Optimisation continue

Farida Meghatria

University of Khemis Miliana

Cours et exercices, Master 1, Analyse Mathématique et Applications

Contents

1. Convexité et optimisation
2. Conditions d'optimalité pour les problèmes sans contrainte
3. Conditions d'optimalité pour les problèmes avec contraintes

Convexité et optimisation

Ensembles convexes

Définition

Un ensemble $C \subset \mathbb{R}^n$ est dit **convexe** si, pour tous $x, y \in C$ et pour tout $\lambda \in [0,1]$, le point $z = \lambda x + (1-\lambda)y$ appartient également à C.

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

Exemple

Les exemples suivants sont convexes :

- L'espace $\mathbb{R}^n$,
- Un sous-espace vectoriel,
- Une boule fermée $\{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}$,
- Un polyèdre défini par des inégalités linéaires $Ax \leq b$.

Fonctions convexes

Définition

Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite **convexe** sur un ensemble convexe $C \subset \mathbb{R}^n$ si, pour tous $x, y \in C$ et tout $\lambda \in [0, 1]$, on a :

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

Définition (Fonction strictement convexe)

La fonction f est **strictement convexe** si l'inégalité est stricte pour $x \neq y$:

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

exemple

Les fonctions suivantes sont convexes :

- $f(x) = x^2$,
- $f(x) = e^x$,
- $f(x) = \|x\|$,
- $f(x) = -\ln(x)$ sur $(0, \infty)$.

Caractérisation par le gradient (fonction différentiable)

Théorème

Si $f : \mathbb{R}^n \to \mathbb{R}$ est de classe $\mathcal{C}^1$ sur un ensemble convexe C, alors f est convexe si et seulement si :

$$\forall x, y \in C, \quad f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

Cette inégalité exprime que le graphe de f est toujours au-dessus de son plan tangent.

Caractérisation par la hessienne (fonction deux fois dérivable)

Théorème

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

- $\surd$ *f est convexe sur un ensemble convexe C si et seulement si sa matrice hessienne $H_f(x)$ est* ***semi-définie positive*** *sur C.*
- $\surd$ *f est strictement convexe si $H_f(x)$ est* ***définie positive*** *sur C.*

Théorème

Soit $f : \mathbb{R}^n \to \mathbb{R}$ convexe sur un ensemble convexe C.

- *Si $x^* \in C$ vérifie $\nabla f(x^*) = 0$, alors x^* est un **minimum global**.*
- *Si f est strictement convexe, alors ce minimum est **unique**.*

Exemple

Considérons $f(x) = x^2 + 1$. On a $\nabla f(x) = 2x$, donc $x^* = 0$ est le seul point critique. Comme $f''(x) = 2 > 0$, f est strictement convexe implique $x^* = 0$ est le minimum global unique.

Combinaisons convexes et opérations

Proposition

Les propriétés suivantes sont vérifiées :

- *La somme de fonctions convexes est convexe.*
- *La composition $g \circ f$ est convexe si f est convexe et g est convexe croissante.*
- *Le maximum de fonctions convexes est convexe :*

$$f(x) = \max\{f_1(x), f_2(x), \ldots, f_k(x)\}$$

Exemples en optimisation

Exemple

Minimiser la fonction $f(x) = \|Ax - b\|_2^2$. Cette fonction est convexe car elle s'écrit :

$$f(x) = (Ax - b)^T (Ax - b) = x^T A^T Ax - 2b^T Ax + b^T b$$

La hessienne est $H_f(x) = 2A^T A$, semi-définie positive $\Rightarrow$ f est convexe.

Exemple

Minimiser $f(x) = -\ln(x)$ sur $(0, \infty)$. On a :

$$f''(x) = \frac{1}{x^2} > 0 \Rightarrow f \text{ est strictement convexe.}$$

Conditions d'optimalité pour les problèmes sans contrainte

Problème général

Soit $f : \mathbb{R}^n \to \mathbb{R}$, une fonction de classe $\mathcal{C}^2$. On considère le problème suivant :

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

On cherche à déterminer les conditions qui permettent de garantir :

- $\surd$ l'existence d'un minimum,
- $\surd$ l'unicité du minimum,
- $\surd$ la caractérisation des points candidats au minimum.

Existence du minimum global

Théorème (Weierstrass)

Si f est continue sur un ensemble fermé et borné $K \subset \mathbb{R}^n$, alors f admet un minimum global sur K.

Remarque

Dans $\mathbb{R}^n$, si f est continue et $\lim_{\|x\| \to \infty} f(x) = +\infty$, alors f admet un minimum global.

Unicité du minimum

Proposition

Si f est strictement convexe sur un ensemble convexe $C \subset \mathbb{R}^n$, alors tout minimum global de f sur C est unique.

Conditions nécessaires du premier ordre

Théorème

*Soit $x^* \in \mathbb{R}^n$ un point **localement optimal** pour f, supposée différentiable. Alors :*

$$\nabla f(x^*) = 0$$

Définition

Un point x^* tel que $\nabla f(x^*) = 0$ est appelé un **point critique** ou **stationnaire**.

Conditions du second ordre

Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois continûment différentiable. Soit x^* un point critique de f, c'est-à-dire $\nabla f(x^*) = 0$.

Condition nécessaire du second ordre

Théorème

Si x^* est un minimum local de f, alors la hessienne de f en x^* est ***semi-définie positive*** *:*

$$\forall y \in \mathbb{R}^n, \quad y^T H_f(x^*) y \geq 0$$

Condition suffisante du second ordre

Théorème

*Si $\nabla f(x^*) = 0$ et si la hessienne $H_f(x^*)$ est **définie positive**, alors x^* est un **minimum local strict** de f.*

Résumé

$\surd$ $\nabla f(x^*) = 0$ est une **condition nécessaire** (1^{er} ordre).

$\surd$ $H_f(x^*) \succeq 0$ est une **condition nécessaire** (2^{e} ordre).

$\surd$ $H_f(x^*) \succ 0$ est une **condition suffisante** pour un minimum local strict.

$\surd$ Si f est convexe, alors tout point critique est un minimum global.

$\surd$ Si f est strictement convexe, le minimum global est unique.

Exemple

Soit $f(x,y) = x^2 + xy + y^2$.

$\surd$ $\nabla f(x,y) = \begin{pmatrix} 2x+y \\ x+2y \end{pmatrix} \Rightarrow \nabla f = 0 \Leftrightarrow x = 0, y = 0$

$\surd$ La hessienne est :

$$H_f = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Elle est définie positive (valeurs propres : $1, 3 > 0$)

Conclusion : $(0,0)$ est un minimum local strict (et même global car f est convexe).

Conditions d'optimalité pour les problèmes avec contraintes

Problème d'optimisation sous contraintes

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction à minimiser, sous des contraintes :

$$\begin{cases} h_i(x) = 0, & i = 1, \ldots, m \quad \text{(contraintes d'égalité)} \\ g_j(x) \leq 0, & j = 1, \ldots, p \quad \text{(contraintes d'inégalité)} \end{cases}$$

On cherche à résoudre :

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{tel que} \quad h_i(x) = 0,\ g_j(x) \leq 0$$

Multiplicateurs de Lagrange (cas sans inégalités)

Dans cette section, nous présentons une condition nécessaire du premier ordre pour les problèmes d'extremum avec contraintes. Le résultat est le théorème de Lagrange, bien connu. Pour mieux comprendre l'idée sous-jacente à ce théorème, nous considérons d'abord les fonctions de deux variables et une seule contrainte d'égalité. Soit h : la fonction de contrainte. Rappelons qu'en tout point x du domaine, le vecteur gradient $\nabla h(x)$ est orthogonal à l'ensemble de niveau passant par ce point. Choisissons en effet un point tel que $h(x^*) = 0$, et supposons que $\nabla h(x^*) \neq 0$.
L'ensemble de niveau passant par le point x^* est l'ensemble $\{x : h(x) = 0\}$. Nous paramétrons alors cet ensemble de niveau au voisinage de x^* par une courbe $\{x(t)\}$, c'est-à-dire une fonction vectorielle continûment différentiable, $x : \mathbb{R} \longrightarrow \mathbb{R}^2$ telle que

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \quad t \in (a,b), \quad x^*(t) = x(t^*), \quad x^{'}(t^*) \neq 0, \ t^* \in (a,b)$$

Nous pouvons maintenant montrer que $\nabla h(x^*)$ est orthogonal à $x^{'}(t^*)$. En effet, puisque h est constante sur la courbe $\{x(t) : t \in (a,b)\}$, on a que pour tout $t \in (a,b)$, $h(x) = 0$.

Donc, pour tout $t \in (a,b)$, $\dfrac{d}{dt}h(x(t)) = 0$.

En appliquant la règle de la chaine (chain rule), on obtient

$$\frac{d}{dt}h(x(t)) = \nabla h(x(t))^T x^{'}(t) = 0$$

par conséquent, $\nabla h(x^*)$ est orthogonal à $x^{'}(t)$.

Supposons maintenant que x^* minimise $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ sur l'ensemble $\{x : h(x) = 0\}$. Nous affirmons que $\nabla f(x^*)$ est orthogonal à $x^{'}(t^*)$. Pour le démontrer, il suffit d'observer que la fonction composée de t donnée par $\phi(t) = f(x(t))$ atteint un minimum en t^*. Par conséquent, la condition nécessaire du premier ordre pour le problème d'extremum sans contrainte implique que $\dfrac{d\phi}{dt}(t^*) = 0$.

L'application de la règle de la chaîne donne:

$$0 = \frac{d\phi}{dt}(t^*) = \nabla f(x(t^*))^T x'(t^*) = \nabla f(x^*)^T x'(t^*).$$

Ainsi, $\nabla f(x^*)$ est orthogonal à $x'(t^*)$. Le fait que $x'(t^*)$ soit tangente à la courbe $\{x(t)\}$ en x^* signifie que $\nabla f(x^*)$ est orthogonal à la courbe en x^* (voir 26)

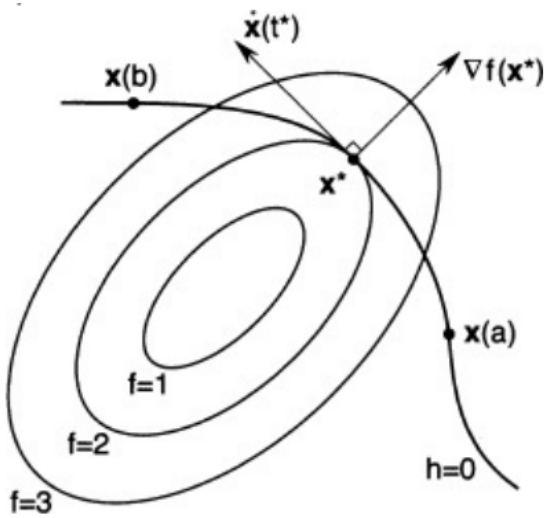


Figure: Le gradient $\nabla f(x^*)$ est orthogonal à la courbe $\{x(t)\}$ au point x^* qui minimise f sur la courbe.

Rappelons que $\nabla h(x^*)$ est également orthogonal à $\{x(t)\}$. Par conséquent, les vecteurs $\nabla h(x^*)$ et $\nabla f(x^*)$ sont parallèles; autrement dit, $\nabla f(x^*)$ est un multiple scalaire de $\nabla f(x^*)$. Ces observations nous permettent maintenant de formuler le théorème de Lagrange pour les fonctions de deux variables avec une contrainte.

Théorème (Théorème de Lagrange pour $n = 2$, $m = 1$.)

Soit le point x^* minimisant $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, sous la contrainte $h(x) = 0$, $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$. Alors, $\nabla f(x^*)$ et $\nabla h(x^*)$ sont parallèles. Autrement dit, si $\nabla h(x^*) \neq 0$, alors il existe un scalaire λ^* tel que

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$$

Dans le théorème 3.1, nous appelons λ^* le multiplicateur de Lagrange. Notons que le théorème est également valable pour les problèmes de maximisation.
Le théorème de Lagrange fournit une condition nécessaire du premier ordre pour qu'un point soit un minimum local. Cette condition, que nous appelons condition de Lagrange, se compose de deux équations

$$\begin{cases} \nabla f(x^*) + \lambda^* \nabla h(x^*) = 0 \\ h(x^*) = 0 \end{cases}$$

On remarque que la condition de Lagrange est nécessaire mais non suffisante. Dans la figure 29, nous illustrons divers points où la condition de Lagrange est satisfaite, y compris un cas où le point n'est pas un extremum (ni un minimum, ni un maximum).

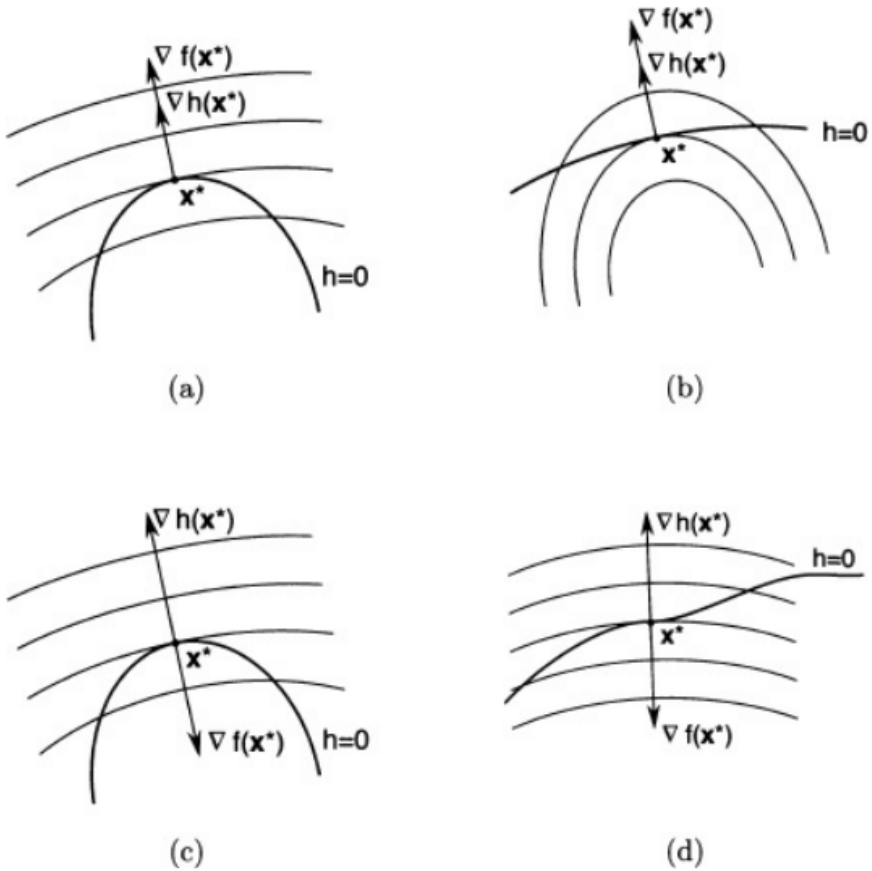


Figure: Quatre exemples où la condition de Lagrange est satisfaite: (a) maximum, (b) minimum, (c) minimum, (d) point non extremum.

Nous généralisons maintenant le théorème de Lagrange au cas où $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m \leq n$

Théorème (Conditions de Lagrange — 1er ordre)

Soit $x^* \in \mathbb{R}^n$ une solution locale du problème :

$$\min f(x) \quad \textit{sous} \quad h_i(x) = 0,\ i = 1, \ldots, m$$

Supposons que les gradients $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ soient linéairement indépendants. Alors il existe des multiplicateurs $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tels que :

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$$

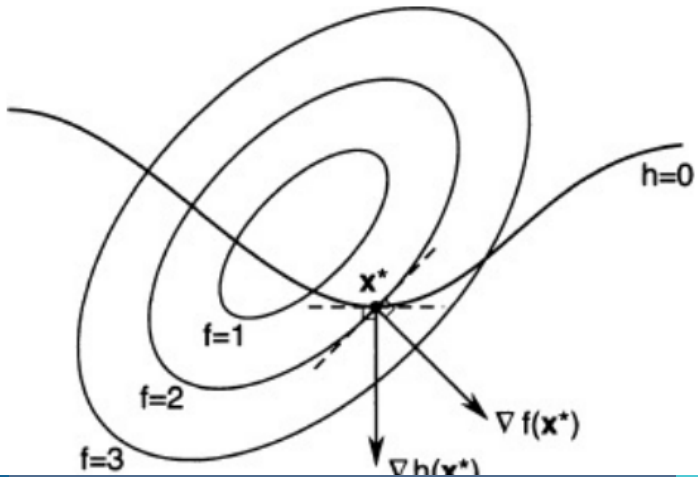
$$h_i(x^*) = 0, \quad \textit{pour tout } i$$

Définition

On appelle **fonction de Lagrange** :

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

Le théorème de Lagrange stipule que si x^* est un extrémum, alors le gradient de la fonction objectif f peut être exprimé comme une combinaison linéaire des gradients des contraintes.
Dans le théorème 3.2, le vecteur λ^* est appelé vecteur multiplicateur de Lagrange, et ses composantes, multiplicateurs de Lagrange.
Si cette condition n'est pas vérifiée, alors x^* ne peut pas être un extrémum. Cette situation est illustrée dans la figure 31.



On remarque que la régularité est une hypothèse du théorème de Lagrange. Cette hypothèse est essentielle, comme le montre l'exemple suivant.

Exemple

Considérons le problème suivant:

$$\begin{cases} \min f(x) \\ S.C \quad h(x) = 0 \end{cases}$$

où $f(x) = x$ et

$$h(x) = \begin{cases} x^2 \quad si \quad x < 0 \\ 0 \quad si \quad 0 \leq x \leq 1 \\ (x-1)^2 \quad si \quad x > 1. \end{cases}$$

L'ensemble admissible est manifestement $[0, 1]$. Clairement, $x^* = 0$ est un minimiseur local. Cependant, $f'(x^*) = 1$ et $h'(x^*) = 0$. Par conséquent, x^* ne satisfait pas la condition nécessaire du théorème de Lagrange.
Remarquons toutefois que x^* n'est pas un point régulier, ce qui explique pourquoi le théorème de Lagrange ne s'applique pas ici.
Il est commode d'introduire la fonction lagrangienne $l : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$, définie par $l(x, \lambda) \triangleq f(x) + \lambda^T h(x)$
La condition de Lagrange pour un minimum local x^* peut être représentée à l'aide de la fonction Lagrangienne comme suit:

$$Dl(x^*, \lambda^*) = 0^T$$

pour un certain λ^*, où la dérivée D est prise par rapport à l'argument entier $[x, \lambda]$. En d'autres termes, la condition nécessaire du théorème de Lagrange est équivalente à la condition nécessaire du premier ordre pour l'optimisation sans contrainte appliquée à la fonction lagrangienne.

Pour illustrer ce qui précède, notons la dérivée de l par rapport à x par $D_x l$ et la dérivée de l par rapport à λ par $D_\lambda l$. Alors,
Remarquons que $D_x l(x, \lambda) = Df(x) + \lambda^T Dh(x)$ et $D_x l(x, \lambda) = h(x)$. Par conséquent,

$$\begin{cases} D_x l(x^*, \lambda^*) = 0^T \\ D_\lambda l(x^*, \lambda^*) = 0^T \end{cases}$$

le théorème de Lagrange pour un minimum local x^* peut s'énoncer comme suit : pour un certain λ^*, ce qui est équivalent à
$Dl(x^*, \lambda^*) = 0^T$
Autrement dit, la condition de Lagrange peut s'exprimer par $Dl(x^*, \lambda^*) = 0$.
La condition de Lagrange est utilisée pour trouver d'éventuels extrémums.

Cela implique la résolution des équations

$$\begin{cases} D_x l(x^*, \lambda^*) = 0^T, \\ D_\lambda l(x^*, \lambda^*) = 0^T. \end{cases}$$

Le système ci-dessus représente $n + m$ équations à $n + m$ inconnues. Il est important de noter que la condition de Lagrange est nécessaire mais non suffisante ; autrement dit, un point x^* satisfaisant les équations ci-dessus n'est pas nécessairement un extremum.

Exemple

Considérons le problème de l'extrémisation de la fonction objectif $f(x) = x_1^2 + x_2^2$ sur l'ellipse,

$$[x_1, x_2]^T : h(x) = x_1^2 + 2x_2^2 - 1 = 0$$

On a,

$\nabla f(x) = [2x_1, 2x_2]^T$
$\nabla h(x) = [2x_1, 4x_2]^T$

Ainsi,

$$D_x l(x, \lambda) = [f(x) + \lambda h(x)] = [2x_1 + 2\lambda x_1, 2x_2 + 4\lambda x_2]$$

et

$$D_\lambda l(x, \lambda) = h(x) = x_1^2 + 2x_2^2 - 1$$

En posant $D_x l(x, \lambda) = 0$ et $D_\lambda l(x, \lambda) = 0$, nous obtenons trois équations à trois inconnues,

$2x_1 + 2\lambda x_1 = 0$
$2x_2 + 4\lambda x_2 = 0$
$x_1^2 + 2x_2^2 = 1$

Tous les points admissibles de ce problème sont réguliers. À partir de la première des équations ci-dessus, nous obtenons soit $x_1 = 0$, soit $\lambda = -1$. Dans le cas où $x_1 = 0$, les deuxième et troisième équations impliquent que $\lambda = \frac{1}{2}$ et $x_2 = \pm\frac{1}{\sqrt{2}}$. Dans le cas où $\lambda = 1$, les deuxième et troisième équations impliquent que $x_1 = \pm 1$ et $x_2 = 0$. Ainsi, les points qui satisfont la condition de Lagrange pour les extrema sont

$$x^{(1)} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad x^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Puisque, $f(x^{(1)}) = f(x^{(2)} = \frac{1}{2}$ et $f(x^{(3)}) = f(x^{(4)} = 1$

nous en concluons que s'il existe des minimums, ils sont situés en $x^{(1)}$ et $x^{(2)}$ et s'il existe des maximums, ils sont situés en $x^{(3)}$ et $x^{(4)}$.
Il s'avère qu'en effet, $x^{(1)}$ et $x^{(2)}$ sont des minimums et $x^{(3)}$ et $x^{(4)}$ sont des maximums. Ce problème peut être résolu graphiquement, comme illustré dans la figure 20.14.

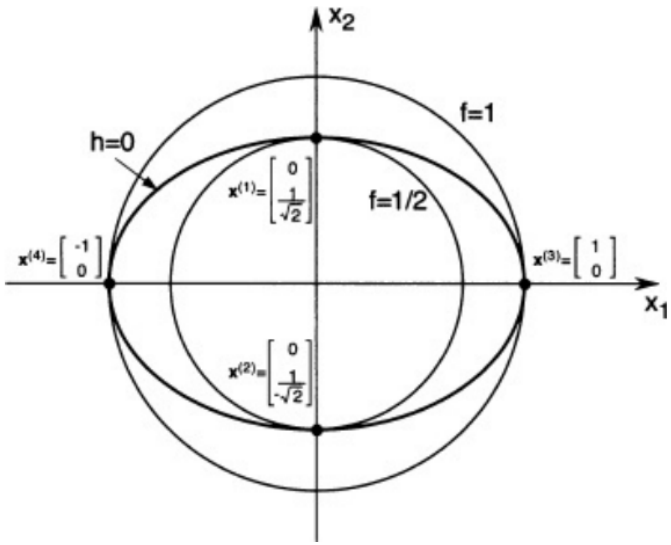


Figure: Solution graphique du problème de l'exemple

Les problèmes avec contraintes d'inégalité peuvent également être traités à l'aide des multiplicateurs de Lagrange.
On considère le problème suivant:

$$Minimiser f(x)\ S.C\ \ h(x) = 0 \qquad\qquad g(x) \leq 0$$

où $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ tel que $m \leq n$, et $g : \mathbb{R}^n \longrightarrow \mathbb{R}^p$. Pour le problème général ci-dessus, nous adoptons les définitions suivantes.

Définition

Une contrainte d'inégalité $g_j(x)$ est dite active en x^* si $g_j(x^*) = 0$. Elle est inactive en x^* si $g_j(x^*) < 0$.
Par convention, nous considérons qu'une contrainte d'égalité $h_i(x) = 0$ est toujours active.

Définition

Soit x^* tel que $h(x^*) = 0$, $g(x^*) \leq 0$, et soit $J(x^*)$ l'ensemble des indices des contraintes d'inégalité actives: $J(x^*) \triangleq \{j : g_j(x^*) = 0\}$.
Alors, nous disons que x^* est un point régulier si les vecteurs $\nabla h_i(x^*) = 0,\ \nabla g_j(x^*) = 0,\quad 1 \leq i \leq m$ et $j \in J(x^*)$ sont linéairement indépendants.

Nous démontrons maintenant une condition nécessaire du premier ordre pour qu'un point soit un minimum local. On appelle cette condition la condition de Karush-Kuhn-Tucker (KKT).
Dans la littérature, cette condition est parfois également appelée condition de Kuhn-Tucker.

Théorème

Soit $f,\ h,\ g \in \mathcal{C}$. Soit x^* un point régulier et un minimum local du problème de minimisation de f sous les contraintes: $h(x) = 0,\ g(x) \leq 0$. Alors, il existe $\lambda^* \in \mathbb{R}^m$ et $\mu^* \in \mathbb{R}^p$ tels que:

- $\mu^* \geq 0$
- $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$
- $\mu^{*T} g(x^*) = 0$

Dans le théorème **??**, λ^* est appelé vecteur des multiplicateurs de Lagrange et μ^* le vecteur des multiplicateurs de Karush-Kuhn-Tucker (KKT). Leurs composantes sont appelées respectivement multiplicateurs de Lagrange et multiplicateurs de Karush-Kuhn-Tucker (KKT).
Avant de démontrer ce théorème, examinons d'abord sa signification. On observe que $\mu_j^* \geq 0$ (d'après la condition 1) et $g_j^(x)* \leq 0$. Par conséquent, la condition implique que si $g_j(x^*) < 0$, alors $\mu_j^* = 0$; c'est-à-dire que pour tout $j \notin J(x^*)$, on a $\mu_j^* = 0$.
Autrement dit, les multiplicateurs KKT μ_j^* correspondant aux contraintes inactives sont nuls. Les autres multiplicateurs KKT, $\mu_i^*,\ i \in J(x^*)$, sont non négatifs; ils peuvent être ou non égaux à zéro.

Exemple

Une illustration graphique du théorème KKT est donnée dans la figure **??**. Dans cet exemple bidimensionnel, nous n'avons que des contraintes d'inégalité $g_j(x) \leq 0, \quad j = 1, 2, 3.$

Figure: Illustration du théorème de Karush-Kuhn-Tucker (KKT).

suit

Remarquons que le point x^* sur la figure est bien un minimum. La contrainte $g_3(x) \leq 0$ est inactive : $g_3(x) < 0$; par conséquent $\mu_3^* = 0$. D'après le théorème de Karush-Kuhn-Tucker, nous avons

$$\nabla f(x^*) + \mu_1^* \nabla g_1(x^*) + \mu_2^* \nabla g_2(x^*) = 0$$

ou, de manière équivalente,

$$\nabla f(x^*) = -\mu_1^* \nabla g_1(x^*) - \mu_2^* \nabla g_2(x^*)$$

où μ_1^* et μ_2^*. Il est facile d'interpréter graphiquement la condition de KKT pour cet exemple. Plus précisément, on peut voir sur la figure **??** que $\nabla f(x^*)$ doit être une combinaison linéaire des vecteurs $-\nabla g_1(x^*)$ et $-\nabla g_2(x^*)$ à coefficients positifs. Ceci se reflète exactement dans l'équation ci-dessus, où les coefficients μ_1^* et μ_2^* sont les multiplicateurs KKT.
On applique la condition de KKT de la même manière que toute condition nécessaire. Plus précisément, nous recherchons les points satisfaisant la condition de KKT et considérons ces points comme des minimums potentiels.

En résumé, la condition de KKT se compose de cinq parties (trois équations et deux inégalités):

- $\mu^* \geq 0$, (complémentarité)
- $Df(x^*) + \lambda^{*T} Dh(x^*) + \mu^{*T} Dg(x^*) = 0^T$, (stationnarité)
- $\mu^{*T} g(x^*) = 0$, (complémentarité)
- $h(x^*) = 0$, (primalité)
- $g(x^*) \leq 0$. (primalité)

Nous supposons que $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, sont deux fois continûment différentiables : $f, h \in \mathcal{C}^2$. Soit
$l(x, \lambda) = f(x) + \lambda^T h(x) = f(x) + \lambda_1 h_1(x) + ... + \lambda_m h_m(x)$ la fonction lagrangienne. Soit $L(x, \lambda)$ la matrice hessienne de $l(x, \lambda)$ par rapport à x:
$L(x, \lambda) = F(x) + \lambda_1 H_1(x) + ... + H_m(x)$, où $F(x)$ est la matrice hessienne de f en x et $H_k(x)$ est la matrice hessienne de h_k en x, $k = 1, ..., m$. Donnée par

$$H_k(x) = \begin{bmatrix} \dfrac{\partial^2 h_k(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 h_k(x)}{\partial x_n \partial x_1} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial^2 h_k(x)}{\partial x_1 \partial x_n} & \cdots & \dfrac{\partial^2 h_k(x)}{\partial x_n^2} \end{bmatrix}$$

Nous introduisons la notation $[\lambda H(x)]$:
$[\lambda H(x)] = \lambda_1 H_1(x), ..., \lambda_m H_m(x)$.

Théorème

Condition nécessaire de deuxième ordre Soit x^* un minimum local de $f : \mathbb{R}^n \to \mathbb{R}$ avec la contrainte $h(x) = 0$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$ et $f, h \in C^2$.
Supposons que x^* est régulière. Alors, il existe $\lambda^* \in \mathbb{R}^m$ tel que:

- $Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$
- pour tout $y \in T(x^*)$, on a $y^T L(x^*, \lambda^*) y \geq 0$

Observons que $H_L(x, \lambda)$ joue un rôle similaire à celui de la matrice hessienne $F(x)$ de la fonction objectif f. fait dans le cas de minimisation sans contraintes. Cependant, nous exigeons maintenant que $L(x^*, \lambda^*) \geq 0$ uniquement sur $T(x^*)$ plutôt que sur $\mathbb{R}^n$.
Les conditions ci-dessus sont nécessaires, mais pas suffisantes, pour qu'un point soit un minimum local.
Nous présentons maintenant, sans preuve, des conditions suffisantes pour qu'un point soit un minimum local strict.

Théorème

Condition d'optimalité suffisante de deuxième ordre
Supposons que $f, h \in C^2$ *et il existe un point* $x^* \in \mathbb{R}^n$, *et* $\lambda^* \in \mathbb{R}^m$ *tel que:*

- $Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$
- *pour tout* $y \in T(x^*)$, $y \neq 0$ *on a* $y^T L(x^*, \lambda^*) y \geq 0$

Remarque

L'explication la plus simple du théorème **??** *est comme suit: [Condition suffisante du second ordre] Soit* x^* *un point stationnaire et* λ *les multiplicateurs associés. On définit l'espace tangent aux contraintes :*

$$T = \{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0,\ i = 1, \ldots, m\}$$

Si :

$$\forall d \in T \setminus \{0\}, \quad d^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda) d > 0$$

alors x^* *est un minimum local strict.*

Exemple

Considérons le problème:

$$maximiser \frac{x^T Q x}{x^T P x}$$

Où $Q = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ et $P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

Comme indiqué précédemment, nous pouvons ce problème sous la forme équivalente:

$$\begin{cases} Minimiser x^T Q x \\ s.c \quad x^T P x = 1 \end{cases}$$

La fonction objectif pour le problème transformé est donné par

$$L(x, \lambda) = x^T Q x + \lambda(1 - x^T P x),$$

Les conditions de Lagrange, donne:

$$(\lambda I - P^{-1} Q)x = 0$$

Où

$$P^{-1} Q = \begin{bmatrix} 2 & 0 \end{bmatrix}$$

Il n'y a que deux valeurs de λ, satisfant $(\lambda I - P^{-1}Q)x = 0$, ce sont, les valeurs propres de $P^{-1}Q^{-1}$, $\lambda = 2, \quad \lambda = 1$.
Nous rappelons de notre discussion précédente de ce problème que le multiplicateur de Lagrange correspondant à la solution est la valeur propre maximale de $P^{-1}Q$, notée: $\lambda^* = \lambda_1 = 2$.
Le vecteur propre est le maximum-la solution de ce problème. Le vecteur propre correspondant au valeur propre $\lambda^* = 2$, satisfont la contrainte $x^T Q x = 1$ est $\pm x^*$ où,

$$x^* = [\frac{1}{\sqrt{2}}, \quad 0]$$

A ce stade, tout ce que nous avons établi est que les paires $(\pm x^*, \lambda^*)$ satisfont la condition de Lagrange.
Nous montrons maintenant que les points $\pm x^*$ sont, en fait, des maximums locaux stricts.
Notons qu'une procédure similaire s'applique à $-x^*$.

Nous calculons d'abord la matrice hessienne de la fonction Lagrangienne. On à:

$$H_L(x^*, \lambda^*) = 2Q - 2\lambda P = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$$

$$T(x^*) = \{y \in \mathbb{R} : \quad x^{*T} P y = 0\} = \{y : \quad (\sqrt{2}, 0)y = 0\}$$

$= \{y = [0, a]^T, \quad a \in \mathbb{R}\}$
L'espace tangent $T(x^*)$ a $\{x : \quad 1 - x^T P x = 0\}$ est

$$y^T L(x^*, \lambda^*) y = (0, a) \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ a \end{bmatrix} = -2a^2 < 0$$

Notons que pour chaque $y \in T(x^*), y \neq 0$

$\sqrt{}$ La stationnarité exprime que le gradient de f est une combinaison linéaire des gradients des contraintes actives.

$\sqrt{}$ La complémentarité signifie que si une contrainte $g_j(x^*) < 0$ est inactive, alors son multiplicateur $\mu_j = 0$.

$\sqrt{}$ Les multiplicateurs λ_i, μ_j mesurent la sensibilité de la fonction objectif à la contrainte associée.