Université Paris-Dauphine
Département MIDO
Chargé de cours : F.-X. Vialard.
Chargés de TD/TP : G. Jankowiak, A. Frouvelle.
DUMI2E troisième année

Optimisation numérique

année 2012/2013

Correction du Contrôle Continu

Remarques générales :

- On ne vous demande pas de recopier un cours par cœur mais de l'avoir compris, et de le montrer…
- Bien lire les questions ! Une large majorité n'a pas lu la fin de la première question de l'exercice 3 par exemple.
- Bien indiquer les résultats que l'on admet, surtout à l'intérieur d'une question.

Question de cours 1.

Énoncé de la règle de Wolfe.

- On se donne deux réels $0 < \alpha < \beta < 1$, une fonction $f \in C^1(\mathbb{R}^n, \mathbb{R})$, un point $x \in \mathbb{R}^n$ et une direction de descente $d \in \mathbb{R}^n$: $\langle \nabla f(x), d\rangle < 0$. La fonction f satisfait la règle de Wolfe pour le pas t si et seulement si
$$\begin{cases} f(x+td) \leqslant f(x) + \alpha t\langle \nabla f(x), d\rangle \\ \langle \nabla f(x+td), d\rangle \geqslant \beta \langle \nabla f(x), d\rangle \end{cases}$$

Proposition : Soit $0 < \alpha < \beta < 1$, et $f \in C^1(\mathbb{R}_+, \mathbb{R})$ une fonction bornée inférieurement et telle que $f'(0) < 0$. Il existe un intervalle ouvert $I \subset \mathbb{R}_+$ tel que pour la direction de descente $d = 1$ et au point $x = 0$, f satisfait la règle de Wolfe pour tout $t \in I$.

Démonstration : Les deux conditions s'écrivent, en dimension 1, pour $d = 1$ et $x = 0$
$$f(t) \leqslant f(0) + \alpha t f'(0) \qquad (1)$$
$$f'(t) \geqslant \beta f'(0) \qquad (2)$$

On pose $t_1 = \inf\{t > 0 | f(t) > f(0) + \alpha t f'(0)\}$.

- t_1 est bien défini. En effet, comme <u>f est bornée inférieurement</u> (par un réel M), on a pour t suffisament grand $f(0) + \alpha t f'(0) < M$, et donc $f(t) \geqslant M > f(0) + \alpha t f'(0)$. L'ensemble sur lequel on prend la borne inférieure est bien non vide (et minoré par 0). De plus, $t_1 > 0$. En effet, dans un voisinage de 0 on a $f(t) = f(0) + tf'(0) + o(t)$ ce qui est strictement plus petit que $f(0) + \alpha t f'(0)$ pour t suffisament petit, puisque <u>$\alpha < 1$ et $f'(0) < 0$</u>.
- Par <u>continuité de f</u>, on a $f(t_1) \geqslant f(0) + \alpha t_1 f'(0)$. On l'obtient en passant à la limite pour une suite $t_n > t_1$ réalisant la borne inférieure.
- Par le théorème des accroissements finis, on obtient $t_2 \in]0, t_1[$ tel que
$$f'(t_2) = \frac{f(t_1) - f(0)}{t_1} \geqslant \alpha f'(0) > \beta f'(0).$$
- Par <u>continuité de f'</u>, il existe ε tel que pour tout $t \in]t_2 - \varepsilon, t_2 + \varepsilon[$, on a $f'(t) > \beta f'(0)$ (l'inégalité <u>(2) est satisfaite</u>). Et quitte à prendre ε suffisamment petit, on peut avoir $]t_2 - \varepsilon, t_2 + \varepsilon[\subset]0, t_1[$, donc l'intervalle $I =]t_2 - \varepsilon, t_2 + \varepsilon[$ convient (on a (1) pour tout $t \in I$ puisque $t < t_1$).

Remarque : l'intérêt de l'existence d'un petit intervalle (et pas seulement du point t_2) est de s'assurer de la terminaison de l'algorithme de recherche d'un tel point.

Question de cours 2.

Convergence de la suite. On se fixe un pas $\varepsilon < \frac{2}{L}$.

- La suite des points avec l'algorithme de descente de gradient est donnée par
$$\underline{x_{k+1} = x_k - \varepsilon \nabla f(x_k)}.$$
- On pose $F(x) = x - \varepsilon \nabla f(x)$. On a alors $x_{k+1} = F(x_k)$. La fonction F est de classe C^1 (de $\mathbb{R}^n$ dans $\mathbb{R}^n$), et sa différentielle est donnée par
$$\underline{F'(x) = \mathrm{Id} - \varepsilon \nabla^2 f(x)},$$
donc on a que $F(x)$ est une matrice symétrique et
$$(1 - \ell\varepsilon)\mathrm{Id} \leqslant F'(x) \leqslant (1 - L\varepsilon)\mathrm{Id},$$
au sens des matrices symétriques. Ainsi, la plus grande valeur propre de $F'(x)$ (en valeur absolue) vérifie $1 - \ell\varepsilon \leqslant \lambda_{max} \leqslant 1 - L\varepsilon$, donc $|\lambda_{max}| \leqslant \max(|1 - \ell\varepsilon|, |1 - L\varepsilon|)$.
- D'après l'inégalité des accroissements finis, on a
$$\|F(x) - F(y)\| \leqslant \sup_{z \in [x,y]} \|F'(z)\|_{L(\mathbb{R}^n, \mathbb{R}^n)} \|x - y\| \leqslant \max(|1 - \ell\varepsilon|, |1 - L\varepsilon|)\|x - y\|. \tag{3}$$
- On pose $m(\varepsilon) = \max(|1 - \ell\varepsilon|, |1 - L\varepsilon|)$, et si $\varepsilon < \frac{2}{L} (\leqslant \frac{2}{\ell})$, on obtient $\underline{0 \leqslant m(\varepsilon) < 1}$.
- En prenant $x = x_k$ et $y = x_{k-1}$ dans (3), on obtient $\|x_{k+1} - x_k\| \leqslant m(\varepsilon)\|x_k - x_{k-1}\|$. Et par récurrence $\|x_{k+1} - x_k\| \leqslant m(\varepsilon)^k \|x_1 - x_0\|$. Comme $m(\varepsilon) < 1$, la série $\sum \|x_{k+1} - x_k\|$ converge, et donc la série $\sum (x_{k+1} - x_k)$ converge également. Comme $\sum_{i=0}^{k-1} (x_{i+1} - x_i) = x_k - x_0$, on en conclut que $(x_k - x_0)$ converge, donc la suite (x_k) converge.

Ordre de convergence.

- On pose x^* la limite de la suite x_k (par continuité la limite de $x_{k+1} = F(x_k)$ est égale à $F(x^*)$, donc $\underline{x^* = F(x^*)}$. On applique (3) à $x = x_k$ et $y = x^*$ et on obtient $\|x_{k+1} - x^*\| \leqslant m(\varepsilon)\|x_k - x^*\|$. La convergence est donc linéaire.
Le taux de convergence asymptotique est $\limsup_k \frac{\|x_{k+1} - x_k\|}{\|x_k - x_{k-1}\|} \leqslant m(\varepsilon)$ (< 1 quand $\varepsilon < \frac{2}{L}$).

Exercice 1.

1. **Faux.** Un contre-exemple est donné par la fonction $f : x \mapsto \ln(1 + e^x)$ en dimension 1, en partant de n'importe quel point. On a $f'(x) = \frac{e^x}{1+e^x} = \frac{1}{e^{-x}+1}$, et $f''(x) = \frac{e^{-x}}{(e^{-x}+1)^2} = \frac{1}{e^x+2+e^{-x}}$. On obtient donc $0 < f''(x) < \frac{1}{2}$ donc f' est bien $\frac{1}{2}$-Lipschitz. La suite des itérées est donnée par $x_{k+1} = x_k - \frac{\varepsilon}{1+e^{-x_k}}$. La suite (x_k) est décroissante. Supposons qu'elle ait une limite x^*, par continuité de l'exponentielle, on obtient en passant à la limite $x^* = x^* - \frac{\varepsilon}{1+e^{-x^*}}$ ce qui est absurde. La suite diverge donc vers $-\infty$. Et on est donc bien dans l'hypothèse de l'exercice.
2. **Faux.** Le contre-exemple est difficile et nécessite de passer en dimension 2 (exemple 8 du cours). Il est possible de montrer que c'est vrai en dimension 1.
3. **Faux.** Le contre-exemple du premier cas est valide puisque la fonction est convexe (et même strictement convexe : $f''(x) > 0$).

Exercice 2.

1. $\nabla f = (x, y^3 - y) = (x, y(y-1)(y+1))$. Il y a 3 points critiques $(0, 1)$, $(0, 0)$, et $(0, -1)$.
2. $H = \begin{pmatrix} 1 & 0 \\ 0 & 3y^2 - 1 \end{pmatrix}$. Aux points critiques $(0, 1)$ et $(0, -1)$, la hessienne $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ est définie positive, les points sont des minima locaux (stricts). Au point $(0, 0)$, la hessienne $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ a deux valeurs propres de signe opposé, c'est un point selle (donc pas un min local).
3. Par récurrence, puisque $x_0 = (1, 0)$ et que $x_k = x_{k+1} - h\nabla f(x_k)$, on a $x_k = ((1-h)^k, 0)$ pour tout k. Et on obtient $x_k \to (0, 0)$ (si $h < 1$). La suite des points converge donc vers le point $(0, 0)$, qui n'est pourtant pas un minimum local. On en conclut que l'algorithme de descente de gradient ne fournit pas nécessairement un minimum local, même s'il converge.

Exercice 3.

1. Pour x fixé, la fonction $y \mapsto \langle \nabla f(x), y \rangle$ est C^1 (et même C^∞ puisqu'elle est linéaire, de la forme $y \mapsto \langle a, y \rangle$, son gradient est a, et les dérivées suivantes sont toutes nulles). Donc F est également C^1 et $\nabla F(y) = \nabla f(y) - \nabla f(x)$. Donc

$$\|\nabla F(y_1) - \nabla F(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leqslant L\|y_1 - y_2\|,$$

et F est bien L-Lipschitz.

La fonction f étant convexe, F l'est aussi (on ajoute une fonction linéaire à f pour obtenir F), et donc tout point critique est un minimiseur local. Comme $\nabla F(x) = 0$, x est un minimiseur local, on obtient

$$F(y) = f(y) - \langle \nabla f(x), y \rangle \geqslant F(x) = f(x) - \langle \nabla f(x), x \rangle.$$

2. La proposition à appliquer est la suivante, puisque ∇F est L-Lipschitz :

$$F(z) \leqslant F(y) + \langle \nabla F(y), z - y \rangle + \frac{L}{2}\|z - y\|^2.$$

On prend $z = y - \frac{1}{L}\nabla F(y)$, on obtient $z - y = -\frac{1}{L}\nabla F(y) = \frac{1}{L}(\nabla f(x) - \nabla f(y))$, et donc

$$\begin{aligned} F(z) &\leqslant F(y) - \langle \nabla f(y) - \nabla f(x), \tfrac{1}{L}(\nabla f(x) - \nabla f(y)) \rangle + \tfrac{L}{2}\|\tfrac{1}{L}(\nabla f(x) - \nabla f(y))\|^2 \\ &= F(y) + (-\tfrac{1}{L} + \tfrac{L}{2L^2})\|\nabla f(x) - \nabla f(y)\|^2. \end{aligned}$$

Comme $F(z) \geqslant F(x)$ d'après la question précédente, on obtient

$$F(x) \leqslant F(y) - \tfrac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2,$$

ce qui donne, en réécrivant la définition de $F(x)$ et $F(y)$,

$$f(x) - \langle \nabla f(x), x \rangle \leqslant f(y) - \langle \nabla f(x), y \rangle - \tfrac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2,$$

qui est l'inégalité demandée.

3. On réécrit l'inégalité de la question précédente en intervertissant le rôle de x et y :

$$\begin{aligned} f(y) &\geqslant f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \\ f(x) &\geqslant f(y) + \langle \nabla f(y), x - y \rangle + \tfrac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2, \end{aligned}$$

et en sommant ces inégalités, les termes $f(x)$ et $f(y)$ se simplifient, on obtient

$$0 \geqslant \langle \nabla f(x) - \nabla f(y), y - x \rangle + \tfrac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2,$$

qui correspond à la première inégalité demandée. En appliquant ce résultat à $y = x_k$ et $x = x^*$ et en utilisant le fait que $\nabla f(x^*) = 0$ (c'est un minimum global), on obtient directement la deuxième inégalité.

4. On a, en utilisant la définition de x_{k+1}, et en développant

$$\begin{aligned} r_{k+1}^2 &= \|x_{k+1} - x^*\|^2 = \|x_k - h\nabla f(x_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - 2h\langle \nabla f(x_k), x_k - x^* \rangle + h^2, k\|\nabla f(x_k)\|^2 \\ &\leqslant r_k^2 - \tfrac{2h}{L}\|\nabla f(x_k)\|^2 + h^2\|\nabla f(x_k)\|^2, \end{aligned}$$

où la dernière inégalité provient de la question précédente, ce qui donne l'inégalité demandée. Cette inégalité implique $r_{k+1}^2 \leqslant r_k^2$ puisque $h < \frac{2}{L}$, et comme les r_k sont positifs, on obtient donc $r_{k+1} \leqslant r_k$, la suite (r_k) est donc décroissante, d'où $r_k \leqslant r_0$.

5. On applique toujours la même proposition, puisque f est L-Lipschitz :

$$f(x_{k+1}) \leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2.$$

   Comme $x_{k+1} - x_k = -h\nabla f(x_k)$, cela devient

$$f(x_{k+1}) \leqslant f(x_k) - h\|\nabla f(x_k)\|^2 + \frac{Lh^2}{2} \|\nabla f(x_k)\|^2.$$

6. Par convexité de f, on a $f(x^*) - f(x_k) \geqslant \langle \nabla f(x_k), x^* - x_k \rangle$, qui devient grâce à l'inégalité de Cauchy-Schwarz

$$f(x_k) - f(x^*) \leqslant \langle \nabla f(x_k), x_k - x^* \rangle \leqslant \|\nabla f(x_k)\| \, \|x_k - x^*\|,$$

   ou encore $f(x_k) - f(x^*) \leqslant r_k \|\nabla f(x_k)\| \leqslant r_0 \|\nabla f(x_k)\|$ d'après la question 4.

7. En soustrayant $f(x^*)$ de part et d'autre de l'inégalité de la question 6, on obtient

$$\Delta_{k+1} \leqslant \Delta_k - w\|\nabla f(x_k)\|^2 \leqslant \Delta_k - \tfrac{w}{r_0^2} \Delta_k^2, \tag{4}$$

   d'après la question 5.
   On obtient, puisque $\Delta_{k+1} \geqslant 0$ (le minimum de f^* est atteint en x^*), que $\Delta_k - \frac{w}{r_0^2}\Delta_k^2 \geqslant 0$, soit $\frac{w}{r_0^2}\Delta_k \leqslant 1$. Ensuite, en prenant l'inverse de (4) (puisque les termes sont positifs), on obtient

$$\frac{1}{\Delta_{k+1}} \geqslant \frac{1}{\Delta_k(1 - \frac{w}{r_0^2}\Delta_k)} \geqslant \frac{1}{\Delta_k}(1 + \tfrac{w}{r_0^2}\Delta_k),$$

   d'après l'indication demandée : $\frac{1}{1-x} \geqslant 1 + x$ (équivalent à $1 \geqslant 1 - x^2$) avec $x = \frac{w}{r_0^2}\Delta_k^2 \in [0, 1]$. En développant le dernier terme on obtient l'inégalité demandée.

8. On a

$$\frac{1}{\Delta_k} = \frac{1}{\Delta_0} + \sum_{i=0}^{k-1} \frac{1}{\Delta_{i+1} - \Delta_i} \geqslant \frac{1}{\Delta_0} + \sum_{i=0}^{k-1} \frac{w}{r_0^2},$$

   ce qui donne le résultat.

9. On prend l'inverse de l'inégalité de la question précédente

$$f(x_k) - f(x^*) = \Delta_k \leqslant \frac{1}{\frac{1}{\Delta_0} + k\frac{w}{r_0^2}} = \frac{\Delta_0 r_0^2}{r_0^2 + \Delta_0 w k} = \frac{2L\Delta_0 r_0^2}{2Lr_0^2 + 2Lw\Delta_0 k}.$$

   En remplaçant Δ_0 par $f(x_0) - f(x^*)$ et r_0 par $\|x_0 - x^*\|$, on obtient directement le résultat demandé dès que $2Lw = 1$. Cela correspond à $Lh(2 - Lh) = 1$, ou encore $(Lh - 1)^2 = 0$, soit $Lh = 1$, c'est à dire $h = \frac{1}{L}$.

10. On applique encore la même proposition entre x_0 et x^* :

$$f(x_0) \leqslant f(x^*) + \langle \nabla f(x^*), x_0 - x^* \rangle + \frac{L}{2} \|x_0 - x^*\|^2,$$

   qui devient directement $\Delta_0 \leqslant \frac{L}{2} r_0^2$ puisque $\nabla f(x^*) = 0$, donc $2Lr_0^2 \geqslant 4\Delta_0$. Lorsque $2Lw = 1$, l'estimation de la question précédente s'écrit

$$\Delta_k \leqslant \frac{2L\Delta_0 r_0^2}{2Lr_0^2 + \Delta_0 k} \leqslant \frac{2L\Delta_0 r_0^2}{4\Delta_0 + \Delta_0 k} = \frac{2Lr_0^2}{4 + k},$$

   qui correspond à l'estimation demandée.