<u>Mathématique pour l'intelligence artificielle (MIA)</u>

Séance 3

Scribe: BRIZZI Anouk & OXISOGLOU Thomas

Faculté des sciences AMU – M2 I3A
2023/2024

Table des matières :

1. Résolution d'exercice de la planche précédente (Planche 2)

Exercice 2 :

1) $[v_1, \dots, v_m][v_1, \dots, v_m]^T$

$[v_1, \dots, v_m]^T[v_1, \dots, v_m]$: produit scalaire entre vecteur orthonormé donne $\rightarrow$ 0 si les vecteurs sont différent

$\hookrightarrow$ 1 sinon

$$V^T V : \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

On peut le montrer par : • AB = Id

• BA = Id

Si on trouve C linéaire et inverse, alors :

BC = Id

$\hookrightarrow$ pour prouver que $C=A$, on peut prendre : $ABC = C = A$

$$V^T V = Id \quad \equiv \quad V V^T = Id$$

2. Optimisation

2.1. Optimisation sans contraintes

Les fonctions de la classe C^1 sont continues dérivables et admettent donc plusieurs dérivées partielles.

Une matrice Jacobienne est définie de $\mathbb{R}^n \longmapsto \mathbb{R}^m \times \mathbb{R}^n$

Rappel : On dit d'une fonction qu'elle est la composée d'autres fonctions quand sa forme est un ensemble de ces fonctions. En générale on note la composition par le signe "$\circ$".

Exemple : On note $h = g \circ f$, on a :
$f : x \to Ax - y$ et $g : z \to \|z\|_2^2$
D'où : $h(x) = \|Ax - y\|^2$

Une chain rule est tel que : $J\, f \circ g(a) = J\, f(\underline{g(a)})\, J\, g(a)$

$\hookrightarrow$ domaine de départ = arrivée de g

2.1.1. Analyse

La matrice Hessienne de f ce défini comme :

$\forall f : U \subset \mathbb{R}^m \longmapsto \mathbb{R}$

$$\nabla^2 f = H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m \partial x_1} & \frac{\partial^2 f}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_m^2} \end{bmatrix}$$

2.1.2. Convexité

Si deux points d'un ensemble sont reliés entre eux, tous les points sur la droite ainsi formée sont dans l'ensemble.

Si f est une fonction strictement convexe définie sur un ensemble convexe $X \subset \mathbb{R}^n$, alors f admet au plus un minimum global. Cependant, de même façon, si cette fois ci f est convexe non stricte, alors il y a plusieurs minimum globaux de f possibles.

2.2. Optimisation sous contraintes

Exemple :

On veut $x_1 + x_2$ à min, avec comme contrainte $x_1 \cdot x_2 = 10$

On peut écrire : $-x_1 \cdot x_2 \geq 10$

$-x_1 \cdot x_2 - 10 \geq 0$

2.2.1. Condition nécessaire d'optimalité : Intuition

Sur une figure donnée, lorsque les gradient calculés préalablement sont affichés, on ne peut se déplacer en partant de chacun d'eux que de façon orthogonale, dans la continuité de f.

Si λ_1^* est une contrainte active, elle est égale à 0 et donc la condition nécessaire d'optimalité est nulle ($\lambda_1^* c_1(x^*) = 0$).

2.3. Dualité

L'idée générale de la dualité est de voir un même problème sous un angle différent. Cela permet souvent de déduire des propriétés différentes.

Le Lagrangien peut s'écrire de deux façon, comme une somme ou une soustraction, en fonction de l'inégalité :

$L(x,\mu) = f(x) - \mu' g(x)$

ou $L(x,\mu) = f(x) + \mu' g(x)$

Dans les problème de dualité, la fonction $q(\mu)$ n'est pas forcément une fonction continue.

On appelle "Duality Cap" la différence entre q* et f*.

Sur la figure a) diapo 86, on cherche la droite $q(\mu)$ qui minimise sa position sur l'axe y, tel que la droite passe par au moins un point de la figure.

3. Exercices

Exercice 1 :

1) $\frac{\partial f_{x,y}}{\partial x} \longmapsto 2xy$

$\frac{\partial f_{x,y}}{\partial y} \longmapsto x^2$

2) $\|x\|_2 = \sqrt{\|x\|_2^2} = \sqrt{x^T \cdot x^*}$

$f : \mathbb{R}^n \longmapsto \mathbb{R}^+$ $\qquad J_f(x) = 2x^*$
$x \longmapsto x^T \cdot x$

$g : \mathbb{R}^+ \longmapsto \mathbb{R}$ $\qquad J_g(y) = \frac{1}{2\sqrt{y}}$
$x \longmapsto \sqrt{x}$

$h : x \longmapsto \|x\|_2$ $\qquad h = \frac{2a}{2\sqrt{a}} = \frac{a}{\|a\|_2}$
$h = g \circ f$

$^* x^T \cdot x = \sum_{i=1}^{m} x_i^2$

$\frac{\partial x^T \cdot x}{x} = [2x_1, 2x_2, \dots, 2x_m]$
$= 2x$

Conditions d'optimalité : La réciproque est non vrai.

Exercice 2 :

1) Le théorème dis que ce doit être ouvert, hors dans ce cas il est fermé donc cela ne peut pas marcher.

$\tilde{f} :]0, \pi[\longmapsto \mathbb{R}$
$x \longmapsto \sin(x)$

Notons x* un minimum globale de f :

Si x* $\in]0, \pi[$:

alors $\nabla \tilde{f}(x^*) = 0$

i.e $\cos(x^*) = 0$

C'est impossible

Si x* = 0, alors min f = 0

Si x * = , alors min f = 0

Donc le minimum globale de f est égale à 0.

Exercice 6 :

1) $f : x, y \longmapsto x^2 + y^2$
$c_1 : x + 1 - 1$
$c_2 : x, y \longmapsto 2 - y$
$c_3 : x, y \longmapsto y^2 - x$

KKT x^*, y^*
$x^* + y^* \geq 1$
$y^* \leq 2$
$y^{2*} \geq x^*$

$L(x, y, \lambda_1, \lambda_2, \lambda_3) = x^2 + y^2 - \lambda_1(x + y - 1) - \lambda_2(2 - y) - \lambda_3(y^2 - x)$
$\nabla_x L(x, y, \lambda) = 2x - \lambda_1 + \lambda_3$
$\nabla_y L(x, y, \lambda) = 2y - \lambda_1 + \lambda_2 - 2\lambda_3 y$

Donc : $x^* = \frac{\lambda_1 - \lambda_3}{2}$ $\qquad y^* = \frac{\lambda_1 - \lambda_2}{2(1 - \lambda_3)}$

$\lambda_1, \lambda_2, \lambda_3 \geq 0$
$\lambda_1(x^* + y^* - 1) = 0$
$\lambda_2(2 - y^*) = 0$
$\lambda_3(y^{2*} - x^*) = 0$

2) Pas de solution compatible si $\lambda_1, \lambda_2, \lambda_3 = 0$

λ_1 et $\lambda_3 \neq 0$ $\quad \lambda_2 = 0$ $\quad$ 2 solutions

Exercice 8 :

1) $f : x \longmapsto \sum_{i=1}^{m} x_i^2$
$c_1 : x \longmapsto \sum_{i=1}^{m} x_i \leq c$
$c_2 : x \longmapsto x_i \geq 0$
$f^* : x \longmapsto -x^T \cdot x$
$g^* : x \longmapsto \sum_{i=1}^{m} x_i - c$
$g_1, \dots, g_m, x \longmapsto -x_1, \dots, -x_m$
$L(x, \mu_1, \dots, \mu_m) = -x^T \cdot x - \sum_{i=1}^{m} \mu_i x_i + \mu \sum_{i=1}^{m} x_i - \mu c$
$= \sum_{i=1}^{m} -x_i^2 - \mu_i x_i + \mu x_i - \frac{\mu c}{m}$

$\inf(f_1(x_1) + f_2(x_2), x_1, x_2 \dots x_m + \dots + f_n(x_n))$
$\Leftrightarrow \inf_x(f_1(x_1)) + \inf_x f_2(x_2)) + \dots + \inf_x(f_n(x_n))$