

IFT 3515  
Fonctions à plusieurs variables  
Optimisation avec contraintes  
Conditions d'optimalité – 2<sup>e</sup> ordre

Fabian Bastin  
DIRO  
Université de Montréal

# Contexte

Notes partiellement basées sur

<http://www.numerical.rl.ac.uk/people/nimg/course/lectures/raphael/lectures/lec10slides.pdf>

Nous considérons à nouveau le problème d'optimisation nonlinéaire (PNL) général

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{t.q.} \quad & g_i(x) = 0, \quad i \in \mathcal{E} \\ & g_i(x) \leq 0, \quad i \in \mathcal{I} \end{aligned}$$

Hypothèse :  $f$  et  $g_i$  ( $i \in \mathcal{E} \cup \mathcal{I}$ ) sont des fonctions deux fois continûment différentiables.

## Chemin de sortie réalisable : rappel

### Définition

Soit  $x^* \in \mathbb{R}^n$  un point réalisable pour le problème de PNL et définissons  $x \in C^2((-\epsilon, \epsilon), \mathbb{R}^n)$  un chemin tel que

$$x(0) = x^*$$

$$d := \frac{d}{dt}x(0) \neq 0$$

$$g_i(x(t)) = 0, \quad i \in \mathcal{E}, \quad t \in (-\epsilon, \epsilon)$$

$$g_i(x(t)) \leq 0, \quad i \in \mathcal{I}, \quad t \in [0, \epsilon)$$

$x(t)$  est *un chemin de sortie réalisable* à partir de  $x^*$  et le vecteur tangent  $d = \frac{d}{dt}x(0)$  est *une direction de sortie réalisable* à partir de  $x^*$ .

## Optimalité au second ordre

L'analyse d'optimalité au second ordre est basée sur l'observation suivante : si  $x^*$  est un minimiseur local du problème de PNL et  $x(t)$  est un chemin de sortie réalisable à partir de  $x^*$ , alors  $x^*$  doit aussi être un minimiseur local pour le problème d'optimisation sous contraintes avec une variable

$$\min_t f(x(t)) \text{ t.q. } t \geq 0.$$

Analysons d'avantage les directions de sortie réalisables à partir de  $x^*$ . La définition implique

$$d^T \nabla g_i(x^*) = \frac{d}{dt} g_i(x(t))|_{t=0} = \begin{cases} \frac{d}{dt} 0 = 0 & i \in \mathcal{E} \\ \lim_{t \rightarrow 0^+} \frac{g_i(x(t)) - 0}{t} & i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{cases}$$

## Conditions nécessaires

Dès lors, les conditions suivantes sont nécessaires pour que  $d \in \mathbb{R}^n$  soit une direction de sortie réalisable à partir de  $x^*$  :

$$d \neq 0$$

$$d^T \nabla g_i(x^*) = 0, \quad i \in \mathcal{E}$$

$$d^T \nabla g_i(x^*) \leq 0, \quad i \in \mathcal{A}(x^*) \cap \mathcal{I}$$

## Conditions suffisantes

D'autre part, si la LICQ tient en  $x^*$ , du Lemme sur la LICQ, nous avons l'existence d'un chemin de sortie réalisable  $x(t)$  à partir de  $x^*$  tel que

$$x(0) = x^*$$

$$\frac{d}{dt}x(0) = d$$

$$g_i(x(t)) = td^T \nabla g_i(x^*), \quad i \in \mathcal{E} \cup \mathcal{A}(x^*), \quad t \in (-\epsilon, \epsilon)$$

Dès lors, lorsque la LICQ tient, les conditions nécessaires sur  $d$  sont aussi suffisantes, et fournissent par conséquent une caractérisation exacte pour que  $d$  soit une direction de sortie réalisable à partir de  $x^*$ .

## Conditions nécessaires d'optimalité au second ordre

Soit  $x^*$  un minimum local tel que la LICQ tienne. Les conditions KKT indiquent qu'il existe un vecteur de multiplicateurs de Lagrange  $\lambda^*$  tel que

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

$$g_i(x^*) = 0, \quad i \in \mathcal{E}$$

$$g_i(x^*) \leq 0, \quad i \in \mathcal{I}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i \in \mathcal{I}$$

$$\lambda_i^* \geq 0, \quad i \in \mathcal{I}$$

où

$$L(x^*, \lambda) = f(x^*) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* g_i(x^*).$$

## Conditions nécessaires d'optimalité au second ordre

Soit  $x(t)$  un chemin de sortie réalisable à partir de  $x^*$  avec la direction de sortie  $d$ , et considérons le problème restreint

$$\begin{aligned} \min f(x(t)) \\ \text{t.q. } t \geq 0. \end{aligned}$$

Puisque  $x^*$  est un minimiseur local du problème de PNL,  $t = 0$  doit aussi être un minimiseur local du problème restreint.

Du développement de Taylor autour de  $t = 0$  et des conditions KKT,

$$\begin{aligned} f(x(t)) &= f(x(0)) + (t - 0) \frac{d}{dt} f(x(t))|_{t=0} + O(t^2) \\ &= f(x^*) + t \left( \frac{d}{dt} x(t)|_{t=0} \right)^T \nabla_x f(x(t))|_{t=0} + O(t^2) \\ &= f(x^*) + t d^T \nabla f(x^*) + O(t^2) \\ &= f(x^*) - t \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* d^T \nabla g_i(x^*) + O(t^2) \end{aligned}$$



## Conditions nécessaires d'optimalité au second ordre

Nous souhaiterions dès lors montrer que pour  $t \geq 0$  assez petit,

$$-t \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* d^T \nabla g_i(x^*) + O(t^2) \geq 0$$

Notons que

$$\lambda_i^* d^T \nabla g_i(x^*) = 0, \quad i \in \mathcal{E} \cup (\mathcal{I} \setminus \mathcal{A}(x^*))$$

de sorte que ces termes peuvent être omis.

Mais qu'en est-il des indices  $j \in \mathcal{A}(x^*) \cap \mathcal{I}$ ? Nous savons déjà que  $\lambda_i^* d^T \nabla g_i(x^*) \leq 0, \quad i \in \mathcal{A}(x^*)$ .

Nous allons distinguer deux cas.

## Conditions nécessaires d'optimalité au second ordre

Cas 1 : il existe un indice  $j \in \mathcal{A}(x^*)$  tel que  $\lambda_j^* d^T \nabla g_j(x^*) < 0$ .

Alors, pour tout  $0 < t \ll 1$ ,

$$\begin{aligned} f(x(t)) &= f(x^*) - t \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* d^T \nabla g_i(x^*) + O(t^2) \\ &\geq f(x^*) - t \lambda_j^* d^T \nabla g_j(x^*) + O(t^2) \\ &> f(x^*) \end{aligned}$$

Dès lors, dans ce cas,  $f$  est strictement croissante le long du chemin  $x(t)$  pour  $t$  suffisamment petit, même si  $\frac{d^2}{dt^2} f(x(0))$  était négatif.

## Conditions nécessaires d'optimalité au second ordre

Cas 2 :

$$\lambda_i^* d^T \nabla g_i(x^*) = 0, \quad i \in \mathcal{E} \cup \mathcal{I}.$$

Dans ce cas, l'argument précédent ne peut garantir que  $f$  est localement croissante le long du chemin  $x(t)$ . Nous savons seulement que  $\frac{d}{dt}f(x(0)) = 0$ , c'est-à-dire que  $x^*$  est un point stationnaire du problème restreint.

Mais il pourrait très bien s'agir d'un maximiseur local ou d'un point selle du problème restreint !

Les dérivées secondes  $\frac{d^2}{dt^2}f(x(0))$  déterminent à présent si  $t = 0$  est un minimiseur local du problème restreint, conduisant à une information supplémentaire nécessaire.

## Conditions nécessaires d'optimalité au second ordre

Nous noterons

$$N^+ = \left\{ d \neq 0 \left| \begin{array}{ll} d^T \nabla g_i(x^*) = 0, & i \in \mathcal{E} \cup \{j \in \mathcal{A}(x^*) \cap \mathcal{I} \mid \lambda_j^* > 0\} \\ d^T \nabla g_i(x^*) \leq 0, & \{i \in \mathcal{A}(x^*) \cap \mathcal{I} \mid \lambda_i^* = 0\} \end{array} \right. \right\}$$

Dès lors,  $d \in N^+$  implique  $\lambda_i^* d^T \nabla g_i(x^*) = 0$ ,  $i \in \mathcal{E} \cup \mathcal{I}$ .

### Théorème (Conditions nécessaires d'optimalité au second ordre)

*Soit  $x^*$  un minimiseur local du PNL où la LICQ tient. Soit  $\lambda^* \in \mathbb{R}^m$  un vecteur de multiplicateurs de Lagrange tel que  $(x^*, \lambda^*)$  satisfaisant les conditions KKT. Alors nous avons*

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0$$

*pour tout  $d \in N^+$ .*

## Preuve

Soit  $d \in N^+$  et soit  $x(\cdot) \in C^2((-\epsilon, \epsilon), \mathbb{R}^n)$  un chemin de sortie réalisable à partir de  $x^*$  correspondant à  $d$ . Alors

$$\begin{aligned} L(x(t), \lambda^*) &= f(x(t)) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* g_i(x(t)) \\ &= f(x(t)) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* t d^T \nabla g_i(x^*) \\ &= f(x(t)) \end{aligned}$$

Le développement de Taylor de  $L(x(t), \lambda^*)$  autour de  $t = 0$  implique

$$\begin{aligned} f(x(t)) &= L(x^*, \lambda^*) + t \nabla_x L(x^*, \lambda^*)^T d \\ &\quad + \frac{t^2}{2} \left( d^T \nabla_{xx}^2 L(x^*, \lambda^*) d + \nabla_x L(x^*, \lambda^*)^T \frac{d^2}{dt^2} x(t) \Big|_{t=0} \right) \\ &\quad + O(t^3) \end{aligned}$$

## Preuve

Les conditions KKT impliquent

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

Dès lors

$$f(x(t)) = L(x^*, \lambda^*) + \frac{t^2}{2} d^T \nabla_{xx}^2 L(x^*, \lambda^*) d + O(t^3)$$

et par les conditions d'optimalité au premier ordre (dans le cas convexe, la dualité forte),  $L(x^*, \lambda^*) = f(x^*)$

Si  $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d < 0$  alors pour  $t$  suffisamment petit,  $f(x(t)) < f(x^*)$ , contredisant l'hypothèse que  $x^*$  est un minimiseur local.

Dès lors, nous devons avoir

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0$$

# Conditions suffisantes d'optimalité au second ordre

## Théorème (Conditions suffisantes d'optimalité au second ordre)

*Soit  $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$  satisfaisant les conditions KKT, et sous une hypothèse de qualification de contrainte, si*

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0$$

*pour toute direction  $d \in N^+$ . Alors  $x^*$  est un minimiseur local strict.*

## Preuve

Supposons par contradiction que  $x^*$  n'est pas un minimiseur local. Il existe alors une séquence de points réalisables  $\{x_k\}$ ,  $k = 1, \dots$ , telle que  $x_k \neq x^*$  pour tout  $k$ ,  $\lim_{k \rightarrow \infty} x_k = x^*$  et

$$f(x_k) \leq f(x^*), \quad \forall k \in \mathbb{N}.$$

La séquence

$$\frac{x_k - x^*}{\|x_k - x^*\|}$$

tient sur la sphère unité, qui est un ensemble compact, et dès lors nous pouvons extraire une sous-séquence  $\{x_{k_i}\}$ ,  $i \in \mathbb{N}$ ,  $k_i < k_j$  ( $i < j$ ), tel qu'il existe une direction limite  $d := \lim_{k \rightarrow \infty} d_{k_i}$  existe, avec

$$d_{k_i} = \frac{x_{k_i} - x^*}{\|x_{k_i} - x^*\|}$$

Puisque  $d$  tient sur la sphère unité, nous avons  $d \neq 0$ . En ne considérant que la sous-séquence ainsi construite, nous pouvons supposer sans perte de généralité que  $k_i = i$ .



## Preuve

Vérifions que  $d \in N^+$ . Considérons tout d'abord les contraintes d'égalité.  $\forall i \in \mathcal{E}$ , en utilisant le développement de Taylor, nous avons

$$\begin{aligned} 0 &= \frac{g_i(x_k)}{\|x_k - x^*\|} \\ &= \frac{1}{\|x_k - x^*\|} \left( g_i(x^*) + \|x_k - x^*\| \nabla g_i(x^*)^T d_k + O(\|x_k - x^*\|^2) \right) \\ &= \nabla g_i(x^*)^T d_k + O(\|x_k - x^*\|) \end{aligned}$$

En faisant tendre  $k$  vers l'infini, nous obtenons

$$\nabla g_i(x^*)^T d_k = 0$$

## Preuve

$\forall i \in \mathcal{A}(x^*) \cap \mathcal{I}$ , en utilisant le développement de Taylor, nous avons de manière similaire

$$\begin{aligned} 0 &\geq \frac{g_i(x_k)}{\|x_k - x^*\|} \\ &= \frac{1}{\|x_k - x^*\|} \left( g_i(x^*) + \|x_k - x^*\| \nabla g_i(x^*)^T d_k + O(\|x_k - x^*\|^2) \right) \\ &= \nabla g_i(x^*)^T d_k + O(\|x_k - x^*\|) \end{aligned}$$

En faisant tendre  $k$  vers l'infini, nous obtenons

$$\nabla g_i(x^*)^T d \leq 0$$

Ainsi,  $d$  est une direction de sortie réalisable. Montrons que  $d \in N^+$ .

## Preuve

Si  $d \notin N^+$ , alors  $\exists j \in \mathcal{A}(x^*) \cap \mathcal{I}$  tel que

$$\lambda_j^* \nabla g_j(x^*)^T d < 0$$

En vertu du développement de Taylor,

$$\begin{aligned} \lambda_j^* g_j(x_k) &= \lambda_j^* g_j(x^*) + \lambda_j^* \|x_k - x^*\| \nabla g_j(x^*)^T d_k + O(\|x_k - x^*\|^2) \\ &= \lambda_j^* \|x_k - x^*\| \nabla g_j(x^*)^T d_k + O(\|x_k - x^*\|^2) \end{aligned}$$

Dès lors

$$\begin{aligned} L(x_k, \lambda^*) &= f(x_k) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* g_i(x_k) \\ &\leq f(x_k) + \lambda_j^* g_j(x_k) \\ &= f(x_k) + \lambda_j^* \|x_k - x^*\| \nabla g_j(x^*)^T d_k + O(\|x_k - x^*\|^2) \end{aligned}$$

## Preuve

D'autre part, le développement de Taylor de  $L(x_k, \lambda^*)$  autour de  $(x^*, \lambda^*)$  donne

$$\begin{aligned} L(x_k, \lambda^*) &= L(x^*, \lambda^*) + \|x_k - x^*\| d_k^T \nabla_x L(x^*, \lambda^*) \\ &\quad + \frac{\|x_k - x^*\|^2}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^3) \end{aligned}$$

De part les conditions KKT,  $\nabla_x L(x^*, \lambda^*) = 0$  et  $L(x^*, \lambda^*) = f(x^*)$ .  
Dès lors, nous obtenons

$$L(x_k, \lambda^*) = f(x^*) + \frac{\|x_k - x^*\|^2}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^3)$$

En combinant ces expressions,

$$\begin{aligned} f(x_k) &\geq L(x_k, \lambda^*) - \lambda_j^* \|x_k - x^*\| \nabla g_j(x^*)^T d_k + O(\|x_k - x^*\|^2) \\ &= f(x^*) - \lambda_j^* \|x_k - x^*\| \nabla g_j(x^*)^T d_k + O(\|x_k - x^*\|^2) \end{aligned}$$

et pour  $k$  assez grand,  $f(x_k) > f(x^*)$  contredisant nos hypothèses de travail.

## Preuve

Ainsi  $d \in N^+$ , et des hypothèses du théorème, nous avons

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0.$$

D'autre part, des conditions KKT et  $d \in N^+$ ,

$$\nabla f(x^*)^T d = -d^T \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla g_i(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i (-d^T \nabla g_i(x^*)) = 0$$

De plus, de nos hypothèses de départ et des conditions KKT,

$$f(x^*) \geq f(x_k) \geq f(x_k) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* g_i(x_k) = L(x_k, \lambda^*)$$

et, reprenant le développement de  $L(x_k, \lambda^*)$

$$f(x^*) \geq f(x^*) + \frac{\|x_k - x^*\|^2}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^3)$$

## Preuve

Il s'ensuit

$$d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k \leq O(\|x_k - x^*\|)$$

en passant à la limite comme  $k \rightarrow \infty$

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \leq 0$$

Ceci contredit l'hypothèse du théorème

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0$$

et donc l'hypothèse d'existence de la séquence  $\{x_k\}$  ne tient pas.  
Dès lors,  $x^*$  est minimiseur local.

## Complémentarité stricte

Il est courant, et commode, de supposer que la condition de complémentarité stricte s'applique. Celle-ci requiert que le multiplicateur de Lagrange ne peut s'annuler pour une contrainte active, i.e.

$$\lambda_j^* \neq 0 \quad \forall j \in \mathcal{A}(x^*) \cap \mathcal{I}$$

La complémentarité stricte permet de réexprimer  $N^+$  de manière très commode. Rappelons que

$$N^+ = \left\{ d \neq 0 \left| \begin{array}{ll} d^T \nabla g_i(x^*) = 0, & i \in \mathcal{E} \cup \{j \in \mathcal{A}(x^*) \cap \mathcal{I} \mid \lambda_j^* > 0\} \\ d^T \nabla g_i(x^*) \leq 0, & \{i \in \mathcal{A}(x^*) \cap \mathcal{I} \mid \lambda_i^* = 0\} \end{array} \right. \right\}$$

sous la complémentarité stricte, nous avons

$$N^+ = \left\{ d \neq 0 \mid d^T \nabla g_i(x^*) = 0, i \in \mathcal{A}(x^*) \right\}$$

## Complémentarité stricte

Autrement dit,  $d$  appartient au noyau du Jacobien des contraintes actives :

$$d \in \text{Ker}(J), \quad J = (\nabla g_i(x^*)^T, i \in \mathcal{A}(x^*))$$

Comme

$$d \in N^+ \Rightarrow \lambda_i^* d^T \nabla g_i(x^*) = 0, \quad i \in \mathcal{E} \cup \mathcal{I},$$

nous avons également

$$d \in N^+ \Leftrightarrow \lambda_i^* d^T \nabla g_i(x^*) = 0, \quad i \in \mathcal{E} \cup \mathcal{I},$$