

CC2 - Optimisation

Durée : 2h30.

Seuls le polycopié de cours et les notes personnelles de cours sont autorisés.

Exercice 1. QUESTIONS DE BASE

1. Soient $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ et $b \in \mathbb{R}^m$. Calculer le gradient et la hessienne de $f : x \mapsto \frac{1}{2} \|Ax - b\|_2^2$ en détaillant le calcul.

Correction : On a $\nabla f(x) = A^T(Ax - b)$. Faire un calcul des variations en utilisant le produit scalaire usuel.

2. Quelle est la différence si $x \in \mathbb{C}^n$, $A \in \mathbb{C}^{m \times n}$ et $b \in \mathbb{C}^m$.

Correction : Ce n'est pas A^T , mais A^* , la transconjuguée qui apparait dans les expressions précédentes.

3. Donner la hessienne de f en tout point dans le cas réel.

Correction : On a $Hf(x) = A^T A$.

4. On rappelle que le simplexe en dimension n est défini par

$$\Delta_n = \{x \in \mathbb{R}^n, x_i \geq 0, \forall i \in \{1, \dots, n\}, \sum_{i=1}^n x_i = 1\}.$$

Dessiner le simplexe en dimension 2 et 3.

Correction : Le segment qui lie (0,1) à (1,0) en dimension 2 et un triangle qui a pour extrémités les points (1,0,0), (0,1,0) et (0,0,1) en dimension 3.

5. Déterminer le cône normal au simplexe pour $x \in \text{ri}(\Delta_n)$.

Correction : C'est l'ensemble $\{\alpha e, \alpha \in \mathbb{R}\}$ où e est le vecteur (1,1,...,1).

6. On souhaite projeter un point $x_0 \in \mathbb{R}^n$ sur Δ_n . Ecrire le problème d'optimisation associé au problème de projection.

Correction : C'est

$$\min_{x \in \Delta_n} \|x - x_0\|_2^2.$$

7. Donner les conditions d'optimalité de ce problème.

Correction : Elles s'écrivent $x - x_0 \parallel e$, c'est-à-dire que $x - x_0$ est orthogonal au simplexe.

Exercice 2. DESCENTES DE SOUS-GRADIENT

Le but de cet exercice est de mieux comprendre les propriétés intrinsèques des descentes de sous-gradient. On considère le problème suivant :

$$\min_{x \in \mathbb{R}^n} f(x)$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est convexe. On considère aussi la descente de sous-gradient suivante :

$$\begin{aligned} x_1 &\in \mathbb{R}^n \\ x_{k+1} &= x_k - \tau_k \frac{\eta_k}{\|\eta_k\|} \quad \text{avec } \eta_k \in \partial f(x_k) \\ \text{Si } 0 &\in \partial f(x_{k+1}) \text{ arrêter, sinon, itérer.} \end{aligned}$$

1. Soit $f(x) = |x|$ avec $x \in \mathbb{R}$. Rappeler l'expression du sous-différentiel de f .

Correction : voir cours.

2. Pour les choix suivants, dire si l'algorithme converge et si c'est le cas, donnez un taux de convergence approximatif.

(a) $x_1 = 1, \tau_k = 1, \forall k \in \mathbb{N}$.

Correction :

convergence en 1 itération.

(b) $x_1 = 1, \tau_k = 2, \forall k \in \mathbb{N}$.

Correction :

Pas de convergence (ca donne la suite $(1, -1, 1, -1, \dots)$).

(c) $x_1 = 10, \tau_k = \frac{1}{k^2}$.

Correction : Ca ne converge pas car la série $\sum \frac{1}{k^2}$ converge et a une somme égale à $\pi^2/6 < 10$.

(d) $x_1 = 2, \tau_k = \frac{1}{\sqrt{k}}$.

Correction : Ca converge d'après le cours avec un taux en $O(1/\sqrt{k})$

Ces exemples montrent qu'en général $(\tau_k)_{k \in \mathbb{N}}$ doit tendre vers 0 pour que la méthode converge. Cependant, il ne doit pas converger trop rapidement vers 0. Dans la suite de l'exercice, on s'intéresse aux propriétés géométriques de l'algorithme et on cherche une condition suffisante pour que la suite $(x_k)_{k \in \mathbb{N}}$ soit bornée.

3. On considère la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f(x, y) = |x| + 2|y|$.

(a) Déterminer le sous-différentiel de f au point $(1, 0)$.

Correction : On a

$$\partial f(1, 0) = \{(1, \alpha), \alpha \in [-2, 2]\}.$$

(b) Dessiner 3 lignes de niveau de f , dont celle de niveau 1.

(c) Montrer sur le même schémas que les directions données par le sous-différentiel ne sont pas nécessairement des directions de descente (i.e. des direction susi-vants lesquelles la fonction décroît).

Correction : A faire.

4. A partir de maintenant $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction convexe arbitraire. Une des étapes cruciales de la preuve de convergence d'un algorithme consiste à montrer que $(x_k)_{k \in \mathbb{N}}$ est bornée. Soit x^* un minimiseur de f et $r_k = \|x_k - x^*\|$. Comme les sous-gradients ne sont pas forcément des directions de descente, on étudie la distance au minimiseur r_k plutôt que $f(x_k) - f(x^*)$.

- (a) Donner une expression de r_{k+1}^2 qui dépend de r_k , η_k , τ_k et du produit scalaire usuel.

Correction :

$$\begin{aligned} r_{k+1} &= \left\| x_k - \tau_k \frac{\eta_k}{\|\eta_k\|_2} - x^* \right\|_2^2 \\ &= r_k^2 + \tau_k^2 - 2 \frac{\tau_k}{\|\eta_k\|_2} \langle \eta_k, x_k - x^* \rangle \end{aligned}$$

- (b) Montrer que $\langle \eta_k, x^* - x_k \rangle \leq f(x^*) - f(x_k)$.

Cette inégalité est une conséquence directe de la convexité de f . On a donc :

$$\langle \eta_k, x^* - x_k \rangle \leq f(x^*) - f(x_k) \leq 0.$$

- (c) En combinant les inégalités précédentes, montrer que $r_{k+1}^2 \leq r_k^2 + \tau_k^2$.

Correction : Direct d'après la question précédente.

- (d) Soit $\tau_k = \frac{1}{k^\alpha}$. Dédurre des questions précédentes une condition suffisante sur α pour que la suite des itérées soit bornée.

Correction : On somme ces inégalités de $k = 1$ à $k = n$. On obtient une suite télescopique et :

$$r_{n+1}^2 \leq r_1^2 + \sum_{k=1}^n \tau_k^2.$$

Une condition suffisante pour que la suite soit bornée est donc que la série $\sum_{k=1}^{\infty} \tau_k^2$ soit convergente. La série $\sum \frac{1}{k^{2\alpha}}$ est convergente si et seulement si $\alpha > \frac{1}{2}$.

Exercice 3. PROJECTIONS ALTERNEES

Soient A et B des ensembles fermés, non vides, convexes de \mathbb{R}^n . On définit :

$$\forall x \in \mathbb{R}^n, d_A(x) = \min_{a \in A} \|x - a\|_2 \quad \text{et} \quad d_B(x) = \min_{b \in B} \|x - b\|_2.$$

$$\forall x \in \mathbb{R}^n, d_A^2(x) = \min_{a \in A} \|x - a\|_2^2 \quad \text{et} \quad d_B^2(x) = \min_{b \in B} \|x - b\|_2^2.$$

La projection d'un point $x \in \mathbb{R}^n$ sur A est notée $p_A(x)$.

Dans cet exercice, on considère le problème d'optimisation non contraint suivant. Minimiser :

$$f(x) = \frac{1}{2} \max(d_A^2(x), d_B^2(x)).$$

1. Comprendre les fonctions distances et f graphiquement.

- (a) On se concentre d'abord sur un problème dans \mathbb{R} . Soit $A = \{0\}$ et $B = \{1\}$. Tracer le graphe de d_A^2 ainsi que ceux de d_B^2 et de f .

- (b) Est-ce que f est différentiable? Convexe?

Correction : Elle est convexe et non différentiable.

- (c) Soient $A = [0, 1]$ et $B = [1, 2]$. Tracer les graphes de d_A^2 , de d_B^2 et de f .

- (d) On se place maintenant dans \mathbb{R}^2 . Soit $A = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$. Dessiner A ainsi que les lignes de niveau 1, 4 et 9 de d_A^2 .
- (e) **A partir de maintenant et jusqu'à la fin, on suppose que $A \cap B \neq \emptyset$.**
Donner les valeurs de $f(x)$ pour $x \in A \cap B$, $x \in A$, $x \in B$. Déduire les solutions globales ainsi que la valeur du minimum du problème : $\min_{x \in \mathbb{R}^n} f(x)$.

Correction : Pour $x \in A \cap B$, $f(x)=0$. Pour $x \in A$, $f(x) = d_B^2(x)$. Pour $x \in B$, $f(x) = d_A^2(x)$.

2. Propriétés analytiques des fonctions distances.

- (a) Montrer que d_A et d_B sont convexes.

Correction :

Soient x_1 et x_2 dans \mathbb{R}^n . On note a_1 et a_2 la projection de x_1 et x_2 sur A . On a donc $d_A(x_1) = \|x_1 - a_1\|_2$ et $d_A(x_2) = \|x_2 - a_2\|_2$. Soit $\lambda \in [0, 1]$. Par convexité de A on a $(\lambda a_1 + (1 - \lambda)a_2) \in A$. On a finalement :

$$\begin{aligned} d_A(\lambda x_1 + (1 - \lambda)x_2) &= \min_{a \in A} \|\lambda x_1 + (1 - \lambda)x_2 - a\|_2 \\ &\leq \|\lambda x_1 + (1 - \lambda)x_2 - (\lambda a_1 + (1 - \lambda)a_2)\|_2 \\ &\leq \|\lambda(x_1 - a_1) + (1 - \lambda)(x_2 - a_2)\|_2 \\ &\leq \lambda \|x_1 - a_1\|_2 + (1 - \lambda) \|x_2 - a_2\|_2 \\ &= \lambda d_A(x_1) + (1 - \lambda) d_A(x_2). \end{aligned}$$

- (b) Montrer que d_A^2 and d_B^2 sont convexes.

Correction : d_A^2 et d_B^2 sont la composition d'une fonction positive, monotone croissante et convexe ($t \mapsto t^2$) avec une fonction convexe (d_A ou d_B). Elles sont donc convexes.

- (c) Montrer que f est convexe.

Correction : Le max de deux fonctions convexes est convexe.

- (d) Montrer que d_A est Lipschitz de constante 1.

Correction : Soient x_1 et x_2 deux éléments de \mathbb{R}^n . On suppose que $d_A(x_1) \geq d_A(x_2)$. On a donc

$$\begin{aligned} |d_A(x_1) - d_A(x_2)| &= \left| \|x_1 - p_A(x_1)\|_2 - \|x_2 - p_A(x_2)\|_2 \right| \\ &= \left| \|x_1 - p_A(x_1)\|_2 - \|x_2 - p_A(x_2)\|_2 \right| \\ &\leq \left| \|x_1 - p_A(x_2)\|_2 - \|x_2 - p_A(x_2)\|_2 \right| \\ &\leq \|x_1 - x_2\|_2 \end{aligned}$$

Le passage de la ligne 1 à 2 est lié à l'hypothèse de départ. Le passage de la ligne 2 à la ligne 3 est lié au fait que $\min_{a \in A} \|x_1 - a\|_2 = \|x_1 - p_A(x_1)\|_2 \leq \|x_1 - p_A(x_2)\|_2$. Le passage de la ligne 3 à la ligne 4 est dû à la deuxième inégalité triangulaire.

On peut procéder de la même façon si $d_A(x_2) \geq d_A(x_1)$.

On obtient donc dans tous les cas $|d_A(x_1) - d_A(x_2)| \leq \|x_1 - x_2\|_2$.

(e) Montrer que pour tout $x \in \text{int}(A)$, d_A est différentiable et que $\nabla d_A(x) = 0$.

Correction : Pour $x \in \text{int}(A)$, $d_A(x) = 0$. Le gradient est donc nul en tout point $x \in \text{int}(A)$, car sur un (petit) ouvert autour de x , d_A est identiquement nulle.

3. **Un algorithme.** On s'intéresse maintenant à la résolution numérique de $\min_{x \in \mathbb{R}^n} f(x)$.

On considère l'algorithme suivant :

Initialisation : $x_0 \in A \setminus B$, $k := 0$;

Tant que $x_k \notin A \cup B$,

i. Calculer la solution x_{k+1} du sous-problème :

$$(P_k) \quad \min_{x \in \mathbb{R}^n} f(x_k) + \langle g_k, x - x_k \rangle + \|x - x_k\|_2^2$$

où $g_k \in \partial f(x_k)$ est un sous-gradient de f au point x_k .

ii. $k \leftarrow k + 1$;

(a) Ecrire les conditions d'optimalité du premier ordre de (P_k) . En déduire l'expression de x_{k+1} en fonction de x_k et de g_k .

Correction : Les conditions d'optimalité sont $g_k + 2(x - x_k) = 0$ soit encore : $x_{k+1} = x_k - g_k/2$.

(b) Donner l'expression du sous-différentiel $\partial f(x)$ de f au point x dans le cas où $x \in A$ et $x \notin B$. Faire de même dans le cas similaire où $x \in B$ et $x \notin A$.

Correction : Dans le premier cas, $\partial f(x) = \{\nabla d_B^2(x)\} = \{2(x - p_B(x))\}$. Dans le second, $\partial f(x) = \{2(x - p_A(x))\}$.

(c) Donner l'expression du sous-différentiel de f en tout point. Où se trouvent les singularités de f (les points de non différentiabilité de f) ?

Correction :

On peut écrire $f(x) = \max(f_1(x), f_2(x))$ avec $f_1 = d_A^2$ et $f_2 = d_B^2$. Le sous-différentiel d'une fonction définie comme un maximum de deux fonctions, est donné par le résultat de cours suivant :

$$\partial f(x) = \text{conv}(\{\nabla f_i(x), i \in I(x), I(x) = \{i, f(x) = f_i(x)\}\}).$$

Ce résultat peut être réécrit comme suit :

Si $d_A^2(x) = d_B^2(x)$

$$\partial f(x) = \{t2(x - p_A(x)) + (1 - t)2(x - p_B(x)), t \in [0, 1]\}.$$

Si $d_A^2(x) > d_B^2(x)$

$$\partial f(x) = \{2(x - p_A(x))\}.$$

Si $d_B^2(x) > d_A^2(x)$

$$\partial f(x) = \{2(x - p_B(x))\}.$$

Les singularités peuvent apparaître uniquement aux points x tels que $d_A(x) = d_B(x)$.

- (d) D eduire de la question (3.b) l'expression de $g_0 \in \partial f(x_0)$ et x_1 . Calculer $g_1 \in \partial f(x_1)$, et x_2 . De fa on plus g n rale, quelles sont les expressions de g_k et de x_{k+1} ?

Correction :

Comme $x_0 \in A \setminus B$, on a $\partial f(x_0) = 2(x_0 - p_B(x_0))$. Donc

$$x_1 = x_0 - (x_0 - p_B(x_0)) = p_B(x_0).$$

Si $p_B(x_0) \in A \cup B$, l'algorithme s'arr te. Sinon, $\partial f(x_1) = 2(x_0 - p_B(x_0))$ et $x_2 = p_A(x_1) = p_A(p_B(x_0))$.

De fa on g n rale, x_k est une succession de projections sur A puis sur B , d'o  le nom projections altern es.

- (e) Montrer la trajectoire de l'algorithme (i.e. les it r es $(x_k)_{k \in \mathbb{N}}$) quand $x_0 = (5, 5)$, A est la droite d' quation $y = x$, et B est la droite d' quation $y = 0$.

Correction :

On observe une convergence lin aire.

Cet algorithme est appel  l'algorithme des projections altern es. C'est un grand standard de l'optimisation.