Rappel mathématiques & statistiques
M2 SOAC université Paul Sabatier
F. Bouttier

Ce cours présente de manière très simplifiée les notions essentielles à connaître pour suivre le cours de M2 SOAC. Il ne prétend pas à la rigueur mathématique.

Calcul matriciel

Matrice A : tableau de nombres réels (ou *coefficients*) a_{ij} (i et j sont respectivement les indices de ligne et de colonne)

Vecteur x : matrice à 1 seule colonne, de dimension $\dim x$.

un nombre réel (ou *scalaire*) est assimilable à une matrice ne contenant qu'un seul coefficient.

multiplication matricielle : étant donné 2 matrices A et B, le produit $C = AB$ est la matrice constituée des coefficients $c_{ij} = \sum_k a_{ik} b_{kj}$ (somme sur toutes les valeurs possibles de k). A doit avoir autant de lignes que B a de colonnes. En général $AB \neq BA$. La multiplication entre une matrice A et un vecteur x est assimilable à une application linéaire $x \mapsto Ax$.

mulitplication de matrice par un scalaire α : αA est la matrice consituée des coefficients de A multipliés chacun par α.

addition de deux matrices : $A + B$ est la matrice des sommes de coefficients 2 à 2 de A et de B : $a_{ij} + b_{ij}$. A et B doivent avoir les mêmes dimensions.

matrice identité I : matrice carrée contenant des 1 sur la diagonale et des zéros partout ailleurs. Pour toute matrice A on a $AI = IA = A$.

matrice inverse A^{-1} : étant donné A, son inverse vérifie $A^{-1}A = AA^{-1} = I$. L'inverse n'existe pas toujours, et son calcul peut être coûteux. Pour calculer l'inverse d'un produit : $(AB)^{-1} = B^{-1}A^{-1}$

matrice transposée A^{T} : étant donné A, sa transposée est la matrice dont les lignes et les colonnes ont été permutées (tournées autour de la diagonale). La transposée existe toujours. Pour calculer la transposée d'un produit : $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$.

matrice symétrique : matrice égale à sa transposée : $A^{\mathrm{T}} = A$.

Les **matrices définies positives** ont des propriétés mathématiques intéressantes :

définition : une matrice A est définie positive si, pour tout vecteur non-nul x, on a $x^{\mathrm{T}}Ax > 0$ ($x \mapsto x^{\mathrm{T}}Ax$ est une fonction du vecteur x, à valeurs réelles)

toute matrice définie positive est forcément carrée, symétrique et inversible

elle est aussi **diagonalisable :** il existe M inversible et D diagonale telles que $A = M^{\mathrm{T}}DM$. Les lignes de M sont les vecteurs propres e_i de A. Les coefficients diagonaux de D sont les carrés des valeurs propres λ_i de A : $Ae_i = \lambda_i e_i$

il existe des librairies scientifiques pour calculer l'inverse d'une matrice définie positive, ou pour résoudre le système d'équations $Ax = b$ (avec b vecteur connu et x vecteur inconnu) beaucoup plus rapidement que si A n'était pas définie positive.

une matrice de covariances est toujours symétrique définie positive (les variables doivent être toutes à variance non-nulle)

Statistiques scalaires

Soit des réels $(a_k)_{k=1...K}$ et $(b_k)_{k=1...K}$: ensembles de K réalisations des 2 variables réelles a et b. On peut imaginer que a et b sont 2 listes de mesures.

moyenne de a : c'est le scalaire $\overline{a} = \frac{1}{K}\sum_{k=1}^{K} a_k$ (moyenne 'arithmétique')

moyenne quadratique de a : c'est le scalaire $\text{rms}(a) = \sqrt{\frac{1}{K}\sum_{k=1}^{K} a_k^2}$, autrement dit la racine de la moyenne des carrés de a.

variance de a : c'est le scalaire $\text{var}(a) = \frac{1}{K}\sum_{k=1}^{K}(a_k - \overline{a})^2$, autrement dit la moyenne des carrés des écarts à la moyenne de a. (Pour avoir un calcul plus précis on peut multiplier la somme par $\frac{1}{K-1}$ au lieu de $\frac{1}{K}$)

écart-type de a : racine carrée de la variance $\sigma(a) = \sqrt{\text{var}(a)}$

covariance de a et b : c'est le scalaire $\text{cov}(a,b) = \frac{1}{K}\sum_{k=1}^{K}(a_k - \overline{a})(b_k - \overline{b})$, autrement dit la moyenne des produits des écarts de a et de b à leurs moyennes.

corrélation de a et b : c'est la covariance normalisée par les écarts-types : $\rho(a,b) = \frac{\text{cov}(a,b)}{\sigma(a)\sigma(b)}$

Statistiques vectorielles

Soit des vecteurs de dimension n : $(x_k)_{k=1...K}$: ensemble de K réalisations du vecteur x. On peut imaginer que x est un champ, ou une série temporelle, par exemple, et que l'on dispose de K réalisations de x qui peuvent être tirées d'une climatologie de x. Chaque composante $x(i)$ de x est un réel à K réalisations $x_k(i)$, sur lequel on peut calculer des statistiques scalaires.

La plupart des opérations statistiques scalaires s'étendent naturellement aux vecteurs, en effectuant les calculs sur chacune des composantes :

moyenne $\overline{x}$: c'est le vecteur x dont chaque composante a été moyennée : $\overline{x}(i) = \overline{x(i)} = \frac{1}{K}\sum_{k=1}^{K} x_k(i)$

$\overline{x}$ est la moyenne d'ensemble de x, à ne pas confondre avec la moyenne spatiale d'un champ qui est $\sum_{i=1}^{n}\frac{x(i)}{n}$

on définit de même la variance, l'écart-type d'un ensemble de vecteurs

Exception : la **covariance entre 2 ensembles de vecteurs x et y** est la matrice des covariances scalaires entre tous les couples possibles de composantes $x(i)$ et $x(j)$:

$$\text{cov}(x,y)_{i,j} = \text{cov}(x(i), y(j)) = \frac{1}{K}\sum_{k=1}^{K}(x_k(i) - \overline{x(i)})(y_k(j) - \overline{y(j)})$$

En notation matricielle :

$$\text{cov}(x,y) = \overline{(x - \overline{x})(y - \overline{y})}$$

Comme la moyenne est une opération linéaire, si M et N sont 2 matrices, on a $\text{cov}(Mx, Ny) = M\text{cov}(x,y)N^{\text{T}}$.

Dérivée vectorielle

Soit f une application de R^n dans R^p (fonction d'un vecteur, à valeurs vectorielles). Par définition, sa dérivée en x, aussi appelée *différentielle* ou *linéaire tangent* de f, est l'opérateur linéaire F_x de R^n dans R^p tel que :

$$\forall x, h \in \mathrm{R}^n \qquad f(x+h) = f(x) + F_x h + \mathcal{O}(\| h \|^2)$$

L'opérateur F_x étant linéaire, on peut l'identifier à une matrice à p lignes et n colonnes. Ses coefficients sont les dérivées partielles de f : $F_{ij} = \frac{\partial F_i}{\partial x_j}(x)$. Ils mesurent la sensibilité de la i-ième sortie de f par rapport à sa j-ième entrée, au voisinage de x.

Définition du gradient : si $p = 1$, f est à valeurs réelles, F est identifiable à un vecteur ∇J appelé **gradient** de f. Les règles de calcul du gradient ressemblent aux règles de dérivation des fonctions scalaires, avec quelques complications (parce que la multiplication matricielle n'est pas commutative). Dans le doute, toujours se ramener à la formule de Taylor ci-dessus.

Exemple avec la fonction coût de l'analyse variationnelle : le gradient de la fonction

$$J(x) = (x - x_b)^{\mathrm{T}} B^{-1} (x - x_b) + (y - Hx)^{\mathrm{T}} R^{-1} (y - Hx)$$

est le vecteur

$$\nabla J(x) = 2B^{-1}(x - x_b) - 2H^{\mathrm{T}} R^{-1} (y - Hx)$$

et sa dérivée seconde, appelée *hessienne* est la matrice carrée de taille $n \times n$:

$$J'' = 2B^{-1} + 2H^{\mathrm{T}} R^{-1} H$$