Optimisation différentiable : Théorie et Algorithmes

Séance 3 : Conditions d'optimalité (suite) dans le cas de contraintes d'égalité/inégalité

Abdelilah HAKIM

30/04/2024

Ecole Centrale Casablanca

Outline

Conditions d'optimalité : Cas sans contraintes

Conditions d'optimalité : Cas avec contraintes

Cas général

Conditions d'optimalité : Cas sans contraintes

Conditions d'optimalité (sans contraintes)

Théorème 1.

Condition nécessaire d'optimalité : Si u est minimum local de J dans V, alors

- Si J est différentiable en u alors $J'(u) = 0$
- Si J est deux fois différentiable en u, on a de plus

$$\forall w \in V, J''(u)(w,w) \geqslant 0$$

.

Preuve:

Soit $v \in V$, on définit la fonction d'une variable réelle $\varphi(t) = J(u+tv)$:

- Pour t réel assez petit, $u+tv$ est dans un voisinage de u et puisque u est minimum local, $J(u+tv) \geqslant J(u)$ ce qui se traduit par $\varphi(t) \geqslant \varphi(0)$.
  Par définition φ est dérivable en 0 et $\varphi'(0) = J'(u) \cdot v$. Écrivons :

$$0 \leq \lim_{t \to 0_+} \frac{\varphi(t) - \varphi(0)}{t} = \varphi'(0) = \lim_{t \to 0_-} \frac{\varphi(t) - \varphi(0)}{t} \leq 0$$

  ce qui prouve que $\varphi'(0) = 0$.
  On a donc pour tout $v, J'(u) \cdot v = 0$, et donc $J'(u) = 0$.
- Dans le deuxième cas
  - on applique la formule de Taylor-Young à l'ordre 2 à la fonction φ :

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\left(\varphi''(0) + \varepsilon(t)\right), \text{ avec } \lim_{t \to 0} \varepsilon(t) = 0$$

  - Si $\varphi''(0)$ est strictement négatif alors $\frac{1}{2}\varphi''(0) + \varepsilon(t)$ l'est aussi.
  - donc $\varphi(t) < \varphi(0)$ ce qui contredit le fait que u est minimum local.
  - Donc $\varphi''(0) \geqslant 0$.
  - Il est facile de calculer que c'est égal à $J''(u)(v,v)$.

Conditions d'optimalité (Equation d'Euler) : Cas sans contraintes)

Théorème 2.

condition suffisante : Soit J une fonction différentiable dans V et u un point de V tel que $J'(u) = 0$.

- Si J est deux fois differentiable dans un voisinage de u et s'il existe un voisinage Ω de u tel que $\forall v \in \Omega, \forall w \in V, J''(v)(w,w) \geqslant 0$, alors u est minimum local de J
- Si J est deux fois differentiable, et s'il existe $\alpha > 0$ tel que

$$\forall w \in V, J''(u)(w,w) \geqslant \alpha \|w\|^2,$$

alors u est minimum local strict pour J.

Preuve:

- Puisque Ω est un voisinage, il contient une boule ouverte de centre u, donc convexe.
- Soit v dans cette boule. Alors tout le segment] u, v[est contenu dans cette boule,
- d'après la formule de Taylor Mac-Laurin à l'ordre 2,

$$J(v) = J(u) + \underbrace{J'(u) \cdot (v-u)}_{0} + \frac{1}{2} \underbrace{J''(u + \theta(u-u))(v-u, v-u)}_{\geqslant 0}$$

- $J(v) \geqslant J(u)$ d'où u est un minimum local.
- Utilisons la formule de Taylor-Young et l'hypothèse :

$$J(v) \geqslant J(u) + \alpha \|(v-u)\|^2 + \varepsilon(v-u)\|v-u\|^2, \quad \lim_{w \to 0} \varepsilon(w) = 0$$

- Pour v suffisamment proche de u

$$\alpha + \varepsilon(v-u) > 0$$

- Donc $J(v) > J(u)$

Conditions d’optimalité : Cas avec contraintes

Condition d'optimalité (Inéquation d'Euler) : cas d'un ensemble de contraintes convexe

- Minimisation sur un ensemble de contrainte K convexe

$$u \in K \ \text{ tel que } \quad J(u) = \inf_{v \in K} J(v) \tag{1}$$

.

- K est fermé non vide
- J est différentiable sur un ouvert contenant K.
- Conditions d'optimalité : on peut tester l'optimalité de u dans la "direction admissible " $(v-u)$ car $u + t(v-u) \in K$ si $t \in [0,1]$.

Théorème 3.

Inéquation d'Euler : cas d'un ensemble de contraintes convexe

- Condition nécessaire : Si u est solution optimale, on a l'inéquation d'Euler suivante :

$$\begin{cases} u \in K \\ \forall v \in K, J'(u) \cdot (v-u) \geqslant 0. \end{cases} \tag{2}$$

- Condition suffisante :Réciproquement si on a l'inéquation d'Euler en u et si de plus J est convexe, alors u est solution optimale.

Condition d'optimalité (Inéquation d'Euler) : cas d'un ensemble de contraintes convexe

Preuve:

1. Condition nécessaire :
   - Soit $v \in K$. Puisque K est convexe, le segment $[u, v]$ est contenu dans K
   - $\varphi(t) = J(u + t(v - u))$ pour $t \in [0, 1]$
   - φ est dérivable puisque J est différentiable
   - On a : $\varphi(t) = \varphi(0) + t\varphi'(0) + t\varepsilon(t), \quad \lim_{t \to 0} \varepsilon(t) = 0$
   - En terme de J :

   $$J\,(u + t(v - u) = J(u) + t\,(J'(u)(v - u) + \varepsilon(t)) \geqslant J(u),$$

   car u est solution optimale, d'où $J'(u)(v - u) + \varepsilon(t) \geqslant 0$.
   - tendre t vers 0, d'où $J'(u)(v - u) \geqslant 0$
2. Condition suffisante :
   - On suppose que J est en plus convexe
   - J satisfait l'inégalité d'Euler $J'(u)(v - u) \geqslant 0$
   - caractérisation de la convexité. Pour tout v dans K,

   $$J(v) \geqslant J(u) + J'(u) \cdot (v - u) \geqslant J(u).$$

Remarque 1.

- Cette condition exprime que la dérivée directionnelle de J au point u dans toutes les directions $(v - u)$, qui sont rentrantes dans K, est positive.
- Il faut aussi remarquer que si u est intérieur à K , cette condition se réduit simplement à l'équation d'Euler $J'(u) = 0$

Conditions d'optimalité : Cas particuliers de K

- $K = V$: Sans contrainte

Théorème 4.

- Si J est convexe différentiable, alors u réalise le minimum de J sur V si et seulement si $J'(u) = 0$
- si J est α-convexe, il existe une unique solution optimale, caractérisée par $J'(u) = 0$.

- **K sous-espace affine engendré par le sous-espace vectoriel fermé E de V, i.e. $K = \{u_0 + v, v \in E\}$**
  - La condition d'optimalité :

$$(2) \Longleftrightarrow \begin{cases} u \in K \\ \forall w \in E, J'(u) \cdot w = 0 \end{cases} \tag{3}$$

  - Ce qui se traduit par : $J'(u) \in E^{\perp}$.

Conditions d'optimalité : Cas particuliers de K

K **cône convexe fermé**

Définition 1.

K_0 est un cone si $w \in K_0 \Longrightarrow tw \in K_0$ pour tout $t \geqslant 0$

Proposition 1.

- Soit $K = u_0 + K_0$
- La condition d'optimalité devient :

$$(2) \Longleftrightarrow \begin{cases} u \in K \\ J'(u) \cdot (u_0 - u) = 0 \\ \forall w \in K_0, J'(u) \cdot w \geqslant 0. \end{cases} \tag{4}$$

Preuve:

- Soit $v \in K, u_0 + t\,(v - u_0) \in K$
- En particulier pour $v = u$ l'inéquation d'Euler s'écrit :

$$\forall t \geqslant 0, \quad J'(u) \cdot (u_0 + t\,(u - u_0) - u) \geqslant 0,$$

$$\forall t \geqslant 0, \quad (1 - t)J'(u) \cdot (u_0 - u) \geqslant 0.$$

- $1 - t$ peut prendre des valeurs positives ou négatives, ceci implique que $J'(u) \cdot (u_0 - u) = 0$
- Soit maintenant $v = u_0 + w, w \in K_0$
- On a $J'(u) \cdot (u_0 + w - u) \geqslant 0$
- Or $J'(u) \cdot (u_0 - u) = 0$
- $\forall w \in K_0, J'(u).w \geqslant 0$

Le cône dual/normal

Le lagrangien

cône dual :

Définition 2.

Soit M un sous ensemble d'un Hilbert V, Le cône dual ou orthogonal ou normal à l'ensemble M au point u est défini par :

$$M_u^\star = \{c \in V, \forall v \in M, (c, v-u) \geqslant 0\}. \tag{5}$$

Il est facile de voir que ceci définit bien un cône, et même un cône convexe fermé comme intersection d'une collection de demi-espaces fermés. On notera M^* pour M_0^*.

Note 1.

- La condition d'optimalité 2 s'écrit sous la forme :

$$\begin{cases} u \in K \\ J'(u) \in K_u^*. \end{cases} \tag{6}$$

- Si M est engendré par un nombre fini de vecteurs, alors on peut décrire $M^\star$ par le Lemme de Farkas :

Lemme 1.

Si $M = \{v \in V, \forall i \in \{1, \dots, m\}, (v, a_i) \leqslant 0\}$, alors $c \in M^\star$ si et seulement si $-c$ appartient au cone convexe engendré par les a_i, i.e. il existe $\{\lambda_1, \dots, \lambda_m\}$ tous $\geqslant 0$ tels que $c = -\sum_{i=1}^{m} \lambda_i a_i$

Preuve:

- Si $c = -\sum_{i=1}^{m} \lambda_i a_i$ avec tous les λ_i positifs ou nuls, alors pour tout v dans M,$(v, c) = -\sum_{i=1}^{m} \lambda_i (a_i, v) \geqslant 0$ donc $c \in M^*$
- La réciproque repose sur le théorème de Hahn-Banach qui est hors programme.

Si $K = u_0 + K_0$ avec K_0 cône convexe fermé engendré par un nombre fini de vecteurs :

$$K_0 = \{w \in V, (a_i, w) \leq 0, 1 \leqslant i \leqslant m\}$$

- La troisième ligne dans (4) exprime que $-J'(u)$ est dans $K_0^\star$
- L'équation (4) se réécrit

$$(2) \Longleftrightarrow \begin{cases} u \in K \\ J'(u) \cdot (u_0 - u) = 0 \\ \exists (\lambda_1, \cdots, \lambda_m) \geqslant 0, \nabla J(u) + \sum_{i=1}^{m} \lambda_i a_i = 0 \end{cases} \tag{7}$$

Lagrangien et point de selle :

Définition 3.

Soient $A \subset \mathbb{R}^n$ (ou d'un Hilbert v) et $B \subset \mathbb{R}^m$ deux ensembles non vide et

$$L : A \times B \subset (V \times \mathbb{R}^m) \to \mathbb{R}$$

une fonction appellée Lagrangien.

Un point $(\bar{x}, \bar{y}) \in A \times B$ est un point de selle du lagrangien L si

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall x \in A, \quad \forall y \in B$$

Exemple 1.

Le point $(0,0)$ est un point de selle de la fonction $L(x,y) = x^2 - y^2$ avec $A = B = \mathbb{R}$ car

$$L(0,y) = -y^2 \leq 0 = L(0,0) \leq x^2 = L(x,0) \quad \forall x, y \in \mathbb{R}.$$

Rapport/TD2/selle-eps-converted-to.pdf

Théorème 5.

Si $(\bar{x}, \bar{y})$ est un point de selle de L, alors

$$\max_{y\in B}\min_{x\in A} L(x,y) = \min_{x\in A}\max_{y\in B} L(x,y) = L(\bar{x},\bar{y})$$

Autrement dit, on peut inverser l'ordre du min et du max.

Preuve:

Montrons en premier que

$$\max_{y\in B}\min_{x\in A} L(x,y) \leq \min_{x\in A}\max_{y\in B} L(x,y).$$

En effet, on a que

$$\min_{x\in A} L(x,y) \leq L(x,y) \leq \max_{y\in B} L(x,y)$$

Le membre de gauche est une fonction de y seulement. De même, pour le membre de droite qui est une fonction de x uniquement. Ceci implique

$$\max_{y\in B}\min_{x\in A} L(x,y) \leq \max_{y\in B} L(x,y), \quad \forall x \in A$$

et, en prenant le minimum par rapport à x,

$$\max_{y\in B}\min_{x\in A} L(x,y) \leq \min_{x\in A}\max_{y\in B} L(x,y),$$

d'où le résultat.

En deuxième, montrons que

$$\min_{x\in A}\max_{y\in B} L(x,y) \leq L(\bar{x},\bar{y}).$$

En effet, selon la définition du point de selle, on a

$$\max_{y\in B} L(x,y) = L(\bar{x},\bar{y}) \Longrightarrow L(\bar{x},\bar{y}) \geq \min_{x\in A}\max_{y\in B} L(x,y).$$

De même, on a

$$\min_{x\in A} L(x,y) = L(\bar{x},\bar{y}) \Longrightarrow L(\bar{x},\bar{y}) \leq \max_{y\in B}\min_{x\in A} L(x,y).$$

En recollant les morceaux, on obtient

$$L(\bar{x},\bar{y}) \leq \max_{y\in B}\min_{x\in A} L(x,y) \leq \min_{x\in A}\max_{y\in B} L(x,y) \leq L(\bar{x},\bar{y}),$$

d'où le résultat final.

Proposition 2.

Soient $U, V \subset \mathbb{R}^n$ deux sous-ensembles convexes et fermés. On fera les hypotheses suivantes :

- Pour y fixé, la fonction $L : x \to L(x, y)$ est convexe.
- Pour x fixé, la fonction $L : y \to L(x, y)$ est concave ($-L$ est convexe).

Au point selle $(\bar{x}, \bar{y})$ du lagrangien L, la condition d'optimalité du minimum $L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall x \in U$ s'écrit

$$\left(\frac{\partial L}{\partial x}(\bar{x}, \bar{y}), x - \bar{x}\right) \geq 0 \quad \forall x \in U$$

De même, la condition d'optimalité du maximum

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \quad \forall y \in V \text{ s'écrit}$$
$$\left(\tfrac{\partial L}{\partial y}(\bar{x}, \bar{y}), y - \bar{y}\right) \leq 0 \quad \forall y \in V$$

où $\frac{\partial L}{\partial x}$ est le gradient de la fonction $L(x, y)$ pour y fixé.

Définition 4.

Soit $K \subset \mathbb{R}^n$ un sous-ensemble quelconque et non vide. La fonction indicatrice de l'ensemble K est définie par

$$I_K(x) = \begin{cases} 0 & \text{si } x \in K \\ \infty & \text{sinon.} \end{cases}$$

Remarque 2.

La fonction indicatrice sert à enlever les contraintes. Pour un problème de minimisation, on a évidemment la relation

$$min_{x \in K} f(x) = min_x f(x) + I_K(x)$$

Pour un problème de maximisation, on a plutôt

$$max_{x \in K} f(x) = max_x f(x) - I_K(x)$$

Cas de contraintes de la forme : $Bx = c$

Appliquons la dualité lagrangienne au problème (primal)

$$\min_{Bx=c} f(x)$$

où B est une matrice de format $m \times n$ et $c \in \mathbb{R}^m$. Ce problème a un sens seulement si $c \in \operatorname{Im} B$ ou encore que le rang de B est maximal ($\lim B = \mathbb{R}^m$).

Lemme 2.

La fonction indicatrice de $K = \{x \in \mathbb{R}^n \mid Bx = c\}$ est donnée par

$$I_K(x) = max_{\lambda \in K^*}(\lambda, Bx - c)$$

Preuve:

en effet, si $Bx \neq c$, on peut prendre

$$\lambda = r(Bx - c) \Longrightarrow (\lambda, Bx - c) = r\|Bx - c\|^2 \to \infty \text{ si } r \to \infty$$

Le maximum sera ∞. Au contraire, si $Bx = c$, on a que $(\lambda, Bx - c) = 0 \quad \forall \lambda \in \mathbb{R}^m$ de maximum 0.

Par conséquent, on peut écrire

$$\min_{Bx=c} f(x) = \min_x f(x) + I_K(x) = \min_x f(x) + \max_{\lambda \in \mathbb{R}^m} (\lambda, Bx - c)$$

Ceci suggère la fonction lagrangienne

$$L(x, \lambda) = f(x) + (\lambda, Bx - c) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m$$

ce qui transforme le problème primal en un problème de point-selle

$$\min_{Bx=c} f(x) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} L(x, \lambda)$$

La condition d'optimalité du Lagrangien sera

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \Longleftrightarrow \nabla f(x) + B^T \lambda = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \Longleftrightarrow Bx = c \end{cases}$$

Cas de contraintes de la forme : $Bx = c$

Note 2.

La condition d'optimalité du problème $f(a) = \min_{Bx=c} f(x)$ s'écrit

$$(\nabla f(a), x - a) \geq 0 \quad \forall x \in U = \{x \mid Bx = c\}$$

Sachant que $c \in ImB$, on a que $U = x_0 + \text{Ker}\, B$ où $Bx_0 = c$. Ce qui implique que U est un sous-espace affine et dans ce cas, la condition d'optimalité s'écrit :

$$\nabla f(a) \in (\text{Ker}\, B)^{\perp} = \text{Im}\, B^T \implies \quad \exists \lambda \in \mathbb{R}^m \quad \nabla f(a) = -B^T \lambda$$

Ceci fournit les conditions en disant que a et λ sont solutions du système :

$$\begin{cases} \nabla f(x) + B^T \lambda = 0 \\ Bx = c \end{cases}$$

Ce sont les mêmes conditions que celles obtenues par la méthode de Lagrange.

Cas de Contraintes de la forme : $Bx \leq c$

Appliquons la dualité lagrangienne au problème (primal)

$$\min_{Bx \leq c} f(x)$$

où B est une matrice de taille $m \times n$ et $c \in \mathbb{R}^m$.

Lemme 3.

La fonction indicatrice de $K = \{x \in \mathbb{R}^n \mid Bx \leq c\}$ est donnée par

$$I_K(x) = \max_{\lambda \geq 0} (\lambda, Bx - c)$$

Preuve:

En effet, si $Bx > c$, on peut prendre :

$$\lambda = r(Bx - c) \geq 0 \Longrightarrow (\lambda, Bx - c) = r\|Bx - c\|^2 \to \infty \text{ si } r \to \infty$$

Le maximum sera ∞. Au contraire, si $Bx \leq c$, on a que $(\lambda, Bx - c) \leq 0 \quad \forall \lambda \geq 0$ car $Bx - c \leq 0$. Donc le maximum sera 0 .

Par conséquent, on peut écrire

$$\min_{Bx \leq c} f(x) = \min_x f(x) + I_K(x) = \min_x f(x) + \max_{\lambda \geq 0} (\lambda, Bx - c)$$

Ceci suggère la fonction lagrangienne

$$L(x, \lambda) = f(x) + (\lambda, Bx - c) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \geq 0 \text{ et } \lambda \in \mathbb{R}^m.$$

ce qui transforme le problème primal en un problème de point-selle
La condition d'optimalité du lagrangien sera

- $\left(\frac{\partial L}{\partial x}(\bar{x}, \lambda), x - \bar{x}\right) \geq 0 \quad \forall x \in \mathbb{R}^n$ ce qui donne
$$\frac{\partial L}{\partial x}(\bar{x}, \lambda) = 0 \Longleftrightarrow \nabla f(\bar{x}) + B^t \lambda = 0$$
- $\left(\frac{\partial L}{\partial \lambda}(x, \bar{\lambda}), \lambda - \bar{\lambda}\right) \leq 0 \quad \forall \lambda \geq 0$. Étant donné que $\{\lambda \geq 0\}$ forme un cône convexe, on obtient les conditions suivantes

$$\left(\frac{\partial L}{\partial \lambda}(x, \bar{\lambda}), \lambda\right) = (B\bar{x} - c, \lambda) \leq 0 \quad \forall \lambda \geq 0 \quad \Longrightarrow \quad B\bar{x} - c \leq 0 \Longleftrightarrow B\bar{x} \leq c$$
$$\left(\frac{\partial L}{\partial \lambda}(x, \bar{\lambda}), \bar{\lambda}\right) = 0 \Longleftrightarrow (B\bar{x} - c, \bar{\lambda}) = 0$$

Conditions KKT pour les contraintes : $Bx \leq c$

A partir des conditions d'optimalité du lagrangien $L(x, \lambda)$, on peut écrire les conditions dites de *Karush – Kuhn – Tucker* (KKT) pour le problème

$$\min_{Bx \leq c} f(x)$$

Observons que $(B\bar{x} - c, \bar{\lambda}) = 0$ avec $\bar{\lambda} \geq 0$ signifie que $\bar{\lambda}_i(B\bar{x} - c)_i = 0$ pour tous les $i = 1, \ldots, m$. Les conditions KKT s'écrivent sous la forme :

$$\begin{cases} \nabla f(\bar{x}) + B^T\bar{\lambda} & = 0, \\ \bar{\lambda}_i & \geq 0, \quad \forall i = 1, \ldots, m, \\ (B\bar{x})_i & \leq c_i, \qquad\qquad\qquad\qquad \forall i = 1, \ldots, m, \\ \bar{\lambda}_i(B\bar{x} - c)_i & = 0 \quad \forall i = 1, \ldots, m. \end{cases}$$

Note 3.

Le problème $f(a) = \min_{Bx \leq c} f(x)$ est équivalent à

$$\min_{-Bx \geq -c} f(x)$$

De nouveau, on fait l'hypothèse que $c \in \operatorname{Im} B$ donc il existe x_0 tel que $Bx_0 = c$. Ceci implique

$$\begin{aligned} U &= \{x \mid Bx \leq c\} = \{x \mid -Bx \geq -c\} \\ &= \{x \mid -B(x - x_0) \geq 0\} = x_0 + \{-Bx \geq 0\} = x_0 + W \end{aligned}$$

où $W = \{-Bx \geq 0\}$ est le cône positif. La condition d'optimalité du problème s'écrit

$$(\nabla f(a), x - a) \geq 0 \quad \forall x \in U \quad \Longrightarrow (\nabla f(a), y) \geq 0 \quad \forall y \in W,$$

autrement dit $\nabla f(a) \in W^*$. Mais $W^* = -B^t \Lambda^+$avec $\Lambda^+ = \{\lambda \geq 0\}$. Ceci montre l'existence du multiplicateur λ vérifiant $\nabla f(a) + B^T \bar{\lambda} = 0$.
Mais il y a une autre condition que l'on peut tirer de la condition $(\nabla f(a), x - a) \geq 0$. Il s'agit de la condition :

$$(\nabla f(a), a - x_0) = 0 \Longrightarrow \left(-B^T \lambda, a - x_0\right) = 0 \Longrightarrow (\lambda, Ba - Bx_0) = (\lambda, Ba - c) = 0.$$

En effet, il suffit de prendre $x = x_0$ et $x = x_0 + 2(a - x_0)$ dans le condition ci-dessus. En combinant toutes les conditions ci-dessus, on obtient exactement le même système KKT que doit vérifier la solution a et le multiplicateur λ

$$\begin{cases} \nabla f(\bar{x}) + B^t \bar{\lambda} = 0, & \\ \bar{\lambda}_i \geq 0, & \forall i = 1, \ldots, m, \\ (B\bar{x})_i \leq c_i, & \forall i = 1, \ldots, m, \\ \bar{\lambda}_i (B\bar{x} - c)_i = 0, & \forall i = 1, \ldots, m. \end{cases}$$

Cas général

Cas général : contraintes $g(x) = 0$

Définition 5.

es contraintes sont régulières en $u \in K$ si les $g_i'(\bar{x})$ sont linéairement indépendantes.
On dit alors que $\bar{x}$ est un point régulier.
Dans la suite, on suppose que $\bar{x}$ est régulier.

Appliquons la méthode de Lagrange au problème primal suivant :

$$\min_{g(x)=0} f(x)$$

où $g = (g_1, g_2, \dots, g_m)$ désigne m contraintes scalaires. En général, l'ensemble des contraintes n'est pas convexe.
On prend la fonction lagrangienne :

$$L(x, \lambda) = f(x) + (\lambda, g(x)) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \in \mathbb{R}^m.$$

Ceci transforme le problème primal en un problème de point selle

$$min_{g(x)=0} f(x) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} L(x, \lambda)$$

La condition d'optimalité du lagrangien sera

$$\begin{cases} \frac{\partial L}{\partial x}(\bar{x}, \lambda) = 0 \Longleftrightarrow \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0, \\ \frac{\partial L}{\partial \lambda}(x, \bar{\lambda}) = 0 \Longleftrightarrow g_i(x) = 0, \quad \forall i = 1, \dots, m. \end{cases}$$

Théorème 6.

i $\bar{x} \in K$, $\bar{x}$ régulier, est minimum local pour J, il existe m réels $\lambda_1, \dots, \lambda_m$ tels que

$$J'(\bar{x}) + \sum_{i=1}^{m} \lambda_i g_i'(\bar{x}) = 0. \tag{8}$$

Si K (ie les g_i) et J sont convexes, alors c'est une condition nécessaire et suffisante.

Cas général : Contraintes $g(x) \leq 0$

Lemme 4.

La fonction indicatrice de $K = \{x \in \mathbb{R} =| \ g(x\} \leq 0\}$ est donnée par :

$$h(x) = \max_{\lambda \in \mathbb{R}_+^m} (\lambda, g(x))$$

Preuve:

En effet, si $g_i(x) > 0$, on peut prendre $\lambda = 0$ sauf pour la composante i

$$\lambda_i = r\,(g_i(x)) > 0 \Longrightarrow (\lambda, g(x)) = r g_i(x)^2 \to \infty \text{ si } r \to \infty.$$

Le maximum sera ∞.
Au contraire, si $g(x) \leq 0$, on a que $(\lambda, g(x)) \leq 0 \quad \forall \lambda \geq 0$ car $g(x) \leq 0$. Donc le maximum sera 0 .

Par conséquent, on peut écrire

$$\min_{g(x) \leq 0} f(x) = \min_x f(x) + I_K(x) = \min_x f(x) + \max_{\lambda \geq 0} (\lambda, g(x))$$

Contraintes de la forme : $g(x) \leq 0$ (suite)

Appliquons la dualité lagrangiene au problème (primal)

Ceci suggère la fonction lagrangienne

$$L(x,\lambda) = f(x) + (\lambda, g(x)) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \quad \forall x \in \mathbb{R}^n, \quad \forall \lambda \geq 0 \text{ et } \lambda \in \mathbb{R}^m$$

ce qui transforme le problème primal en un problème de point-selle

$$\min_{g(x)\leq 0} f(x) = \min_{x\in\mathbb{R}^n} \max_{\lambda\geq 0} L(x,\lambda)$$

La condition d'optimalité du lagrangien sera :

- $\left(\frac{\partial L}{\partial x}(\bar{x},\lambda), x - \bar{x}\right) \geq 0 \quad \forall x \in \mathbb{R}^n$ ce qui donne

$$\frac{\partial L}{\partial x}(\bar{x},\lambda) = 0 \Longleftrightarrow \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0$$

- $\left(\frac{\partial L}{\partial \lambda}(x,\bar{\lambda}), \lambda - \bar{\lambda}\right) \leq 0 \quad \forall \lambda \geq 0$. Étant donné que $\{\lambda \geq 0\}$ forme un cône convexe, on obtient les conditions suivantes

$$\left(\frac{\partial L}{\partial \lambda}(x,\bar{\lambda}), \lambda\right) = (g(\bar{x}),\lambda) \leq 0 \quad \forall \lambda \geq 0 \quad \Longrightarrow \quad g(\bar{x}) \leq 0$$
$$\left(\frac{\partial L}{\partial \lambda}(x,\bar{\lambda}), \bar{\lambda}\right) = 0 \Longleftrightarrow (g(\bar{x}),\bar{\lambda}) = 0$$

Conditions KKT pour les contraintes : $g(x) \leq 0$

A partir des conditions d'optimalité du lagrangien $L(x, \lambda)$, on peut écrire les conditions dites de Karush-Kuhn-Tucker (KKT) pour le problème

$$\min_{g(x) \leq 0} f(x)$$

Observons que $(g(\bar{x}), \bar{\lambda}) = 0$ avec $\bar{\lambda} \geq 0$ signifie que $\bar{\lambda}_i g_i(\bar{x}) = 0$ pour tous les $i = 1, \ldots, m$. Les conditions KKT s'écrivent sous la forme :

$$(KKT) \begin{cases} \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = 0, & \\ \bar{\lambda}_i \geq 0, & \forall i = 1, \ldots, m, \\ g_i(\bar{x}) \leq 0, & \forall i = 1, \ldots, m, \\ \bar{\lambda}_i g_i(\bar{x}) = 0 & \forall i = 1, \ldots, m. \end{cases}$$

Définition 6.

Soit $\bar{x}$ tel que $g(\bar{x}) \leq 0$.

- On appelle ensemble des contraintes actives en $\bar{x}$ $I(\bar{x}) = \{i \in \{1, \ldots, m\}, g_i(\bar{x}) = 0\}$.
- Les contraintes sont dites qualifiées en $\bar{x}$ si

$$\exists \bar{w} \in V, \forall i \in I(\bar{x}), (g_i'(\bar{x}), \bar{w}) < 0 \,(\leq 0 \text{ si } g_i \text{ est affine }) \tag{9}$$

Remarque 3.

Si toutes les contraintes sont affines, alors $\bar{w} = 0$ convient : les contraintes sont qualifiées en tout point.
Si la famille $(g_i'(\bar{x}))_{i \in I(\bar{x})}$ est libre alors les contraintes sont qualifiées.

Théorème 7.

Si $\bar{x} \in K$, où les contraintes sont qualifiées, est minimum local pour f, il existe m réels $\bar{\lambda}_1, \ldots, \bar{\lambda}_m \geqslant 0$ tels que le système KKT est vérifié.

Contraintes d'égalité et d'inégalité (Conditions mêlées)

- On peut bien sûr coupler les deux types de contraintes.
- On suppose donc que K est donné par

$$K = \{v \in V, \quad G(v) = 0, \quad F(v) \leq 0\},$$

- $G(v) = (G_1(v), \dots, G_N(v))$ et $F(v) = (F_1(v), \dots, F_M(v))$ sont deux applications de V dans $\mathbb{R}^N$ et $\mathbb{R}^M$.

Définition 7.

Soit :

$$I(u) = \{i \in \{1, \dots, M\}, F_i(u) = 0\}$$

On dit dans ce cas que les contraintes sont qualifiées en $u \in K$ si et seulement si les vecteurs $(G_i'(u))_{1 \leq i \leq N}$ sont linéairement indépendants.

On a alors les conditions nécessaires d'optimalité suivantes

Théorème 8.

Soit $u \in K$ où K est . On suppose que J et F sont dérivables en u, que G est dérivable dans un voisinage de u, et que les contraintes sont qualifiées en u . Alors, si u est un minimum local de J sur K, il existe des multiplicateurs de Lagrange $\mu_1, \dots, \mu_N$, et $\lambda_1, \dots, \lambda_M \geq 0$, tels que

$$J'(u) + \sum_{i=1}^{N} \mu_i G_i'(u) + \sum_{i=1}^{M} \lambda_i F_i'(u) = 0, \quad \lambda \geq 0, F(u) \leq 0, \lambda \cdot F(u) = 0.$$