Correction de l'examen Master SICOM, Cours Optimisation
Luc Pronzato, janvier 2007

PARTIE 1

1) *Expliquer le fonctionnement de la méthode à l'aide d'un dessin représentant l'allure de la fonction $h_k(t)$.*

Voir le cours.

2) *On suppose que $h_k(t)$ est quadratique en t avec un minimum en t^*.*

- *Exprimer $h_k(t^*)$ en fonction de $h_k(0)$, $h_k'(0)$ et t^*.*
- *Expliquer pourquoi il est souhaitable de choisir $m_1 < 1/2 < m_2$ dans la méthode de Goldstein et Price.*
- *On pose $\Delta h_k(t^*) = h_k(0) - h_k(t^*)$. Interpréter la partie droite de la condition (a) de la méthode de Goldstein et Price à partir de $\Delta h_k(t)$, la diminution optimale du critère lorsqu'il est quadratique.*

On peut écrire $h_k(t) = (1/2)at^2 + bt + h_k(0)$, avec une dérivée nulle en $t = t^*$, c'est-à-dire $at^* + b = 0$, et une dérivée en 0 qui vaut $h_k'(0) = b$. Ceci donne directement

$$h_k(t^*) = (1/2)h_k'(0)t^* + h_k(0) .$$

Choisir m_1 et m_2 tels que $m_1 < 1/2 < m_2$ permet de ne pas considérer la valeur t^* comme trop grande ou trop petite.

On a $\Delta h_k(t) = -(1/2)h_k'(0)t$, et la partie droite de la condition (a) de la méthode de Goldstein et Price s'écrit donc $h_k(t) - h_k(0) \leq -2m_1\Delta h_k(t)$. On demande donc d'avoir au moins $2m_1$ fois la diminution idéale obtenue si f est quadratique.

PARTIE 2

1) L'algorithme utilise $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k\mathbf{d}_k$. Exprimer $t_k h_k'(0)$ en fonction de $\|\mathbf{g}_k\|$, $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ et

$$c_k = -\frac{\mathbf{g}_k^\top \mathbf{d}_k}{\|\mathbf{g}_k\| \, \|\mathbf{d}_k\|} .$$

A partir de la partie droite de la condition (a) de la méthode de Goldstein et Price, donner une borne inférieure pour $f_k - f_{k+1}$ en fonction de $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ (et aussi $\|\mathbf{g}_k\|$ et c_k).

On a $t_k h_k'(0) = t_k \mathbf{g}_k^\top \mathbf{d}_k = -t_k c_k \|\mathbf{g}_k\| \, \|\mathbf{d}_k\| = -c_k \|\mathbf{g}_k\| \, \|\mathbf{x}_{k+1} - \mathbf{x}_k\|$. La partie droite de la condition (a) de la méthode de Goldstein et Price donne

$$h_k(0) - h_k(t_k) = f_k - f_{k+1} \geq -t_k m_1 h_k'(0) = m_1 c_k \|\mathbf{g}_k\| \, \|\mathbf{x}_{k+1} - \mathbf{x}_k\| .$$

2) On utilise à présent la partie gauche de la condition (a). Ré-écrire l'inégalité sous la forme d'une borne inférieure pour $(f_{k+1} - f_k)/t_k - \mathbf{g}_k^\top \mathbf{d}_k$ (on exprimera cette borne en fonction de m_2, c_k, $\|\mathbf{g}_k\|$ et $\|\mathbf{d}_k\|$). Utiliser un développement limité de $f(\cdot)$ au premier ordre pour exprimer $(f_{k+1} - f_k)/t_k$ en fonction de $\mathbf{d}_k$ et $\mathbf{g}(\hat{\mathbf{x}}_k)$, avec $\hat{\mathbf{x}}_k$ entre $\mathbf{x}_k$ et $\mathbf{x}_{k+1}$ (c'est-à-dire, $\hat{\mathbf{x}}_k = (1-\alpha)\mathbf{x}_k + \alpha\mathbf{x}_{k+1}$ pour un α dans (0,1)).

On obtient pour la partie gauche

$$\frac{f_{k+1} - f_k}{t_k} - \mathbf{g}_k^\top \mathbf{d}_k \geq (m_2 - 1)\mathbf{g}_k^\top \mathbf{d}_k = (1 - m_2)c_k \|\mathbf{g}_k\| \, \|\mathbf{d}_k\| .$$

Le développement limité de $f(\cdot)$ au premier ordre au voisinage de $\mathbf{x}_k$ donne

$$\frac{f_{k+1} - f_k}{t_k} = \mathbf{g}^\top(\hat{\mathbf{x}}_k)\mathbf{d}_k$$

et donc

$$[\mathbf{g}(\hat{\mathbf{x}}_k) - \mathbf{g}_k]^\top \mathbf{d}_k \geq (1 - m_2)c_k \|\mathbf{g}_k\| \, \|\mathbf{d}_k\| .$$

3) On suppose que le gradient de $f(\cdot)$ est K-Lipschitzien, c'est-à-dire, qu'il existe $K > 0$ tel que

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{z})\| \leq K\|\mathbf{x} - \mathbf{z}\| \quad \forall \, \mathbf{x} \, , \mathbf{z} \in \mathbb{R}^d .$$

En utilisant l'inégalité de Cauchy-Schwarz ($\mathbf{x}^\top \mathbf{z} \leq \|\mathbf{x}\| \, \|\mathbf{z}\| \; \forall \, \mathbf{x} \, , \mathbf{z} \in \mathbb{R}^d$), déduire de l'inégalité obtenue en 2) une borne inférieure sur $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ (en fonction de m_2, c_k, $\|\mathbf{g}_k\|$ et K).

On peut écrire

$$K\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \geq K\|\hat{\mathbf{x}}_k - \mathbf{x}_k\| \geq \|\mathbf{g}(\hat{\mathbf{x}}_k) - \mathbf{g}_k\|$$

et donc d'après Cauchy-Schwarz

$$K\|\mathbf{x}_{k+1}-\mathbf{x}_k\|\,\|\mathbf{d}_k\| \geq [\mathbf{g}(\hat{\mathbf{x}}_k)-\mathbf{g}_k]^\top \mathbf{d}_k \geq (1-m_2)c_k\|\mathbf{g}_k\|\,\|\mathbf{d}_k\|\,,$$

c'est-à-dire

$$\|\mathbf{x}_{k+1}-\mathbf{x}_k\| \geq (1-m_2)c_k\|\mathbf{g}_k\|/K\,.$$

4) Utiliser les résultats de 1) et 3) pour donner une borne inférieure sur $f_k - f_{k+1}$ en fonction de c_k, $\|\mathbf{g}_k\|$ et

$$r = \frac{m_1(1-m_2)}{K}\,.$$

Avec les résultats qui précèdent, on obtient directement l'inégalité

$$f_k - f_{k+1} \geq r c_k^2 \|\mathbf{g}_k\|^2\,.$$

PARTIE 3

Supposons que la suite des points $\mathbf{x}_k$ proposés par l'algorithme satisfait

$$r c_k^2\|\mathbf{g}_k\|^2 \leq f_k - f_{k+1}$$

avec r constant et $\sum_{k=1}^\infty c_k^2 = \infty$.

1) Montrer que la propriété ci-dessus implique que

- *soit $f_k \to -\infty$ quand $k \to \infty$,*
- *soit $\liminf_{k\to\infty} \|\mathbf{g}_k\| = 0$ (c'est-à-dire, $\forall\delta > 0$, $\forall k_0 \geq 1$, $\exists k > k_0$ tel que $\|\mathbf{g}_k\| < \delta$).*

Si $f_k \to -\infty$ c'est terminé. Sinon, il existe f^* qui borne $f(\cdot)$ inférieurement. On peut écrire pour tout $N \geq k$

$$\sum_{k=1}^{N} r c_k^2\|\mathbf{g}_k\|^2 \leq f_1 - f_{N+1} \leq f_1 - f^*$$

et $\sum_{k=1}^\infty r c_k^2\|\mathbf{g}_k\|^2$ est donc borné. Supposons que la proposition ne soit pas vérifiée. Cela impliquerait l'existence d'un $\delta > 0$ tel que $\|\mathbf{g}_k\| > \delta$ pour tout $k > k_0$ avec k_0 assez grand, soit

$$r\delta^2 \sum_{k=k_0}^{\infty} c_k^2 \leq \sum_{k=k_0}^{\infty} r c_k^2 \|\mathbf{g}_k\|^2 < \infty\,,$$

ce qui contredirait $\sum_{k=1}^\infty c_k^2 = \infty$. Donc $\liminf_{k\to\infty}\|\mathbf{g}_k\| = 0$.

2) Déduire de l'inégalité obtenue en 4) de la partie 2 la raison pour laquelle il est primordial de choisir $m_2 < 1$ dans la méthode de Goldstein et Price.

$m_2 < 1$ assure $r > 0$ dans la question 4 de la partie 2. C'est le même r qui est utilisé ci dessus pour établir la convergence.

3) On implémente la méthode de Goldstein et Price dans un algorithme tel que

- *$\mathbf{d}_k$ est une direction de descente à chaque itération ($c_k > 0$ pour tout k)*
- *et $\mathbf{d}_k = -\mathbf{g}_k$ pour $k = im$, $i = 1,2,3\ldots$ et m fixé, $m \geq 1$ (c'est-à-dire, la direction de recherche est celle de l'algorithme du gradient toutes les m itérations).*

On suppose que la fonction à minimiser est continûment dérivable, de gradient K-Lipschitzien et bornée inférieurement. Que peut on dire de la convergence de l'algorithme ?

D'après ce qui précède, il suffit d'établir que $\sum_{k=1}^\infty c_k^2 = \infty$ pour obtenir la convergence. Or, on a ici

$$\sum_{k=1}^{\infty} c_k^2 \geq \sum_{i=1}^{\infty} c_{im}^2$$

avec $\mathbf{d}_{im} = -\mathbf{g}_{im}$ et donc $c_{im} = 1$ pour tout i.

4) On l'implémente à présent dans l'algorithme de Newton. On suppose que $f(\cdot)$ est deux fois continûment dérivable, que la matrice $\mathbf{H}(\mathbf{x})$ de ses dérivées secondes est définie positive avec des valeurs propres minimale et maximale $\lambda_{\min}$ et $\lambda_{\max}$ qui satisfont

$$\rho(\mathbf{x}) = \frac{\lambda_{\max}[\mathbf{H}(\mathbf{x})]}{\lambda_{\min}[\mathbf{H}(\mathbf{x})]} \leq \bar{\rho} < \infty\,,\ \forall\, \mathbf{x} \in \mathbb{R}^d\,.$$

Donner une borne inférieure pour c_k^2 en fonction de $\rho(\mathbf{x}_k)$ puis de $\bar{\rho}$. On rappelle que d'après l'inégalité de Kantorovich

$$(\mathbf{u}^\top \mathbf{H}^{-1}\mathbf{u})\,(\mathbf{u}^\top \mathbf{H}\mathbf{u}) \leq \frac{(\lambda_{\max}(\mathbf{H}) + \lambda_{\min}(\mathbf{H}))^2}{4\lambda_{\max}(\mathbf{H})\lambda_{\min}(\mathbf{H})}\,(\mathbf{u}^\top \mathbf{u})^2\,,\ \forall\,\mathbf{u} \in \mathbb{R}^d\,.$$

Conclure quant à la convergence de l'algorithme.

Comme précédemment, il suffit d'étudier le comportement de $\sum_{k=1}^{\infty} c_k^2$. On a $\mathbf{d}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k$, avec $\mathbf{H}_k = \mathbf{H}(\mathbf{x}_k)$, et donc

$$c_k = \frac{\mathbf{g}_k^\top \mathbf{H}_k^{-1}\mathbf{g}_k}{(\mathbf{g}_k^\top \mathbf{H}_k^{-2}\mathbf{g}_k)^{1/2}(\mathbf{g}_k^\top \mathbf{g}_k)^{1/2}}$$

ou encore en posant $\mathbf{u}_k = \mathbf{H}_k^{-1/2}\mathbf{g}_k$ (ce qui est légitime puisque la matrice $\mathbf{H}_k$ est définie positive)

$$c_k = \frac{\mathbf{u}_k^\top \mathbf{u}_k}{(\mathbf{u}_k^\top \mathbf{H}_k^{-1}\mathbf{u}_k)^{1/2}(\mathbf{u}_k^\top \mathbf{H}_k\mathbf{u}_k)^{1/2}}\,.$$

D'après l'inégalité de Kantorovich on obtient donc

$$c_k^2 \geq \frac{4\rho_k}{(1+\rho_k)^2}$$

où $\rho_k = \rho(\mathbf{x}_k)$. Le terme de droite est une fonction décroissante de ρ_k, et donc

$$c_k^2 \geq \frac{4\bar{\rho}}{(1+\bar{\rho})^2}\,.$$

Par conséquent $\sum_{k=1}^{\infty} c_k^2 = \infty$ et l'algorithme converge (au sens de la partie 3.1).