

Conditions nécessaires d'optimalité pour les problèmes d'optimisation différentiable

Guillaume Lecué¹

Soit U est un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$. On s'intéresse aux deux problèmes suivants :

— trouver une solution au problème d'optimisation **sans contrainte**

$$\min_{x \in U} f(x) \tag{0.1}$$

— trouver une solution au problème d'optimisation **sous contrainte**

$$\min_{x \in K} f(x) \tag{0.2}$$

où $K \subset U$ est un ensemble fermé de \mathbb{R}^n .

Dans ce cours, on traite seulement le cas où f est différentiable et K est une contrainte définie par des fonctions différentiables. C'est sous cette hypothèse de différentiation que les approximations locales de f et K sont les plus faciles à donner.

Dans ce chapitre, on aborde d'abord la question de l'existence d'une solution aux problèmes (0.1) et (0.2). Ensuite, on identifie des conditions du premier ordre que les solutions de (0.1) et (0.2) doivent nécessairement satisfaire. Ces conditions ne sont que nécessaires et non suffisantes car, dans ce chapitre, on ne fait pas d'hypothèse de convexité. Elles deviendront des CNS dans le chapitre suivant si on suppose la convexité en plus. Dans le cas du problème d'optimisation sans contrainte (0.1), on identifiera une condition du second ordre suffisante pour qu'un point soit un minimum local.

Le but de ce chapitre est de rassembler tous les théorèmes utiles en optimisation différentiable (non nécessairement convexe) et la méthodologie associée pour les appliquer. On donnera les preuves plus tard. On fera de même pour les problèmes d'optimisation différentiables et convexes.

1 Existence de solutions

On donne ici quelques conditions suffisantes assurant l'existence de solution aux problèmes (0.1) et (0.2). Ces conditions sont dérivées du théorème de Weierstrass.

Définition 1.1 Soit $\|\cdot\|$ une norme sur \mathbb{R}^n . Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est **coercive** quand $\lim_{\|x\| \rightarrow \infty} f(x) = +\infty$.

Exemple 1.2 Soit $A \in \mathbb{R}^{n \times n}$ une matrice symétrique, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$. Soit $f : x \in \mathbb{R}^n \rightarrow c + \langle b, x \rangle + x^\top A x$. Si A est définie positive alors f est coercive.

En effet, on considère la décomposition en valeur singulière (SVD) de A donnée par $A = PDP^\top$ où $P \in \mathcal{O}(n)$ (où $\mathcal{O}(n)$ est le groupe des matrices orthogonales, i.e. telles que $PP^\top =$

1. CREST, ENSAE. Bureau 3029, 5 avenue Henry Le Chatelier. 91 120 Palaiseau. Email : guillaume.lecue@ensae.fr.

$P^\top P = I_n$ et I_n est la matrice identité de $\mathbb{R}^{n \times n}$ et $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ où $\lambda_1 \geq \dots \geq \lambda_n > 0$ (on a $\lambda_n > 0$ car on a supposé que A est définie positive). On a

$$f(Px) = x^\top Dx + \langle Px, b \rangle + c$$

et comme $x^\top Dx \geq \lambda_n \|x\|_2^2$ et $|\langle Px, b \rangle| \leq \|x\|_2 \|b\|_2$, on en déduit que $\lim_{\|x\|_2 \rightarrow \infty} f(Px) = +\infty$ et comme $\|Px\|_2 = \|x\|_2$, on a bien que f est coercive.

Théorème 1.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et K un sous-ensemble non vide de \mathbb{R}^n . Le problème (0.2) admet une solution si l'une des conditions suivante est satisfaite :

1. K est un ensemble compact de \mathbb{R}^n (théorème de Weierstrass)
2. f est coercive et K est fermée.

Preuve. Quand K est compact, on note par $m = \inf_{x \in K} f(x)$ et $(x_n)_n \subset K$ telle que $f(x_n) \rightarrow m$. Par compacité, on peut extraire une sous-suite de $(x_n)_n$ qui converge dans K : on note $x_{\varphi_n} \rightarrow x^*$ pour un certain $x^* \in K$. Par continuité, on a $f(x_{\varphi_n}) \rightarrow f(x^*)$ et comme $f(x_{\varphi_n})$ est une sous-suite de la suite $f(x_n)$ qui converge vers m , on a aussi $f(x_{\varphi_n}) \rightarrow m$ et donc $m = f(x^*)$ pour un certain $x^* \in K$ donc l'infimum est atteint et donc (0.2) admet une solution.

Dans le deuxième cas, on note $m = \inf_{x \in K} f(x)$ et on définit

$$K_1 = \begin{cases} \{x \in K : f(x) \leq m + 1\} & \text{si } m \in \mathbb{R} \\ \{x \in K : f(x) \leq 0\} & \text{si } m = -\infty. \end{cases}$$

Comme f est coercive, K_1 est bornée. Comme f est continue K_1 est un ensemble fermé de K et comme K est fermé, c'est aussi un ensemble fermé de \mathbb{R}^n . Alors K_1 est un fermé borné de \mathbb{R}^n , il est donc compact dans \mathbb{R}^n . Par Weierstrass, f atteint $\inf_{x \in K_1} f(x)$, on note $x^* \in K_1$ tel que $f(x^*) = \inf_{x \in K_1} f(x)$. Montrons que x^* est aussi un minimum de f sur K . Soit $x \in K$, si $x \in K_1$, on a bien $f(x^*) \leq f(x)$. Si $x \notin K_1$ alors soit $m \in \mathbb{R}$ et alors $f(x) \geq m + 1 \geq f(x^*)$, soit $m = -\infty$ et alors $f(x^*) \leq 0 \leq f(x)$. Dans tous les cas, $f(x) \geq f(x^*)$. ■

On peut étendre la notion de coercivité aux fonctions définies seulement sur un ouvert U de \mathbb{R}^n plutôt que sur tout \mathbb{R}^n . Pour cela, on rappelle que le bord de U est $\partial U = \bar{U} \setminus U$ (ici U est ouvert, il est donc égal à son intérieur). On dit alors qu'une fonction $f : U \rightarrow \mathbb{R}$ est coercive quand

$$\lim_{x \rightarrow \partial U} f(x) = +\infty. \quad (1.1)$$

Dans ce cas, on peut aussi appliquer le théorème précédent : si $f : U \rightarrow \mathbb{R}$ est continue et coercive sur son domaine U au sens de (1.1) et que $K \subset U$ est un fermé de \mathbb{R}^n alors (0.2) admet une solution.

Pour les problèmes d'optimisation sans contraintes, c'est-à-dire de la forme $\min_{x \in U} f(x)$, où U est un ouvert de \mathbb{R}^n . On peut utiliser le résultat d'existence suivant.

Théorème 1.4 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$. On suppose qu'il existe c tel que $\{x \in U : f(x) \leq c\}$ est un ensemble non vide et compact de \mathbb{R}^n . Alors le problème d'optimisation sans contrainte (0.1) admet une solution.

Preuve. On note $K_1 = \{x \in U : f(x) \leq c\}$. Comme K_1 est non vide et compact et que f est continue, f admet un minimum x^* sur K_1 . Montrons que x^* est solution de $\min_{x \in U} f(x)$. Soit $x \in U$. Si $x \in K_1$ alors $f(x^*) \leq f(x)$ par minimalité de x^* sur K_1 ; sinon $f(x) > c$. Or $x^* \in K_1$, donc $f(x^*) \leq c$ et donc $f(x^*) \leq f(x)$. On a donc bien pour tout $x \in U$, $f(x) \geq f(x^*)$ et donc x^* est un minimum de f sur U . ■

2 Condition d'optimalité du 1-er ordre pour les problèmes d'optimisation sans contrainte

Si U est un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$, on dit que le problème $\min_{x \in U} f(x)$ est un **problème d'optimisation sans contrainte** car U est un ensemble ouvert. On n'a donc pas de problème d'atteinte de solution au bord.

Définition 2.1 Soit $f : U \rightarrow \mathbb{R}$ où U est un ouvert de \mathbb{R}^n . Soit $x^* \in \mathbb{R}^n$. On dit que

1. x^* est un **minimum local** quand il existe $\epsilon > 0$ tel que pour tout $x \in U \cap B_2(x^*, \epsilon)$, $f(x^*) \leq f(x)$ où on note $B_2(x^*, \epsilon) = \{x \in \mathbb{R}^n : \|x - x^*\|_2 \leq \epsilon\}$
2. on dit que x^* est un **minimum global** quand pour tout $x \in U$, $f(x) \geq f(x^*)$.

En particulier, un minimum global, c'est une solution au problème $\min(f(x) : x \in U)$ est forcément un minimum local. On donne dans le résultat suivant une condition nécessaire satisfaite par les minima locaux.

Théorème 2.2 Soit $f : U \rightarrow \mathbb{R}$ où U est un ouvert de \mathbb{R}^n . Soit $a \in U$. On suppose que f est différentiable en a . Si a est un minimum local de f sur U alors $\nabla f(a) = 0$.

Preuve. On veut montrer que $\nabla f(a) = 0$ quand a est un minimum local de f . Pour cela, il suffit de montrer que $\langle \nabla f(a), v \rangle = 0$ pour tout $v \in \mathbb{R}^n$. Soit $v \in \mathbb{R}^n$ tel que $v \neq 0$. Comme f est différentiable en a , on a quand, $t \rightarrow 0$,

$$f(a + tv) = f(a) + \langle \nabla f(a), tv \rangle + o(t). \quad (2.1)$$

Or a est un minimum local donc il existe $\epsilon > 0$ tel que pour tout $y \in B_2(a, \epsilon)$, $f(y) \geq f(a)$. En particulier, pour tout $0 < t \leq \epsilon / \|v\|_2$, on a $f(a + tv) \geq f(a)$. Alors quand $t \rightarrow 0$ et $0 < t \leq \epsilon / \|v\|_2$, on a d'après (2.1) que $\langle \nabla f(a), tv \rangle + o(t) \geq 0$. Alors, par linéarité, on a aussi $\langle \nabla f(a), v \rangle + o(1) \geq 0$ et donc $\langle \nabla f(a), v \rangle \geq 0$. Ceci étant vrai pour tout $v \in \mathbb{R}^n$, on conclut que c'est aussi vrai pour $-v$ alors $-\langle \nabla f(a), v \rangle \geq 0$ et donc $\langle \nabla f(a), v \rangle = 0$. ■

Remarque 2.3 — La réciproque est fausse comme on peut le voir avec $f : x \in \mathbb{R} \rightarrow x^3$ qui est tel que $f'(0) = 0$ mais 0 n'est pas un minimum local de f .

Définition 2.4 Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction différentiable. Soit $a \in U$. On dit que a est un **point critique de f** quand $\nabla f(a) = 0$.

Les points critiques jouent un rôle central en optimisation car d'après le Théorème 2.2, c'est parmi eux qu'on doit chercher les solutions des problèmes d'optimisation sans contrainte. La condition " $\nabla f(x^*) = 0$ " est une condition du premier ordre (car elle ne fait intervenir que le gradient – qui donne l'approximation du premier ordre de f en x^*) et c'est uniquement une condition nécessaire (c'est-à-dire, si x^* est solution de $\min f(x)$ alors nécessairement $\nabla f(x^*) = 0$). Cette condition deviendra une CNS quand on supposera de plus que f est convexe.

3 Condition du second ordre pour les problèmes d'optimisation sans contrainte

Quand on cherche une solution à un problème d'optimisation sans contrainte, on sait que nécessairement, cette solution se trouve parmi les minima locaux de f qui eux sont forcément des points critiques : c'est la condition (nécessaire) du premier ordre donné par le Théorème 2.2. Il se trouve qu'on peut trouver une condition suffisante pour qu'un point critique soit en fait un minimum local. Pour cela, on regarde la forme quadratique qui approche le mieux f en un point critique. L'idée est que si localement, f ressemble à un "bol" de sommet x^* alors x^* sera un minimum local.

La meilleure approximation quadratique d'une fonction s'obtient en déterminant sa matrice Hessienne. On va donc introduire des conditions faisant apparaître la matrice Hessienne. De telle condition sont appelées des conditions du second ordre.

On commence par rappeler ici, ce qu'on appelle "décomposition en valeur singulière". On s'en est déjà servi avant. On dit qu'une matrice $A \in \mathbb{R}^{n \times n}$ est **symétrique** quand $A^\top = A$. Une matrice symétrique est dite **positive** quand pour tout $x \in \mathbb{R}^n$, $\langle x, Ax \rangle \geq 0$ et que A est **définie positive** quand $\langle x, Ax \rangle > 0$ pour tout $x \in \mathbb{R}^n \setminus \{0\}$. On note

$$A \succeq 0 \text{ quand } A \text{ est positive;} \quad A \succ 0 \text{ quand } A \text{ est définie positive.}$$

On sait que toute matrice symétrique est diagonalisable dans une base orthonormée. En terme d'écriture matricielle cela signifie qu'il existe une matrice orthonormale P – c'est-à-dire $PP^\top = P^\top P = I_n$ – et une matrice diagonale $D = \text{diag}(\sigma_1, \dots, \sigma_n)$ telles que $A = PDP^\top$. Les éléments diagonaux de D , $\sigma_1, \dots, \sigma_n$, sont appelés les **valeurs singulières** de A qui sont ici aussi les valeurs propres de A , car A est symétrique. Si on note par p_1, \dots, p_n les vecteurs colonnes de P alors (p_1, \dots, p_n) forme une base orthonormale de \mathbb{R}^n et $Ap_j = \sigma_j p_j, j = 1, \dots, n$. C'est l'écriture $A = PDP^\top$ qu'on appelle **décomposition en valeurs singulières de A** . On voit que $A \succeq 0$ si et seulement si $\sigma_j \geq 0, j = 1, \dots, n$ et $A \succ 0$ si et seulement si $\sigma_j > 0, j = 1, \dots, n$.

Théorème 3.1 Soit U un ouvert de \mathbb{R}^n . Soit $f : U \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 . Soit $a \in U$. On a les deux propriétés suivantes :

(CN) Si a est un minimum local de f alors $\nabla f(a) = 0$ et $\nabla^2 f(a) \succeq 0$

(CS) Si $\nabla f(a) = 0$ et $\nabla^2 f(a) \succ 0$ alors a est un minimum local.

Preuve. On montre (CN). Soit a un minimum local de f . On sait d'après la condition du premier ordre que nécessairement $\nabla f(a) = 0$. Par ailleurs, d'après la formule de Taylor-Young, on a, quand $h \rightarrow 0$,

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} h^\top \nabla^2 f(a) h + o(\|h\|_2^2). \quad (3.1)$$

Mais $\nabla f(a) = 0$ et a est un minimum local. Donc si on prends $v \in \mathbb{R}^n$ alors pour un certain $t_0 > 0$, on aura, pour $0 < t \leq t_0$ et $t \rightarrow 0$, $\frac{1}{2}(tv)^\top \nabla^2 f(a)(tv) + o(\|tv\|_2^2) \geq 0$ c'est-à-dire $v^\top \nabla^2 f(a)v \geq 0$. Donc $\nabla^2 f(a) \succeq 0$.

On montre (CS). Comme $\nabla f(a) = 0$, on a, d'après la formule de Taylor, quand $h \rightarrow 0$,

$$f(a+h) - f(a) = \langle \nabla f(a), h \rangle + \frac{1}{2} h^\top \nabla^2 f(a) h + o(\|h\|_2^2) \geq \|h\|_2^2 \left(\frac{1}{2} \left(\frac{h}{\|h\|_2} \right)^\top \nabla^2 f(a) \left(\frac{h}{\|h\|_2} \right) + o(1) \right) \quad (3.2)$$

Par ailleurs, $v^\top \nabla^2 f(a) v > 0$ pour tout $v \in \mathcal{S}_2^{n-1}$ où $\mathcal{S}_2^{n-1} = \{v \in \mathbb{R}^n : \|v\|_2 = 1\}$ est la sphère Euclidienne de \mathbb{R}^n . Or \mathcal{S}_2^{n-1} est compact (c'est un fermé borné de \mathbb{R}^n) et $v \rightarrow v^\top \nabla^2 f(x) v$ est continue donc $\inf_{v \in \mathcal{S}_2^{n-1}} v^\top \nabla^2 f(x) v = \sigma_0 > 0$. Alors d'après (3.2), quand $h \rightarrow 0$,

$$f(a+h) - f(a) \geq \|h\|_2^2 (\sigma_0/2 + o(1)).$$

Par définition de $o(1)$, il existe $\epsilon > 0$ tel que pour tout $h \in \mathbb{R}^n$ tel que $\|h\|_2 \leq \epsilon$, on a $|o(1)| \leq \sigma_0/4$. Alors, pour tout $h \in B_2(0, \epsilon)$, on a $f(a+h) - f(a) \geq 0$. Donc a est bien un minimum local de f . ■

Le théorème 3.1 ne donne pas une condition nécessaire et suffisante (CNS) pour qu'un point soit un minimum local mais seulement une condition nécessaire et une condition suffisante qui ne coïncident pas vu que l'une demande $\nabla^2 f(a) \succ 0$ alors que l'autre ne suppose que $\nabla^2 f(a) \succeq 0$. Les cas pathologiques ont lieu quand $\nabla^2 f(a) \succeq 0$ mais pas $\nabla^2 f(a) \succ 0$. C'est-à-dire quand la plus petite valeur singulière de $\nabla^2 f(a)$ est nulle. C'est-à-dire quand le noyau de $\nabla^2 f(a)$ est non trivial. Cela indique qu'il y a des directions, où la fonction f est "très plate", soit qu'elle croît et ensuite décroît en passant par a (on parle de point-selle dans ce dernier cas).

Aucune des deux conditions " $\nabla f(a) = 0$ et $\nabla^2 f(a) \succeq 0$ " et " $\nabla f(a) = 0$ et $\nabla^2 f(a) \succ 0$ " est une CNS pour avoir un minimum local. Pour la première condition, on a déjà vu que $f : x \in \mathbb{R} \rightarrow x^3$ donne un contre-exemple et pour la deuxième condition, on peut voir que $f : x \in \mathbb{R} \rightarrow x^4$ a bien un minimum local en 0 mais $f''(0) = 0$.

Remarque 3.2 En dimension $n = 2$, on peut caractériser le fait que $\nabla^2 f(a) \succeq 0$ ou $\nabla^2 f(a) \succ 0$ à partir du calcul de la trace et du déterminant de $\nabla^2 f(a)$. En effet, si on note par σ_1 et σ_2 les 2 valeurs singulières de $\nabla^2 f(a)$, on voit que $\det(\nabla^2 f(a)) = \sigma_1 \sigma_2$ et $\text{Tr}(\nabla^2 f(a)) = \sigma_1 + \sigma_2$. Comme $\nabla^2 f(a) \succeq 0$ (resp. $\nabla^2 f(a) \succ 0$) si et seulement si $\sigma_1, \sigma_2 \geq 0$ (resp. $\sigma_1, \sigma_2 > 0$) et que $\sigma_1, \sigma_2 \geq 0$ (resp. $\sigma_1, \sigma_2 > 0$) si et seulement si $\det(\nabla^2 f(a)) \geq 0$ et $\text{Tr}(\nabla^2 f(a)) \geq 0$ (resp. $\det(\nabla^2 f(a)) > 0$ et $\text{Tr}(\nabla^2 f(a)) > 0$), on obtient bien la caractérisation :

i) $\nabla^2 f(a) \succeq 0$ si et seulement si $\det(\nabla^2 f(a)) \geq 0$ et $\text{Tr}(\nabla^2 f(a)) \geq 0$

ii) $\nabla^2 f(a) \succ 0$ si et seulement si $\det(\nabla^2 f(a)) > 0$ et $\text{Tr}(\nabla^2 f(a)) > 0$.

Ces dernières conditions sur la trace et le déterminant de $\nabla^2 f(a)$ peuvent s'écrire en fonction des dérivées partielles d'ordre 2 de f :

$$\det(\nabla^2 f(a)) = rs - t^2 \text{ et } \text{Tr}(\nabla^2 f(a)) = r + s$$

où $r = \partial_x^2 f(a)$, $s = \partial_y^2 f(a)$ et $t = \partial_{xy}^2 f(a)$

Exemple : calcul de l'estimateur du maximum de vraisemblance pour des données Gaussiennes. Pour tout $\sigma > 0$ et $\theta \in \mathbb{R}$, la densité d'une variable Gaussienne sur \mathbb{R} est donnée par

$$\varphi_{\theta, \sigma} : x \in \mathbb{R} \rightarrow \frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{(x - \theta)^2}{2\sigma^2} \right).$$

On se donne n points X_1, \dots, X_n différents de \mathbb{R} . Le problème est de maximiser en $(\theta, \sigma) \in \mathbb{R}_+^* \times \mathbb{R}$, la fonction

$$(\theta, \sigma)^\top \rightarrow \prod_{i=1}^n \varphi_{\theta, \sigma}(X_i). \quad (3.3)$$

On montre que si $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ est la moyenne empirique et $\hat{\sigma}_n^2 = (1/n) \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est la variance empirique alors

$$\left\{ \begin{pmatrix} \bar{X}_n \\ \hat{\sigma}_n \end{pmatrix} \right\} = \underset{(\sigma, \theta) \in \mathbb{R}_+^* \times \mathbb{R}}{\text{argmax}} \prod_{i=1}^n \varphi_{\theta, \sigma}(X_i).$$

Preuve. On voit que maximiser (3.3) est équivalent à minimiser $(\theta, \sigma)^\top \rightarrow -\sum_{i=1}^n \log \varphi_{\theta, \sigma}(X_i)$ et pour tout $(\theta, \sigma) \in \mathbb{R}_+^* \times \mathbb{R}$, on a

$$-\sum_{i=1}^n \log \varphi_{\theta, \sigma}(X_i) = \frac{n}{2} \log(\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta)^2 + \frac{1}{2} \log(2\pi) = (1/2)(G(\theta, \sigma^2) + \log(2\pi))$$

où

$$G(\theta, v) = n \log(v) + \frac{1}{v} \sum_{i=1}^n (X_i - \theta)^2.$$

Il suffit alors de minimiser G sur $(\theta, v) \in U$ où $U = \mathbb{R} \times \mathbb{R}_+^*$.

Existence d'une solution : On voit que G est coercive sur son domaine car si $|\theta| \rightarrow +\infty$ ou $v \rightarrow 0^+$ ou $v \rightarrow +\infty$ alors $G(\theta, v) \rightarrow +\infty$. Donc les ensembles de la forme $\{(\theta, v) \in \mathbb{R} \times \mathbb{R}_+^* : G(\theta, v) \leq c\}$ sont soit vides soit compacts. Alors G admet un minimum sur $\mathbb{R}_+^* \times \mathbb{R}$.

Identification des points critiques : $(\theta, v) \in U$ est un point critique de G quand $\nabla G(\theta, v) = 0$. On a

$$\nabla G(\theta, v) = \begin{pmatrix} \partial_\theta G(\theta, v) \\ \partial_v G(\theta, v) \end{pmatrix} = \begin{pmatrix} -\frac{2}{v} \sum_{i=1}^n (X_i - \theta) \\ -\frac{n}{v} - \frac{1}{v^2} \sum_{i=1}^n (X_i - \theta)^2 \end{pmatrix}.$$

Ainsi, $\nabla G(\theta, v) = 0$ si et seulement si $\theta = \bar{X}_n$ et $v = \hat{\sigma}_n^2$. Donc G admet un unique point critique sur U . Comme on sait que $\min_{(\theta, v) \in U} G(\theta, v)$ admet une solution c'est forcément $(\bar{X}_n, \hat{\sigma}_n^2)$. On peut donc s'arrêter ici et conclure le résultat annoncé. Vérifions quand même la condition du second ordre dans ce cas.

Conditions du second ordre au point critique : On a

$$\nabla^2 G(\theta, v) = \begin{pmatrix} \frac{2n}{v} & \frac{2}{v^2} \sum_{i=1}^n (X_i - \theta) \\ \frac{2}{v^2} \sum_{i=1}^n (X_i - \theta) & \frac{-n}{v^2} + \frac{2}{v^3} \sum_{i=1}^n (X_i - \theta)^2 \end{pmatrix}$$

et donc

$$\nabla^2 G(\bar{X}_n, \hat{\sigma}_n^2) = \begin{pmatrix} \frac{2n}{\hat{\sigma}_n^2} & 0 \\ 0 & \frac{n}{\hat{\sigma}_n^4} \end{pmatrix} \succ 0.$$

Donc le point critique $(\bar{X}_n, \hat{\sigma}_n^2)$ est un minimum local et comme il est unique, c'est bien un minimum global et donc une solution au problème $\min_{(\theta, v) \in U} G(\theta, v)$.

Finalement, le lien entre G et le problème initial donne que $(\bar{X}_n, \hat{\sigma}_n)$ est l'unique solution au problème initial. ■

4 Théorème des extremas liés

Le but de cette section est de donner un premier exemple de théorème d'optimisation sous contrainte. Ici on souhaite donner une condition nécessaire au problème d'optimisation

$$\min_{x \in K} f(x) \tag{4.1}$$

où $f : U \rightarrow \mathbb{R}$, U est un ouvert de \mathbb{R}^n et K est une contrainte de la forme

$$K = \{x \in U : g_1(x) = \dots = g_r(x) = 0\} \tag{4.2}$$

pour $g_1, \dots, g_r : U \rightarrow \mathbb{R}$.

Les solutions de (4.1) sont de minima globaux de f sur K . On précise ici cette notion.

Définition 4.1 Soit $f : U \rightarrow \mathbb{R}$ où U est un ouvert de \mathbb{R}^n . Soit $K \subset U$ un fermé de \mathbb{R}^n . Soit $x^* \in K$. On dit que

1. x^* est un **minimum local de f sur K** quand il existe $\epsilon > 0$ tel que pour tout $x \in K \cap B_2(x^*, \epsilon)$, $f(x^*) \leq f(x)$ où on note $B_2(x^*, \epsilon) = \{x \in \mathbb{R}^n : \|x - x^*\|_2 \leq \epsilon\}$
2. on dit que x^* est un **minimum global de f sur K** quand pour tout $x \in K$, $f(x) \geq f(x^*)$.

On donne maintenant une condition suffisante pour qu'un point soit un minimum local de f sur K .

Théorème 4.2 (extrema liés / Lagrange) On suppose que f, g_1, \dots, g_r sont de classe \mathcal{C}^1 . Soit $a \in K$. On suppose que

$$(\nabla g_1(a), \dots, \nabla g_r(a)) \text{ est une famille libre de } \mathbb{R}^n. \quad (4.3)$$

Si a est un minimum local de f sur K alors il existe des nombres réels $\lambda_1, \dots, \lambda_r$ tels que

$$\nabla f(a) + \sum_{i=1}^r \lambda_i \nabla g_i(a) = 0. \quad (4.4)$$

Quelques commentaires sur le Théorème des extrema liés :

1) Le Théorème des extrema liés donne une condition nécessaire pour qu'un point soit un minimum local de f sur K . Comme une solution au problème (4.1) est aussi un minimum local de f sur K , toute solution du problème (4.1) vérifie aussi la condition (4.4). On peut donc restreindre l'espace de recherche des solutions au problème (4.1) vérifiant (4.3) aux points vérifiant cette condition : Si a est une solution du problème (4.1) alors il existe des nombres réels $\lambda_1, \dots, \lambda_r$ tels que

$$\nabla f(a) + \sum_{i=1}^r \lambda_i \nabla g_i(a) = 0.$$

Cette dernière condition est nécessaire pour toute solution au problème (4.1). Elle n'est pas suffisante. C'est l'équivalent de la condition du premier ordre "si a est solution d'un problème d'optimisation sans contrainte alors nécessairement c'est un point critique".

2) Les théorèmes en optimisation sous contraintes du premier ordre s'énoncent presque toujours sous la même forme :

- i) On commence par donner le cadre du problème en donnant les hypothèses sur f et les fonctions définissant la contrainte, ici, on a : "On suppose que f, g_1, \dots, g_r sont de classe \mathcal{C}^1 ". On est donc dans un problème d'optimisation différentiable (OD). On sait alors que les conditions qu'on obtiendra seront uniquement nécessaires mais pas suffisantes en général. Elles le deviendront quand on supposera de plus la convexité (plus tard).
- ii) On donne ensuite une hypothèse de qualification de la contrainte au point a . Ici, pour le théorème des extrema liés cette condition est donnée en (4.3) : " $(\nabla g_1(a), \dots, \nabla g_r(a))$ est une famille libre de \mathbb{R}^n ." On reviendra plus en détails (plus tard) sur la condition de qualification et différentes conditions qui l'impliquent.
- iii) Finalement, on fini avec la condition KKT qui est ici donnée en (4.4) : "il existe des nombres réels $\lambda_1, \dots, \lambda_r$ tels que $\nabla f(a) + \sum_{i=1}^r \lambda_i \nabla g_i(a) = 0$ ".

3) On verra la preuve du théorème des extrema liés plus tard dans le cadre plus général du théorème de KKT.

Exemple : Trouver une solution au problème

$$\min (2x + y : x^2 + y^2 = 1). \quad (4.5)$$

La fonction objectif est $f : (x, y) \in \mathbb{R}^2 \rightarrow 2x + y$ et la contrainte est $K = \{(x, y) \in \mathbb{R}^2 : g_1(x, y) = 0\}$ où $g_1(x, y) = x^2 + y^2 - 1$.

existence d'une solution : f est continue et K est compact donc d'après Weierstrass, le problème (4.5) admet une solution.

qualification : K est définie par une seule contrainte d'égalité. Il suffit alors de vérifier que $(\nabla g_1(x, y))$ est libre pour assurer que K est qualifiée en (x, y) . Or une famille d'un seul vecteur est libre si et seulement si ce vecteur est non nul. Or $\nabla g_1(x, y) = (2x, 2y) \neq 0$ pour tout $(x, y) \in K$. Donc K est bien qualifiée en tout point de K (autrement dit K est qualifiée).

Condition KKT : D'après le théorème des extrema lié, si $(x, y) \in K$ est solution au problème (4.5) alors nécessairement il existe $\lambda_1 \in \mathbb{R}$ tel que $\nabla f(x, y) + \lambda_1 \nabla g_1(x, y) = 0$. On a

$$\nabla f(x, y) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ et } \nabla g_1(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

L'équation KKT s'écrit donc ici par : il existe λ tel que

$$\begin{cases} 2 + \lambda(2x) = 0 \\ 1 + \lambda(2y) = 0 \end{cases}$$

en particulier, $\lambda \neq 0$ et comme, par ailleurs, $(x, y) \in K$ (càd $x^2 + y^2 = 1$), on obtient le système suivant d'équations

$$\begin{cases} 2y = x \\ x^2 + y^2 = 1 \end{cases}$$

qui a pour solutions $(2, 1)/\sqrt{5}$ et $-(2, 1)/\sqrt{5}$. On sait donc que les solutions de (4.5) sont parmi ces deux points grâce au théorème des extrema liés. On calcul ensuite les valeurs prises par la fonction objectif en ces deux points identifiés par la condition KKT :

$$f((2, 1)/\sqrt{5}) = \sqrt{5} \text{ et } f(-(2, 1)/\sqrt{5}) = -\sqrt{5}.$$

Il n'y a donc qu'une seule solution au problème (4.5) qui est $-(2, 1)/\sqrt{5}$. ■

5 Théorème et condition de KKT

Dans cette section, on donne le théorème de KKT qui est une généralisation du théorème des extrema liés au cas d'un ensemble de contrainte K défini par des contraintes d'égalité et des contraintes d'inégalité. C'est-à-dire on étudie les problèmes de la forme

$$\min_{x \in K} f(x) \tag{5.1}$$

où $f : U \rightarrow \mathbb{R}$, U est un ouvert de \mathbb{R}^n et K est une contrainte de la forme

$$K = \left\{ x \in U : \begin{array}{l} g_1(x) = \dots = g_r(x) = 0 \\ h_1(x) \leq 0, \dots, h_l(x) \leq 0 \end{array} \right\} \tag{5.2}$$

pour $g_1, \dots, g_r : U \rightarrow \mathbb{R}$ définissant les contraintes d'égalité et $h_1, \dots, h_l : U \rightarrow \mathbb{R}$ définissant les contraintes d'inégalité.

Théorème 5.1 (Théorème de Karush-Kuhn-Tucker (KKT)) *On suppose que les fonctions $f, g_1, \dots, g_r, h_1, \dots, h_l$ sont de classe \mathcal{C}^1 . Soit $a \in K$. On suppose que K est qualifiée en a . Si a est un minimum local de f restreint à K alors il existe $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ et $\mu_1, \dots, \mu_l \in \mathbb{R}$ tels que :*

a)

$$\nabla f(a) + \sum_{i=1}^r \lambda_i \nabla g_i(a) + \sum_{j=1}^l \mu_j \nabla h_j(a) = 0$$

b) $\mu_j \geq 0$ pour tout $j = 1, \dots, l$

c) $\mu_j h_j(a) = 0$ pour tout $j = 1, \dots, l$.

Quelques commentaires sur le théorème de KKT :

1) L'hypothèse “ K est qualifiée en a ” n’a pas encore été définie. On reviendra sur cette hypothèse plus tard. On donne juste ci-dessous deux conditions suffisantes assurant la qualification d’une contrainte en un de ses points.

Définition 5.2 On dit que la contrainte K définie en (5.2) satisfait la **condition de Mangasarian-Fromovitz** en un point $a \in K$ quand :

i) la famille $(\nabla g_1(a), \dots, \nabla g_r(a))$ est une famille libre de vecteurs de \mathbb{R}^n

ii) On note $J(a) = \{j \in \{1, \dots, l\} : h_j(a) = 0\}$. Si $J(a) \neq \emptyset$ alors il existe un $v \in \mathbb{R}^n$ tel que 1) $\langle \nabla g_i(a), v \rangle = 0$ pour tout $i = 1, \dots, r$ et 2) $\langle \nabla h_j(a), v \rangle < 0$ pour tout $j \in J(a)$.

Le résultat suivant dit que la condition de Mangasarian-Fromovitz en a implique la qualification en a . On ne montre pas ce résultat pour le moment, étant donné qu’on n’a pas encore défini formellement la qualification d’une contrainte (car elle fait intervenir des outils qu’on n’a pas encore vu et qui ne sont pas nécessaire pour l’application pratique de KKT).

Proposition 5.3 Soit K la contrainte définie en (5.2). Soit $a \in K$. On suppose que K satisfait la condition de Mangasarian-Fromovitz en a . Alors K est qualifiée en a .

Il suffit alors de vérifier la condition de Mangasarian-Fromovitz pour assurer la qualification de K en a et ainsi de pouvoir appliquer le théorème de KKT. On voit que la condition de Mangasarian-Fromovitz coïncide avec la condition de qualification du théorème des extrema liés dans le cas où K est seulement définie à partir de contraintes d’égalités. Il existe (plein) d’autres conditions impliquant la qualification d’une contrainte. On en donne une autre maintenant qui est assez utile, par exemple, en programmation linéaire.

Définition 5.4 On dit que la contrainte K définie en (5.2) satisfait la **condition QC-A** en un point $a \in K$ quand :

i) $(g_i)_{i=1}^r$ et $(h_j)_{j \in J(a)}$ sont affines dans un voisinage ouvert de a ,

ii) $(h_j)_{j \notin J(a)}$ sont continues en a

où on rappelle que $J(a) = \{j \in \{1, \dots, l\} : h_j(a) = 0\}$.

Tout comme la condition de Mangasarian-Fromovitz, la condition QC-A est suffisante pour assurer la qualification en un point.

Proposition 5.5 Soit K la contrainte définie en (5.2). Soit $a \in K$. On suppose que K satisfait la condition QC-A en a . Alors K est qualifiée en a .

Nous reviendrons plus en détails dans les sections suivantes sur les problèmes de qualification de contrainte.

2) Les trois conditions a), b) et c) du théorème 5.1 sont appelées les “conditions KKT”.

Définition 5.6 Soit K une contrainte définie comme dans (5.2) et $a \in K$. On dit que a vérifie les **conditions KKT** quand il existe $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ et $\mu_1, \dots, \mu_l \in \mathbb{R}$ tels que :

a)

$$\nabla f(a) + \sum_{i=1}^r \lambda_i \nabla g_i(a) + \sum_{j=1}^l \mu_j \nabla h_j(a) = 0$$

b) $\mu_j \geq 0$ pour tout $j = 1, \dots, l$

c) $\mu_j h_j(a) = 0$ pour tout $j = 1, \dots, l$.

La condition b) est appelée condition de **dual feasibility** et la condition c) est appelée **complementary condition**.

Tout comme le théorème des extrema liés, le théorème de KKT donne une condition nécessaire pour qu'un point a soit solution de (5.1). En effet, si $a \in K$ est tel que K est qualifié en a et a est solution de (5.1) alors nécessairement a vérifie les conditions KKT. La réciproque n'est pas vraie.

Méthodologie pour la résolution d'un problème d'optimisation différentiable grâce au Théorème de KKT : On peut systématiser l'application du théorème de KKT pour la résolution de problèmes sous contrainte de la forme (5.1). On donne ici les étapes :

- 0) On vérifie que les fonctions $f, g_1, \dots, g_r, h_1, \dots, h_l$ sont bien \mathcal{C}^1 .
- 1) Preuve de l'existence d'une solution (par exemple, par Weierstrass ou coercivité)
- 2) On cherche les points de K où la contrainte n'y est pas qualifiée. On note par $E_1 \subset K$ cet ensemble.
- 3) On cherche les points de K vérifiant la condition KKT. On note par $E_2 \subset K$ cet ensemble.
- 4) Par le théorème de KKT, si (5.1) a une solution alors nécessairement elle est dans $E_1 \cup E_2$.
- 5) On évalue f en tous les points de $E_1 \cup E_2$ pour identifier celui ou ceux qui minimisent f .

Cette méthodologie montre qu'on peut très bien appliquer KKT sans connaître sa démonstration et même sans vraiment comprendre ce théorème. On peut "bachoter" sur KKT pour savoir l'appliquer de manière systématique. C'est l'objectif de cette section que d'extraire une méthode d'application de KKT. La compréhension de KKT nécessite des outils de géométrie différentielle qu'on introduira plus tard (ces outils nous permettront de donner une approximation locale de K).

On propose maintenant de donner un exemple d'application de KKT suivant la méthodologie ci-dessus. On encourage le lecteur à "s'entraîner" à appliquer le théorème de KKT sur d'autres exemples.

Exemple : Trouver les solutions au problème d'optimisation sous contrainte

$$\min (x + 2y + 3z : x^2 + y^2 + z^2 = 1; x + y + z \leq 0). \quad (5.3)$$

On pose des notations afin d'identifier la fonction objectif et la contrainte : pour tout $(x, y, z) \in \mathbb{R}^3$,

$$f(x, y, z) = x + 2y + 3z; g_1(x, y, z) = x^2 + y^2 + z^2 - 1 \text{ et } h_1(x, y, z) = x + y + z.$$

La fonction objectif est donnée par f et la contrainte est $K = \{a \in \mathbb{R}^3 : g_1(a) = 0, h_1(a) \leq 0\}$.

0) Les fonctions f, g_1 et h_1 sont de classe \mathcal{C}^1 car ce sont des polynômes de \mathbb{R}^3 .

1) K est un compact (en tant qu'intersection de deux fermés dont l'un est borné) et f est continue donc (5.3) admet au moins une solution.

2) Soit $a = (x, y, z) \in K$. Étudions la qualification de K en a . On peut par exemple vérifier la condition de Mangasarian-Fromovitz. On a $\nabla g_1(a) = (2x, 2y, 2z) \neq 0$ car $a \in K$ et donc $(\nabla g_1(a))$ est libre – donc la première condition de Mangasarian-Fromovitz est satisfaite. Passons à la deuxième condition : on rappelle que $J(a) = \{j : h_j(a) = 0\}$. Si $J(a) = \emptyset$ alors la deuxième condition de Mangasarian-Fromovitz est vérifiée. Sinon $J(a) \neq \emptyset$ et comme il n'y a qu'une seule contrainte d'inégalité, on a forcément $J(a) = \{1\}$ et donc $h_1(a) = 0$, càd $x + y + z = 0$. Pour $v = (-1, -1, -1)$, on a

$$\langle \nabla g_1(a), v \rangle = -2(x + y + z) = 0 \text{ et } \langle \nabla h_1(a), v \rangle = -3 < 0.$$

Donc la deuxième condition de Mangasarian-Fromovitz est vérifiée dans tous les cas. On en déduit que K est qualifiée en a et comme ceci est vraie pour tout $a \in K$, on a montré que K est qualifiée. Autrement dit, l'ensemble E_1 introduit dans la méthodologie ci-dessus est vide.

3) On cherche les points $a = (x, y, z) \in K$ vérifiant les conditions KKT. Soit $a = (x, y, z) \in K$. Si a vérifie les conditions KKT alors il existe $\lambda_1, \mu_1 \in \mathbb{R}$ tels que

a)

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

b) $\mu_1 \geq 0$

c) $\mu_1 h_1(a) = 0$.

En ajoutant la contrainte que $a \in K$ aux conditions KKT, on veut résoudre le système d'équations suivant :

$$\begin{cases} 1 + 2\lambda_1 x + \mu_1 = 0 \\ 2 + 2\lambda_1 y + \mu_1 = 0 \\ 3 + 2\lambda_1 z + \mu_1 = 0 \\ \mu_1 \geq 0 \\ \mu_1(x + y + z) = 0 \\ x^2 + y^2 + z^2 = 1 \\ x + y + z \leq 0 \end{cases}$$

On a forcément $\lambda_1 \neq 0$ car $\mu_1 \geq 0$ et sinon les 3 premières équations ne pourraient pas être vérifiées. On a alors en réécrivant ces 3 équations :

$$x = \frac{-\mu_1 - 1}{2\lambda_1}, y = \frac{-\mu_1 - 2}{2\lambda_1} \text{ et } z = \frac{-\mu_1 - 3}{2\lambda_1}. \quad (5.4)$$

D'après la complementary condition, on a $\mu_1(x + y + z) = 0$. Mais d'après (5.4), on a $x + y + z = (-3\mu_1 - 6)/(2\lambda_1)$ donc $\mu(-3\mu_1 - 6) = 0$ alors soit $\mu_1 = 0$ soit $\mu_1 = -2$ mais comme $\mu_1 \geq 0$, on a nécessairement $\mu_1 = 0$. Ainsi, dans (5.4), on obtient

$$x = \frac{-1}{2\lambda_1}, y = \frac{-2}{2\lambda_1} \text{ et } z = \frac{-3}{2\lambda_1}.$$

et comme $x^2 + y^2 + z^2 = 1$, on en déduit que $\lambda_1 \in \{\pm\sqrt{14}/2\}$. Mais pour $\lambda_1 = -\sqrt{14}/2$, on a $x + y + z > 0$ donc forcément $\lambda_1 = \sqrt{14}/2$. Alors le seul point vérifiant la condition de KKT est

$$\left(\frac{-1}{\sqrt{14}}, \frac{-2}{\sqrt{14}}, \frac{-3}{\sqrt{14}} \right)^\top.$$

5) Comme on sait que le problème admet au moins une solution et qu'il n'y a qu'un seul point vérifiant les conditions KKT et que la contrainte est qualifiée, c'est forcément cet unique point qui est solution au problème (5.1). Cette solution est unique vu qu'il n'y a qu'un seul point satisfaisant KKT et pas de problème de qualification.