Troisième cours

- DM1 en ligne ce soir (thomas.schatz.cogserver.net/teaching), à rendre pour le 6 Octobre (en main propre ou par email: thomas.schatz@univ-amu.fr

- Scribe ? (rendu = fichier .tex)

- Aujourd'hui

  - Fin de l'algèbre linéaire

    - Théorème de Cayley-Hamilton

    - Algèbre linéaire numérique

  - Début de l'Optimisation

Plan du cours

1. Introduction générale
2. Preuves (revue)
3. Algèbre linéaire (revue)
4. Optimisation (revue)
5. Optimisation sous contraintes
6. Probabilités (revue)
7. Statistique
8. **Théorie de l'apprentissage**
9. **Optimisation pour l'apprentissage**

Propriétés du produit matriciel

$$A \times (B \times C) = (A \times B) \times C$$

$$A \times (B + C) = A \times B + A \times C$$

$$(A + B) \times C = A \times C + B \times C$$

$$c(A \times B) = (cA) \times B = A \times (cB)$$

Espaces de fonctions linéaires

Espace de matrice, muni de l'addition et la multiplication matricielle et de la multiplication par un scalaire:

$$\mathcal{M}_{m,n}(\mathbf{R})$$

Espace de fonctions linéaires muni de l'addition et de la composition de fonctions linéaires et de la multiplication par un scalaire:

$$\mathcal{L}(E, F)$$

Déterminant et trace

A matrice carrée $n \times n$

$$\det(A) = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i} \right),$$

$\operatorname{sgn}(\sigma)$ est la parité du nombre d'élement dans une décomposition de σ en une séquence de transpositions (échange de deux éléments).

$$\det(AB) = \det(A)\det(B)$$

$$\operatorname{Tr}(A) = \sum_{i=1}^{n} a_{i,i} \quad \operatorname{Tr}(\mathrm{AB}) = \operatorname{Tr}(\mathrm{BA})$$

$$\operatorname{Tr}(A_1 A_2 \dots A_k) = \operatorname{Tr}(A_2 A_3 \dots A_k A_1) = \cdots = \operatorname{Tr}(A_k A_1 A_2 \dots A_{k-1})$$

Théorème de Cayley-Hamilton

Cayley-Hamilton theorem: for any $A \in \mathbf{R}^{n\times n}$ we have $\mathcal{X}(A) = 0$, where $\mathcal{X}(s) = \det(sI - A)$

Corollaire : pour tout entier naturel p, $A^p \in \mathrm{Vect}(I, A, A^2, \ldots, A^{n-1})$

Algèbre linéaire numérique

Exemple : produit matrice-matrice $A\colon\ m \times n,\ B\colon\ n \times p.$

$$\text{Produit matriciel } C = AB \ :$$

$$C = 0$$
$$\text{Boucle } i = 1..m,\ j = 1..p,\ k = 1..n \ :$$
$$c_{ij} = c_{ij} + a_{ik} b_{kj}$$

Ordre des trois boucles ?

i en premier $C = \begin{bmatrix} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{bmatrix} = AB = \begin{bmatrix} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{bmatrix}$

j en premier $C = \begin{bmatrix} c_1 \cdots c_p \end{bmatrix} = AB = \begin{bmatrix} Ab_1 \cdots Ab_p \end{bmatrix}$

k en premier $C = AB = \sum_{k=1}^{n} a_k \tilde{b}_i^T$

Algèbre linéaire numérique

Stabilité numérique

Normes de vecteurs

A function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is called a *vector norm* if it has the following properties:

1. $\|\mathbf{x}\| \geq 0$ for any vector $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for any vector $\mathbf{x} \in \mathbb{R}^n$ and any scalar $\alpha \in \mathbb{R}$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$$

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

$$\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n} |x_i|$$

Si Q est une matrice orthogonale,
$\|Qx\|_2 = \|x\|_2$

Application de la notion de norme: lien valeurs propres, valeurs singulières

$$\sigma_r \leq |\lambda| \leq \sigma_1$$

Application de la notion de norme: lien valeurs propres, valeurs singulières

$$\sigma_r \leq |\lambda| \leq \sigma_1$$

Éléments propres de $A^T A$ et AA^T ?

Normes de matrices

A matrix norm is a function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ that has the following properties:

- $\|A\| \geq 0$ for any $A \in \mathbb{R}^{m\times n}$, and $\|A\| = 0$ if and only if $A = 0$
- $\|\alpha A\| = |\alpha|\|A\|$ for any $m \times n$ matrix A and scalar α
- $\|A + B\| \leq \|A\| + \|B\|$ for any $m \times n$ matrices A and B

$$\|A\| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$$

$$\|A\|_2 = \sigma_1$$

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}.$$

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

Algèbre linéaire numérique

Stabilité numérique

Supposons qu'un processus nous donne la solution estimée $\hat{x}$ d'un système linéaire $Ax = b$ en effectuant toutes les opérations matricielles de manière exacte mais qu'il y a des erreurs d'arrondi dans le stockage en nombre flottants de A et de b dans la mémoire (en pratique les opérations matricielles effectuées pour trouver x introduisent davantage d'erreurs d'arrondis, ce modèle donne donc une borne supérieure sur la qualité possible d'un algorithme numérique pour résoudre des systémes linéaires).

Alors,

$$\text{si } \mathbf{u}\kappa_\infty(A) \le .5 \qquad \frac{\|x - \hat{x}\|_\infty}{\|x\|_\infty} \le 4\mathbf{u}\kappa_\infty(A)$$

$\mathbf{u}$ est le unité d'arrondi, égale à la moitié de lécart entre 1 et le plus petit nombre flottant strictement supérieur à 1. Pour les nombres flottants IEEE single précision, $\mathbf{u}$ est d'environ 10^{-7}. Il est d'environ 10^{-16} pour les nombres flottants IEEE double précision.

$\kappa_\infty(A) := \|A\|_\infty \|A^{-1}\|_\infty$ est le conditionnement de A pour la norme ∞

Plan du cours

1. Introduction générale
2. Preuves (revue)
3. Algèbre linéaire (revue)
4. Optimisation (revue)
5. Optimisation sous contraintes
6. Probabilités (revue)
7. Statistique
8. **Théorie de l'apprentissage**
9. **Optimisation pour l'apprentissage**

Partie 4 —Optimisation (revue)

Optimisation sans contraintes

$$f : \mathbf{R}^n \mapsto \mathbf{R}$$

$$\min_x f(x) \ ?$$

x^* est un minimum global de f ssi pour tout $x \in \mathbf{R}^n$, $f(x^*) \leq f(x)$

x^* est un minimum local de f ssi il existe un ensemble ouvert $U \subset \mathbf{R}^n$ contenant x^* tel que pour tout $x \in U$, $f(x^*) \leq f(x)$

Analyse

Concept central $\lim_{x \to a} f(x) = l$?

Notions de base de topologie

Dans $\mathbf{R}^n$, La boule fermée $B_f(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x - y\|_2 \leq \epsilon$

Dans $\mathbf{R}^n$, La boule ouverte $B_o(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x - y\|_2 < \epsilon$

U est un ouvert de $\mathbf{R}^n$ ssi $U \subset \mathbf{R}^n$ et pour tout x dans U, il existe $\epsilon > 0$, tel que $B_f(x, \epsilon) \subset U$.

U est un fermé de $\mathbf{R}^n$ ssi son complément dans $\mathbf{R}^n$ est un ouvert de $\mathbf{R}^n$

Jacobienne, gradient

$$f : U \subset \mathbf{R}^n \to \mathbf{R}^m, \text{ de classe C}^1$$

Dérivée partielle $f : x \mapsto (f_1(x), f_2(x), \ldots, f_m(x))$

$$\frac{\partial f_i}{\partial x_j} : (a_1, \ldots, a_n) \mapsto \lim_{h \to 0} \frac{f_i(a_1, \ldots, a_{j-1}, a_j + h, a_{j+1}, \ldots a_n) - f_i(a_1, \ldots, a_j, \ldots, a_n)}{h}$$

Matrice Jacobienne (transposée du gradient si m=1)

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

"Chain rule"

$$J_{f \circ g}(\mathbf{a}) = J_f(g(\mathbf{a})) J_g(\mathbf{a}),$$

Crédits

Certaines figures de

- Wikipedia