Chapitre I — Représentation matricielle pour la fouille de donnees, l'apprentissage automatique, ou l'extraction de connaissances

I) Introduction

L'exploration de données, ou encore fouille de données ("data mining") a pour objet l'extraction de connaissances, à partir de grandes quantités de données, par des techniques automatiques ou semi-automatiques.

Les outils d'observation modernes (satellites, capteurs embarqués, biopuces, ...) ou de communication (internet, objets connectés, ...) contribuent à la collection de quantités de plus en plus importantes d'informations, plus ou moins structurées, voire même hétérogènes suivant les cas, dans lesquelles peuvent apparaître des structures intéressantes, des motifs, selon des critères fixés, permettant d'en extraire de la connaissance, ou de valider des hypothèses et construire des modèles à partir des données.

Les applications sont nombreuses, telles que
- la reconnaissance d'écriture manuscrite
- la bioinformatique (séquencement ADN, expressions des gènes, ...)
- les moteurs de recherche sur internet
- la classification (supervisée ou non)
- ...

La fouille des données est une science à la croisée de plusieurs disciplines, telles que l'informatique, les statistiques et l'analyse des données, l'algèbre linéaire, l'optimisation ...
L'objectif de ce cours est de présenter les outils ou méthodes spécifiques en algèbre linéaire numérique et les problématiques qu'ils adressent.

II) Représentation matricielle

L'utilisation des matrices / vecteurs est classique dans la représentation des données, et dans les processus d'analyse et d'extraction de connaissances.

Exemple ① : La matrice "Mot clé / Document" en recherche d'information

Document 1:	The **Google**$^{\text{TM}}$ **matrix** P is a model of the **Internet**.
Document 2:	P_{ij} is nonzero if there is a **link** from **Web page** j to i.
Document 3:	The **Google matrix** is used to **rank** all **Web pages**.
Document 4:	The **ranking** is done by solving a **matrix eigenvalue** problem.
Document 5:	**England** dropped out of the top 10 in the **FIFA** **ranking**.

Si on compte la fréquence d'apparition des termes dans chaque document, on obtient le tableau suivant :

Term	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5
eigenvalue	0	0	0	1	0
England	0	0	0	0	1
FIFA	0	0	0	0	1
Google	1	0	1	0	0
Internet	1	0	0	0	0
link	0	1	0	0	0
matrix	1	0	1	1	0
page	0	1	1	0	0
rank	0	0	1	1	1
Web	0	1	1	0	0

Par conséquent, chaque document est représenté par un vecteur (de $\mathbb{R}^{10}$, dans cet exemple), et on peut regrouper les informations relatives à l'ensemble des documents dans une matrice

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

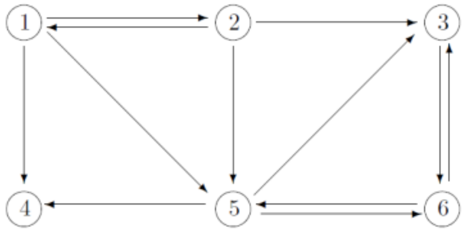
Maintenant, imaginons que l'on souhaite extraire tous les documents pertinents relativement à la question "ranking of Web Pages".

Cette question peut elle même être representé comme un document à part entière, avec les fréquences de mots clé suivantes :

$$q = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \in \mathbb{R}^{10}.$$

L'extraction d'information (recherche des documents pertinents) peut alors être formulée comme un problème mathématique, à savoir déterminer les colonnes de A qui sont "les plus proches" du vecteur q.

Exemple ② : les matrices peuvent aussi parfois servir à représenter les liens qui existent entre certaines données. Par exemple dans les moteurs de recherche, tels que google, un des points clé réside dans la détermination d'une mesure de qualité (sur l'ensemble des pages Web par exemple), et cela implique la construction d'une matrice représentant l'ensemble des liens qui existent entre toutes les pages Web.
Cette matrice est mise à jour et exploitée de manière récurrente afin de calculer cette mesure.
Imaginons par exemple qu'on ait 6 pages Web avec des référencements croisés suivants.



La matrice d'adjacence du graphe ci dessus est

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

où dans chaque colonne i on indique les liens sortants de "i vers j".
Cette matrice dite "matrice d'adjacence du graphe orienté" ci dessus est ensuite utilisée pour classer les pages Web entre elles, à l'aide de la résolution d'un problème aux valeurs propres que nous décrirons par la suite.

En recherche d'information (ex google) il est classique que les dimensions soient très grandes ($\sim 10^6$). Comme les documents ne contiennent qu'un faible nombre de l'ensemble des termes, seul peu d'éléments (en relatif de la dimension) sont en fait non nuls dans chacune des lignes de la matrice de représentation. Ces matrices sont donc souvent extrêmement creuses
Cependant ceci n'est pas toujours le cas, comme par exemple en bio-informatique, avec une matrice exprimant par exemple l'expression de diverses protéines par individu, en réponse à un contexte précis.
Dans ces cas là, les matrices sont denses. De même, les images d'observations de la terre peuvent être représentées par des tableaux de pixels (2D), mais avec de multiples valeurs

associés à chaque pixel (spectre visible, IR, UV, ...)
Il conviendra donc d'investiguer des méthodes et outils numériques adaptés suivant les cas, et surtout de réfléchir à leur "passage à l'échelle" en fonction du volume des données et de la puissance des calculateurs (parallèles) disponibles.

II) Compléments / Rappels d'algèbre linéaire

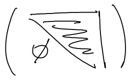
① notation : A^T = transposée de A

$$A^H = \overline{A^T} = (\overline{A})^T \quad \text{pour } A \in \mathcal{M}_{m,n}(\mathbb{C})$$

Si A est réelle, $A^H = A^T$; et $(AB)^H = B^H A^H$

Pour des matrices carrées on rappelle que :

1. A est dite hermitienne ssi $A^H = A$
2. si A est réelle, A est symétrique ssi $A^T = A$
3. A est semi-définie-positive ssi $x^H A x \geq 0, \forall x$
4. A est définie positive ssi $x^H A x > 0, \forall x \neq 0$
5. A est unitaire ssi $A^H A = A A^H = I$
6. A, réelle, est orthogonale ssi $A^T A = A A^T = I$
7. A est normale ssi $A^H A = A A^H$
8. A est triangulaire supérieure ssi pour $i > j$ $a_{ij} = 0$

   $A =$ 

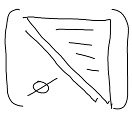
9. De même A triangulaire inférieure pour $i < j$ $a_{ij} = 0$

   $A =$ 

10. A est diagonale ssi $\forall i \neq j$ $a_{ij} = 0$

    on notera $A = \mathcal{D}(\alpha)$ pour indiquer que $a_{ii} = \alpha_i$ pour $i = 1 \dots n$

11. A est de forme Hessenberg supérieure ssi pour $i > j+1$, $a_{ij} = 0$

    $A =$ 

12. A est tridiagonale ssi $a_{ij} = 0$ pour $i > j+1$ et pour $i < j-1$

Dans le cas des matrices rectangulaires, on rappelle aussi que $A \in \mathbb{C}^{m \times n}$, $m > n$ est orthonormale ssi $A^H A = I_n$

(si $m = n$ une matrice orthonormale est en fait dite unitaire)

③ Transformations orthogonales élémentaires :

Rotations de Givens :

elles traduisent des rotations dans un plan

→ la matrice 2×2 de rotation $G = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}$ avec $c = \cos\theta$, $s = \sin\theta$

est unitaire (vérification immédiate, sachant que $c^2+s^2=1$)

La multiplication d'un vecteur x par G fait pivoter x dans le sens des aiguilles d'une montre de l'angle θ.

Dans $\mathbb{R}^n$, une matrice de rotation de Givens se généralise par

$$G = \begin{array}{c} \\ p \\ \\ q \\ \\ \end{array} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & c & \cdots & s & & \\ & & \vdots & \ddots & \vdots & & \\ & & -s & \cdots & c & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

et réalise une rotation élémentaire dans le plan formé par les vecteurs canoniques $\{e_p, e_q\}$.

Les rotations de Givens peuvent être utilisées en particulier pour mettre à zéro le deuxième élément d'un vecteur x non nul en posant $c = \frac{x_1}{\sqrt{x_1^2+x_2^2}}$ et $s = \frac{x_2}{\sqrt{x_1^2+x_2^2}}$

on a en effet : $\frac{1}{\sqrt{x_1^2+x_2^2}} \begin{pmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{x_1^2+x_2^2} \\ 0 \end{pmatrix}$

Avec une séquence de rotations de Givens, on peut mettre à zéro une partie d'un vecteur x non nul.

Le produit de rotations de Givens étant un produit de matrices unitaires est lui aussi unitaire, et conserve en particulier la norme euclidienne du vecteur x.

Reflexions de Householder :

elles traduisent des symétries orthogonales au travers d'un hyperplan

Soit un vecteur $v \neq 0$ donné. L'endomorphisme de $\mathbb{R}^n$ défini par

$$P = I - 2\frac{vv^T}{v^Tv} \qquad \left(\text{ou } I - 2\frac{vv^H}{v^Hv} \text{ dans } \mathbb{C}^n\right)$$

est symétrique et unitaire. Il vérifie aussi $P^2 = I_n$,

De plus, pour tout vecteur w orthogonal à v on a :

De plus, pour tout vecteur w orthogonal à v on a :

$$Pw = w - 2\frac{v}{\|v\|_2^2}\underbrace{(v^T w)}_{(=0)} = w$$

et
$$Pv = v - 2v\frac{v^T v}{v^T v} = -v$$

Soient $x \neq y$ deux vecteurs de même norme.
On peut construire une reflexion de Householder telle que $Px = y$:

En effet, il suffit de poser $v = x - y$ $(\neq 0)$; on a alors

$$v^T v = x^T x - 2x^T y + y^T y = 2(x^T x - x^T y) \quad (\text{car } \|x\|_2 = \|y\|_2)$$

et $v^T x = x^T x - y^T x = \frac{1}{2}v^T v$

d'où $Px = x - 2v\frac{v^T x}{v^T v} = x - v = x - (x-y) = y$

Cas particulier : pour $x \neq 0$, posons $y = \pm\|x\|_2 e_1$ $(\|y\|_2 = \|x\|_2)$

et $v = x - y$; alors pour $P = I - 2\frac{vv^T}{\|v\|_2^2}$

$Px = \pm\|x\|_2 e_1$ n'aura plus qu'une composante nonnulle.

<u>Rque</u> : afin d'éviter tout risque de "cancellation"
on prendra $y = -\text{signe}(x_1)\|x\|_2 e_1$

Transformations élémentaires en arithmétique finie :

Elles sont très stables : par exemple une reflexion de Householder $\hat{P}$ calculée avec des unités de calcul verifiant la norme (ieee) - sera une approximation d'une reflexion exacte P à O de la précision machine :

$$\|\hat{P} - P\|_2 = O(\text{eps})$$

On a aussi le résultat d'erreur inverse suivant :

$$fl(\hat{P}A) = P(A+E) \quad \text{avec } \|E\|_2 = O(\text{eps} \times \|A\|_2)$$

3 Propriétés fondamentales

On rappelle que : • le noyau d'une matrice $A \in \mathbb{R}^{m\times n}$ est un sous-espace vectoriel de $\mathbb{R}^n$ défini par

$$\operatorname{Ker} A = \{x \in \mathbb{R}^n \,/\, Ax = 0\}$$

• l'image d'une matrice $A \in \mathbb{R}^{m\times n}$ est un sous espace vectoriel de $\mathbb{R}^m$ défini par

$$\operatorname{Im} A = \{y \in \mathbb{R}^m \text{ tq } y = Ax, x \in \mathbb{R}^n\}$$

Le rang d'une matrice $A \in \mathbb{R}^{m\times n}$, noté $rg(A)$, est est donné par la dimension de $\operatorname{Im}(A)$ ($rg(A) = \dim(\operatorname{Im} A)$)

On a :

$$rg(A) = rg(A^T) \quad \text{et} \quad rg(A) + \dim(\operatorname{Ker} A) = n$$

On rappelle les propriétés suivantes :

1) Pour $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{n\times m}$

$$\operatorname{trace}(AB) = \operatorname{trace}(BA)$$

2) Pour A et B 2 matrices carrés de même taille

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

3) Pour $A \in \mathbb{R}^{m\times n}$, $\operatorname{Ker} A = \operatorname{Ker}(A^TA)$

et $\operatorname{Im}(A^T) = \operatorname{Im}(A^TA)$

4) Pour $A \in \mathbb{R}^{m\times n}$, la matrice $A^TA \in \mathbb{R}^{n\times n}$ est symétrique semi définie positive.

On appelle matrice de Gram la matrice A^TA.
Elle est inversible ssi les colonnes de A sont linéairement indépendantes ($\operatorname{Ker} A = \{0\}$)

5) Pour $A \in \mathbb{R}^{m\times n}$, $\operatorname{Ker} A$ et $\operatorname{Im} A^T$ forment deux sous espaces supplémentaires orthogonaux dans $\mathbb{R}^n$

De même, $\operatorname{Im}(A)$ et $\operatorname{Ker}(A^T)$ forment deux sous espaces supplémentaires orthogonaux dans $\mathbb{R}^m$

On note :

$$\mathbb{R}^n = \operatorname{Ker} A \overset{\perp}{\oplus} \operatorname{Im} A^T$$
$$\mathbb{R}^m = \operatorname{Im} A \overset{\perp}{\oplus} \operatorname{Ker} A^T$$

④ Procédé d'orthogonalisation de Schmidt

- C'est un procédé récurrent, qui à partir de tout ensemble de p vecteurs $\{a_1 \dots a_p\}$ dans $\mathbb{R}^n$, $p \leq n$, linéairement indépendants, permet de construire une famille de p vecteurs $\{q_1 \dots q_p\}$ orthogonaux 2 à 2 et normés, vérifiant :

$$\forall j, \quad 1 \leq j \leq p, \qquad \mathrm{Vect}\{a_1, \dots a_j\} = \mathrm{Vect}\{q_1, \dots, q_j\}$$

- Il se décrit très facilement par l'algorithme suivant :

Initialisation : $k = 1$, poser $q_1 = \dfrac{a_1}{\|a_1\|_2}$ et $Q_1 = [q_1]$

Pour $k = 2 : p$

$$v_k = a_k - Q_{k-1} Q_{k-1}^T a_k \; ; \qquad (v_k \perp Q_{k-1})$$

$$q_k = \frac{v_k}{\|v_k\|_2} \; ; \quad Q_k = [Q_{k-1}, q_k] \; ;$$

Fin Pour

- Le caractère récurrent - emboîté des espaces générés :

$$\mathrm{Vect}\{a_1 \dots a_j\} = \mathrm{Vect}\{q_1 \dots q_j\} \quad \forall j$$

permet d'établir le résultat suivant :

Soit $A \in \mathbb{R}^{m \times n}$, $m \geq n$, de rang n.
La matrice A peut être factorisée en $A = QR$
où $Q \in \mathbb{R}^{m \times n}$ est orthogonale et
$R \in \mathbb{R}^{n \times n}$ est triangulaire supérieure à éléments positifs sur la diagonale

On a d'ailleurs, pour $1 \leq j \leq n$,

$$\begin{cases} r_{i,j} = q_i^T a_j & \forall i < j \\ \text{et } r_{ii} = \|v_i\|_2 & (= q_i^T a_i) \end{cases}$$

⑤ Normes matricielles / rayon spectral

Norme : $\nu : \mathbb{R}^n \longrightarrow \mathbb{R}$ est une norme sur $\mathbb{R}^n$ ssi

a) $\forall x \neq 0$, $\nu(x) > 0$

b) $\nu(\alpha x) = |\alpha| \, \nu(x) \quad \forall \alpha \in \mathbb{R}$

c) $\nu(x+y) \leq \nu(x) + \nu(y) \quad \forall x, y \in \mathbb{R}^n$

Propriétés :

a) $|\nu(x) - \nu(y)| \leq \nu(x-y) \quad \forall x, y \in \mathbb{R}^n$

b) en dimension finie toutes les normes sont équivalentes

Par exemple, on a : $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$

$$\frac{1}{\sqrt{n}} \|x\|_2 \leq \|x\|_\infty \leq \|x\|_2$$

$$\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty$$

Norme matricielle : On appelle norme matricielle une norme définie sur l'ensemble des matrices carrées d'ordre n (n donné) qui vérifie, en plus des trois propriétés précédentes définissant une norme, la relation :

$$\|AB\| \leq \|A\| \, \|B\|$$

Norme matricielle induite par une norme vectorielle :

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|u\|=1} \|Au\|$$

Ces notions se généralisent aisément au cas des matrices rectangulaires.

Norme de Frobénius : Pour une matrice $A \in \mathbb{C}^{m \times n}$ on définit la norme de Frobénius par

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Propriétés :

a) $\|A\|_1 = \max_{\|x\|_1 \neq 0} \dfrac{\|Ax\|_1}{\|x\|_1} = \max_{1 \leq j \leq n} \left(\sum_{i=1}^{n} |a_{ij}| \right)$

b) $\|A\|_\infty = \max_{\|x\|_\infty \neq 0} \dfrac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1 \leq i \leq n} \left(\sum_{j=1}^{n} |a_{ij}| \right)$

c) $\|A\|_2 = \max_{\|x\|_2 \neq 0} \dfrac{\|Ax\|_2}{\|x\|_2} = \sqrt{\rho(A^T A)}$

où $\rho(A^T A)$ est la plus grande valeur propre en module de $A^T A$, ou encore appelé "rayon spectral" de $A^T A$.

d) la norme induite par la norme euclidienne $(\|\cdot\|_2)$ ainsi que la norme de Frobénius $(\|\cdot\|_F)$ sont unitairement invariantes

c'est à dire que $\forall Q$ et Q' unitaires, alors $\|A\| = \|QA\| = \|AQ'\| = \|QAQ'\|$.

Norme induite et rayon spectral :

on rappelle les résultats suivants :

a) Soit A une matrice carré d'ordre n.
Pour toute norme matricielle, induite ou non,
$$\rho(A) \leq \|A\|$$ (exo)

b) Si A est diagonalisable, il existe une norme induite (dépendant de A) telle que
$$\|A\| = \rho(A)$$ (exo)

c) (Householder - Ostrowski) Pour toute matrice $A \in \mathbb{R}^{n\times n}$, et pour tout $\varepsilon > 0$, il existe une norme induite dépendant de A et ε telle que
$$\|A\| \leq \rho(A) + \varepsilon$$ (admis)

d) Soit A une matrice carré. Les propriétés suivantes sont équivalentes :
$$\left(\lim_{k\to+\infty} A^k x = 0, \ \forall x\right)$$
$$\Leftrightarrow \left(\rho(A) < 1\right)$$
$$\Leftrightarrow \left(\|A\|_* < 1 \text{ pour au moins une norme induite } \|\cdot\|_*\right)$$ (exo)

e) Une condition suffisante pour que la matrice $(I - A)$ soit inversible est que $\rho(A) < 1$ et on a alors $(I-A)^{-1} = \sum_{k=0}^{\infty} A^k$ (exo)

f) Pour toute matrice carré et toute norme matricielle (induite ou non)
on a : $\rho(A) = \lim_{k\to+\infty} \|A^k\|^{1/k}$ (admis)

Systèmes linéaires

dimanche 31 mars 2019
22:59

III) Systèmes linéaires

1) Factorisations de type "Gauss" pour les matrices carrées :

- factorisation LU sans pivotage numérique.

  $A = LU$ (n'existe que sous certaines conditions)

  avec L triangulaire inférieure à diagonale unité

  et U triangulaire supérieure

- fatorisation LU avec pivotage partiel

  $PA = LU$ (existe pour toute matrice non singulière)

  avec P matrice de permutation.

- factorisation LU avec pivotage total

  $PAQ^T = LU$ (existe toujours)

  avec P et Q deux matrices de permutation

- factorisation LDL^T dans le cas symétrique ($A = A^T$)

  $A = LDL^T$ (existe dans le cas où $A = LU$ existe aussi)

  avec L triangulaire inférieure à diagonale unité

  D diagonale

- factorisation LDL^T avec pivotage

  $PAP^T = LDL^T$ (existe pour A symétrique non singulière)

  avec D bloc diagonale formée de blocs 1×1 ou 2×2 sur la diagonale

  et P matrice de permutation.

  L triangulaire inférieure à diagonale unité

- factorisation de Choleski, dans le cas où A est symétrique définie positive.

  $A = LL^T$ (ou R^TR) (existe $\forall A > 0$)

  avec L triangulaire inférieure

  ($R = L^T$ triangulaire supérieure)

② Factorisation QR

On a déjà vu, avec le procédé d'orthogonalisation de Schmidt, qu'une matrice $A \in \mathbb{R}^{m \times n}$ avec $m \geq n$ de rang maximal égal à n admet une factorisation de type

$$A = QR$$

avec $Q \in \mathbb{R}^{m \times n}$ (orthogonale $Q^T Q = I_n$)
et R triangulaire supérieure

- On a, de manière générale, pour toute matrice $A \in \mathbb{R}^{m \times n}$, avec $m \geq n$, et de rang maximal n, l'existence d'une matrice $Q \in \mathbb{R}^{m \times m}$ unitaire et $R \in \mathbb{R}^{m \times n}$ triangulaire supérieure, telles que :

$$A = Q \begin{bmatrix} R_1 \\ \emptyset \end{bmatrix}$$

- La factorisation réduite s'obtient en sélectionnant les n 1ères colonnes de Q : $Q = [Q_1 | Q_2]$ et à réduire R aux n 1ères lignes, de sorte que

$$A = Q_1 R_1$$

- La factorisation QR peut être obtenue à l'aide de réflexions de Householder, ou bien des rotations de Givens appliquées récursivement de façon à annuler la partie triangulaire inférieure dans la matrice A.

- On vérifie aussi que $A^T A = R^T Q^T Q R = R^T R = R_1^T R_1$ et R_1 correspond aux facteurs de Cholesky de $A^T A$.

- De même, pour $A \in \mathbb{R}^{m \times n}$ avec $m \leq n$, et de rang maximal égal à m, on a la factorisation LQ :

$$\underset{m \times n}{A} = \underset{m \times n}{\begin{bmatrix} L & \emptyset \end{bmatrix}} \; \underset{n \times n}{Q}$$

(qui revient simplement à factoriser A^T sous forme QR)

- Dans le cas des matrices $A \in \mathbb{R}^{m\times n}$, $m \geq n$, qui ne sont pas de rang maximal égal à n, on peut aussi construire une factorisation QR à condition de considérer une permutation des colonnes de A de façon à "rejeter à la fin" les colonnes qui se déduisent linéairement des 1ères colonnes :

$$AP = m\underbrace{\begin{bmatrix} \\ \quad Q \quad \\ \\ \end{bmatrix}}_{m} \begin{array}{c} \begin{array}{cc} r & (n-r) \end{array} \\ \left[\begin{array}{cc} R_1 & R_2 \\ \hline \multicolumn{2}{c}{0} \end{array}\right] \end{array} \begin{array}{l} \} \, r \\ \\ \end{array}$$

avec $Q \in \mathbb{R}^{m\times m}$ unitaire, $r = \text{rang}\, A$
$P \in \mathbb{R}^{n\times n}$ matrice de permutation
$R_1 \in \mathbb{R}^{r\times r}$ triangulaire supérieure à diagonale strictement positive et $R_2 \in \mathbb{R}^{r\times(n-r)}$.

- A cause des pertes d'orthogonalité dues à l'accumulation des erreurs d'arrondi en arithmétique finie, la factorisation QR à base de transformée de Householder ou de Givens permet d'obtenir des matrices $\tilde{Q}$ calculées qui sont de bien meilleure qualité (du point de vue des pertes d'orthogonalité) que celle obtenue avec Gram Schmidt.

Pour compenser en partie ce défaut, on peut modifier l'algorithme de Gram Schmidt de la façon suivante :

Algorithme de Gram Schmidt modifié :

Initialisation : $k = 1$, $q_1 = \dfrac{a_1}{\|a_1\|_2}$; $Q_1 = [q_1\ a_2 \cdots a_n]$

Pour $k = 2:n$

$$Q_{k-1}(:, k:n) = Q_{k-1}(:, k:n) - q_{k-1}\, q_{k-1}^T\, Q_{k-1}(:, k:n)$$

$$q_k = Q_{k-1}(:,k) / \| Q_{k-1}(:,k) \|_2$$

$$Q_k = Q_{k-1} \quad \text{et} \quad Q_k(:,k) = q_k$$

$$k = k+1$$

fin Pour

Notons $\omega = \| I - \tilde{Q}^T \tilde{Q} \|_2$ une mesure de la perte d'orthogonalité dans une matrice $\tilde{Q}$ calculée :

Pour GS classique, $\omega \sim \text{cond}_2(A)^2 \times \text{eps}$

pour GS modifié, $\omega \sim \text{cond}_2(A) \times \text{eps}$

③ Décomposition en valeurs singulières (SVD)

SVD : Soit $A \in \mathbb{R}^{m\times n}$ quelconque, de rang $= r$

Il existe $U \in \mathbb{R}^{m\times m}$ unitaire

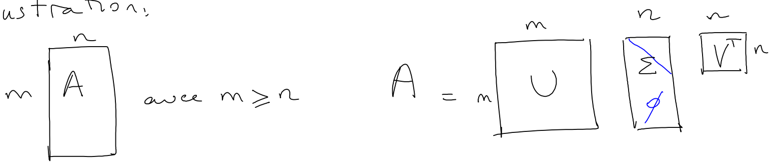
$V \in \mathbb{R}^{n\times n}$ unitaire

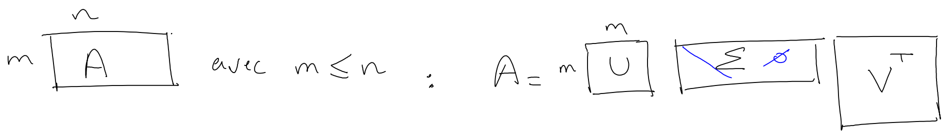
et $\Sigma \in \mathbb{R}^{m\times n}$ avec $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$ complétée par des $\emptyset$

et avec $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$,

tel que $A = U \Sigma V^T$

Illustration :





Dans tous les cas peut écrire :

$$\Sigma = \begin{array}{cc} \left(\begin{array}{cc} \Sigma_r & 0 \\ 0 & 0 \end{array} \right) & \begin{array}{l} r \\ m-r \end{array} \\ \begin{array}{cc} r & n-r \end{array} & \end{array}$$

avec $\Sigma_r = \mathrm{diag}(\sigma_1 \dots \sigma_r) \in \mathbb{R}^{r\times r}$,

Propriétés :

- on a $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

et $\{u_1, \dots, u_r\}$ forme une base orthonormée de $\mathrm{Im}\, A$

$\{u_{r+1}, \dots, u_m\}$ forme une base orthonormée de $(\mathrm{Im}\, A)^\perp$

$\{v_{r+1}, \dots, v_n\}$ forme une base orthonormée de $\mathrm{Ker}\, A$

$\{v_1, \dots, v_r\}$ forme une base orthonormée de $(\mathrm{Ker}\, A)^\perp$

Sachant aussi que $(\mathrm{Im}\, A)^\perp = \mathrm{Ker}(A^T)$
et $(\mathrm{Ker}\, A)^\perp = \mathrm{Im}(A^T)$

- $A^TA = V\Sigma^T\Sigma V^T \Rightarrow$ les colonnes de V forment une famille de vecteurs propres de A^TA, et les σ_i^2 (éventuel^t $\emptyset$) sont les valeurs propres associées.

De même, $AA^T = U\Sigma\Sigma^T U$, et les colonnes de U sont des vecteurs propres de AA^T.

- On a aussi $\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2$ (exo ?)

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

- Pour finir, si on pose $U_r = [u_1 \cdots u_r]$ et $V_r = [v_1 \cdots v_r]$ les matrices formées à partir des r premiers vecteurs singuliers à gauche et à droite respectivement (ceux associés aux r valeurs singulières non nulles)

on a alors : $A = U_r \Sigma_r V_r^T$ appelée aussi SVD réduite (ou "thin SVD")

SVD et meilleure approximation de A de rang $k < r$:

On rappelle les résultats d'approximation suivant, liés à ce qu'on est aussi appelé "SVD tronquée" :

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T = U_k \Sigma_k V_k^T = \boxed{U_k}\,\boxed{\Sigma_k}\,\boxed{V_k^T}$$

où $U_k = [u_1 \cdots u_k]$

$V_k = [v_1 \cdots v_k]$

$\Sigma_k = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}$

réalise la meilleure approximation de rang $k < r = \text{rang } A$ à la fois en norme 2 et en norme de Frobenius,

càd que $A_k = \underset{\text{rang } Z = k}{\arg\min} \|A - Z\|_2$

et $A_k = \underset{\text{rang } Z = k}{\arg\min} \|A - Z\|_F$

En outre, $\|A - A_k\|_2 = \sigma_{k+1}$ et $\|A - A_k\|_F^2 = \sum_{j=k+1}^{r} \sigma_j^2$

SVD et pseudo inverse :

A partir de la décomposition SVD : $A = U \Sigma V^T$

où $\Sigma = \begin{array}{cc|l} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} & & \begin{array}{l} r \\ m-r \end{array} \\ r \quad n-r & & \end{array}$ et U et V unitaires

posons $\Sigma^+ = \begin{array}{cl} \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} & \begin{array}{l} r \\ n-r \end{array} \\ r \quad m-r & \end{array} \in \mathbb{R}^{n \times m}$

(obtenue en transposant Σ et en inversant Σ_r)

et considérons alors

$$A^+ = V \Sigma^+ U^T$$ appelé "pseudo inverse de A"

On a alors le résultat suivant :

La pseudo inverse A^+ de A est l'unique matrice X vérifiant les équations de Moore-Penrose :

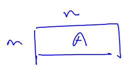
$$\begin{cases} XAX = X \\ AXA = A \\ (AX)^T = AX \\ (XA)^T = XA \end{cases}$$

A^+ est parfois aussi appelé Moore-Penrose inverse de A.

Quelques propriétés :

- si A est carrée inversible, alors $A^+ = A^{-1}$
- La pseudo inverse d'une matrice nulle est sa transposée.
- $(A^+)^+ = A$
- $(A^T)^+ = (A^+)^T$, $(A^H)^+ = (A^+)^H$
- $(\alpha A)^+ = \frac{1}{\alpha} A^+$ (pour $\alpha \neq 0$)
- Soit $A \in \mathbb{R}^{m \times n}$, $m \geq n$, t.q. $rg(A) = n$ $\begin{array}{cc} & n \\ m & \boxed{A} \end{array}$

  alors $A^+ = (A^T A)^{-1} A^T$ et $A^+ A = I_n$ (exo)

- Soit $A \in \mathbb{R}^{m \times n}$, $m \leq n$, tq $rg(A) = m$ 

alors $A^+ = A^T (A A^T)^{-1}$ et $A A^+ = I_m$ (exo)

- $P = A A^+$ est le projecteur orthogonal sur $\mathrm{Im}\, A$ (exo)

  $Q = A^+ A$ est le projecteur orthogonal sur $\mathrm{Im}(A^T)$ (exo)

  A quoi correspondent $(I - AA^+)$ et $(I - A^+A)$? (exo)

- On vérifie aussi que (admis)

$$A^+ = \lim_{\delta \to 0^+} (A^T A + \delta I_n)^{-1} A^T$$
$$= \lim_{\delta \to 0^+} A (A A^T + \delta I_m)^{-1}$$

limites qui existent même si $A^T A$ ou $A A^T$ ne sont pas inversibles

(Voir "régularisation de Tikhonov")

4 Angles principaux et vecteurs principaux

Définition :

Soient $\mathcal{F}$ et $\mathcal{G}$ deux sous-espaces de $\mathbb{R}^m$ tels que $p = \dim(\mathcal{F}) \geq \dim(\mathcal{G}) = q \geq 1$

Les "angles principaux" entre $\mathcal{F}$ et $\mathcal{G}$, notés $\{\theta_i, i = 1 \dots q\}$ et "vecteurs principaux" associés $\{f_i \; i = 1 \dots q, \; g_i \; i = 1 \dots q\}$ sont définis récursivement par :

$$\cos \theta_k = f_k^T g_k = \max_{\substack{f \in \mathcal{F}, \|f\|_2 = 1 \\ [f_1 \dots f_{k-1}]^T f = 0}} \; \max_{\substack{g \in \mathcal{G}, \|g\|_2 = 1 \\ [g_1 \dots g_{k-1}]^T g = 0}} f^T g$$

Il est à noter que $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_q \leq \pi/2$.

La détermination des angles principaux et vecteurs principaux entre deux sous-espaces est aussi appelée "corrélation canonique entre sous-espaces"

SVD et corrélation canonique :

On peut considérer que $\mathcal{F} = \mathrm{Im}(Q_1)$ et $\mathcal{G} = \mathrm{Im}(Q_2)$ avec $Q_1 \in \mathbb{R}^{m\times p}$, et $Q_2 \in \mathbb{R}^{m\times q}$ dont les colonnes forment une base orthonormée de $\mathcal{F}$ et de $\mathcal{G}$ respectivement.

Soit
$$p\underbrace{\boxed{\;Q_1^T Q_2\;}}_{q} = p\underbrace{\boxed{\;U\;}}_{q}\;\boxed{\Sigma}\;\underbrace{\boxed{\;V^T\;}}_{q}q$$
la SVD réduite de $Q_1^T Q_2$

Comme $\|Q_1^T Q_2\|_2 \leq \|Q_1^T\|_2 \, \|Q_2\|_2 = 1$

$$\Sigma = \mathrm{diag}(\sigma_i) \quad \text{avec } 0 \leq \sigma_i \leq 1 \quad \text{pour } i = 1 \dots q$$

les σ_i étant classées dans l'ordre décroissant

on peut noter $\quad \theta_i = \arccos \sigma_i \quad i = 1 \dots q$

$$\left(\sigma_i = \cos \theta_i \;, \; \theta_i \in [0, \pi/2] \right)$$

Introduisons alors $F = Q_1 U$ et $G = Q_2 V$ toutes deux dans $\mathbb{R}^{m\times q}$

Ces deux matrices sont formées de colonnes orthonormées car en effet :

$$F^T F = U^T Q_1^T Q_1 U = U^T I_p U = U^T U = I_q$$
$$\text{et} \quad G^T G = V^T Q_2^T Q_2 V = V^T I_q V = V^T V = I_q$$

Les vecteurs principaux entre $\mathcal{F}$ et $\mathcal{G}$ sont donnés par les colonnes de F et G, et les angles principaux θ_k correspondent aux $\arccos \sigma_k$ dans la SVD de $Q_1^T Q_2$

(exo)

Definition : l'angle principal le plus large est associé à la notion de distance entre deux sous-espaces de même dimension $\mathcal{F}$ et $\mathcal{G}$, définie par :

$$\mathrm{dist}(\mathcal{F}, \mathcal{G}) = \|P_{\mathcal{F}} - P_{\mathcal{G}}\|_2$$

avec $P_{\mathcal{F}}$ le projecteur orthogonal sur $\mathrm{Im}(\mathcal{F})$

et $P_{\mathcal{G}}$ ——————— $\mathrm{Im}(\mathcal{G})$

On peut vérifier en effet que :

$$\mathrm{dist}(\mathcal{F},\mathcal{G}) = \sqrt{1-\cos^2\theta_p} = \sin\theta_p$$

<u>Application</u> : la CS décomposition $(m \geq n)$

Soit une matrice $Q \in \mathbb{R}^{m\times n}$, orthogonale $(Q^TQ = I_n)$ partitionnée de la façon suivante :

$$Q = \begin{array}{c} m_1 \\ m_2 \end{array}\begin{bmatrix} Q_1 \\ \hline Q_2 \end{bmatrix} \quad \text{où } Q_1 \in \mathbb{R}^{m_1\times n},\ Q_2 \in \mathbb{R}^{m_2\times n}$$
$$\qquad\qquad\qquad n \qquad\qquad \text{avec } m_1 \geq n \text{ et } m_2 \geq n$$

Vérifiez qu'il existe deux matrices orthogonales

$U_1 \in \mathbb{R}^{m_1\times n}$, $U_2 \in \mathbb{R}^{m_2\times n}$, et une matrice unitaire $V \in \mathbb{R}^{n\times n}$

telles que
$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{array}{c} m_1 \\ m_2 \end{array}\underbrace{\begin{bmatrix} \boxed{U_1} & 0 \\ 0 & \boxed{U_2} \end{bmatrix}}_{n\quad n} \underbrace{\begin{bmatrix} C \\ \hline S \end{bmatrix}}_{n}\begin{array}{c} n \\ n \end{array} \underbrace{\boxed{V^T}}_{n} n = \begin{bmatrix} U_1 C V^T \\ U_2 S V^T \end{bmatrix}$$

avec $C = \mathrm{diag}(c_i)_{i=1\ldots n}$

$S = \mathrm{diag}(s_i)_{i=1\ldots n}$

et $c_i^2 + s_i^2 = 1$

5 Quelques identités remarquables

Vérifier que (en supposant l'inversibilité acquise quand nécessaire)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\left(D - CA^{-1}B\right)^{-1}CA^{-1} & -A^{-1}B\left(D-CA^{-1}B\right)^{-1} \\ -\left(D - CA^{-1}B\right)^{-1}CA^{-1} & \left(D-CA^{-1}B\right)^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1} & -\left(A-BD^{-1}C\right)^{-1}BD^{-1} \\ -D^{-1}C\left(A-BD^{-1}C\right)^{-1} & D^{-1} + D^{-1}C\left(A-BD^{-1}C\right)^{-1}BD^{-1} \end{bmatrix}$$

$$B^{-1} = A^{-1} - B^{-1}(B-A)A^{-1}$$

$$\left(A + UBV\right)^{-1} = A^{-1} - A^{-1}U\left(B^{-1} + VA^{-1}U\right)^{-1}VA^{-1}$$

(Formule de Sherman Morrison Woodbury)

Analyse d'erreur

samedi 6 avril 2019

20:27

IV | Analyse d'erreur

① Analyse de sensibilité pour les systèmes linéaires

Considérons le système paramétré suivant :

$$(A + \varepsilon F)\, x(\varepsilon) = b + \varepsilon f, \qquad x(0) = x$$

avec $F \in \mathbb{R}^{n\times n}$, $f \in \mathbb{R}^n$.

Si A est inversible, il existe un voisinage de 0 dans lequel $A+\varepsilon F$ est inversible, et $x(\varepsilon)$ existe et est unique.

On a :

$$x(\varepsilon) = A^{-1}\left(b + \varepsilon f - \varepsilon F\, x(\varepsilon)\right)$$

$$x(0+\varepsilon) = A^{-1}b + \varepsilon A^{-1}\left(f - F x(\varepsilon)\right)$$

$$= A^{-1}b + \varepsilon A^{-1}\left(f - F A^{-1}b - \varepsilon x(\varepsilon)\right)$$

$$= x(0) + \varepsilon A^{-1}\left(f - FA^{-1}b\right) + O(\varepsilon^2)$$

d'où

$$\frac{\partial x}{\partial \varepsilon}(0) = A^{-1}\left(f - FA^{-1}b\right) = A^{-1}\left(f - Fx\right)$$

Pour n'importe quelle norme matricielle induite par une norme vectorielle, on peut alors écrire :

$$\frac{\| x(\varepsilon) - x\|}{\|x\|} \leq |\varepsilon|\, \|A^{-1}\| \left(\frac{\|f\|}{\|x\|} + \|F\|\right) + O(\varepsilon^2)$$

Définissons **le conditionnement de A**, noté $K(A)$, par

$$K(A) = \|A\|\, \|A^{-1}\| \quad (\text{avec } K(A) = +\infty \text{ pour } A \text{ singulière})$$

On obtient alors :

$$\boxed{\frac{\| x(\varepsilon) - x\|}{\|x\|} \leq K(A)\, |\varepsilon| \left(\frac{\|f\|}{\|b\|} + \frac{\|F\|}{\|A\|}\right) + O(\varepsilon^2)}$$

car $\|b\| = \|Ax\| \leq \|A\|\,\|x\|$

d'où $\dfrac{\|f\|}{\|x\|} \leq \dfrac{\|f\|}{\|b\|} \cdot \|A\|$.

On voit donc que le conditionnement de A quantifie la sensibilité des variations sur la solution en fonction des variations relatives sur les données du système linéaire.

Remarques: en norme 2 : $K_2(A) = \|A\|_2 \|A^{-1}\|_2 = \dfrac{\sigma_{max}(A)}{\sigma_{min}(A)}$

- On a aussi le résultat suivant, dû à Kahan (1966) :

en norme p $\left(K_p(A) = \|A\|_p \|A^{-1}\|_p\right)$ on a :

$$\frac{1}{K_p(A)} = \min_{(A+\delta A) \text{ singulière}} \frac{\|\delta A\|_p}{\|A\|_p}$$

ce qui indique que le conditionnement mesure aussi la distance de A à la singularité.

- Pour toute norme p : $K_p(A) \geq 1$

En norme 2, une matrice unitaire est parfaitement conditionnée, en ce sens que

$$K_2(Q) = 1 \quad (\text{pour } Q \text{ unitaire})$$

Sans travailler au 1^er ordre comme précédemment, on peut aussi établir le résultat suivant :

Théorème : Supposons que $Ax = b$ avec A inversible et $b \neq 0$

Soit $(A + \delta A)y = b + \delta b$ avec

$\|\delta A\| \leq \varepsilon \|A\|$ et $\|\delta b\| \leq \varepsilon \|b\|$.

Si $\varepsilon K(A) = \rho < 1$ ($K(A)$ pour la norme matricielle subordonnée à la norme vectorielle)

alors $$\frac{\|y - x\|}{\|x\|} \leq \frac{2\varepsilon}{1-\rho} K(A)$$

(démo en exo)

② Analyse d'erreur à posteriori — erreur inverse

Définition : Erreur inverse :

Soit $\tilde{x}$ une solution calculée (avec erreurs numériques) du système $Ax = b$.
On appelle erreur inverse associée à $\tilde{x}$ la quantité

$$\eta(\tilde{x}) = \min \left[\max\left(\frac{\|\delta A\|}{\|A\|}, \frac{\|\delta b\|}{\|b\|} \right) \right]$$

avec δA et δb tels que $(A+\delta A)\tilde{x} = b + \delta b$

En d'autres termes, $\eta(\tilde{x})$ est égal au minimum des valeurs $\omega > 0$ pour lesquelles il existe δA et δb telles que

$$\|\delta A\| \leq \omega \|A\| \quad \text{et} \quad \|\delta b\| \leq \omega \|b\|$$

et pour lesquelles $\tilde{x}$ est une solution exacte du système perturbé

$$(A + \delta A)\tilde{x} = b + \delta b$$

Théorème

Soit $\tilde{x}$ solution approchée du système $Ax = b$
et $r = b - A\tilde{x}$ le résidu associé.
On a alors :

$$\eta(\tilde{x}) = \frac{\|r\|}{\|A\| \|\tilde{x}\| + \|b\|}$$

(demo en exo)

Les majorations en norme sont souvent suffisantes pour qualifier l'ordre de grandeur des erreurs sur la solution, mais il est parfois aussi utile d'analyser ces erreurs au niveau de chaque composante.

Pour ce faire, on peut citer le résultat suivant :

Théorème : Oettli et Prager (1964)

Soit $\tilde{x}$ une solution approchée du système $Ax = b$ et soient $E \geq 0$ une matrice et $f \geq 0$ un vecteur donnés avec des valeurs toutes positives ou nulles.

Alors, le plus petit scalaire $\omega > 0$ tel que il existe δA et δb vérifiant

$|\delta A| \leq \omega E$, $|\delta b| \leq \omega f$ (inégalité composante à composante)

et tels que $(A + \delta A)\tilde{x} = b + \delta b$

est donné par la quantité :

$$\omega_{min} = \max_{1 \leq i \leq n} \frac{|A\tilde{x} - b|_i}{(E|\tilde{x}| + f)_i}$$

Rque : si $\omega_{min} = +\infty$, $\tilde{x}$ n'est solution d'aucun système avec des perturbations structurés par E et f considérés.

Ce théorème permet en particulier de structurer l'erreur en imposant par exemple qu'elle respecte un certain "pattern".

Exemple : considérer $E = |A|$, $f = |b|$ qui seront donc nuls partout où $a_{ij} = 0$ et $b_i = 0$