Optimisation sans contrainte : conditions d'optimalité

- Problème :

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

- Soit x^* un minimum local
- Comment caractériser x^* ?

Conditions nécessaires

Conditions nécessaires d'optimalité *Soit x^* un minimum local d'une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.*
Si f est différentiable dans un voisinage ouvert V de x^*, alors,

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

Il s'agit de la condition nécessaire du premier ordre*.*
Si, de plus, f est deux fois différentiable sur V, alors

$$\nabla^2 f(x^*) \text{ est semi définie positive.}$$

et f est localement convexe en x^*.
Il s'agit de la condition nécessaire du second ordre*(p. 127).*

Preuve

Premier ordre:

- Par l'absurde, supposons que $\nabla f(x^*) \neq 0$.
- Comme $-\nabla f(x^*)$ est direction de descente,

$$f(x^* - \alpha \nabla f(x^*)) < f(x^*) \quad \forall 0 < \alpha \leq \eta.$$

- Cela contredit l'optimalité de x^*.

Second ordre

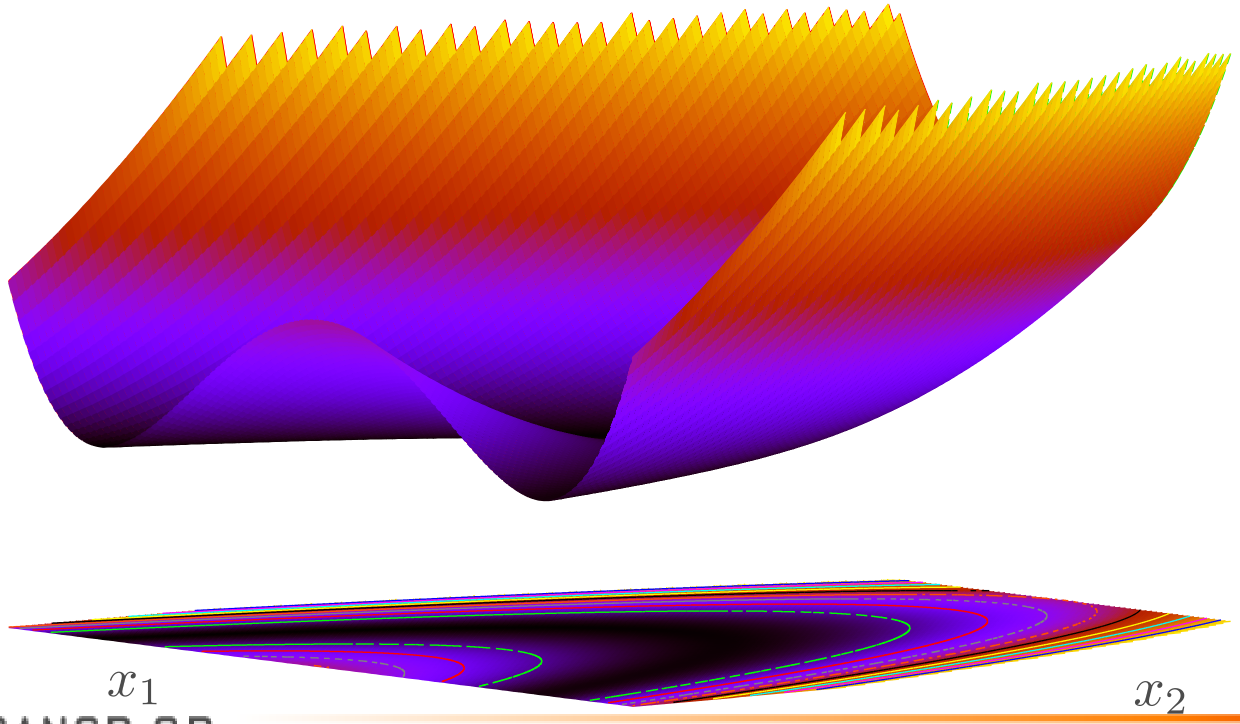
- Taylor:

$$f(x^*+\alpha d)-f(x^*) = \alpha d^T \nabla f(x^*)+\frac{1}{2}\alpha^2 d^T \nabla^2 f(x^*) d+o(\|\alpha d\|^2) \geq 0.$$

- $\nabla f(x^*) = 0$, $o(\|\alpha d\|^2)$ est négligeable. Donc $d^T \nabla^2 f(x^*) d \geq 0$.

Conditions nécessaires

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Conditions nécessaires

$x^* = (1,1)$ minimum local

$$\nabla f(x_1,x_2) = \begin{pmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{pmatrix}$$

et $\nabla f(1,1) = 0$.

$$\nabla^2 f(x_1,x_2) = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

et

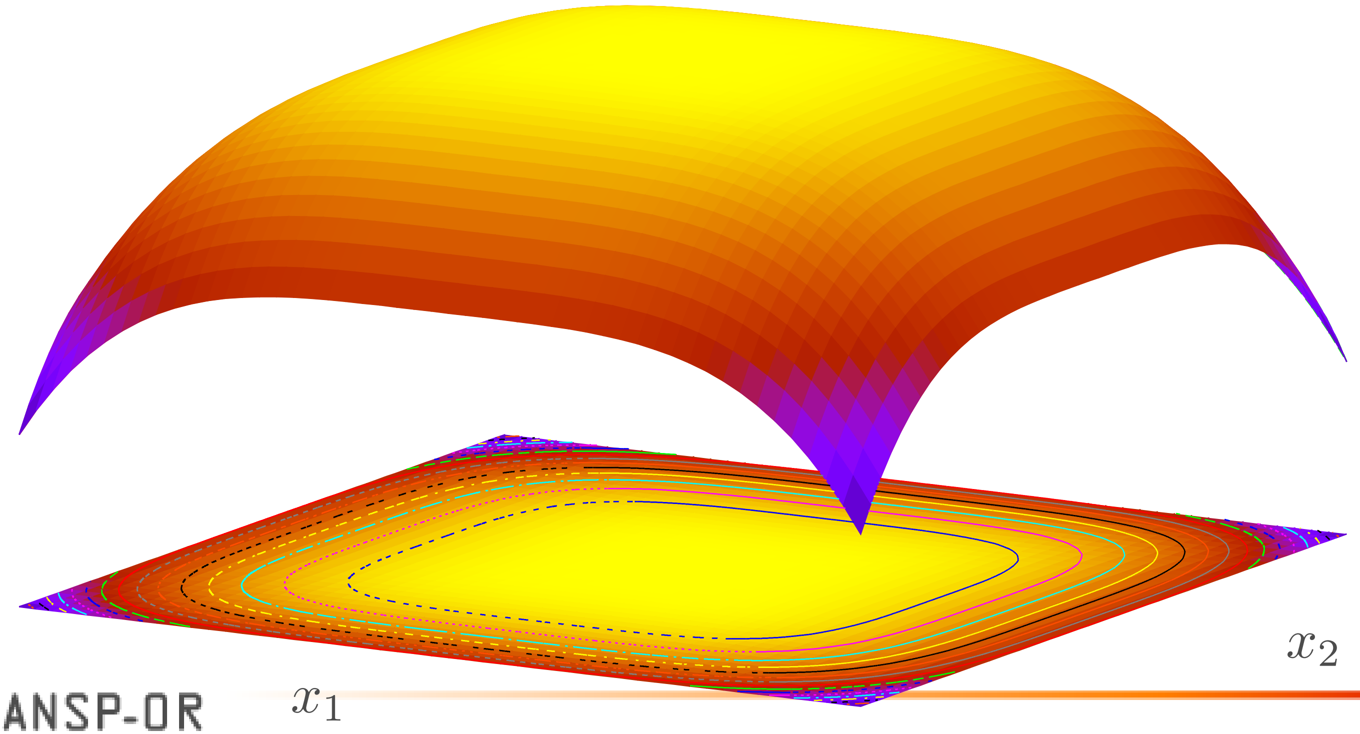
$$\nabla^2 f(1,1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

Conditions nécessaires

- Valeurs propres positives : 0.39936 et 1001.6
- Mauvais conditionnement : 2508

Conditions nécessaires

$$f(x_1, x_2) = -x_1^4 - x_2^4$$



Conditions nécessaires

$x^* = (0,0)$ vérifie les cond. néc. d'optimalité

$$\nabla f(x_1, x_2) = \begin{pmatrix} -4x_1^3 \\ -4x_2^3 \end{pmatrix}$$

et $\nabla f(0,0) = 0$

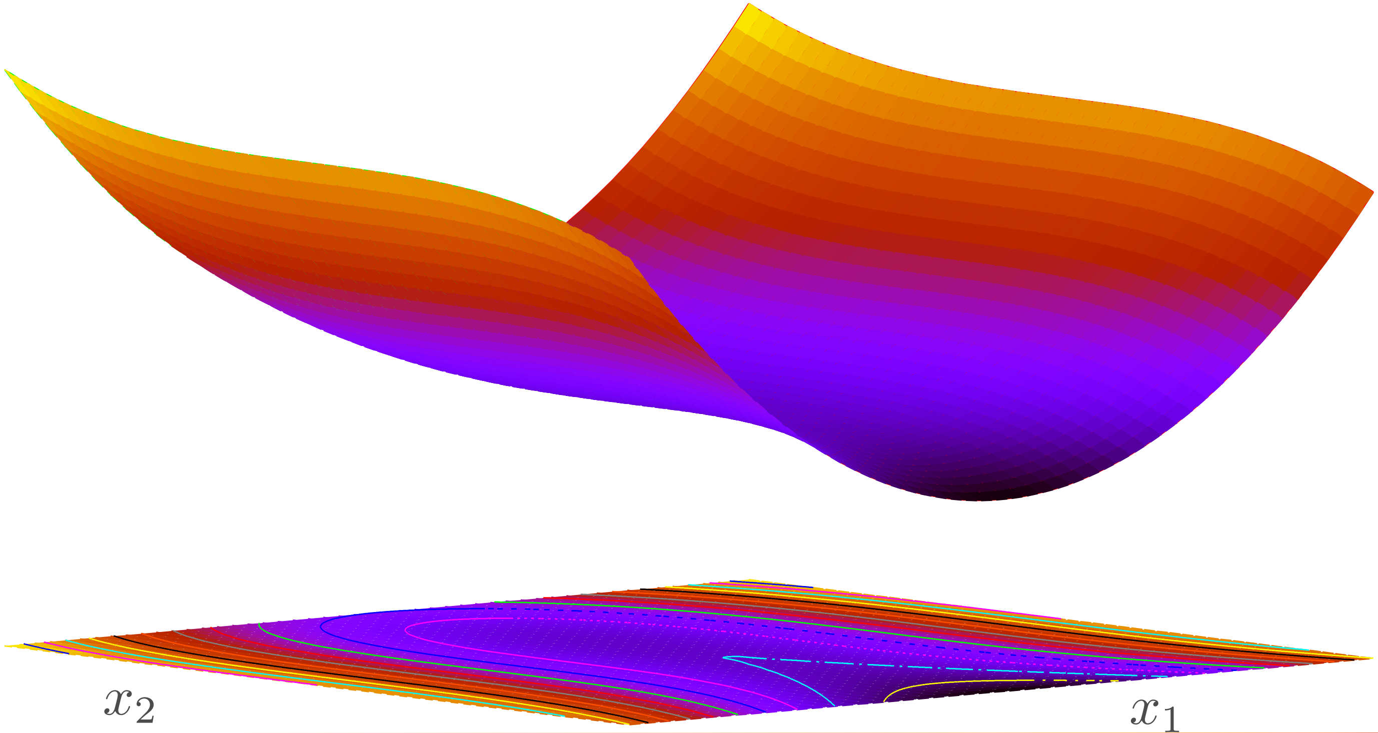
$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} -12x_1^2 & 0 \\ 0 & -12x_2^2 \end{pmatrix}$$

et $\nabla^2 f(0,0)$ est semi défini positif.

$(0,0)$ **n'est pas un minimum local**

Conditions nécessaires

$$f(x_1, x_2) = 50x_1^2 - x_2^3$$



Conditions nécessaires

$x^* = (0,0)$ vérifie les cond. néc. d'optimalité

$$\nabla f(x_1, x_2) = \begin{pmatrix} 100x_1 \\ -3x_2^2 \end{pmatrix}$$

et $\nabla f(0,0) = 0$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 100 & 0 \\ 0 & -6x_2 \end{pmatrix}$$

et $\nabla^2 f(0,0)$ est semi déf. pos.

Conditions nécessaires

- Considérons la direction $d = (0,1)^T$ en $(0,0)$.
- Avançons d'un pas α. On obtient

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$$

et

$$f\begin{pmatrix} 0 \\ \alpha \end{pmatrix} = -\alpha^3 < 0 = f\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$(0,0)$ **n'est pas un minimum local**

Conditions nécessaires

- Considérons la direction $d = (0, -1)^T$ en $(0, 0)$.
- Avançons d'un pas α. On obtient

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -\alpha \end{pmatrix}$$

et

$$f\begin{pmatrix} 0 \\ -\alpha \end{pmatrix} = \alpha^3 > 0 = f\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$(0, 0)$ **est un point de selle**

Conditions nécessaires

Point critique

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable.*
Tout vecteur $x \in \mathbb{R}^n$ *tel que* $\nabla f(x) = 0$ *est appelé* point critique *ou* point stationnaire *de* f

- En pratique les algorithmes cherchent les points critiques
- Ils seront conçus pour se diriger vers les minima, et pas vers les autres points critiques

Conditions suffisantes

Conditions suffisantes d'optimalité *Soit une fonction* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *deux fois différentiable dans un sous-ensemble ouvert* V *de* $\mathbb{R}^n$*, et soit* $x^* \in V$ *qui vérifie les conditions*

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

et

$$\nabla^2 f(x^*) \text{ est définie positive.}$$

Alors, x^* *est un minimum local strict de* f*.(p. 132)*

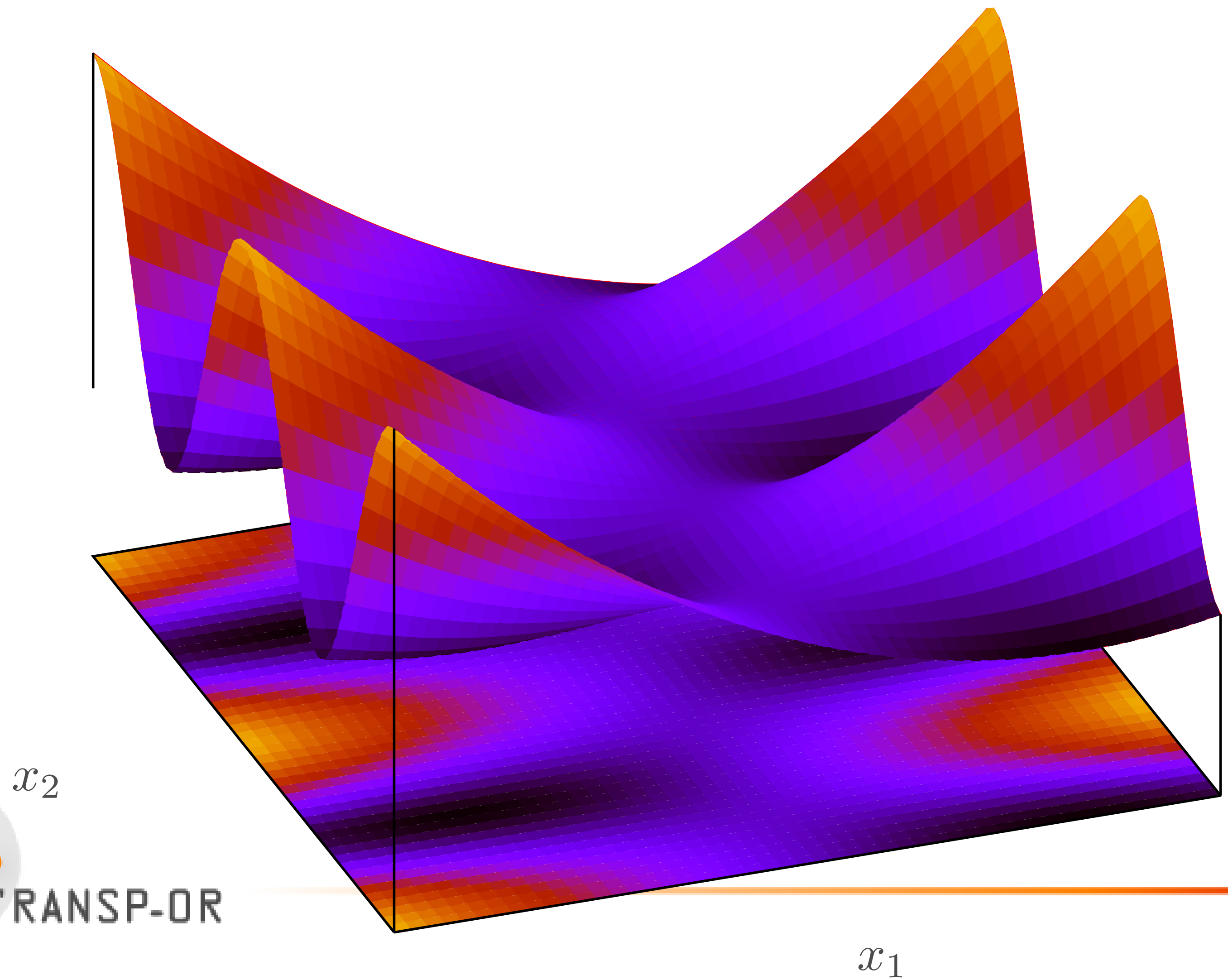
Preuve

Par l'absurde...

- S'il existe une direction de descente,
- alors $\nabla f(x^*) d < 0$, ce qui contredit $\nabla f(x^*) = 0$.
- Par Taylor, $d^T \nabla^2 f(x^*) d \leq 0$, ce qui contredit $\nabla^2 f(x^*)$ définie positive.

Conditions suffisantes

$$f(x_1, x_2) = \frac{1}{2}x_1^2 + x_1 \cos x_2,$$

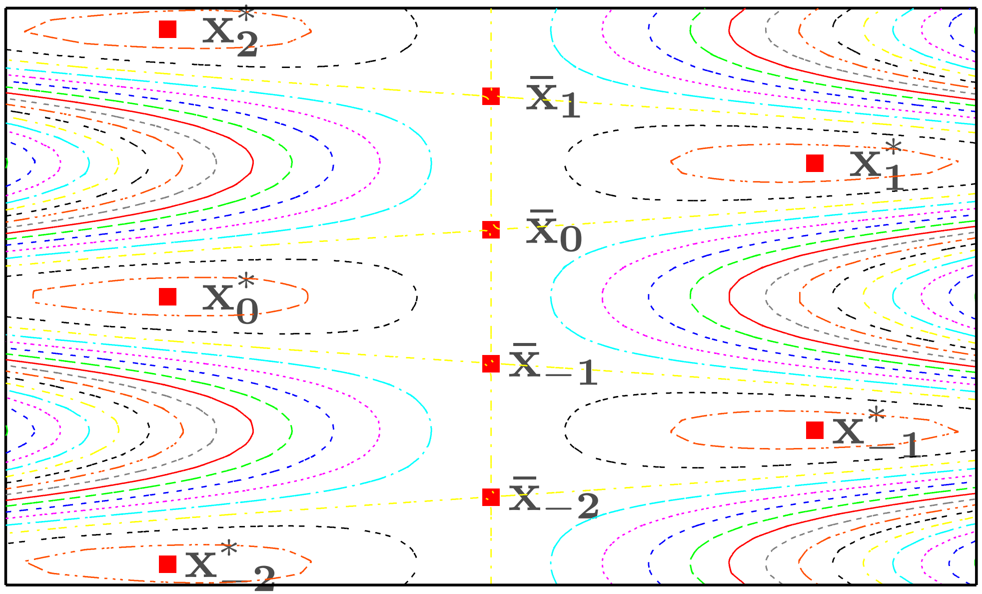


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conditions suffisantes

$$f(x_1, x_2) = \frac{1}{2}x_1^2 + x_1 \cos x_2,$$



Utilisons les cond. d'opt. pour identifier les minima locaux

Conditions suffisantes

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_1 + \cos x_2 \\ -x_1 \sin x_2 \end{pmatrix}.$$

Le gradient s'annule pour

- $x_k^* = ((-1)^{k+1}, k\pi)^T$, $k \in \mathbb{Z}$,
- $\bar{x}_k = (0, \frac{\pi}{2} + k\pi)^T$, $k \in \mathbb{Z}$.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conditions suffisantes

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 1 & -\sin x_2 \\ -\sin x_2 & -x_1 \cos x_2 \end{pmatrix}.$$

$$\nabla^2 f(x_k^*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\nabla^2 f(\bar{x}_k) = \begin{pmatrix} 1 & (-1)^{k+1} \\ (-1)^{k+1} & 0 \end{pmatrix}.$$

Conditions suffisantes

- x_k^* vérifie les conditions suffisantes d'optimalité pour tout k
- $\bar{x}_k$ ne vérifie les conditions nécessaires d'optimalité pour aucun k.

Conditions suffisantes

Conditions suffisantes d'optimalité globale *Soit une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ deux fois différentiable dans $\mathbb{R}^n$, et soit $x^* \in \mathbb{R}^n$ un minimum local de f.*

Si f est une fonction convexe, alors x^* est un minimum global de f.

Si de plus f est strictement convexe, x^* est l'unique minimum global de f.(p. 134)

Preuve

- Si f est convexe, la fonction est toujours “au-dessus” de la tangente.
- En x^*, la tangente est horizontale,
- donc la fonction ne peut pas aller plus bas.
- Si f est strictement convexe, la tangente ne rencontre la fonction qu’en x^*, et tout autre point est forcément “plus haut”.

Problèmes quadratiques

Conditions d'optimalité pour les problèmes quadratiques

Condidérons le problème

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T Q x + g^T x + c$$

où Q est une matrice symétrique $n \times n$, $g \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

1. *Si Q n'est pas semi définie positive, alors le problème ne possède pas de solution, c'est-à-dire qu'il n'existe aucun $x \in \mathbb{R}^n$ qui soit un minimum local.*
2. *Si Q est définie positive, alors $x^* = -Q^{-1} g$ est l'unique minimum global.*

(p. 136)

Preuve

Proposition 2.

- $\nabla f(x^*) = Qx^* + g = -QQ^{-1}g + g = 0$.
- $\nabla^2(x^*) = Q$.
- Si Q est définie positive, les conditions suffisantes s'appliquent en x^*.

Proposition 1.

- Si x^* est minimum local, alors Q semi définie positive. Contradiction.