MAT 3741 : Algèbre linéaire appliquée, Hiver 2017
Devoir #3

À remettre ~~le vendredi le 3 mars~~ mardi le 7 mars à 10h.

Cette fois, tout le devoir est ensemble dans une partie mais vous êtes priés d'utiliser MATLAB pour faire les calculs (questions notés *), et pouvez remettre les résultats en-ligne au lieu de les recopier sur papier (espéciallement si Matlab ne vous donne que des résultats arrondis).

1. Soit A une matrice de taille $n \times m$ telle que $\text{rang}(A) = m$.
   (a) Démontrer que $\overline{A^T}A$ est une matrice inversible, en utilisant la décomposition Q-R de A.
   Solution: Soit $A = QR$ la décomposition Q-R non-normalisée de A. Alors Q est de taille $n \times k$ et R de taille $k \times m$, de rang k, où $\text{rang}(A) = k$. Ici, ceci implique que R est carrée et inversible, or $\overline{R^T}$ l'est aussi. Donc
   $$\overline{A^T}A = \overline{R^T}\,\overline{Q^T}QR = \overline{R^T}R$$
   est inversible, de taille $m \times m$.
   (b) Démontrer que la projection orthogonale de $\mathbb{C}^n$ sur l'espace engendré par les colonnes de A est donnée par
   $$P = A(\overline{A^T}A)^{-1}\overline{A^T}.$$
   Il suit qu'on peut calculer la projection orthgonale sans faire Gram-Schmidt !
   Solution: Il faut démontrer que pour tout $x \in \mathbb{C}^n$, $Px \in \text{Col}(A)$ et $x - Px$ est orthogonale à $\text{Col}(A)$. Puisque
   $$Px = A\left((\overline{A^T}A)^{-1}\overline{A^T}x\right)$$
   il suit que la réponse est dans l'image de A, donc dans $\text{Col}(A)$. Un vecteur v est orthogonale à $\text{Col}(A)$ ssi pour toute colonne u de A, nous avons $\langle u, v\rangle = 0$, qui veut dire $\overline{u^T}v = 0$; donc ssi $\overline{A^T}v = 0$ (qui vérifie toutes les colonnes en même temps). On calcule
   $$\overline{A^T}(x - Px) = \overline{A^T}x - \overline{A^T}A(\overline{A^T}A)^{-1}\overline{A^T} = \overline{A^T}x - \overline{A^T}x = 0.$$
   Donc c'est bien la projection orthogonale sur $\text{Col}(A)$.
   (c) * Soit $A = \begin{bmatrix} 1 & i & 1 \\ 0 & i & -1 \\ i & 1 & -i \\ 0 & i & 1 \end{bmatrix}$. Calculer $(\overline{A^T}A)^{-1}$ et puis calculer P de deux façons : avec la formule en (b), et en calculant $A = QR$ une décomposition Q-R normalisée, or $P = Q\overline{Q^T}$. Quelques remarques de Matlab :

```
A'   est la conjuguée complexe de la transposée de A
inv(A) est l'inverse de A
[Q,R]=qr(A,0) donne notre décomposition QR, si rang(A)=#col(A)
norm(Pn-P,1)  donne la norme de la matrice Pn-P.  Quand il s'agit
              d'erreurs d'arrondissements, c'est mieux de calculer
              la norme de la différence que d'espérer que la différence
              sera la matrice nulle
```

Solution:

```
>> A=[1 i 1; 0 i -1 ; i 1 -i; 0 i 1]

>> Z=inv(A'*A)

Z =

  0.5000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
```

```
   0.0000 + 0.0000i   0.3333 + 0.0000i   0.0000 + 0.1667i
   0.0000 + 0.0000i   0.0000 - 0.1667i   0.3333 + 0.0000i

>> Pf=A*Z*A'   --- notre formule en 1(b)

Pf =

   0.8333 + 0.0000i   0.0000 + 0.0000i   0.0000 - 0.1667i   0.3333 + 0.0000i
   0.0000 + 0.0000i   1.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.1667i   0.0000 + 0.0000i   0.8333 + 0.0000i   0.0000 - 0.3333i
   0.3333 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.3333i   0.3333 + 0.0000i

>> [Q,R]=qr(A,0)

Q =

  -0.7071 + 0.0000i   0.0000 - 0.5000i  -0.2887 + 0.0000i
   0.0000 + 0.0000i   0.0000 - 0.5000i   0.8660 + 0.0000i
   0.0000 - 0.7071i  -0.5000 + 0.0000i   0.0000 + 0.2887i
   0.0000 + 0.0000i   0.0000 - 0.5000i  -0.2887 - 0.0000i


R =

  -1.4142 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i  -2.0000 + 0.0000i   0.0000 + 1.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i  -1.7321 + 0.0000i

>> Q'*Q    ---- faudra être I si normalisé

ans =

   1.0000 + 0.0000i   0.0000 - 0.0000i  -0.0000 + 0.0000i
   0.0000 + 0.0000i   1.0000 + 0.0000i  -0.0000 + 0.0000i
  -0.0000 - 0.0000i  -0.0000 - 0.0000i   1.0000 + 0.0000i

>> Pqr = Q*Q'

Pqr =

   0.8333 + 0.0000i  -0.0000 + 0.0000i   0.0000 - 0.1667i   0.3333 - 0.0000i
  -0.0000 - 0.0000i   1.0000 + 0.0000i   0.0000 - 0.0000i  -0.0000 - 0.0000i
   0.0000 + 0.1667i   0.0000 + 0.0000i   0.8333 + 0.0000i  -0.0000 - 0.3333i
   0.3333 + 0.0000i  -0.0000 + 0.0000i  -0.0000 + 0.3333i   0.3333 + 0.0000i

>> norm(Pf-Pqr)

ans =

   6.1666e-16
```

donc c'est la même matrice, jusqu'à arrondissement.

2. Soit A une matrice de taille $n \times m$ de rang k quelconque, telle que $A = QR$ est sa décomposition Q-R normalisée. La pseudo-inverse de Moore-Penrose est la matrice

$$A^+ = \overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T}.$$

(a) Démontrer, avec l'aide de (1), que cette formule est bien définie, et donner la taille de A^+.
Solution: On a choisi la décomposition Q-R normalisée, alors Q est de taille $n \times k$ et R est de taille $k \times m$, et les deux ont rang k. Il suit que $\overline{R^T}$ est de taille $m \times k$ et de rang k, et donc par (1a) (avec $\overline{R^T}$ à la place de "A") nous avons que $\overline{(\overline{R^T})}^T \overline{R^T} = R\overline{R^T}$ est inversible (donc qu'on a le droit de prendre son inverse !) de taille $k \times k$. Finalement la taille de A^+ est $m \times n$.

(b) Démontrer que A^+A est la projection orthgonale sur $\mathrm{Col}(\overline{A^T})$.
Solution: On a $A = QR$ donc $\overline{A^T} = \overline{R^T}\,\overline{Q^T}$. Donc $\mathrm{Col}(\overline{A^T}) \subseteq \mathrm{Col}(\overline{R^T})$; mais puisque $\mathrm{rang}(A) = \mathrm{rang}(R)$, il suit que $\mathrm{rang}(\overline{A^T}) = \mathrm{rang}(\overline{R^T})$ et donc ces espaces (qui ont donc la même dimension) sont égaux.
Alors, avec la décomposition QR nous avons :

$$A^+A = \overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T}QR = \overline{R^T}(R\overline{R^T})^{-1}R.$$

Puisque R est de taille $k \times m$ et de rang k, $\overline{R^T}$ est de taille $m \times k$ de rang k, or par (1b) (avec $\overline{R^T}$ à la place de "A") nous pouvons conclure que $\overline{R^T}(R\overline{R^T})^{-1}R$ est la projection sur $\mathrm{Col}(\overline{R^T}) = \mathrm{Col}(\overline{A^T})$.
Ou bien, on pourrait procéder directement : soit $x \in \mathbb{C}^m$. Puisque $A^+Ax = \overline{R^T}z$ pour un certain z, il suit que $A^+Ax \in \mathrm{Col}(\overline{R^T}) = \mathrm{Col}(\overline{A^T})$.
De plus, on veut voir que $x - A^+Ax$ est orthogonale à $\mathrm{Col}(\overline{A^T})$, qui est équivalent à dire que $A(x - A^+Ax) = 0$. On calcule

$$Ax - AA^+Ax = Ax - QR\overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T}Ax = Ax - Q\overline{Q^T}Ax.$$

Mais puisque $Q\overline{Q^T}$ est la projection sur $\mathrm{Col}(A)$, on a que $Q\overline{Q^T}Ax = Ax$, or cette différence est égale à 0.

(c) Démontrer que AA^+ est la projection orthogonale sur $\mathrm{Col}(A)$.
Solution: Avec la décomposition $A = QR$ on obtient

$$AA^+ = (QR)(\overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T}) = Q(R\overline{R^T})(R\overline{R^T})^{-1}\overline{Q^T} = Q\overline{Q^T}$$

qui est bien la projection orthogonale sur $\mathrm{Col}(A)$.
Ou bien, on peut démontrer le résultat directement. Soit $x \in \mathbb{C}^n$. Alors $AA^+x = Az$ pour un certain $z = A^+x$, or le produit est dans l'image de A, $\mathrm{Col}(A)$. On calcule

$$\overline{A^T}AA^+ = \overline{A^T}AA^+ = \overline{A^T}QR\overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T} = \overline{A^T}Q\overline{Q^T} = \overline{R^T}\,\overline{Q^T}Q\overline{Q^T} = \overline{R^T}\,\overline{Q^T} = \overline{A^T}$$

et donc $\overline{A^T}x = \overline{A^T}AA^+x$, qui démontre que $x - AA^+x$ est orthogonale à chaque colonne de A.

(d) Dans quel circonstances est-ce que A^+ est une inverse à gauche, ou à droite, de A ?
Solution: La projection orthogonale sur un espace $U \subset \mathbb{C}^t$ sera l'identité ssi $U = \mathbb{C}^t$. Donc par (2b), A^+ est une inverse à gauche de A ssi $\mathrm{Col}(\overline{A^T}) = \mathbb{C}^m$, qui est ssi $\mathrm{rang}(\overline{A^T}) = \mathrm{rang}(A) = m$. Et par (2c), A^+ est une inverse à droite de A ssi $\mathrm{Col}(A) = \mathbb{C}^n$, qui est ssi $\mathrm{rang}(A) = n$.

(e) * Soit A comme en (1c). Déterminer A^+ et comparer avec le résultat de la fonction pinv(A) de Matlab (en norme). [1]

3. Soit A une matrice de taille $n \times m$ de rang k quelconque.

(a) Soit $y \in \mathrm{Col}(A)$ et posons $S = \{x \in \mathbb{C}^m \mid Ax = y\}$. Démontrer que l'élément de S ayant la plus petite norme est (l'image de) la projection de S sur $\mathrm{Col}(\overline{A^T})$. Indice : décrire S en termes de $\mathrm{Noy}(A)$ et puis identifier $\mathrm{Noy}(A)^\perp$. (Indice : Si les conjuguées complexes vous mêlent, oubliez-les en premier lieu.)

---

1. Par contre, avec MATLAB [Q0,R0]=qr(A) donne une matrice Q0 carrée, et de plus, il le fait d'une manière plus stable que ce qu'on fait à la main. Si vous choisissez une matrice A avec $\mathrm{rang}(A) = \dim(Col(A))$ alors vous pouvez récupérer notre matrice Q des premières $\mathrm{rang}(A)$ colonnes de $Q0$; mais si A n'est **pas** de plein rang, vous ne pouvez pas en générale récuperer notre décomposition Q-R normalisée. Points bonis si vous trouvez, ou produisez, une fonction Matlab qui retournera notre décomposition Q-R dans le cas générale de $n \times m$ de rang k.

Solution: Alors soit x_0 une solution particulière du système $Ax = y$. Par la théorie de solutions de systèmes linéaires, il suit que

$$S = \{x_0 + p \mid p \in \text{Noy}(A)\}.$$

De plus, on a vu en MAT1741 que $\text{Noy}(A)$ et $\text{Col}(\overline{A^T})$ sont des compléments orthogonaux, au moins dans le cas réel. (J'avais inclus une preuve de ce fait sur la page web de ressources qui accompagne le devoir.)

Donc le vecteur le plus court dans S sera la projection orthogonale P de x_0 sur $\text{Col}(\overline{A^T})$. (Ou, plus précisement, c'est la tête de cette projection qui vit dans S.)

Si ce ne vous semble pas évident, utilisez l'argument du cours pour comparer la longueur de x_0 et celle de Px_0, étant donnez que leur différence soit en $\text{Noy}(A)$.

Par (2b) l'élément le plus courte sera A^+Ax_0, pour n'importe quel choix de $x_0 \in S$.

(b) Démontrer (avec les résultats ci-haut) que pour tout $b \in \mathbb{C}^n$, si on pose $x = A^+b$ alors c'est un choix idéale : (i) la norme de $Ax - b$ est le minimum possible,[2] et (ii) la norme de x est le minimum possible tel que (i) soit vrai.

Solution: On pose $x = A^+b$. Alors $Ax = AA^+b$, qui est par (2c) la projection orthgonale de b sur $\text{Col}(A)$, donc par ce qu'on a vu au cours, $\|Ax - b\|$ est le minimum possible, donc (i). De plus, si $y = AA^+b$, alors x est une solution du système $Ax = y$. La solution de ce système de longueur minimale, par (a), est A^+Ax, sa projection sur $\text{Col}(\overline{A^T})$ (par (2b)). Mais on dévéloppe (avec $\overline{Q^T}Q = I$) :

$$\begin{aligned}
A^+Ax &= A^+A(A^+b) \\
&= A^+(AA^+)b \\
&= \overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T}(Q\overline{Q^T})b \quad \text{par (2c)} \\
&= \overline{R^T}(R\overline{R^T})^{-1}\overline{Q^T}b \\
&= A^+b = x.
\end{aligned}$$

Donc x satisfait aussi à (ii).

4. La regression statistique peut être défini comme suit. Soit

$$S = \{(x_{j,1}, x_{j,2}, \cdots, x_{j,n}, y_j) = (\vec{x_j}, y_j) \in \mathbb{R}^{n+1} \mid 1 \le j \le m\}$$

un ensemble de données. On veut trouver des scalaires $a_i \in \mathbb{R}$ tels que la fonction

$$g(x_1, \cdots, x_n) = \sum_{i=1}^{n} a_i x_i$$

ait la propriété que si on pose $y = (y_1, \cdots, y_m)$ et $z = (g(\vec{x_1}), \cdots, g(\vec{x_m}))$ alors

$$\|y - z\|$$

est minimale[3] parmis tous les choix possibles de scalaires a_i.

(a) Un cas particulier, pour fixer ses idées : Vos données sont

x_1	x_2	y
.5	.6	.5
.3	.4	.4
.9	.9	.8
.6	.6	.7

Vous voulez trouver a et b tel que $ax_1 + bx_2 = y$, pour chacune des quatres triplets (x_1, x_2, y) donnée dans ce tableau. Veuillez écrire ce système linéaire. Veuillez vous assurer que ce système est incohérent. Donc on veut une solution moindre carrées ; trouvez-la.[4] On appelle

2. Donc, en particulier c'est une solution moindre carrées si le système est incohérent, et une solution exacte autrement.

3. par rapport à la 2-norme

4. On a vu au cours que si $Bx = b$ est incohérent, alors $\overline{B^T}Bx = \overline{B^T}b$ est cohérent, et sa solution x_0 est dite la solution moindre carrées du système originale, puisque $\|Bx_0 - b\|_2$ est le minimum possible.

les valeurs $ax_1 + bx_2$ (pour (x_1, x_2) dans notre tableau) nos *prédictions* ; elles ne seront pas égale à nos *observations* qui sont les y. Leur différences sont les résiduels. Déterminer la norme du vecteur (taille 4×1) de résiduels dans ce cas. C'est une mesure de l'efficacité de notre prédiction ; en (d) on mentionne une meilleure.

Solution: Soient

$$A = \begin{bmatrix} .5 & .6 \\ .3 & .4 \\ .9 & .9 \\ .6 & .6 \end{bmatrix} \quad \text{et} \quad y = \begin{bmatrix} .5 \\ .4 \\ .8 \\ .7 \end{bmatrix}$$

alors le système est $Ax = y$ avec $x = (a, b)$. Avec MATLAB, on trouve que

```
>> rref([A y])
ans =
     1     0     0
     0     1     0
     0     0     1
     0     0     0
```

donc c'est un système incohérent, dont la solution moindres carrées est la solution de $\overline{A^T}Ax = \overline{A^T}y$:

```
>> x = inv(A'*A)*(A'*y)
x =
   0.331932773109261
   0.634453781512590
```

On calcule nos prédictions $z = Ax$:

```
>> z=A*x
z=
   0.546638655462185
   0.353361344537814
   0.869747899159666
   0.579831932773111
```

Donc le vecteur de résiduels est $y - z$, et sa (2-)norme est (selon MATLAB) 0.1538, qu'on pourrait comparer avec la moyenne 0.6 des observations ; l'erreure c'est d'ordre un quart !

Donc en générale (deux questions très rapides (b) et (c)) :

(b) Quelle est la matrice A telle que si on pose $u = (a_1, \cdots, a_n)$ alors

$$z = Au?$$

Solution: A est la matrice ayant $\{\vec{x_1}, \cdots, \vec{x_m}\}$ comme ses lignes, pour que $g(\vec{x_j}) = \sum_{i=1}^{n} a_i x_{j,i}$.

(c) Et : quel est le système linéaire dont la solution moindre carrrées est le vecteur de constantes u qu'on cherche ?

Solution: $Au = y$, qui est un système incohérent.

(d) * On vous donne un fichier excel avec des données. Vous pouvez facilement copier-coller ce que vous voulez dans MATLAB pour créer une matrice, genre

```
>> A= [ (paste data
from excel)
 ]
>> y =[ (paste data
from excel)
 ]
```

Trouvez la fonction linéaire $g(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3 + b$ pour laquelle $\|\vec{y} - g(\vec{x})\|$ est minimale, en trouvant la solutions moindres carrées du système comme en (a) ci-haut. Indice : ajouter une colonne de 1 comme un x_4 pour obtenir la possibilité d'une constante. Puis calculer le *coéfficient de détermination* R^2 : Soit $\overline{y}$ la moyenne des valeurs $y_1, \cdots, y_n$ et poser $\vec{Y} = (\overline{y}, \overline{y}, \cdots, \overline{y})$ (n composantes). Alors R^2 est lié au rapport entre la norme de la différence entre les observations et les prédictions, et la variance totale S_{tot} des observations :[5]

$$R^2 = 1 - \frac{S_{res}}{S_{tot}} = 1 - \frac{\|g(\vec{x}) - \vec{y}\|^2}{\|\vec{y} - \vec{Y}\|^2}.$$

Typiquement $0 \leq R^2 \leq 1$; le plus proche que le résultat est à 1, le plus que les données sont bien modélisés par la fonction g.[6] Qu'en pensez-vous de ces données ?

Quelques commandes utiles :

```
A(:,4) = 1                  :  cré une quatrième colonne ayant que des 1
ybar = mean(y)              :  la moyenne
Stot = norm(y(:)-ybar)^2    :  y(:)-ybar soustrait ybar de chaque composante de y
```

Solution: Donc notre matrice A est la matrice formée des trois colonnes de données, puis une colonne de 1. Donc après avoir copié-collé les données x_1, x_2, x_3 pour former la matrice A, on utilise la commande `A(:,4)=1` pour ajouter une colonne de 1 et puis on copie-colle la colonne des y dans une variable y en MATLAB. On veut la solution moindre carrées du système $Ax = y$, où $x = (a_1, a_2, a_3, b)$ (pardon, il y avait une faute de frappe dans la question, je n'ai pas voulu $a(x_1 + x_2 + x_3) + b$, qui représente une autre regression — mais par contre puisqu'il s'agit d'une droite, vous pouvez le faire avec le petit tutoriel sur la page de ressources). Donc comme ci-haut,

$$x = (\overline{A^T}A)^{-1}\overline{A^T}y = \begin{bmatrix} 0.2085 \\ 0.3834 \\ 0.5491 \\ -0.2080 \end{bmatrix}$$

Donc la solution est $g(x_1, x_2, x_3) = 0.2085x_1 + 0.3834x_2 + 0.5491x_3 - 0.2080$. C'est le troisième variable, x_3, qui a le plus grand effet sur les observations y.

On a $\overline{y} = mean(y) = 0.6063$, et la commande $Y = y; Y(:) = \overline{y}$ est une façon de produire le Y qu'on veut. Puis on a

```
>> 1-(norm(A*x - y)/norm(y-Y))
ans =
    0.4416
```

qui est le R^2.

5. The coefficient of determination of a linear regression is related to the quotient of the norm-squared of the residues (difference of data and prediction) and the variance of the observations, and is one measure of the goodness of fit of a regression.

6. Si vous étudiez en statistique, vous auriez des formules plus sophistiquées à appuyer ici ; veuillez comparer !