PROSEMINAIRE 2009
Optimisation Numérique : Analyse de Convergence des Méthodes Quasi-Newton BFGS & SR1
Chapitre 8.4*

MUELLER Didier
Bel-Horizon 1
1860 Aigle

26 Novembre 2009

Table des matières

*Jorge Nocedal and Stephen J. Wright : Numerical Optimization, Springer Series in Operations Research, Springer 1999.

1 Introduction

Dans ce proséminaire, nous allons nous intéresser à la convergence global et local des méthodes Quasi-Newton BFGS et SR1. La partie concernant la méthode BFGS sera beaucoup plus détaillée car son étude est plus général et aisée que la méthode SR1.
Il est important de noter que ces deux méthodes sont connues pour leur robustesse dans la pratique. Malgré cela, nous ne sommes pas capable d'établir vraiment la convergence global pour n'importe quelle fonction « objectif »non-linéaire. En effet nous ne pouvons pas prouver que les itérées de ces méthodes approche d'un point stationnaire quelque soit le point de départ de notre algorithme et quelque soit l'approximation initiale de l'hessienne. En fait, à l'heure actuelle, nous ne savons pas encore si ces méthodes possèdent ces propriétés. Ainsi dans notre analyse nous allons supposer que notre fonction est convexe ou que les itérées satisfont certaines propriétés.

2 Méthode BFGS

2.1 Rappel

Toutes les méthodes Quasi-newton fonctionnent sur la même idée de base. Ces méthodes fonctionnent sur le modèle des méthode de Newton. La différence réside dans le fait qu'au lieu de prendre la véritable hessienne (très longue à calculer), nous cherchons une approximation de cette dernière qui ne dépende que du gradient de la fonction. Voici quelques résultats importants pour la compréhension des différentes méthodes.
Nous avons pour chaques x_k :

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p \tag{1}$$

Comme approximation de l'hessienne, nous avons B_k defini comme suit :

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \tag{2}$$

Cette matrice à pour propriétés d'une part d'être symétrique et définie positive et d'autre part de satisfaire l'équation de la secante :

$$B_{k+1} s_k = y_k \tag{3}$$

où $s_k = x_{k+1} - x_k$ et $y_k = \nabla f_{k+1} - \nabla f_k = \bar{G}_k s_k$
avec la matrice hessienne moyenne

$$\bar{G}_k = \left[\int_0^1 \nabla^2 f(x_k + \tau s_k) d\tau \right] \tag{4}$$

2.2 Convergence globale de la méthode BFGS

Pour étudier la convergence de cette méthode nous devons au préalable poser un certain nombre d'hypothèse sur la fonction « objectif ». Pour la suite nous noterons $G(x) := \nabla^2 f(x)$

Hypothèse 1.

1. la fonction f est deux fois continûment différentiable.

2. l'ensemble de niveau

$$\Omega = \{x \in \mathbb{R} \mid f(x) \leq f(x_0)\} \tag{5}$$

est convexe de plus $\exists M, m > 0$, telles que

$$m\|z\|^2 \leq z^T G(x) z \leq M\|z\|^2 \qquad \forall z \in \mathbb{R} \& x \in \Omega \tag{6}$$

Pour la suite, nous allons à l'aide de (6) faire quelques petites observations.

$$m \leq \frac{s_k^T \bar{G}_k s_k}{s_k^T s_k} = \frac{y_k^T s_k}{s_k^T s_k} \tag{7a}$$

$$M \geq \frac{z_k^T \bar{G}_k z_k}{z_k^T z_k} = \frac{y_k^T y_k}{y_k^T s_k} \tag{7b}$$

où

$$z_k = \bar{G}_k^{-} \frac{1}{2} s_k \tag{8}$$

Pour la preuve du théorème sur la convergence globale, nous aurons besoin de notions déjà vues dans d'autres proséminaires. Nous allons donc nous limiter à l'énoncer de ces différents résultats.
Tout d'abord les conditions de Wolfe :

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k, \tag{9a}$$

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k, \tag{9b}$$

Théorème 2 (Condition de Zoutendijk)**.**
Soit une itération de la forme $x_{k+1} = x_k + \alpha_k p_k$,où p_k est une direction de descente et α_k satisfait les conditions de wolfe (9a,9b).On suppose que f est borné inférieurement dans $\mathbb{R}^n$ et que f est continûment différentiable sur un ouvert N qui contient Ω (5). Il faut encore que ∇f soit lipschiptz sur N.Alors :

$$\sum_{k \geq 0} \cos^2 \theta_k \|\nabla f_k\|^2 \leq \infty \tag{10}$$

Une conclusion importante de ce théorème est que la suite $\cos^2 \theta_k \|\nabla f_k\|^2 \to 0$. Donc si nous pouvons montrer que la suite $\cos^2 \theta_k$ ne converge pas vers 0, alors nous avons $\|\nabla f_k\| \to 0$. Ce qui signifie que notre algorithme converge globalement.
Nous somme à présent prêt pour le théorème suivant

Théorème 3 (Convergence global de BFGS)**.**
Soit B_0 la matrice symétrique definie positive(SDP)associée au point de départ x_0 de l'algorithme BFGS qui verifie toutes les hypothèses(Hypothèse 1.)
alors la suite $\{x_k\}$ générée par l'algorithme converge vers x^*.

Démonstration.
Posons :

$$m_k = \frac{y_k^T s_k}{s_k^T s_k}, \qquad M_k = \frac{y_k^T y_k}{y_k^T s_k} \tag{11}$$

On remarque alors avec (7a)et (7b) que

$$m_k \geq m, \qquad M_k \leq M. \tag{12}$$

de plus,

$$\begin{aligned} tr(B_{k+1}) &= tr(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}) \\ &= tr(B_k) - tr(\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}) + tr(\frac{y_k y_k^T}{y_k^T s_k}) \\ &= tr(B_k) - \frac{1}{s_k^T B_k s_k} tr(B_k s_k s_k^T B_k) + \frac{1}{y_k^T s_k} tr(y_k y_k^T) \\ &= tr(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k} \end{aligned}$$

On peut encore montrer (preuve purement technique) que :

$$\det(B_{k+1}) = \det(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k} \tag{13}$$

Posons encore :

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\|s_k\|\,\|B_k s_k\|}, \qquad q_k = \frac{s_k^T B_k s_k}{s_k^T s_k} \tag{14}$$

En fait, (comme nous avons déjà pu le voir) θ_k est l'angle entre s_k (direction de recherche) et $B_k s_k$ (steepest descente). Il est important de noté le résultats suivant.

$$\frac{q_k}{\cos^2\theta_k} = \frac{\|s_k\|\,\|B_k s_k\|^2}{s_k^T B_k s_k} \tag{15}$$

ainsi que : (on utilise(11) et (13))

$$\det(B_{k+1}) = \det(B_k)\frac{m_k}{q_k} \tag{16}$$

Nous allons à présent definir une fonction dont la non-negativité nous permettra de parvenir à une contradiction à l'aide de Zoutendijk (2). soit

$$\Psi(B) = tr(B) - \ln(\det(B)) \tag{17}$$

On montre aisément part le fait que les B_k sont SDP que $\Psi(B) \geq 0$. Nous allons donc à présent l'appliquer à B_{k+1}.

$$\begin{aligned} \Psi(B_{k+1}) &= tr(B_{k+1}) - \ln(\det(B_{k+1})) \\ &= tr(B_k) - \frac{q_k}{\cos^2\theta_k} + M_k - \ln(\det(B_k)\frac{m_k}{q_k}) \\ &= tr(B_k) - \ln(\det(B_k)) - \ln(m_k) + \ln(q_k) - \frac{q-k}{\cos^2\theta_k} + M_k \\ &= \Psi(B_k) + (M_k - \ln(m_k) - 1) + \left[1 - \frac{q_k}{\cos^2\theta_k} + \ln(\frac{q_k}{\cos^2\theta_k})\right] + \ln(\cos^2\theta_k) \end{aligned}$$

On montre facilement la non-positivité des fonctions de la forme :

$$h(t) = 1 - t + \ln(t) \tag{18}$$

On remarque que c'est justement la forme du terme $[\cdots]$ dans (18). On en déduit les inégalités suivantes

$$\begin{aligned}
0 < \Psi(B_{k+1}) &\leq \Psi(B_k) - (M - \ln(m) - 1) + \ln(\cos^2\theta_k) \\
\text{on continue récursivement} & \\
&\leq \Psi(B_1) + k\,(M - \ln(m) - 1) + \sum_{j=1}^{k} \ln(\cos^2\theta_j)
\end{aligned}$$

Pour la suite, On pose $c = (M - \ln(m) - 1)$. De plus, on peut s.r.l.g. supposer $c \geq 0$.

Nous allons maintenant mettre en relation ces différents résultats avec ceux de prescédents proséminaires.
Avec la Condition de Zoutendijk (2), nous savons que la convergence de $\|\nabla f_k\| \to 0$ n'est pas assurée si $\cos\theta_k \to 0$.Nous allons donc procéder par contraposition et supposer que

$$\cos\theta_k \to 0 \tag{19}$$

alors, $\exists k_1$ tel que $\forall j > k_1$:

$$\ln(\cos^2\theta_j) < -2c \tag{20}$$

Nous obtenons donc en replaçant ces résultats dans (19)

$$\begin{aligned}
0 < \Psi(B_{k+1}) &< \Psi(B_1) + ck + \sum_{j=1}^{k_1} \ln(\cos^2\theta_j) + \sum_{j=k_1+1}^{k} (-2c) \\
&= \Psi(B_1) + \sum_{k_1}^{j=1} \ln(\cos^2\theta_j) + 2ck_1 - ck
\end{aligned}$$

Cette dernière expression (21)et n' est constituée que d'éléments bornés à l'exeption de $-ck$. Il s'en suit que pour un k suffisamment grand nous parvenons à une contradiction. Nous avons donc montrer la non-convergence de $\cos\theta_k$ vers 0. Ce qui implique par Zoutendijk (2) que la suite $\|\nabla f_k\| \to 0$ et ainsi sous les hypothèses générales (1) que $x_k \to x^*$. □

On peut généraliser ce théorème à l'entier des classes restreintes de Broyden, à l'exeption de la méthode DFP.On peut en outre montrer (sans preuve) que la suite $\|x_k - x^*\|$ converge suffisamment vite pour que

$$\sum_{k\geq 1} \|x_k - x^*\| < \infty \tag{21}$$

2.3 Convergence superlinéaire de BFGS

Dans ce chapitre, nous allons nous intéresser à la superlinéarité de la convergence de la méthode BFGS. Pour ce faire, nous avons besoin d'une part de notions déjà rencontrées (caractérisation de Dennis-Moré), d'autre part d'une hypothése supplémentaire.

Théorème 4 (Caractérisation de Dennis-Moré(sans preuve)).
Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *trois fois continûment différentiable. Soit l'itération* $x_{k+1} = x_k + \alpha_k p_k$*, où* p_k *est une direction de descente et* α_k *satisfait les conditions de Wolfe (9a)et (9b) avec* $c_1 \leq \frac{1}{2}$*.Si la suite* $\{x_k\}$ *converge vers un point* x^* *tel que* $\nabla f(x^*) = 0$*et que* $\nabla^2 f(x^*)$*et definie positive, et si la direction de recherche(ici le cas quasi-Newton) satisfait*

$$\lim_{k\to\infty} \frac{\left\|\left(B_k - \nabla^2 f(x^*)\right) p_k\right\|}{\|P_k\|} = 0, \tag{22}$$

alors

1. $\exists k_0$ *tel que* $\forall k > k_0$ *la longeur du pas* $\alpha_k = 1$ *et admissible.*
2. *si* $\alpha_k = 1 \forall k > k_0$ *alors* $\{x_k\} \to x^*$ *superlinéairement*

Hypothèse 5.
La matrice hessian G *est Lipschitz-continue en* x^*.
i.e. $\exists L > 0$ *constante telle que*

$$\|G(x) - G(x^*)\| \leq L \|x - x^*\| \qquad \forall x \text{ suffisamment proche de } x^* \tag{23}$$

En vue du prochain théorème portant sur la convergence superlinéaire de BFGS, nous devons à nouveau faire quelques petites préparations qui sont pour la plupart essentiellement technique. Posons

$$\tilde{s_k} = G_*^{\frac{1}{2}} s_k, \qquad \tilde{y_k} = G_*^{-\frac{1}{2}} y_k, \qquad \tilde{B_k} = G_*^{-\frac{1}{2}} B_k G_*^{-\frac{1}{2}} \tag{24}$$

où $G_* = G(x^*)$ avec x^* un minimum de f.

De plus, notons que $G_*^{\frac{1}{2}}$ et $G_*^{-\frac{1}{2}}$ existe et sont symétriques car G_* est SDP (ce qui garantit que ses valeurs propres sont réelles et positives)donc diagonalisable sous forme de Jordan.

Nous allons maintenant utiliser les mêmes outils et idées que pour le théorème précédant (Thm.3). Mais nous devons les transformés un petit peu. On pose

$$\cos\tilde{\theta_k} = \frac{\tilde{s_k}^T \tilde{B_k} \tilde{s_k}}{\|\tilde{s_k}\| \left\|\tilde{B_k}\tilde{s_k}\right\|}, \qquad \tilde{q_k} = \frac{\tilde{s_k}^T \tilde{B_k} \tilde{s_k}}{\|\tilde{s_k}\|^2} \tag{25}$$

nous définissons encore

$$\tilde{M_k} = \frac{\|\tilde{y_k}\|^2}{\tilde{y_k}^T \tilde{s_k}}, \qquad \tilde{m_k} = \frac{\tilde{y_k}^T \tilde{s_k}}{\tilde{s_k}^T \tilde{s_k}} \tag{26}$$

On trouve facilement la formule récursive pour les $\tilde{B_k}$.

$$\tilde{B_{k+1}} = \tilde{B_k} - \frac{\tilde{B_k}\tilde{s_k}\tilde{s_k}^T\tilde{B_k}}{\tilde{s_k}^T\tilde{B_k}\tilde{s_k}} + \frac{\tilde{y_k} y \tilde{-} k^T}{\tilde{y_k}^T\tilde{s_k}} \tag{27}$$

Nous reprenons le même Ψ que précédemment (18) et nous avons

$$\Psi(\tilde{B_{k+1}}) = \Psi(\tilde{B_k}) + \left(\tilde{M_k} - \ln(\tilde{m_k}) - 1\right) + \left[1 - \frac{\tilde{q_k}}{\cos^2\tilde{\theta_k}} + \ln(\frac{\tilde{q_k}}{\cos^2\tilde{\theta_k}})\right] + \ln(\cos^2\tilde{\theta_k}) \tag{28}$$

Comme nous avons par définition $y_k = \bar{G}_k s_k$. nous obtenons donc

$$y_k - G_* s_k = (\bar{G}_k - G_*) s_k \tag{29}$$

en mutltipliant se résultat par la gauche ave $G_*^{-\frac{1}{2}}$ on obtient :

$$\tilde{y_k} - \tilde{s_k} = G_*^{-\frac{1}{2}} (\bar{G}_k - G_*) \tilde{s_k} \tag{30}$$

Par l'hypothèse posée en préparation (Hypothèse5)et la définition de $\bar{G}_k$ (4),nous avons

$$\|\tilde{y_k} - \tilde{s_k}\| \leq \left\| G_*^{-\frac{1}{2}} \right\|^2 \|\tilde{s_k}\| \left\| \bar{G}_k - G_* \right\| \leq \left\| G_*^{-\frac{1}{2}} \right\|^2 \|\tilde{s_k}\| L \epsilon_k \tag{31}$$

Où L est la constante de lipschitz et $\epsilon_k = \max \{\|x_{k+1} - x^*\|, \|x_k - x^*\}\|$. Nous pouvons dès lors montrer

$$\frac{\|\tilde{y_k} - \tilde{s_k}\|}{\|\tilde{s_k}\|} \leq \bar{c}\epsilon_k \tag{32}$$

Avec $\bar{c}$ une constante positive.
Nous sommes maintenant prêts pour le théorème suivant.

Théorème 6 (Convergence superlinéaire de BFGS)**.**
*Soient f fonction deux fois continûment différentiable et la suite $\{x_k\}$ générée par l'algorithme BFGS avec $x_k \to x^*$.Où x^* est un minimum de f et accepte l'hypothèse de G lipschitz(**??**), Alors la convergense de x_k vers x^* est superlinéaire.*

Démonstration. Par le résultat (32) et par l'inégalité triangulaire, nous obtenons

$$\|\tilde{y_k}\| - \|\tilde{s_k}\| \leq \bar{c}\epsilon_k \|\tilde{s_k}\|, \qquad \|\tilde{s_k}\| - \|\tilde{y_k}\| \leq \bar{c}\epsilon_k \|\tilde{s_k}\| \tag{33}$$

Nous pouvons donc borné $\|\tilde{y_k}\|$ avec

$$(1 - \bar{c}\epsilon_k) \|\tilde{s_k}\| \leq \|\tilde{y_k}\| \leq (1 + \bar{c}\epsilon_k) \|\tilde{s_k}\| \tag{34}$$

Après quelques manipulations techniques, nous obtenons

$$\tilde{m_k} = \frac{\tilde{y_k}^T \tilde{s_k}}{\|\tilde{s_k}\|^2} \geq 1 - \bar{c}\epsilon_k \tag{35a}$$

$$\tilde{M_k} = \frac{\|\tilde{y_k}\|^2}{\tilde{y_k}^T \tilde{s_k}} \leq \frac{1 + \bar{c}\epsilon_k}{1 - \bar{c}\epsilon_k} \tag{35b}$$

Nous avons par hypothèse que $x_k \to x^*$. Ce qui implique que $\epsilon_k \to 0$ et donc qu'il existe un $c > \bar{c}$ tel que pour tout k suffisamment grands

$$\tilde{M_k} \leq 1 + \frac{2\bar{c}}{1 - \bar{c}\epsilon_k} \epsilon_k \leq 1 + c\epsilon_k \tag{36}$$

Nous allons à présent à nouveau utiliser la non-positivité des equation de la forme (18) vue précédemment et l'appliquer pour $t = \frac{1}{1-x}$. Alors

$$h(\frac{1}{1-x}) = 1 - \frac{1}{1-x} + \ln(1) - \ln(1-x) = \frac{-x}{1-x} - \ln(1-x) \leq 0 \tag{37}$$

On pose maintenant $x = \bar{c}\epsilon_k$ que l'on remplace dans (37). On obtient

$$\ln(1-\bar{c}\epsilon_k) \geq \frac{-\bar{c}\epsilon_k}{1-\bar{c}\epsilon_k} \geq -2\bar{c}\epsilon_k \tag{38}$$

alors avec (35a)

$$\ln(\tilde{m_k}) \geq \ln(1-\bar{c}\epsilon_k) \geq -2\bar{c}\epsilon_k \geq -2c\epsilon_k \tag{39}$$

Avec toutes ces préparations, nous pouvons transformer (28)et écrire

$$0 \leq \Psi(\tilde{B_{k+1}} \leq \Psi(\tilde{B_k}) + 3c\epsilon_k + \ln(\cos^2\tilde{\theta_k})\left[1 - \frac{\tilde{q_k}}{\cos^2\tilde{\theta_k}} + \ln(\frac{\tilde{q_k}}{\cos^2\tilde{\theta_k}})\right] \tag{40}$$

Nous sommons cette expression et on obtient :

$$\sum_{j\geq 0}\left(\ln(\cos^2\tilde{\theta_j})\left[1 - \frac{\tilde{q_j}}{\cos^2\tilde{\theta_j}} + \ln(\frac{\tilde{q_j}}{\cos^2\tilde{\theta_j}})\right]\right) \leq \Psi(\tilde{B_0}) + 3c\sum_{j\geq 0}\epsilon_j < \infty \tag{41}$$

Alors comme le terme $[\cdots] \leq 0$, et que $\ln(\frac{1}{\cos^2\tilde{\theta_j}}) \geq 0\ \forall j$ alors

$$\lim_{j\to\infty} ln(\frac{1}{\cos^2\tilde{\theta_j}}) = 0, \qquad \lim_{j\to\infty}\left[1 - \frac{\tilde{q_j}}{\cos^2\tilde{\theta_j}} + \ln(\frac{\tilde{q_j}}{\cos^2\tilde{\theta_j}})\right] = 0 \tag{42}$$

cela implique

$$\lim_{j\to\infty}\cos\tilde{\theta_j} = 1, \qquad \lim_{j\to\infty}\tilde{q_j} = 1 \tag{43}$$

Nous avons déjà vu en (15) que $\frac{q_k}{\cos^2\theta_k} = \frac{\|s_k\|\|B_k s_k\|^2}{s_k^T B_k s_k}$. nous pouvons alors écrire :

$$\begin{aligned}
\frac{\left\|G_*^{-\frac{1}{2}}(B_k - G_*)s_k\right\|^2}{\left\|G_*^{\frac{1}{2}}s_k\right\|^2} &= \frac{\left\|(\tilde{B_k} - I)\tilde{s_k}\right\|^2}{\|\tilde{s_k}\|^2} \\
&= \frac{\left\|\tilde{B_k}\tilde{s_k}\right\|^2 - 2\tilde{s_k}^T\tilde{B_k}\tilde{s_k} + \tilde{s_k}^T\tilde{s_k}}{\tilde{s_k}^T\tilde{s_k}} \\
&= \frac{q_k}{\cos^2\theta_k} - 2\tilde{q_k} + 1
\end{aligned}$$

Nous vérifions donc par ce résultat ainsi que (43)que la partie de droite converge vers 0. et donc nous pouvons conclure grâce à l'equivalence des normes euclidienne et des « G_* »-normes que :

$$\lim_{k\to\infty}\frac{\|(B_k - G_*)s_k\|}{\|s_k\|} = 0 \tag{44}$$

Nous avons donc par le théorème de Dennis-Moré(Thm. 4) que suffisamment proche de la solution x^* nous pouvons prendre des $\alpha_k = 1$ qui satisfont les conditions de Wolfe. Ce qui est synonyme de convergence superlinéaire. □

3 Méthode SR1

Les propriétés de cette méthode ne sont pas aussi bien connue et comprise que celles de la méthode BFGS. En effet, il n'existe à ce jour aucun théorème aussi généraux sur la convergence de SR1 que ceux étudiés dans le chapitre précédent pour BFGS. Il existe malgré tout un résultat intéressant concernant l'algorithme de la région de confiance SR1.

3.1 Convergence globale de la méthode SR1

Nous allons voir dans ce chapitre que si la fonction « objetif »possède un unique point stationnaire et que la condition

$$\left|s_k^T(y_k - B_k s_k)\right| \geq r \,\|s_k\|\, \|y_k - B_k s_k\| \tag{45}$$

est vérifiée à chaque itérations et que l'approximation B_k est bornée supérieurement alors l'aglorithme de région de confiance SR1 converge $(n+1)$-pas superlinéairement (nous allons voir se que cela signifie un peu plus loin).

Théorème 7.
Soit x_k la suite des itérées générée par l'algorithme SR1. On suppose que les conditions suivantes vraies

1. *la suite d'itérées n'est pas finie, mais est inclue dans un ensemble fermé, borné et convexe D sur lequel la fonction f est deux fois continûment différentiable et dans lequel elle possède un unique point stationnaire x^**
2. *La matrice $\nabla^2 f(x^*)$ est définie positive, et lipschitz dans un voisinage de x^**
3. *La suite de matrice B_k est bornée en norme*
4. *la condition (45) est vérifiée à chaque itération, où $r \in (0,1)$ une constante*

Alors premièrement

$$\lim_{k\to\infty} x_k = x^* \tag{46}$$

et deuxièmement nous avons la $(n+1)$-pas superlinéarité de la convergence.
i.e.

$$\lim_{k\to\infty} \frac{\|x_{k+n+1} - x^*\|}{\|x_k - x^*\|} = 0 \tag{47}$$

Références

[1] Jorge Nocedal and Stephen J. Wright : Numerical Optimization, Springer Series in Operations Research, Springer 1999.

[2] Page web du Proséminaire : `http://perso.unifr.ch/ales.janka/numeroptim`.