive et convexe, le problème d'optimisation est équivalent à trouver la solution de l'équation d'Euler, c'est-à-dire $\nabla J(X) = 0$.

On peut donc utiliser une linéarisation de $\nabla J(X)$, et poser autour d'un point $X_n \in \mathbb{R}^n$,

$$\nabla J(Y) \simeq \nabla J(X_n) + \nabla^2 J(X_n) \cdot (Y - X_n).$$

L'équation approchée au voisinage de X_n est :

$$\nabla J(X_n) + \nabla^2 J(X_n) \cdot (Y - X_n) = 0.$$

Cela conduit à poser pour J elliptique, à

$$X_{n+1} = X_n - [\nabla^2 J(X_n)]^{-1} \cdot \nabla J(X_n), \quad \text{avec } Y = X_{n+1},$$
$$\Longrightarrow\ Y = X_n - \left[\nabla^2 J(X_n)\right]^{-1} \cdot \nabla J(X_n).$$

C'est l'algorithme de Newton de la fonction ∇J.

Théorème 2.11

Soit $J \in \mathcal{C}^2(\mathbb{R}^n)$, *$\alpha$-fortement convexe, et* X^* *son minimiseur. Supposons* $\nabla^2 J$ *M-lipschitzienne sur une boule* $B(X^*, r)$, *avec* $M > 0$ *et* $r > 0$.
Si $\|X_0 - X^*\| < \min(r, \alpha/M)$, *Newton converge vers* X^*, *reste dans cette boule et vérifie la majoration quadratique*

$$\|X_{n+1} - X^*\| \leq \frac{M}{2\alpha}\|X_n - X^*\|^2.$$

2.2.6 Théorèmes généraux de minimalité

Condition de minimalité du premier ordre

Soit O un ouvert de $\mathbb{R}^n$, et f définie de O dans $\mathbb{R}$. Si $\overline{x}$ est le minimum local de f et si f est deux fois différentiable en $\overline{x}$, alors $\nabla f(\overline{x}) = 0$ et $\nabla^2 f(\overline{x})$ est semi-définie positive.

Si $\overline{x}$ est un point où f est deux fois différentiable et si $\nabla f(\overline{x}) = 0$ et $\nabla^2 f(\overline{x})$ est définie positive, alors $\overline{x}$ est un minimum local strict de f.

Chapitre 3

Optimisation avec contraintes

3.1 Résultats d'existence

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ différentiable (sur un ouvert $O \subset \mathbb{R}^n$), et soit S un ensemble de contraintes décrit par m égalités et p inégalités.

$$S = \left\{ x \in \mathbb{R}^n : \begin{cases} h_1(x) = h_2(x) = \cdots = h_m(x) = 0 \\ g_1(x) \leq 0, \ \ldots, g_p(x) \leq 0 \end{cases} \right\}.$$

Les $m + p$ fonctions h_i et g_i sont supposées de classe $\mathcal{C}^1$.

Théorème 3.1 (*F. John*)

Si $\overline{x} \in O \cap S$ est un minimum local de f sur S, alors il existe $\overline{\lambda_1}, \ldots, \overline{\lambda_m}, \overline{\alpha_0}, \ldots, \overline{\alpha_p}$ non tous nuls, tels que

1. $\overline{\alpha_0} \nabla f(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda_i} \nabla h_i(\overline{x}) + \sum_{j=1}^{p} \overline{\alpha_j} \nabla g_j(\overline{x}) = 0$;
2. $\overline{\alpha_j} \geq 0$ *pour* $j = 0, \ldots, p$;
3. $\overline{\alpha_j} = 0$ *si* $g_j(\overline{x}) < 0$.

On considère les conditions de Mangasarian–Fromovitz suivantes

1. *il existe* d_0 *tel que*
$$\langle \nabla h_i(\overline{x}) \ , \ d_0 \rangle = 0, \qquad \forall i = 1, \ldots, m;$$
2. $\langle \nabla g_j(\overline{x}) \ , \ d_0 \rangle < 0$ *pour tout* j *vérifiant* $g_j(\overline{x}) = 0$;
3. *les vecteurs* $\nabla h_1(\overline{x}), \ldots, \nabla h_m(\overline{x})$ *sont linéairement indépendants.*

Théorème 3.2 (*Karush-Kuhn-Tucker*)

Si $\overline{x} \in O \cap S$ est un minimum local de f sur S et si les conditions de Mangasarian–Fromovitz précédentes ont lieu, il existe

$$\overline{\lambda} = \left(\overline{\lambda_1}, \ldots, \overline{\lambda_m}\right) \in \mathbb{R}^m \qquad et \qquad \overline{M} = \left(\overline{M_1}, \ldots, \overline{M_p}\right) \in \mathbb{R}^p$$

tels que

1. $\nabla f(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda_i} \nabla h_i(\overline{x}) + \sum_{j=1}^{p} \overline{M_j} \nabla g_j(\overline{x}) = 0$;
2. $\overline{M_j} \geq 0, \quad \forall j = 1, \ldots, p$;
3. $\overline{M_j} = 0$ *si* $g_j(\overline{x}) < 0$.

Les $\overline{\lambda_i}$ et $\overline{M_j}$ mis en évidence dans ce Théorème de KKT sont appelés multiplicateurs de Lagrange. Une hypothèse de qualification de contrainte en $\overline{x}$ plus forte que les conditions de Mangasarian–Fromovitz est la suivante :

$(Q)_{\overline{x}}$: les vecteurs $\nabla h_1(\overline{x}), \ldots, \nabla h_m(\overline{x}), \nabla g_j(\overline{x})$, $j \in J(\overline{x})$ sont linéairement indépendants, où $J(\overline{x}) = \{j : \ g_j(\overline{x}) = 0\}$.

$J(\overline{x})$ est l'ensemble des indices de contraintes (de type inégalités) actives. Si les h_i et g_j sont affines, on peut se dispenser de toute hypothèse de qualification des contraintes et déduire quand même les conditions nécessaires de Karush-Kuhn-Tucker.

Les conditions du Théorème précédent (appelées conditions nécessaires de KKT) deviennent suffisantes en présence de convexité dans les domaines. Considérons à cet effet, le contexte suivant

- f est convexe sur l'ouvert O ;
- les $h_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ sont affines ;
- les $g_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ sont convexes.

Théorème 3.3

Dans la situation décrite ci-dessus, les conditions de KKT sont suffisantes pour que $\overline{x}$ soit un minimum local sur $O \cap S$.

3.2 Lagrangien du problème

Soient $\lambda \in \mathbb{R}^m$ et $\alpha \in (\mathbb{R}_+)^p$, on définit

$$\mathcal{L}(x, \lambda, \alpha) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \alpha_j g_j(x)$$
$$\mathcal{L}(x, \lambda, \alpha) = f(x) + \langle \lambda \, , \, h(x) \rangle + \langle \alpha \, , \, g(x) \rangle \, ,$$

avec $h(x) = (h_1(x), \ldots, h_m(x))$.

$\mathcal{L}$ est le Lagrangien de $(\mathcal{P}) : \min_{x \in S} f(x)$, avec

$$S = \left\{ x \in \mathbb{R}^n : \ \begin{cases} h(x) = 0_{\mathbb{R}^m}, \\ g_1(x) \leq 0, \ldots, g_p(x) \leq 0 \end{cases} \right\}.$$

Les paramètres λ et α sont appelés multiplicateurs de Lagrange. L'intérêt du Lagrangien est de ramener le problème de minimisation sous contraintes à un problème de minimisation sans contrainte.

Proposition 3.1. *Le point x^* est un minimiseur global de f sur S si et seulement s'il est aussi solution de*

$$\inf_{x\in\mathbb{R}^n} \sup_{\substack{\lambda\in\mathbb{R}^m \\ \alpha\in\mathbb{R}^p_+}} \mathcal{L}(x,\lambda,\alpha).$$

3.2.1 Points selles du Lagrangien

L'idée générale de l'introduction du Lagrangien est que l'on doit maximiser $\mathcal{L}$ par rapport à sa deuxième variable et le minimiser par rapport à sa première variable. Cela ressemble à la recherche de points selles.

Définition 3.1

Soit $\mathcal{L} : U \times V \longrightarrow \mathbb{R}$ une fonction à deux variables, le point $(x^*, p^*) \in U \times V$ est dit ***point selle*** *de $\mathcal{L}$ sur $U \times V$ si et seulement si*

$$\mathcal{L}(x^*, p) \leq \mathcal{L}(x^*, p^*) \leq \mathcal{L}(x, p^*), \qquad \forall (x, p) \in U \times V.$$

Théorème 3.4

Considérons le problème sous-contraintes fonctionnelles défini par

$$(\mathcal{P}) : \min_{x\in S} f(x), \qquad (S \text{ est l'ensemble des contraintes}),$$

où f, g et h sont supposées différentiables. On note $\mathcal{L}$ le Lagrangien associé à f aux contraintes S.
Si $(x^*, \lambda^*, \alpha^*)$ est un point selle de $\mathcal{L}$ sur $\mathbb{R}^n \times (\mathbb{R}^m \times {\mathbb{R}_+}^p)$, alors x^* est un minimum global de f sur S.
De plus, on a

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \alpha_j^* \nabla g_j(x^*) = 0, \\ \alpha_j^* g_j(x^*) = 0. \end{cases}$$

Théorème 3.5 (*Conditions nécessaires d'optimalité*)

Considérons le problème

$$(\mathcal{P}) : \min_{x\in S} f(x),$$

où f, g et h sont supposées différentiables, S le domaine de contrainte et $\mathcal{L}$ le Lagrangien associé à f et aux contraintes S.
Si x^* est un minimum local de f sur S et si les contraintes sont qualifiées en x^*, alors il existe des multiplicateurs $(\lambda^*, \alpha^*) \in \mathbb{R}^m \times \mathbb{R}^p_+$ tels que

$$\begin{cases} \nabla_x \mathcal{L}(x^*, \lambda^*, \alpha^*) = \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \alpha_j^* \nabla g_j(x^*) = 0, \\ \forall j = 1, \ldots, p, \quad \alpha_j^* g_j(x^*) = 0. \end{cases}$$

Le vecteur (λ^*, α^*) est appelé ***vecteur de multiplicateurs*** du problème $(\mathcal{P})$.

On appelle ***point de Karush-Kuhn-Tucker*** du problème $(\mathcal{P})$ tout point $\bar{x} = (x^*, \lambda^*, \alpha^*)$ vérifiant les conditions d'optimalité du premier ordre suivantes

$$\begin{cases} \nabla_x \mathcal{L}(x^*, \lambda^*, \alpha^*) = 0, \\ \alpha_j^* g_j(x^*) = 0, \quad \forall j = 1, \dots, p, \\ h_i(x^*) = 0, \quad \forall i = 1, \dots, m, \ \alpha_j^* \geq 0, \ g_j(x^*) \leq 0, \ \forall j = 1, \dots, p. \end{cases}$$

3.3 Algorithme

Théorème 3.6

Supposons que f, g et h soient des fonctions de classe $\mathcal{C}^1$ et que le triplet $(x^*, \lambda^*, \alpha^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ soit un point selle de $\mathcal{L}$. Alors ce triplet vérifie les conditions de KKT.
Dans le cas convexe, on a une caractérisation des points selles grâce aux conditions KKT.

Théorème 3.7

Supposons f et g convexes de classe $\mathcal{C}^1$ et h affine. Alors le triplet $(x^*, \lambda^*, \alpha^*)$ est un point selle de $\mathcal{L}$ sur $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ si et seulement s'il vérifie les conditions KKT.

Le Théorème précédent indique une procédure algorithmique de recherche de minimum. Nous allons chercher un triplet $(x^*, \lambda^*, \alpha^*)$ vérifiant les conditions KKT comme suit :

1. Pour (λ^*, α^*) fixé dans $\mathbb{R}^m \times \mathbb{R}_+^p$, on cherche le minimum sur $\mathbb{R}^n$ de la fonction $x \longmapsto \mathcal{L}(x, \lambda^*, \alpha^*)$.
2. Pour x^* fixé dans $\mathbb{R}^n$, on cherche le maximum sur $\mathbb{R}^m \times \mathbb{R}_+^p$ de la fonction $(\lambda, \alpha) \longmapsto \mathcal{L}(x^*, \lambda, \alpha)$.

On alterne minimisation en x et ascension en multiplicateurs dans l'algorithme d'USAWA

- Initialisation : $k = 0$
  - choisir $\lambda^0 \in \mathbb{R}^m$ et $\alpha_0 \in \mathbb{R}_+^p$.
- Tant que le critère d'arrêt n'est pas satisfait :
  - calculer $x_k \in \mathbb{R}^n$ solution de

$$(\mathcal{P}_k) : \begin{cases} x_k \in \mathbb{R}^n, \\ \mathcal{L}\big(x_k, \lambda^k, \alpha_k\big) = \min_{x \in \mathbb{R}^n} \mathcal{L}\big(x, \lambda^k, \alpha_k\big); \end{cases}$$

  - calculer λ^{k+1} et α_{k+1} avec

$$\begin{cases} \lambda_i^{k+1} & = \lambda_i^k + \rho h_i\big(x^k\big), \quad i = 1, \dots, m \\ \alpha_j^{k+1} & = \max\big(0, \alpha_j^k + \rho g_j\big(x^k\big)\big), \quad j = 1, \dots, p, \end{cases}$$

    où $\rho > 0$ est un réel fixé.

La première étape revient à résoudre

$$\nabla_x \mathcal{L}\big(x, \lambda^k, \alpha^k\big) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x) + \sum_{j=1}^{p} \alpha_j^k \nabla g_j(x) = 0.$$

La seconde étape est immédiate. On a alors le Théorème de convergence suivant.

Théorème 3.8

On suppose que f est de classe $\mathcal{C}^1$ elliptique, h affine et g convexe de classe $\mathcal{C}^1$, et que ∇f et g sont lipschitziennes. On suppose de plus que le Lagrangien possède un point selle $(x^*, \lambda^*, \alpha^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$. Alors, il existe ρ_1 et ρ_2 avec $0 < \rho_1 < \rho_2$, tels que, pour $\rho \in [\rho_1, \rho_2]$, la suite (x_k) générée par l'algorithme d'USAWA converge vers x^*.

3.3.1 Algorithme d'USAWA

On considère le cas de m contraintes égalités affines. Soit

$$K = \{x \in \mathbb{R}^n : \ Cx = d\}, \quad \text{où} C \in \mathcal{M}_{m,n}(\mathbb{R}) \text{ et } d \in \mathbb{R}^m.$$

Dans ce cas, les contraintes d'optimalité s'écrivent

$$\exists \lambda \in \mathbb{R}^m : \begin{cases} \nabla f(u) + C^t\lambda = 0, \\ Cu - d = 0. \end{cases}$$

On réécrit ces équations pour un $\rho > 0$ sous la forme

$$\exists \lambda \in \mathbb{R}^m : \begin{cases} \nabla f(u) + C^t\lambda = 0, \\ \lambda = \lambda + \rho(Cu - d). \end{cases}$$

Cette dernière forme suggère

Algorithme d'USAWA (Contraintes égalités affines)

On prend un multiplicateur de départ $\lambda_0 \in \mathbb{R}^m$. On fixe un pas $\rho > 0$, puis on itère pour $k \geq 0$:

- Calculer u^k tel que $\nabla f\big(u^k\big) + C^t\lambda^k = 0$.
- Calculer $\lambda^{k+1} = \lambda^k + \rho\left(Cu^k - d\right)$.
- Passer à $k + 1$.

Théorème 3.9

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, α-convexe et différentiable, et un ensemble convexe

$$K = \{x \in \mathbb{R}^n : \ Cx - d = 0\}, \quad \textit{avec } C \in \mathcal{M}_{m,n}(\mathbb{R}) \textit{ et } d \in \mathbb{R}^m,$$

supposé non vide, avec $C \neq 0$.

1. *Il existe un unique minimiseur u de f sur K.*
2. *Pour tous $\rho \in \left]0, \frac{2\alpha}{\|C\|^2}\right[$ et $\lambda_0 \in \mathbb{R}^m$ de départ, l'algorithme d'USAWA converge :*

$$\lim_{n \to +\infty} x_n = u^*, \quad \textit{avec } f(u^*) = \min_{x \in K} f(x).$$

3. *Si de plus, C est surjective et si ∇f est continue, on a alors aussi la convergence de λ^n vers une unique solution de*

$$\nabla f(u^*) + C^t\lambda^* = 0.$$

Pour rappel, on a

$$\|C\|^2 = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Cx\|^2}{\|x\|^2}.$$