Module A4
Fonctions régulières

Sauf mention contraire, $\mathcal{X}$ (respectivement $\mathcal{Y}$...) est un espace vectoriel réel de dimension finie, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\| \cdot \|$ la norme euclidienne associée. Pour toutes les autres notations et définitions, le lecteur est invité à se reporter aux cours précédents.

1 Fonctions régulières

1.1 Définition et exemples

Définition 1 (Fonction lipschitzienne)

Soit $f : \mathcal{X} \to \mathcal{Y}$ une fonction. On dit que f est *lipschitzienne*, de constante de Lipschitz $L \geq 0$, si

$$\forall (x,z) \in \mathcal{X}, \qquad \|f(x) - f(z)\| \leq L \, \|x - z\|$$

Remarque : Si f est lipschitzienne de constante de Lipschitz L, alors elle est lipschitzienne de constante de Lipschitz L' pour tout $L' \geq L$.

Définition 2 (Fonctions régulières)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction différentiable. Soit $L \geq 0$. On dit que J est L-régulière si son gradient $\nabla J : \mathcal{X} \to \mathcal{X}$ est lipschitzien de constante de Lipschitz L. Autrement dit, J est L-régulière si

$$\forall (x,z) \in \mathcal{X}, \qquad \|\nabla J(x) - \nabla J(z)\| \leq L \, \|x - z\|$$

D'après les remarques précédentes, on peut affirmer que

- si J est L-régulière, alors J est L'-régulière pour $L' \geq L$;
- si J est L-régulière, alors J est de classe $\mathcal{C}^2$. La réciproque est évidemment fausse.

Lorsque la fonction est deux fois différentiable, on a la proposition suivante :

Proposition 1

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction deux fois différentiable. On suppose qu'il existe $M \geq 0$ tel que

$$\forall x \in \mathcal{X}, \qquad \|\nabla^2 J(x)\| \leq M$$

Alors J est M-régulière.

Démonstration : Il s'agit d'une conséquence de l'inégalité des accroissements finis dans le cas vectoriel. ■

Considérons deux exemples qui apparaissent fréquemment dans les applications.

Exemple

Fonction de Huber. Soit $\alpha > 0$. On considère la fonction $J : \mathbb{R} \to \mathbb{R}$ définie de la manière suivante :

$$\forall x \in \mathbb{R}, \qquad J(x) = \begin{cases} \dfrac{x^2}{2} & \text{si } |x| \leq \alpha \\ \alpha\,|x| - \dfrac{\alpha^2}{2} & \text{sinon} \end{cases}$$

Elle est visiblement dérivable sur $]-\infty\,; -\alpha\,[$, $]-\alpha\,; \alpha\,[$ et $]\,\alpha\,; \infty\,[$, de dérivée

$$J'(x) = \begin{cases} x & \text{si } |x| < \alpha \\ \alpha & \text{si } x > \alpha \\ -\alpha & \text{si } x < -\alpha \end{cases}$$

Il est aisé de vérifier que J' est continue en α et $-\alpha$, de sorte que J est différentiable sur tout $\mathbb{R}$. Soit $(x,z) \in \mathbb{R}^2$. En calculant $|J'(x) - J'(z)|$ pour tous les cas de figure possibles, on démontre que

$$\forall (x,z) \in \mathbb{R}^2, \qquad |J'(x) - J'(z)| \leq |x - z|$$

ce qui assure que J est 1-régulière. L'allure de la courbe représentative de quelques unes de ces fonctions est présentée en figure 1.

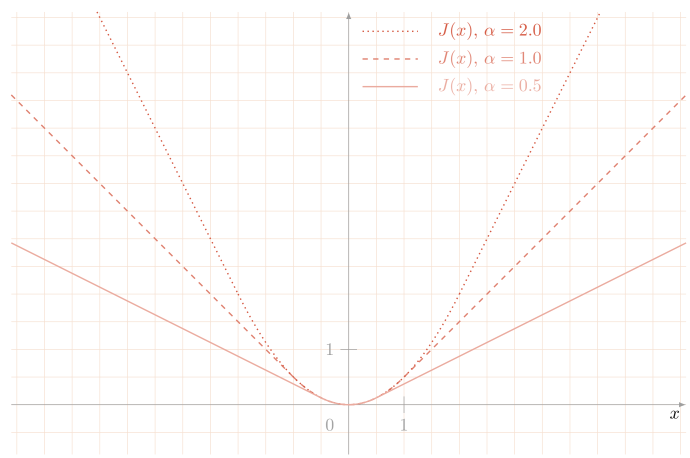

Figure 1 – Exemples de fonctions de Huber.

On considère dans l'exemple suivant le cas d'une fonction deux fois différentiable.

Exemple

Régularisation de la valeur absolue. Soit $\alpha > 0$. On considère la fonction suivante :

$$J : \left\{ \begin{array}{ccl} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & |x| - \alpha \ln\left(1 + \dfrac{|x|}{\alpha}\right) \end{array} \right.$$

En distinguant les deux intervalles $]-\infty\,;0\,[$ et $]\,0\,;+\infty\,[$, on peut vérifier que J est dérivable et que J' est donnée par

$$\forall\, x \in \mathbb{R}, \qquad J'(x) = \frac{x}{\alpha + |x|}$$

De même, on peut démontrer que J est deux fois différentiable, et que

$$\forall\, x \in \mathbb{R}, \qquad |J''(x)| = \left| \frac{\alpha}{(\alpha + |x|)^2} \right| \leq \frac{1}{\alpha}$$

de sorte que J est $(1/\alpha)$-régulière. Comme on peut le voir dans la figure 2, ces fonctions peuvent être utilisées pour approcher de manière régulière la valeur absolue. En effet, elles ont mêmes valeur en 0 et, asymptotiquement, un comportement semblable, puisque

$$\lim_{x \to +\infty} J'(x) = 1 \qquad \text{et} \qquad \lim_{x \to -\infty} J'(x) = -1$$

Elles sont donc utiles dans les applications où apparaît cette fonction.

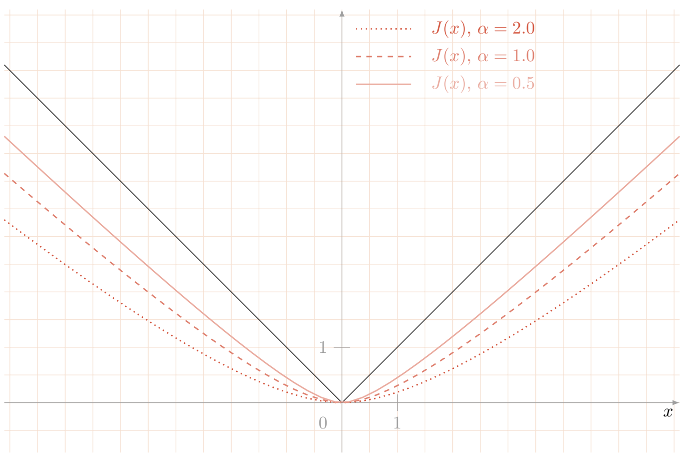

Figure 2 – Exemples de régularisation de la valeur absolue (la courbe représentative de cette dernière apparaît en noir).

1.2 Composition avec un opérateur affine

Commençons par démontrer que l'ensemble des fonctions régulières est stable par combinaison linéaire.

Proposition 2

Soit $J_1 : \mathcal{X} \to \mathbb{R}$ une fonction L_1-régulière et $J_2 : \mathcal{X} \to \mathbb{R}$ une fonction L_2-régulière. Soit $(\lambda, \mu) \in \mathbb{R}^2$. Alors la fonction $\lambda\, J_1 + \mu\, J_2$ est $(|\lambda|\, L_1 + |\mu|\, L_2)$-régulière.

Démonstration : Il suffit d'appliquer la linéarité du gradient, puis l'inégalité triangulaire. ■

On peut interpréter le résultat précédent comme la régularité de la composition extérieure par une fonction linéaire (affine) d'une fonction régulière. La composition interne par une fonction affine préserve également la régularité. Pour cela, on commence par rappeler la définition d'un opérateur linéaire borné :

Définition 3 (Norme d'opérateur)

Soit $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire. On dit que A est *borné* si la quantité suivante

$$|||A||| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

est finie. On appelle alors *norme (d'opérateur)* de A la valeur $|||A|||$.

Remarque : Dans ce cours, étant en dimension finie, on identifie les opérateurs linéaires et l'application linéaire qu'ils définissent. Cette convention nous évite de préciser les dimensions des matrices rencontrées, ce qui permet d'alléger les notations.

> Dans la définition précédente, la norme de A dépend du choix des normes sur les espaces $\mathcal{X}$ et $\mathcal{Y}$. Il est en effet possible de considérer d'autres normes. Puisqu'elles sont équivalentes en dimension finie, le caractère borné d'un opérateur linéaire ne dépend cependant pas de ce choix.

Par linéarité et A et par 1-homogénéité des normes, la norme d'un opérateur A vaut également

$$|||A||| = \sup_{\|x\|=1} \frac{\|Ax\|}{\|x\|} = \sup_{x \in \mathcal{B}(0,1)\setminus\{0\}} \frac{\|Ax\|}{\|x\|}$$

où $\mathcal{B}(0,1)$ est la boule unité (fermée). On peut interpréter le caractère borné d'un opérateur linéaire de diverses manières, parmi lesquelles :

- l'image par A de la boule unité est bornée ;
- l'application A est lipschitzienne, de constante de Lipschitz $|||A|||$.

En dimension finie, tous les opérateurs linéaires sont bornés ; lorsque la norme d'opérateur est définie à l'aide de norme euclidienne, on a de plus les identités suivantes :

Lemme 1

Soit $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire. Alors A et A^* sont bornés et on a

$$|||A|||^2 = |||A^*|||^* = |||A^* A||| = |||A\, A^*|||$$

DÉMONSTRATION : Laissé au lecteur.

Proposition 3

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction L-régulière. Soit $A : \mathcal{Y} \to \mathcal{X}$ un opérateur linéaire borné et $b \in \mathcal{X}$. Alors la fonction $x \mapsto J(A\,x + b)$ est $(|||A|||^2 L)$-régulière.

DÉMONSTRATION : La fonction considérée est différentiable de gradient

$$x \mapsto A^* \, \nabla J(A\,x + b)$$

puis on utilise la régularité de J pour écrire que, pour tout $(x, z) \in \mathcal{X}^2$,

$$\|A^* \, \nabla J(A\,x + b) - A^* \, \nabla J(A\,z + b)\| \leq |||A^*|||\, L\, \|(A\,x + b) - (A\,z + b)\|$$

On obtient le résultat désiré après simplification. ■

1.3 Cas d'une fonction convexe

Dans le cas d'une fonction convexe, on peut démontrer le résultat suivant :

Lemme 2 (BAILLON–HADDAD)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction différentiable. Soit $L > 0$. Alors les deux affirmations suivantes sont équivalentes :

(i) J est **convexe** et L-régulière.

(ii) $\forall\, (x, z) \in \mathcal{X}^2, \qquad \langle \nabla J(x) - \nabla J(z), x - z \rangle \geq \dfrac{1}{L} \, \|\nabla J(x) - \nabla J(z)\|^2$

Dans ce cas, on dit que $(1/L)\nabla J$ est un opérateur *fermement non-expansif* ou que ∇J est un opérateur $(1/L)$-*co-coercif*.

DÉMONSTRATION : On démontre séparément les deux sens de l'équivalence.

- **(i) $\Longrightarrow$ (ii) :** Commençons par remarquer que, pour tout $(x, z) \in \mathcal{X}^2$ et pour tout $p \in \mathcal{X}$, on a

$$\langle \nabla J(z) - \nabla J(x), p - x \rangle - \frac{L}{2} \, \|p - x\|^2 = \frac{1}{2\,L} \, \|\nabla J(z) - \nabla J(x)\|^2 - \frac{L}{2} \left\| \frac{\nabla J(z) - \nabla J(x)}{L} - (p - x) \right\|^2$$

On passe à la borne supérieure sur $p \in \mathcal{X}$; on obtient que

$$\sup_{p \in \mathcal{X}} \left\{ \langle \nabla J(z) - \nabla J(x), p - x \rangle - \frac{L}{2} \, \|p - x\|^2 \right\} = \frac{1}{2\,L} \, \|\nabla J(z) - \nabla J(x)\|^2$$

de sorte que

$$\sup_{p \in \mathcal{X}} \left\{ \langle \nabla J(z) - \nabla J(x), p \rangle - \frac{L}{2} \, \|p - x\|^2 \right\} = \langle \nabla J(z) - \nabla J(x), x \rangle + \frac{1}{2\,L} \, \|\nabla J(z) - \nabla J(x)\|^2$$

Puisque J est L-régulière, le lemme de descente (proposition 4, qui se démontre de manière indépendante) assure que :

$$\langle \nabla J(z) - \nabla J(x), p \rangle - \frac{L}{2} \, \|p - x\|^2 \leq J(x) - J(p) + \langle \nabla J(z), p \rangle - \langle \nabla J(x), x \rangle$$

On passe à nouveau à la borne supérieure sur $p \in \mathcal{X}$:

$$\langle \nabla J(z) - \nabla J(x), x \rangle + \frac{1}{2\,L} \, \|\nabla J(z) - \nabla J(x)\|^2 \leq -\langle \nabla J(x), x \rangle + J(x) + \sup_{p \in \mathcal{X}} \Big\{ \langle \nabla J(z), p \rangle - J(p) \Big\}$$

Après réarrangement, on obtient que

$$\langle \nabla J(x), x \rangle - J(x) \leq \langle \nabla J(x) - \nabla J(z), x \rangle - \frac{1}{2\,L} \, \|\nabla J(z) - \nabla J(x)\|^2 + \sup_{p \in \mathcal{X}} \Big\{ \langle \nabla J(z), p \rangle - J(p) \Big\}$$

La fonction J est convexe et différentiable ; on a donc pour tout $p \in \mathcal{X}$

$$J(p) \geq J(z) + \langle \nabla J(z), p - z \rangle \qquad \text{soit} \qquad \langle \nabla J(z), z \rangle - J(z) \geq \langle \nabla J(z), p \rangle - J(p)$$

Autrement dit,

$$\langle \nabla J(z), z \rangle - J(z) = \sup_{p \in \mathcal{X}} \Big\{ \langle \nabla J(z), p \rangle - J(p) \Big\}$$

On a donc démontré que, pour tout $(x, z) \in \mathcal{X}^2$,

$$\langle \nabla J(x), x \rangle - J(x) \leq \langle \nabla J(x) - \nabla J(z), x \rangle - \frac{1}{2\,L} \, \|\nabla J(z) - \nabla J(x)\|^2 + \langle \nabla J(z), z \rangle - J(z)$$

En réécrivant cette inégalité en inversant le rôle de x et de z, et en sommant les deux inégalités, on démontre finalement le résultat annoncé.

- **(ii) $\Longrightarrow$ (i) :** Il suffit d'utiliser l'inégalité de CAUCHY–SCHWARZ pour majorer le produit scalaire et ainsi obtenir la définition de la régularité. La convexité découle alors de la monotonie du gradient. ■

Enfin, dans le cas d'une fonction J différentiable et fortement convexe de module α, on rappelle qu'on a la relation suivante :

$$\forall\, (x, z) \in \mathcal{X}^2, \qquad J(z) \geq J(x) + \langle \nabla J(x), z - x \rangle + \frac{\alpha}{2} \, \|x - z\|^2$$

On en déduit qu'une fonction régulière et fortement convexe est encadrée par deux fonctions quadratiques. La distance entre ces deux fonctions dépend du rapport entre la constante de LIPSCHITZ du gradient et le module de forte convexité de la fonction.

Définition 4 (Conditionnement d'une fonction régulière fortement convexe)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction L-régulière et fortement convexe de module α. On définit le *conditionnement* de J par

$$\kappa = \frac{L}{\alpha}$$

REMARQUE : D'après ce qui précède cette définition, le conditionnement est nécessairement un nombre positif supérieur à 1.

Le conditionnement d'une fonction régulière et fortement convexe est une quantité qui apparaît parfois dans les preuves de convergence d'algorithmes d'optimisation impliquant de telles fonctions.

> Plus le conditionnement de J est petit (c'est-à-dire, proche de 1), plus les deux fonctions quadratiques encadrant J sont proches. *A contrario*, plus il est grand, et plus ces deux fonctions quadratiques sont éloignées. Or, si ces deux fonctions sont proches, le comportement de J est *a priori* mieux contrôlé (l'espace dont elle dispose est plus réduit). C'est pourquoi un faible conditionnement est intuitivement préférable à un grand conditionnement.

EXEMPLE

Conditionnement d'une forme quadratique convexe. Considérons la forme quadratique suivante :

$$J : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \dfrac{1}{2} \langle A^* A\, x, x \rangle \end{array} \right.$$

avec A une matrice. Cette fonction est différentiable, de gradient $\nabla J(x) = A^* A\, x$. En utilisant la base orthonormale des vecteurs propres de la matrice symétrique $A^* A$, on prouve que

$$\forall\, (x,z) \in \mathcal{X}^2, \qquad \langle A^* A\, x - A^* A\, z, x - z \rangle \geq \lambda_{\min} \|x - z\|^2$$

où $\lambda_{\min}$ la plus petite valeur propre de $A^* A$ (il est aisé de vérifier que les valeurs propres de $A^* A$ sont toutes positives). On en déduit que J est une fonction fortement convexe, de module $\lambda_{\min}$ (cf. Corollaire 1 du module A1). Par ailleurs, ∇J est linéaire et on a donc

$$\forall\, (x,z) \in \mathcal{X}^2, \qquad \|\nabla J(x) - \nabla J(z)\| \leq |||A^* A|||\, \|x - z\| = \lambda_{\max} \|x - z\|$$

avec $\lambda_{\max}$ la plus grande valeur propre de $A^* A$. Ainsi, J est $|\lambda_{\max}|$-régulière. Il s'ensuit que le conditionnement de J vaut

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$$

On reconnaît là la définition usuelle du conditionnement de la matrice $A^* A$.

2 Lemme de descente

Terminons ce module avec un résultat qui explique pourquoi les fonctions régulières jouent un rôle central dans certaines méthodes d'optimisation. Le résultat est connu sous le nom de *lemme de descente* et apparaît dans de nombreuses preuves de convergence d'algorithmes d'optimisation du premier ordre.

Proposition 4 (Lemme de descente)

Soit $J : \mathcal{X} \to \mathbb{R}$ une fonction L-régulière. Alors on a

$$\forall\, (x,z) \in \mathcal{X}^2, \qquad J(z) \leq J(x) + \langle \nabla J(x), z - x \rangle + \frac{L}{2} \|x - z\|^2$$

DÉMONSTRATION : Soit $(x,z) \in \mathcal{X}^2$. Considérons la fonction définie sur $[\,0\,;1\,]$ par

$$\forall\, t \in [\,0\,;1\,], \qquad f(t) = J\big(x + t\,(z-x)\big) - \big\langle \nabla J(x), x + t\,(z-x) \big\rangle$$

Cette fonction est dérivable ; il est alors aisé de vérifier que

$$\begin{aligned} J(z) - J(x) - \langle \nabla J(x), z - x \rangle &= \Big[f(t) \Big]_0^1 = \int_0^1 f'(t)\, \mathrm{d}t \\ &= \int_0^1 \big\langle \nabla J\big(x + t\,(z-x)\big) - \nabla J(x), z - x \big\rangle\, \mathrm{d}t \end{aligned}$$

Or, on a
$$\begin{aligned} \left| \int_0^1 \big\langle \nabla J\big(x + t\,(z-x)\big) - \nabla J(x), z - x \big\rangle\, \mathrm{d}t \right| & \\ \leq \int_0^1 \left| \big\langle \nabla J\big(x + t\,(z-x)\big) - \nabla J(x), z - x \big\rangle \right| \mathrm{d}t & \end{aligned}$$

En utilisant l'inégalité de CAUCHY–SCHWARZ et la régularité de J, on obtient alors

$$\int_0^1 \left| \big\langle \nabla J\big(x + t\,(z-x)\big) - \nabla J(x), z - x \big\rangle \right| \mathrm{d}t \leq L \int_0^1 t\, \|z - x\|^2\, \mathrm{d}t$$

et on en déduit le résultat souhaité après calcul de l'intégrale. ■

Pour aller plus loin

Descentes de gradient explicite. L'hypothèse de régularité apparaît dans de nombreux algorithmes d'optimisation du premier ordre, dès lors que les itérations sont définies à l'aide du gradient de tout ou partie de la fonction objectif (dans le second cas, seule la partie concernée est supposée régulière). Parmi ces algorithmes figurent les méthodes de gradient explicite (module B2), les algorithmes utilisant l'éclatement explicite-implicite (*Forward-Backward Splitting*, module B4) tels que les algorithmes PALM et ASAP (module B5. Dans ces cas, les preuves de convergence reposent fortement sur le lemme de descente que nous avons établi dans ce module.

Conjuguée convexe. On verra dans le module A8 que la régularité d'une fonction convexe se traduit par la forte convexité de sa conjuguée convexe. La démonstration de ce résultat repose en partie sur le lemme de Baillon–Haddad. La conjuguée convexe apparaît d'ailleurs furtivement dans la preuve du lemme de Baillon–Haddad.