

Corrigé du TD 3 — exercices 1 et 2

Calculs euclidiens et profils d'expression, vérifiés avec R

Christophe Ambroise

2026-10-02

Note

Fiche autonome : elle rappelle l'énoncé de chaque exercice, puis déroule la correction. Les résultats numériques sont recalculés par les blocs R affichés ci-dessous (le code R est visible et exécuté à la compilation).

1 Exercice 3.1 — Calculs euclidiens

1.1 Énoncé

Soient $\mathbf{u} = (1, -2, 2)$ et $\mathbf{v} = (2, 1, 0)$.

1. Calculer $\|\mathbf{u}\|$, $\|\mathbf{v}\|$ et $\mathbf{u} \cdot \mathbf{v}$.
2. Calculer la distance $d(\mathbf{u}, \mathbf{v})$.
3. Les deux vecteurs sont-ils orthogonaux ? Vérifier sur cet exemple la relation $\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v}$.
4. Donner le vecteur unitaire de même direction et de même sens que $\mathbf{u}$.

1.2 Correction

On définit les vecteurs et une fonction norme.

```
u <- c(1, -2, 2)
v <- c(2, 1, 0)

norme <- function(x) sqrt(sum(x^2))
```

1.2.1 1. Normes et produit scalaire

```
norme(u)          # norme de u

[1] 3

norme(v)          # norme de v

[1] 2.236068

sum(u * v)        # produit scalaire u . v

[1] 0
```

$\|\mathbf{u}\| = 3$, $\|\mathbf{v}\| = \sqrt{5} \approx 2,236$ et $\mathbf{u} \cdot \mathbf{v} = 0$.

1.2.2 2. Distance

```
sqrt(sum((u - v)^2))          # distance euclidienne
```

```
[1] 3.741657
```

```
as.matrix(dist(rbind(u, v)))  # meme resultat via la matrice de distances
```

```
      u      v  
u 0.000000 3.741657  
v 3.741657 0.000000
```

$\mathbf{u} - \mathbf{v} = (-1, -3, 2)$, donc $d(\mathbf{u}, \mathbf{v}) = \sqrt{1 + 9 + 4} = \sqrt{14} \approx 3,742$.

1.2.3 3. Orthogonalité et relation de Pythagore

```
sum(u * v) == 0
```

```
[1] TRUE
```

```
norme(u - v)^2
```

```
[1] 14
```

```
norme(u)^2 + norme(v)^2 - 2 * sum(u * v)
```

```
[1] 14
```

Le produit scalaire est nul : $\mathbf{u} \perp \mathbf{v}$. La relation donne $\|\mathbf{u} - \mathbf{v}\|^2 = 9 + 5 - 0 = 14$, conforme au calcul direct. C'est le théorème de Pythagore, appliqué à $\mathbf{u}$ et $-\mathbf{v}$.

1.2.4 4. Vecteur unitaire

```
u / norme(u)
```

```
[1] 0.3333333 -0.6666667 0.6666667
```

$\frac{\mathbf{u}}{\|\mathbf{u}\|} = \left(\frac{1}{3}, -\frac{2}{3}, \frac{2}{3}\right)$, de norme $\sqrt{\frac{1+4+4}{9}} = 1$.

2 Exercice 3.2 — Profils d'expression

2.1 Énoncé

Trois gènes sont mesurés à quatre instants successifs :

$$\mathbf{g}_1 = (1, 2, 3, 4), \quad \mathbf{g}_2 = (2, 4, 6, 8), \quad \mathbf{g}_3 = (4, 3, 2, 1).$$

1. Calculer les trois distances euclidiennes entre ces profils.
2. Calculer les trois similarités cosinus, puis les angles correspondants.
3. Quel gène est le plus proche de $\mathbf{g}_1$ selon la distance ? Selon le cosinus ? Lequel de ces deux critères repère des gènes régulés de la même façon ?
4. Centrer les trois profils, puis recalculer le cosinus entre $\mathbf{g}_1$ et $\mathbf{g}_3$. Interpréter.

2.2 Correction

Le point de départ reprend le code utilisé en séance : on empile les profils en lignes d'une matrice, puis on calcule les distances au carré.

```
g1 <- c(1, 2, 3, 4)
g2 <- c(2, 4, 6, 8)
g3 <- c(4, 3, 2, 1)

G <- rbind(g1, g2, g3)
d2 <- as.matrix(dist(G)^2)      # carres des distances euclidiennes
round(d2, 2)
```

```
      g1 g2 g3
g1  0 30 20
g2 30  0 70
g3 20 70  0
```

2.2.1 1. Distances euclidiennes

```
round(d2, 2)      # carres
```

```
      g1 g2 g3
g1  0 30 20
g2 30  0 70
g3 20 70  0
```

```
round(sqrt(d2), 2)      # distances
```

```
      g1 g2 g3
g1 0.00 5.48 4.47
g2 5.48 0.00 8.37
g3 4.47 8.37 0.00
```

$d(g_1, g_2) = \sqrt{30} \approx 5,48$, $d(g_1, g_3) = \sqrt{20} \approx 4,47$, $d(g_2, g_3) = \sqrt{70} \approx 8,37$.

2.2.2 2. Similarités cosinus et angles

On normalise chaque ligne (norme 1), puis le cosinus deux à deux est le produit scalaire des lignes normalisées.

```
Gn <- G / sqrt(rowSums(G^2))      # profils ramenés à la norme 1
cos_mat <- Gn %*% t(Gn)           # cosinus entre profils
round(cos_mat, 3)
```

```
      g1 g2 g3
g1 1.000 1.000 0.667
g2 1.000 1.000 0.667
g3 0.667 0.667 1.000
```

```
round(acos(cos_mat) * 180 / pi, 1)      # angles en degrés
```

```
      g1 g2 g3
g1 0.0 0.0 48.2
g2 0.0 0.0 48.2
g3 48.2 48.2 0.0
```

Les normes valent $\|\mathbf{g}_1\| = \|\mathbf{g}_3\| = \sqrt{30}$ et $\|\mathbf{g}_2\| = 2\sqrt{30}$; les produits scalaires, $\mathbf{g}_1 \cdot \mathbf{g}_2 = 60$, $\mathbf{g}_1 \cdot \mathbf{g}_3 = 20$, $\mathbf{g}_2 \cdot \mathbf{g}_3 = 40$. D'où

$$s(\mathbf{g}_1, \mathbf{g}_2) = 1, \quad s(\mathbf{g}_1, \mathbf{g}_3) = \frac{20}{30} = \frac{2}{3} \approx 0,67, \quad s(\mathbf{g}_2, \mathbf{g}_3) = \frac{40}{60} = \frac{2}{3} \approx 0,67.$$

Angles : 0° entre $\mathbf{g}_1$ et $\mathbf{g}_2$; $\arccos \frac{2}{3} \approx 48,2^\circ$ pour les deux autres couples.

Astuce

Le cosinus se déduit des distances au carré et des normes, ce qui relie le code ci-dessus au premier calcul :

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - d(\mathbf{x}, \mathbf{y})^2}{2 \|\mathbf{x}\| \|\mathbf{y}\|}.$$

```
i <- 1; j <- 3 # couple g1 / g3
(norme(G[i, ])^2 + norme(G[j, ])^2 - sqrt(d2)[i, j]^2) /
(2 * norme(G[i, ]) * norme(G[j, ])) # doit valoir 2/3
[1] 0.6666667
```

2.2.3 3. Distance ou cosinus pour comparer les profils

```
d <- sqrt(d2); diag(d) <- NA
apply(d, 1, which.min) # plus proche au sens de la distance
```

```
g1 g2 g3
3 1 1
```

```
cs <- cos_mat; diag(cs) <- NA
apply(cs, 1, which.max) # plus proche au sens du cosinus
```

```
g1 g2 g3
2 1 1
```

Selon la distance, $\mathbf{g}_3$ est le plus proche de $\mathbf{g}_1$ ($4,47 < 5,48$). Selon le cosinus, c'est $\mathbf{g}_2$ ($s = 1$) : $\mathbf{g}_2 = 2\mathbf{g}_1$, même forme, amplitude double. Pour repérer des gènes régulés de la même façon, c'est la forme qui compte : le cosinus est le bon critère, car la distance est dominée par les niveaux d'expression.

2.2.4 4. Profils centrés

```
Gc <- G - rowMeans(G) # moyenne de chaque profil ramenee a 0
round(Gc, 2)
```

```
 [,1] [,2] [,3] [,4]
g1 -1.5 -0.5  0.5  1.5
g2 -3.0 -1.0  1.0  3.0
g3  1.5  0.5 -0.5 -1.5
```

```
Gcn <- Gc / sqrt(rowSums(Gc^2))
round(Gcn %*% t(Gcn), 3) # cosinus apres centrage
```

```
g1 g2 g3
```

$\mathbf{g}_1 \quad 1 \quad 1 \quad -1$
 $\mathbf{g}_2 \quad 1 \quad 1 \quad -1$
 $\mathbf{g}_3 \quad -1 \quad -1 \quad 1$

Les moyennes valent 2,5, 5 et 2,5. Profils centrés :

$$\mathbf{g}_1^c = (-1,5; -0,5; 0,5; 1,5), \quad \mathbf{g}_3^c = (1,5; 0,5; -0,5; -1,5) = -\mathbf{g}_1^c.$$

Le cosinus vaut -1 : angle de 180° . Sans centrage, les valeurs toutes positives donnaient un cosinus positif malgré des tendances opposées ; après centrage, le cosinus révèle que $\mathbf{g}_3$ décroît quand $\mathbf{g}_1$ croît. C'est la corrélation, étudiée au C5.