Complément : Régularisation

Problème bien posé :

- la solution existe
- la solution est unique
- la solution dépend continûment des données.

Régulariser un problème mal posé, c'est le remplacer par un autre, bien posé, de sorte que l' "erreur commise" soit compensée par un gain de stabilité.

Méthode de Tikhonov

Si le problème $A.X = y$ est mal posé on le régularise par une (des) contrainte(s) $F(x) = 0$:

$$\min \frac{1}{2}\|A.x - b\|^2 + \frac{\epsilon^2}{2}\|F(x)\|^2 \qquad \min \frac{1}{2}\|A.x - b\|^2 + \frac{\epsilon^2}{2}\|x - x_0\|^2$$

Le choix "optimal" du paramètre de régularisation est "délicat".

Méthode de Levenberg-Marquardt

L'algorithme de Levenberg-Marquardt peut être vu comme une régularisation de l'algorithme de Gauss-Newton, en particulier lorsque la jacobienne de F n'est pas de rang plein. Mais il existe un lien étroit avec les méthodes dites de région de confiance.

Idée : comme dans la méthode de Gauss-Newton, on remplace le problème initial au voisinage de x_k par :

$$\min_{y\in\mathbb{R}^n} \frac{1}{2}\|r(x_k) + J(x_k)(y - x_k)\|_2^2$$

sous la contrainte $\|y - x_k\|_2^2 \leq \Delta_k$ (Δ_k rayon de la région de confiance).

On cherche à minimiser :

$$\min_{y\in\mathbb{R}^n} \frac{1}{2}\|r(x_k) + J(x_k)(y - x_k)\|_2^2 + \frac{1}{2}\lambda(\|y - x_k\|_2^2 - \Delta_k^2)$$

(Lagrangien - section ??)

Les CN1 d'optimalité cette nouvelle fonction sont :

- $(J(x_k)^t.J(x_k) + \lambda I).d_k = -J(x_k)^t.r(x_k)$
- $\lambda = 0$ ou $\|y - x_k\| = \Delta_k$
- $\lambda \geq 0$

- Si $\|y - x_k\| < \Delta_k$ alors $\lambda = 0$, la contrainte est inactive, on n'atteint pas le bord de Δ_k alors on a une itération de Gauss-Newton.
- Sinon la solution est au bord de la région de confiance, $\|y - x_k\| = \Delta_k$. La matrice $(J(x-k)^t.J(x_k) + \lambda I)$ est maintenant définie positive et le pas effectué est bien un pas de descente de la fonction $f(x_k)$.
- $\lambda \geq 0$ peut être choisi fixe variable suivant qualité du pas.

Algorithm 2 : Algorithme de Levenberg-Marquardt

```
k = 0
while Critère d'arrêt do
    calculer d_k solution de (J_k(x_k)^t.J_k(x_k) + λ.I).d_k = −J_k^t.r_k
    x_{k+1} = d_k + x_k
    Mise à jour de λ
    k = k + 1
end
output : x_k
```

Part II - Optimisation non linéaire avec contraintes
A/ Conditions théoriques

Soit le problème d'optimisation $(P1)$ suivant

$$\min_{x \in \mathbb{X}} f(x)$$

$$h_i(x) = 0 \qquad i = 1, \ldots, m$$

$$g_j(x) \leq 0 \qquad j = 1, \ldots, p$$

où f, g et h sont différentiables sur $\mathcal{X}$.

Exemple : $\min_{x \in \mathcal{R}^n} \; f(x) = x^2$ sous $x \geq 1$
Condition sur le gradient ?
Directions de descente => directions admissibles

Résultats théoriques

Un point $x \in \mathbb{R}^n$ est admissible s'il vérifie toutes les contraintes.

Soit x un point admissible pour le problème général d'optimisation $(P1)$.
Une direction d est dite admissible en x s'il existe $\epsilon > 0$ tel que $x + \alpha d$ est admissible pour $0 \leq \alpha \leq \epsilon$

Conditions nécessaires d'optimalité pour des contraintes générales

Soit x^* un minimum local du problème général d'optimisation $(P1)$ alors
$\nabla f(x^*)^t.d \geq 0$
pour toute direction d admissible (à la limite) en x^*.

Contraintes actives

Pour $1 \leq j \leq p$, une contrainte $g_j(x) \leq 0$ est dite active en x^* si $g_j(x^*) = 0$ et inactive en x^* si $g_j(x^*) < 0$
Intérêt ?

Soit x^+ un point admissible pour le problème général d'optimisation $(P1)$.

Cône des directions

On appelle cône des directions en x^+, $CD(x^+)$, l'ensemble constitué des directions d (et de leurs multiples) telles que :
$d^t.\nabla g_i(x^+) \leq 0, \;\; \forall i = 1, ..., p$ tel que $g_i(x^+) = 0$
et $d^t.\nabla h_i(x^+) = 0, \;\; \forall i = 1, ..., m$.

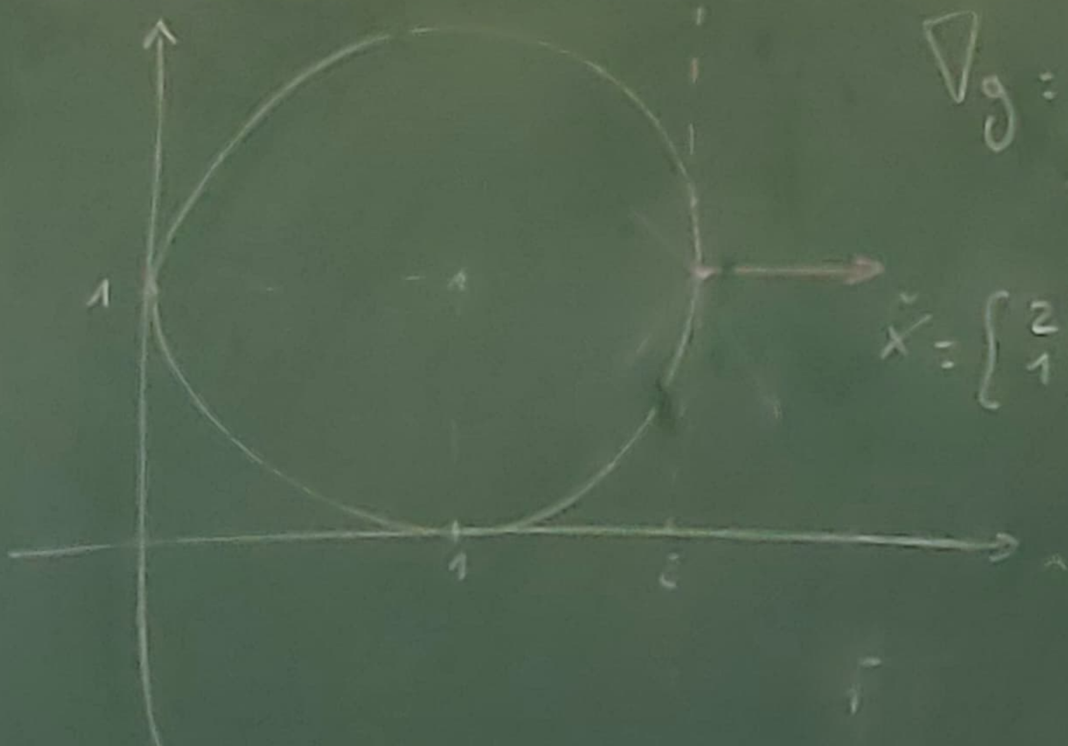
Indépendance linéaire des contraintes

Les contraintes sont linéairement indépendantes en x^+ si les gradients des contraintes égalités, $\nabla h(x^+)$, et les gradients des contraintes d'inégalités actives en x^+, $\nabla g_i(x^+)$, sont linéairement indépendants.

Qualification des contraintes

La qualification des contraintes est vérifiée si $\forall d \in CD(x^+)$ d est une direction admissible à la limite en x^+.
Les directions admissibles à la limite sont une extension de la notion de direction admissible difficile à calculer. Nous supposerons que les contraintes sont qualifiées si elles sont linéairement indépendantes.

$$g(x) = \frac{1}{2}(x-1)^2 + \frac{1}{2}(y-1)^2 - \frac{1}{2} \le 0$$



$$\nabla g = \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$$

$$\tilde{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \to \nabla g(\tilde{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Optimisation avec contrainte égalité

Soit $(P2)$ le problème d'optimisation suivant

$$\min_{x \in \mathcal{X}} f(x)$$

$$h(x) = 0$$

où f et h sont différentiables sur $\mathcal{X}$.

Fonction lagrangienne L

La fonction L définie par

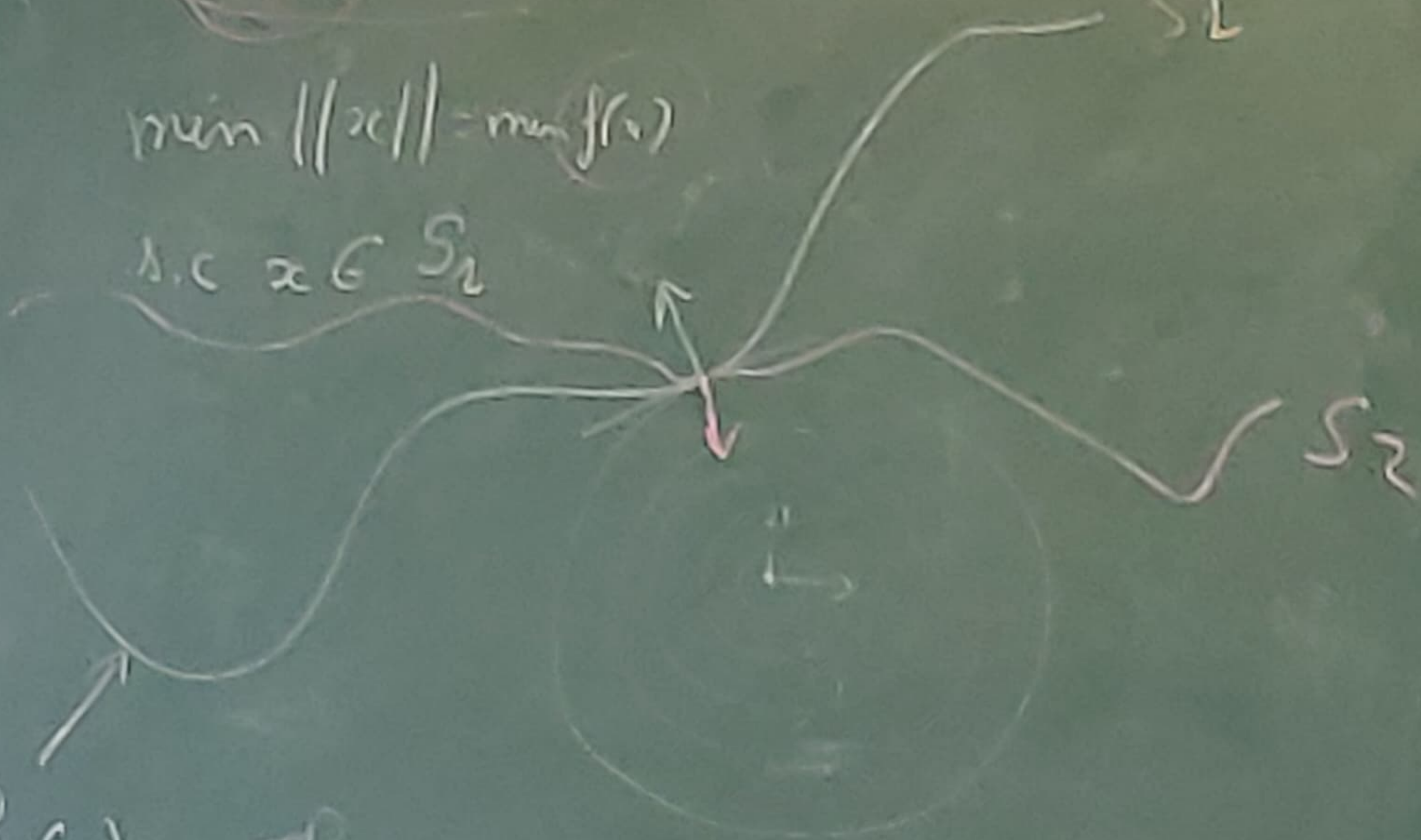
$$L(x, \lambda) = f(x) + \lambda^t h(x)$$

est appelée lagrangien ou fonction lagrangienne du problème $(P2)$.
avec λ multiplicateurs de Lagrange.

$h_1(x) = 0$

$\min \|x\| = \min f(x)$

s.c $x \in S_1$



$h_1(x) \to \nabla h_1(x)$

$\nabla f \pm \lambda \nabla h = 0$

Théorème (condition nécessaire de Karush-Kuhn-Tucker)

Soit x^* un minimum local du problème $(P2)$ alors

$$\begin{cases} \nabla_x L(x^*, \lambda) & = & 0 \\ \nabla_\lambda L(x^*, \lambda) & = & 0 \end{cases}$$

On peut écrire ce système :

$$\nabla_{x,\lambda} L(x^*, \lambda) = 0$$

ou encore plus simplement

$$\nabla L(x^*, \lambda) = 0$$

Si le système comporte n inconnus et p contraintes, le théorème précédent nous fournit $n + p$ équations (et on a $n + p$ contraintes d'égalité).

Théorème

Soit x^* un minimum local du problème $(P2)$:
Si les contraintes sont linéairement indépendantes en x^* alors il existe un vecteur unique λ^* tel que $\nabla L(x^*, \lambda) = 0$.
Si f et h sont deux fois différentiables alors

$$y^t.\nabla^2_{xx} L(x^*, \lambda^*).y \geq 0, \forall y \in \mathcal{D}(x^*)$$

avec $\mathcal{D}$ cône des directions admissibles.
De plus on a :

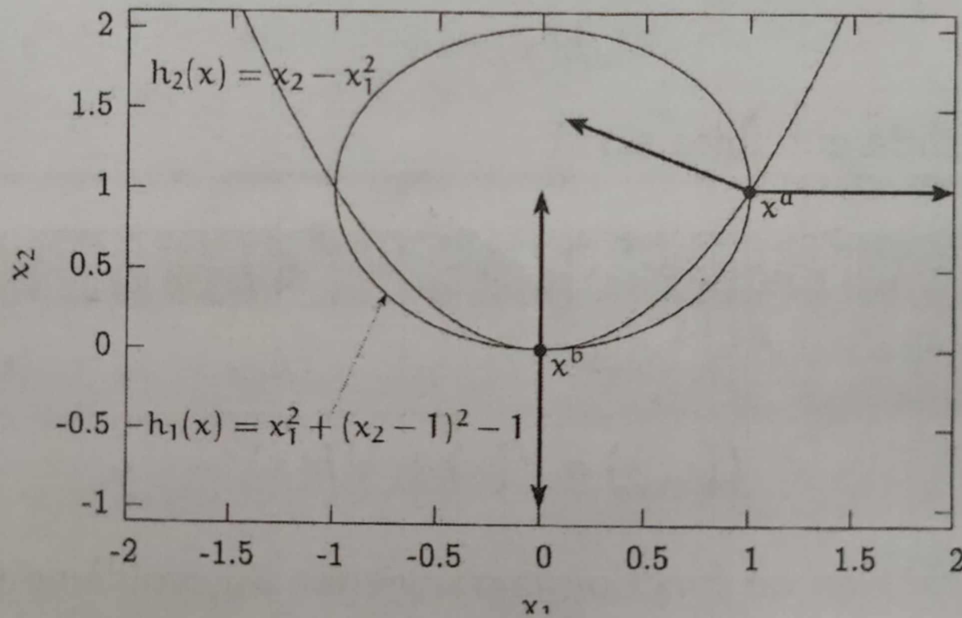
$$\lambda^* = -(\nabla h(x^*)^t.\nabla h(x^*))^{-1}.\nabla h(x^*)^t.\nabla f(x^*)$$

(tableau)

Exemple

$$f(x_1, x_2) = x_1 + x_2$$

$$\text{s.c. } x_1^2 + (x_2 - 1)^2 = 1 \text{ et } -x_1^2 + x_2 = 0$$



$$L = x_1 + x_2 + \lambda_1 \big[\overbrace{x_1^2 + (x_2-1)^2 - 1}^{h_1(x)=0}\big] + \lambda_2 \big[\overbrace{-x_1^2 + x_2}^{h_1(x)=0}\big]$$

$$\begin{cases} \nabla_x L = 0 \\ \nabla_\lambda L = 0 \end{cases} \Rightarrow \begin{cases} 1 + 2\lambda_1 x_1 - 2\lambda_2 x_1 = 0 \\ 1 + 2\lambda_1 [x_2 - 1] + \lambda_2 = 0 \\ x_1^2 + (x_2^2 + 1 - 2x_2) - 1 = 0 \\ -x_1^2 + x_2 = 0 \Rightarrow x_2 = x_1^2 \end{cases}$$

$$x_2(x_2-1)=0$$

① $$\bar{X} = \begin{pmatrix} x_1 = \pm 1 \\ x_2 = 1 \end{pmatrix}$$

$$\xrightarrow{x_1 = 1} \quad \lambda = \begin{pmatrix} -3/2 \\ -1 \end{pmatrix}$$

$$\xrightarrow{x_1 = -1} \quad \lambda = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}$$

$$\nabla h_1 = \begin{bmatrix} 2x_1 \\ 2(x_2-1) \end{bmatrix} \rightarrow \nabla h \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{bmatrix} 2 & -2 \\ 0 & 1 \end{bmatrix}$$

$$\nabla h_2 = \begin{bmatrix} -2x_1 \\ 1 \end{bmatrix}$$

$$\bar{X} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \longrightarrow \nabla h\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{bmatrix} -2 & 2 \\ 0 & 1 \end{bmatrix}$$

direct admissible $\quad d^t \cdot \nabla h = 0$

$$d = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \longrightarrow \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} -2y_1 & 2y_1 + y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$\boxed{y_1 = y_2 = 0}$$

$$\nabla^2_{xx} L = \left[\begin{array}{c|c} 2(\lambda_1 - \lambda_2) & 0 \\ \hline 0 & 2\lambda_1 \end{array}\right]$$

$$\nabla^2_{xx} L = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & -1 \end{array}\right]$$

$$y^T \nabla^2_{xx} L \, y = 0 \quad \forall y \in D$$

Exemple 4

$$f(x) = -x_1x_2 - x_1x_3 - x_2x_3$$

$$\text{s.c. } x_1 + x_2 + x_3 = 3.$$

Les conditions de KKT, se prêtent très bien aux fonctions objectifs quadratiques et aux contraintes linéaires

$$\min f(x) = \frac{1}{2} x^t Q x + g^t x$$

$$\text{s.c. } Ax = b.$$

$$\begin{aligned} \nabla_x f(x) &= Qx + g \\ \nabla_x h(x) &= -A^t \end{aligned}$$

$$\begin{aligned} \nabla_x L(x, \lambda) &= Qx + g - A^t \lambda &= 0 \\ \nabla_\lambda L(x, \lambda) &= b - A.x &= 0. \end{aligned}$$

On résout un système linéaire.
Donner les valeurs de x^* et λ^*.

Optimisation avec contrainte égalité et inégalité

Soit $(P3)$ le problème d'optimisation suivant

$$\min_{x \in \mathcal{X}} \; f(x)$$

$$h_i(x) = 0 \qquad i = 1, \ldots, m$$

$$g_j(x) \leq 0 \qquad j = 1, \ldots, p$$

où f, g et h sont différentiables sur $\mathcal{X}$.

Le lagrangien s'écrit désormais

$$L(x, \lambda, \mu) = f(x) + \lambda^t h(x) + \mu^t g(x)$$

Théorème(conditions nécessaires de Karush-Kuhn-Tucker)

Soit x^* un minimum local du problème $(P3)$:
Si les contraintes sont linéairement indépendantes en x^* alors il existe un vecteur unique λ^* et un vecteur unique μ^* tels que :

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

$$\begin{array}{ll} \mu_j \geq 0 & j = 1, \ldots, p \\ \mu_j g_j(x^*) = 0 & j = 1, \ldots, p \end{array}.$$

Si f et h sont deux fois différentiables alors

$$y^t.\nabla^2_{xx} L(x^*, \lambda^*, \mu^*).y \geq 0, \quad \forall y \neq 0 \text{ tel que}$$

$$\begin{array}{ll} y^t.\nabla h_i(x^*) = 0 & i = 1, \ldots, m \\ y^t.\nabla g_i(x^*) = 0 & i = 1, \ldots, p \mid g_i(x^*) = 0 \ (\textit{contraintes actives}) \end{array}.$$

Théorème(conditions suffisantes)

Soient f et g des fonctions deux fois différentiables.

Soient x^*, λ^* et μ^* tels que :

$$\begin{array}{ll}
\nabla_x L(x^*, \lambda^*, \mu^*) = 0 & \\
h(x^*) = 0 & \\
g(x^*) \leq 0 & \\
\mu^* \geq 0 & j = 1, \ldots, p \\
\mu_j^* g_j(x^*) = 0 & j = 1, \ldots, p \\
\mu_j^* > 0 & \forall \textit{ contrainte } j \textit{ active}
\end{array}$$

$$y^t.\nabla^2_{xx} L(x^*, \lambda^*, \mu^*).y > 0, \quad \forall y \neq 0 \text{ tel que}$$

$$y^t.\nabla h_i(x^*) = 0 \quad i = 1, \ldots, m$$

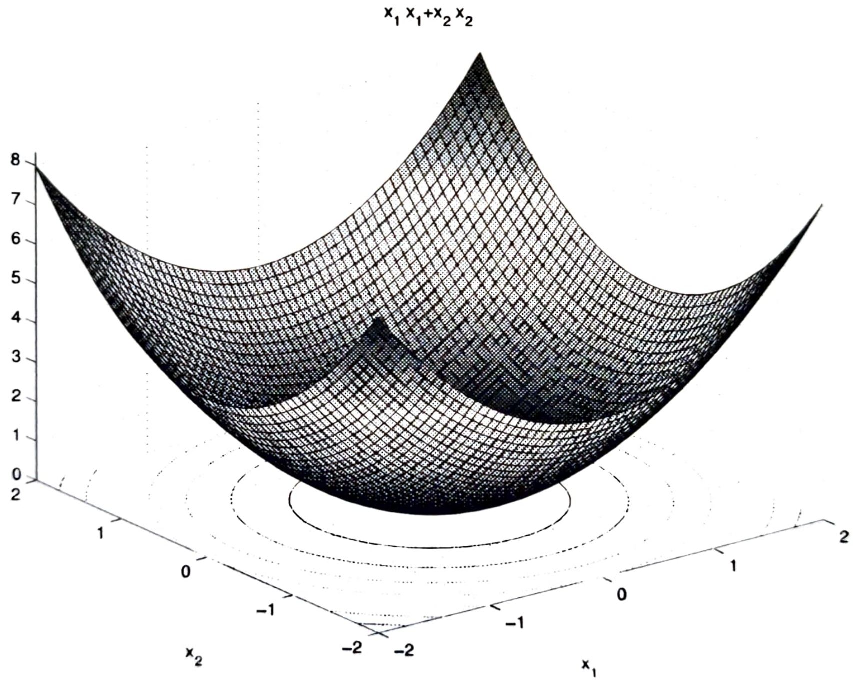
$$y^t.\nabla g_i(x^*) = 0 \quad i = 1, \ldots, p \qquad \textit{tel que } \; g_i(x^*) = 0$$

Alors x^* est un minimum local.

Exemple 5

$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$\text{s.c. } 2x_1 + x_2 \leq -4.$$



Exemple 5

$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$\text{s.c. } 2x_1 + x_2 \leq -4.$$