

# Master de Chemoinformatique M1S1

## Examen de mathématiques pour la chimie\*

MARCOU Gilles

Décembre 2010

### Résumé

*Documents autorisés. Durée : 2h. La copie de l'étudiant sera constituée d'un support papier traditionnel et d'un ou plusieurs fichiers de calcul pour Maple. Toute réponse doit être motivée ou sera considérée comme nulle.*

## Exercice 1

L'objectif de cet exercice est d'utiliser l'Analyse en Composantes Principales (PCA, *Principal Components Analysis*) pour représenter un petite chimiothèque de  $N = 149$  composés. Ces composés sont représentés par des vecteurs d'entiers représentant le nombre d'occurrence de 253 fonctions chimiques. En d'autres termes, ils appartiennent à un espace chimique figuré par un espace vectoriel  $E$  de  $d = 253$  dimensions. Ces données sont stockées dans le fichier `db.dat`.

La PCA consiste à identifier une nouvelle base de l'espace vectoriel dans lequel se trouvent les molécules. Cette base est construite de façon à ce que les vecteurs de la base soient ordonnés : les premiers vecteurs de la base doivent expliquer la majeure partie des variations initiales du jeu de donnée. Celle-ci est mesurée à l'aide de la variance, qui est la trace de la matrice de covariance obtenue à partir d'une matrice contenant les 149 vecteurs de 253 dimensions représentant la chimiothèque.

**Vous répondrez dans la feuille de travail `ExamMath2010Exo1.mw` contenant une mise en page préparatoire.**

### Question 1.1

Expliquez ce qu'est la dimension d'un espace vectoriel.

---

Votre session commence par charger une matrice  $M_0$  en mémoire. Chaque ligne de cette matrice représente une molécule et les colonnes correspondent aux descripteurs. La librairie Maple `LinearAlgebra` est chargée.

La première étape consiste à travailler avec des données centrées. Vous voulez donc calculer le centre géométrique  $G$  de votre jeu de données. Les composantes  $i$  de  $G$ ,  $x_G^i$ , sont données par la formule (1), en fonction des composantes  $i$  de la molécule  $j$ ,  $x_j^i$ , dans l'espace à  $E$ .

$$x_G^i = \frac{1}{N} \sum_{j=1}^N x_j^i \quad (1)$$

Les composantes de  $G$  sont donc les moyennes arithmétiques de chaque colonne de votre matrice  $M_0$ .

---

\*Enseignants: G. Marcou, P. Jost, *Université Louis Pasteur, Institut de Chimie, 4, rue Blaise Pascal, 67000 Strasbourg*

## Question 1.2

Calculez  $G$ . Vous pouvez utiliser la commande `Statistics [Mean]` qui calcule la moyenne d'une liste de nombre.

*Vous pouvez utiliser la commande `read "G.m"` afin de charger dans la mémoire de Maple le vecteur  $G$  pour continuer l'exercice si cela s'avère nécessaire.*

---

## Question 1.3

Créez une matrice  $M_G$  telle que chaque ligne corresponde à une molécule dont les descripteurs sont centrés. Il est possible, par exemple, de construire une matrice de  $N$  lignes et  $d$  colonnes, dont les colonnes seraient constantes et contiendrait les composantes de  $G$ , puis de soustraire cette matrice à la matrice  $M_0$ .

*Vous pouvez utiliser la commande `read "MG.m"` afin de charger dans la mémoire de Maple la matrice  $M_G$  pour continuer l'exercice si cela s'avère nécessaire.*

---

Il faut ensuite calculer la matrice de covariance,  $\Sigma$ , donnée par la formule 2.

$$\Sigma = M_G^T M_G \quad (2)$$

## Question 1.4

Calculez  $\Sigma$ . Définissez ce qu'est le rang d'une matrice et indiquez en quoi il vous renseigne sur la « véritable » dimension de l'espace chimique que vous considérez. Quelle est le rang de la matrice  $\Sigma$  ?

*Vous pouvez utiliser la commande `read "Sigma.m"` afin de charger dans la mémoire de Maple la matrice  $\Sigma$  pour continuer l'exercice si cela s'avère nécessaire.*

---

Les composantes principales sont les vecteurs propres de la matrice  $\Sigma$  et les valeurs propres correspondantes se réfèrent à la variance expliquée par celle-ci. La variance totale est la trace de la matrice  $\Sigma$ .

## Question 1.5

Calculez les valeurs propres et les vecteurs propres de  $\Sigma$  et rangez-les dans des variables  $v, P$ .

*Vous pouvez utiliser la commande `read "vP.m"` afin de charger dans la mémoire de Maple le vecteur  $v$  des valeurs propres et la matrice  $P$  des vecteurs propres.*

---

## Question 1.6

Expliquez ce qu'est la trace d'une matrice et estimez la variance totale du jeu de données.

---

## Question 1.7

Tracez une courbe représentant la variance expliquée par les  $i$  premières composantes principales quand  $i$  varie de 1 à 10. Quel est le pourcentage de la variance totale qui est expliquée par les deux premières composantes principales ?

*Vous décidez de projeter vos molécules dans le plan définis par les deux premières composantes principales en utilisant l'équation (3), où  $P_1^j, P_2^j$  sont les composantes de la première, respectivement de la seconde, composante principale et les  $x_j^i$  sont les composantes  $j$  de la molécule  $i$ .*

$$\begin{cases} x_1^{P,i} = \sum_{j=1}^d P_1^j x_j^i \\ x_2^{P,i} = \sum_{j=1}^d P_2^j x_j^i \end{cases} \quad (3)$$

### Question 1.8

Représentez vos composés dans l'espace des deux premières composantes. Il sera nécessaire d'utiliser le style `point` pour votre graphique.

---

### Question 1.9

Donnez vos conclusions.

---

## Exercice 2

Cet exercice propose d'étudier la décomposition d'un composé hypothétique A, en B puis en C, avec les constantes de vitesse  $k_1$  et  $k_2$ . L'équation différentielle qui gouverne ce processus est donné par le système (4) :

$$\begin{cases} \frac{dA(t)}{dt} = -k_1 A(t) \\ \frac{dB(t)}{dt} = -k_2 B(t) + k_1 A(t) \\ \frac{dC(t)}{dt} = k_2 C(t) \end{cases} \quad (4)$$

Vous répondrez dans la feuille de travail `ExamMath2010Exo2.mw` contenant une mise en page préparatoire.

### Question 2.1

Entrer dans Maple le système d'équations différentielles (4) avec les conditions initiales (5).

---

$$\begin{cases} A(t=0) = 1 \\ B(t=0) = 0 \\ C(t=0) = 0 \end{cases} \quad (5)$$

### Question 2.2

Résoudre le problème différentiel formé par (4) et (5).

---

### Question 2.3

Posez  $k_1 = 10$  et  $k_2 = 5$  et tracer l'évolution des concentrations des espèces A, B et C à partir de  $t = 0$ . Vous choisirez un interval de temps adapté à l'évolution des espèces. Vous pourriez avoir besoin de la commande `assign`.

---

## **Barème**

- Question 1.1 : 1 point
- Question 1.2 : 2 points
- Question 1.3 : 2 points
- Question 1.4 : 2 points
- Question 1.5 : 2 points
- Question 1.6 : 1 point
- Question 1.7 : 2 points
- Question 1.8 : 2 points
- Question 1.9 : 1 point
- Question 2.1 : 2 points
- Question 2.2 : 2 points
- Question 2.3 : 2 points