Solutionnaire 1

Frederic Boileau

28th February 2019

J'introduis quelques notations. En analyse convexe on dénote l'ensemble des fonctions convexes *propres* sur un espace euclidien $\mathbb{E}$ (dans notre cas simplement $\mathbb{R}^n$) par $\Gamma(\mathbb{E})$.

$$\Gamma(\mathbb{E}) \triangleq \{f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \text{f propre et convexe}\}$$

Une fonction *propre* est une fonction pour laquelle il existe au moins un x t.q. $f(x) < +\infty$ et pour laquelle pour tout x $f(x) > -\infty$. On écarte ainsi certains cas pathologiques.
J'utilise la notation de Evans pour la dérivation. Ainsi $D_x f$ est la dérivée de f par rapport à x, si c'est un nombre, un vecteur (comme pour le gradient) ou une matrice (jacobien) dépend du contexte (de f et de x) dans tous les cas D est toujours une approximation linéaire de f.
Soit A une matrice. Si A est définie positive (resp. semi-définies positives) on écrit $A \succ 0$ (resp $A \succeq 0$).

1 Direction de Descente

Soit f une fonction convexe différentiable sur $\mathbb{R}^n$ et x et y deux points de $\mathbb{R}^n$ t.q. $f(y) < f(x)$. Montrer que $y - x$ est une direction de descente de f en x. Notons qu'aucun ordre n'est imposé sur x et y.

La preuve découle directement de la définition de la convexité donnée en cours:

$$\begin{aligned}
f(\alpha y + (1-\alpha)x) &\leq \alpha f(y) + (1-\alpha) f(x) \\
f(x + \alpha(y - x)) &\leq \alpha f(y) + (1-\alpha) f(x) \\
f(x + \alpha(y - x)) &< \alpha f(x) + (1-\alpha) f(x) \\
f(x + \alpha(y - x)) &< f(x)
\end{aligned}$$

2 Système linéaire et optimisation

Prouvez que tout système linéaire $Ax = b$, avec A définie positive, peut se reformuler comme un problème d'optimisation sans contrainte.

Pour la reformulation nous avons plusieurs critères; premièrement on veut que le minimum global corresponde à la solution, dans un second temps si le problème est convexe nous avons des garanties quant à la recherche de ce minimiseur global. Il y a deux formulations simples, je traite les deux.

2.1 Forme Quadratique

Le reformulation la plus simple est de prendre la forme quadratique suivante:

$$f(x) = \frac{1}{2}x^\intercal Ax - x^\intercal b \tag{1}$$

Le calcul de dérivés par rapport à des vecteurs ou des matrices peut être laborieux mais l'identité suivante est très importante:

$$D_x(x^\intercal Ax) = x^\intercal(A + A^\intercal) \tag{2}$$

Ainsi nous avons:

$$D_x(f) = Ax - b \qquad D_x^2(f) = A \succ 0 \tag{3}$$

La première équation montre que le gradient de la fonction est nul ssi le x est une solution du système linéaire. La deuxième équation montre que la deuxième dérivée (le hessien) de f par rapport est *définie positive* (par supposition sur A). La fonction est donc strictement convexe et atteint son minimum ssi le gradient est nulle, ce qui n'arrive que lorsque $Ax = b$.

Notons que pour une fonction à une variable $D_x^2(f) > 0$ implique que f est strictement convexe en x. Pour une fonction à plusieures variables, $f : \mathbb{R}^n \to \mathbb{R}$, nous avons que $D_x^2(f) = \nabla_x^2(f) \succ 0$ implique f strictement convexe. Nous voyons donc que la notation est appropriée puisque l'on généralise la notion de 'plus grand que zéro.'

2.2 Norme euclidienne

Une autre fonction objectif qui a été choisi par certains étudiants est la suivante:

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 \tag{4}$$

Cette fonction est aussi convexe et d'ailleurs C^2. De plus la convexité ne dépend *pas* du fait que la matrice A en tant que telle soit pos-def (definie positive). La preuve est cependant moins directe.

Soit $h(x) = \frac{1}{2}\|x\|_2^2$ et $g(x) = Ax - b$ et donc $f(x) = h(g(x))$.

D'abord nous calculons le gradient de la norme euclidienne au carré.

$$D_{x_i}h(g(x)) = D_{x_i}\frac{1}{2}\|g(x)\|_2^2 = D_{x_i}\frac{1}{2}\langle g(x), g(x)\rangle = D_{x_i}\frac{1}{2}\sum g(x_i)^2 = g'(x_i)g(x_i) \tag{5}$$

Nous avons donc $D\frac{1}{2}\|x\|_2^2 = x^\intercal$ (g est la fonction qui ne fait rien, la fonction identité.) La transpose vient du fait que Dh est une application linéaire sur $\mathbb{R}^n$ qui approxime la fonction h. Ainsi une notation plus claire serait $Dh(x) = \langle x, \cdot\rangle = x^\intercal$

Maintenant nous pouvons appliquer la dérivation en chaine pour trouver la dérivée de la fonction(4).

$$f'(x) = h'(g(x))g'(x) = g'(x)^\intercal g(x) = A^\intercal(Ax - b) \tag{6}$$

Finalement nous avons donc

$$D^2 f = A^\intercal A \tag{7}$$

Or c'est un fait connu en algèbre linéaire que pour toute matrice X invertible $X^\intercal X > 0$. Nous pouvons le démontrer rapidement:

$$x^\intercal AA^\intercal x = (A^\intercal x)^\intercal(A^\intercal x) = \|Ax\|_2^2 \tag{8}$$

Si A est invertible alors clairement $\|Ax\|_2^2 > 0$ pour tout $x \neq 0$. Au final nous avons donc que $D^2 f(x) > 0$ est la fonction objectif est donc strictement convexe avec comme minimum global la solution du système linéaire.

Notes:

Cette petite preuve montre l'approche la plus directe pour trouver le gradient d'une fonction, c'est à dire trouver les dérivées partielles pour un indice arbitraire et utiliser les regles de dérivation en chaine.

De plus; $x^\intercal$ est fondamentalement un autre objet que x[1]

$x^\intercal$ est une *fonction* qui prend un vecteur et retourne un nombre, x est un vecteur.

[1] pour ceux que ca interesse le premier est un element de l'espace des applications lineaires sur $\mathbb{R}^n$, espace que l'on appelle le duel