

# Apprentissage et prévision de la dynamique temporelle des stations de vélos en libre service

---

**Adrien Richou**

[Adrien.Richou@u-bordeaux.fr](mailto:Adrien.Richou@u-bordeaux.fr)

Dans ce projet, il est proposé de s'intéresser à des observations qui sont collectées sous la forme du nombre de vélos disponibles au cours du temps dans l'ensemble des stations de vélos en libre service dans deux villes qui proposent ce type de service : Paris et Toulouse.

Dans un premier temps, on étudiera l'intérêt du principe de l'analyse en composantes principales (ACP) adapté au cadre des données spatio-temporelles afin de pouvoir analyser les comportements urbains des usagers de vélos en libre service. Dans un second temps, il pourra être envisagé de s'intéresser à la problématique de la prévision du nombre de vélos disponible dans une station à différents horizons temporels à l'aide de modèles de régression linéaire ou de séries chronologiques. Ce projet s'appuiera sur les packages disponibles en Python.

# Gestion optimale d'une forêt

---

Adrien Richou

[Adrien.Richou@u-bordeaux.fr](mailto:Adrien.Richou@u-bordeaux.fr)

Supposons que l'on gère l'exploitation d'une forêt de manière durable. À un instant  $t > 0$ , on note  $X_t$  la quantité de bois exploitable dans cette forêt. Sachant que toute opération de coupe induit des frais fixes, on ne peut pas se permettre de couper du bois en permanence et on cherche donc à optimiser les instants où l'on effectue des coupes dans cette forêt. Pour répondre à cette question, on commence par modéliser la dynamique de  $X_t$  par ce que l'on appelle une équation différentielle stochastique, à savoir une équation différentielle à laquelle on ajoute du bruit. Ainsi, on suppose que

$$X_t = \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s,$$

avec  $(W_t)_{t \geq 0}$  un mouvement brownien. L'intégrale  $\int_0^t \sigma(X_s)dW_s$ , appelée Intégrale d'Itô, étant un objet mathématique difficile à appréhender, on se contentera de considérer des approximations en temps discret de  $X_t$  :

$$\tilde{X}_{(k+1)h} = b(\tilde{X}_{kh})h + \sigma(\tilde{X}_{kh})\sqrt{h}\varepsilon_{k+1},$$

avec  $(\varepsilon_k)_{k \geq 1}$  des variables indépendantes de loi gaussienne centrée réduite et  $h$  un pas de temps petit.

Trouver les instants optimaux de coupe consiste à trouver un seuil optimal à partir duquel on choisit d'effectuer la coupe lorsque celui-ci est atteint. Il est possible de calculer ce seuil optimal théoriquement. Néanmoins celui-ci dépend de  $b$ , or en pratique on ne connaît pas parfaitement la dynamique de  $X$ , et donc  $b$ , ce qui rend cette solution inexploitable.

Une première solution pour contourner ce problème consiste à se donner une fenêtre de temps  $[0, T]$  où l'on n'agit pas sur la forêt mais où l'on utilise l'observation de  $(X_t)_{t \in [0, T]}$  pour estimer  $b$ . On utilise ensuite cette estimation de  $b$  pour avoir une estimation du seuil optimal que l'on exploite alors tout le reste du temps, après  $T$ . L'inconvénient de cette approche est que l'erreur d'estimation initiale aura un impact sur le long terme. Une autre approche consiste à faire de l'apprentissage par renforcement, en alternant des phases d'apprentissage (on observe juste la dynamique de  $X$  pour améliorer l'estimation de  $b$ ) et des phases d'exploitation (on utilise notre estimation de  $b$  pour avoir une estimation du seuil optimal de coupe, et on applique cette stratégie de coupe approchée). Le but est alors de déterminer le temps optimal que l'on doit passer dans chaque phase.

Toute l'étude mathématique de ce problème se trouve dans l'article suivant (disponible en ligne, en accès libre) : *Nonparametric learning for impulse control problems*, Søren Christensen, Claudia Strauch, arXiv :1909.09528.

Le but de ce projet est d'étudier numériquement le modèle proposé ci-dessus et d'illustrer l'optimalité des solutions proposées dans l'article de Christensen et Strauch. En particulier, on pourra s'intéresser aux points suivants.

1. Simuler des trajectoires de  $\tilde{X}$  pour différents  $b$  et  $\sigma$ , ainsi que sa loi invariante. Étudier l'influence de  $h$ .
2. En supposant  $b$  connu, illustrer l'optimalité du seuil de coupe optimale.
3. En supposant  $b$  inconnu, illustrer l'optimalité de la stratégie d'apprentissage par renforcement proposée dans l'article.

# Étude de l'équation de la chaleur

4TTV606U : Groupe de travail applicatif

---

Afaf Bouharguane

[Afaf.Bouharguane@u-bordeaux.fr](mailto:Afaf.Bouharguane@u-bordeaux.fr)

## Contexte.

L'équation de la chaleur est l'un des modèles fondamentaux de la physique mathématique. Elle permet de décrire la distribution de la température dans un milieu et intervient dans de nombreuses applications industrielles telles que le thermo-formage, la métallurgie ou le refroidissement de composants électroniques.

Dans ce travail, on s'intéresse à l'équation de la chaleur *stationnaire* en dimension deux, qui s'écrit :

$$-\Delta T(x, y) := -\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)(x, y) = f(x, y) \quad \text{dans } \Omega,$$

avec  $\Omega \subset \mathbb{R}^2$ ,  $T$  la température à l'équilibre thermique et  $f$  source de chaleur.

L'opérateur  $\Delta$  est appelé Laplacien.

## Problème académique.

On souhaite calculer la température  $T(x, y)$  d'équilibre d'une plaque carrée  $[0, 1] \times [0, 1]$  non soumise à une source de chaleur et dont les côtés  $x = 0$ ,  $x = 1$  et  $y = 1$  sont maintenus à une température nulle et dont le quatrième côté  $y = 0$  est maintenu à la température  $T(x, 0) = \sin(\pi x)$ .

1. En utilisant la méthode de **séparation des variables**, calculer une solution analytique de ce problème.
2. Déterminer une solution numérique en utilisant la **méthode des différences finies**.
3. Comparer la solution numérique à la solution analytique (visualisation et erreur).

## Application : moulage par thermo-formage

On considère un domaine  $\Omega$  représentant une section plane d'un moule chauffant pour thermo-formage de plastique (voir Figure ci-dessous) dont à chercher à calculer la répartition de la température.

Le moule est chauffé par quatre résistances électriques internes localisées dans des sous-domaines rectangulaires :

$$D_1, D_2, D_3, D_4 \subset \Omega.$$

On suppose de plus que le moule est maintenu à une température de 20 degrés sur son bord inférieur  $\Gamma_1$ , de 50 degrés sur son bord supérieur  $\Gamma_2$  et isolée thermiquement sur ses bords latéraux  $\Gamma_0$ .

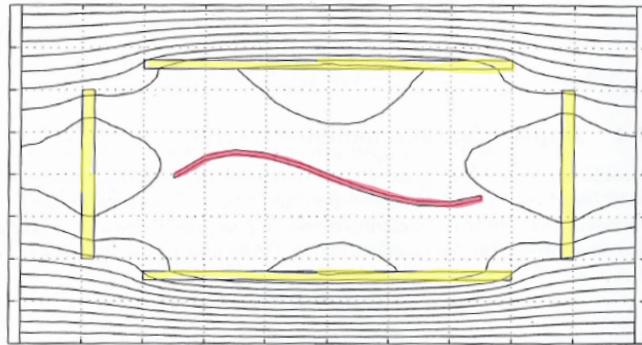


FIGURE 1 – Moule thermique

La modélisation de ce problème s'écrit :

$$\begin{cases} -\Delta T = \sum_{k=1}^4 f_k \mathbf{1}_{D_k} & \text{dans } \Omega, \\ T = 20 & \text{sur } \Gamma_1, \\ T = 50 & \text{sur } \Gamma_2, \\ \frac{\partial T}{\partial n} = 0 & \text{sur } \Gamma_0 \end{cases}$$

où  $f_k$  représente la puissance de la résistance  $D_k$ .

Par une **méthode des moindres carrés**, nous souhaitons déterminer les valeurs à donner aux puissances  $f_k$  dans les 4 résistances pour obtenir la température la plus proche possible de 150 degrés sur un profil  $\gamma$  donné (illustré par exemple en rouge sur la figure). Pour cela,

1. Montrer que ce problème se ramène à un problème des moindres carrés.
2. Écrire le système linéaire vérifié par les coefficients  $f_k$ .
3. Pour des domaines  $D_k$  et  $\gamma$  donnés, donner les résultats numériques obtenus.

# Optimisation de l'allocation de véhicules électriques à des réservations et de leur recharge

Aurélien Froger  
aurelien.froger@math.u-bordeaux.fr  
bâtiment A33 - bureau 264

## Projet

Une entreprise possède une flotte de véhicules électriques qu'elle met à disposition de ses employés par le biais d'une plateforme de réservations. On s'intéresse au problème de sélection des réservations à satisfaire par des véhicules électriques (les autres seront satisfaites par des véhicules thermiques ou par une location de voitures à une compagnie externe) et de planification des recharges de ces véhicules afin de minimiser les coûts opérationnels de l'entreprise.

Le problème à résoudre possède les caractéristiques suivantes. L'entreprise possède un contrat d'approvisionnement en électricité où le coût dépend des moments de la journée. L'électricité peut aussi venir de panneaux photovoltaïques installés sur le parking de l'entreprise et sur les toits des bâtiments. Les véhicules électriques de l'entreprise sont tous identiques (même marque, et même modèle). Chaque véhicule peut être chargé avec une puissance ne devant pas dépasser la limite maximale autorisée par le constructeur (liée aux spécifications de la batterie).

Les hypothèses de modélisation du problème de base sont les suivantes. La version statique du problème est considérée, c'est à dire qu'on prend en compte uniquement les réservations actuellement effectuées sur la plateforme. L'horizon de temps est uniformément discrétisé en périodes de temps. Pour chaque réservation, on connaît sa période de début, sa période de fin et sa quantité d'énergie requise (on a confiance dans l'estimation calculée à partir du trajet à effectuer). Le coût de chaque réservation non satisfaite est composé d'un coût fixe et d'un coût qui dépend linéairement de la quantité d'énergie requise (corrélé très fortement au trajet à effectuer). Chaque véhicule peut être chargé avec une puissance quelconque (respectant seulement les limites physiques) à chaque période de temps. Il possède aussi ses propres périodes de disponibilité (maintenance, réservations déjà validées). Pour chaque période de temps, le prix de l'électricité tiré du réseau est connu. On suppose aussi que l'on connaît une prévision relativement fiable de la production des panneaux photovoltaïques à chaque période de temps. Afin de considérer implicitement les réservations futures, un coût doit être payé pour chaque véhicule électrique en fin d'horizon de temps pour chaque unité d'énergie d'écart avec la charge maximale.

Le travail à réaliser consiste

- définir le problème statique de manière formelle et à expliciter différentes variantes du problème de base (puissance maximale totale pouvant être tirée du réseau, nombre de chargeurs limités, plusieurs types de chargeurs, décalage possible des réservations avec pénalité,...)
- modéliser mathématiquement le problème (et ses variantes) en utilisant la programmation linéaire en nombres entiers et résoudre le modèle à l'aide d'un solveur
- proposer un ou plusieurs algorithmes permettant de résoudre (heuristiquement) le problème et les implémenter
- effectuer des tests numériques sur les données fournies
- si l'avancement le permet, traiter la version dynamique du problème dans laquelle les réservations sont connues uniquement 4 heures avant.

## Données

Les données ont été générées aléatoirement à partir d'un cas d'application chez Honda. Elles ont les caractéristiques suivantes. Chaque véhicule électrique a une batterie de capacité 20kWh et peut être chargé avec une puissance maximale égale à 3,3 kW. Les données de production photovoltaïque et le prix de l'électricité ont été générés aléatoirement de façon à mimer les données fournies par Honda.

La taille des instances est résumée dans le tableau ci-dessous :

Nombre $n$ de véhicules	Horizon de temps	Granularité temporelle	Nombre de réservations
{1, 2, 5}	8 heures	15min	
{5, 10, 20}	2 jours	15min	{4n, 8n, 16n}
{20, 50, 100}	8 jours	15min	

La pénalisation de chaque réservation non satisfaite est égal à un coût fixe de 200€ auquel s'ajoute un coût variable de 150€/kWh. Chaque unité d'écart entre la charge maximale de la batterie et la charge du véhicule à la fin de l'horizon est pénalisée par une valeur égale à 75€/kWh.

## Implémentation

Le langage d'implémentation sera à choisir entre C++, Python et Java. Le code devra être versionné en utilisant Git. Pour résoudre les programmes linéaires en nombres entiers, vous utiliserez un solveur parmi Gurobi (commercial), SCIP (open-source) ou Highs (open-source).

# Calibration de modèles EDO d'activité mitochondriale par algorithme génétique.

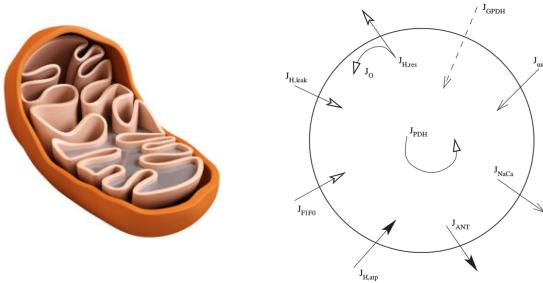
Michael LEGUÈBE, michael.leguebe@inria.fr

## Introduction

Les mitochondries sont les organelles fournissant la source principale d'énergie des cellules biologiques. Elles sont notamment activées par le calcium et le glucose présents dans la cellule. Les modèles de mitochondries transcrivent cette activité sous forme d'équations, qui sont généralement des équations différentielles ordinaires, et produisent des résultats qui, dans l'idéal, se rapprochent de données expérimentales. L'objectif de ce projet est d'implémenter un de ces modèles et de déterminer les paramètres d'entrée de celui-ci en utilisant un algorithme d'optimisation stochastique, ici un algorithme génétique.

## Modèles de mitochondrie

Il existe de nombreux modèles de mitochondries, écrits pour la plupart par des biologistes, mais souvent trop complexes à appréhender par des mathématiciens. Dans ce projet nous étudierons un modèle proposé par Richard Bertram *et al.* en 2006, qui a procédé à de nombreuses simplifications dans le but de faciliter la résolution des équations, ainsi que de rendre possible la calibration de ses paramètres. Ce modèle décrit notamment l'évolution des concentrations en ions calcium  $\text{Ca}^{2+}$  et en ATP (la molécule nécessaire à la production d'énergie) à l'intérieur et à l'extérieur de la mitochondrie. On a ainsi une représentation de la production de celle-ci en fonction de son environnement.



**Figure 1:** Gauche: schéma de mitochondrie. Droite: transferts de molécules entre la mitochondrie et la cellule autour, retenus dans le modèle de Bertram [1].

## Travaux

Pour mener à bien ce projet il faudra dans un premier temps étudier l'article de Bertam *et al.* [1] qui présente le modèle et en extraire les équations qu'il faut résoudre. Il faudra également se renseigner sur CellML, qui est une base de données qui collecte les codes de nombreux modèles biologiques décrivant des phénomènes à l'échelle cellulaire. Un travail de conversion du code téléchargé sera nécessaire, vers un langage qui permette une utilisation aisée par tous (à décider, préférentiellement Python). Il faudra ensuite écrire ou utiliser un solveur d'EDO adapté, qui puisse notamment reproduire les figures de l'article.

La seconde partie du travail, qui peut être commencée en parallèle, consiste à se renseigner sur les algorithmes génétiques et les tester sur des exemples simples, avant de l'appliquer au modèle de Bertram. On déterminera enfin s'il est possible de calibrer ce modèle sur des données synthétiques, et éventuellement expérimentales.

## References

- [1] Richard Bertram, Morten Gram Pedersen, Dan S. Luciani, and Arthur Sherman. A simplified model for mitochondrial atp production. *Journal of Theoretical Biology*, 243(4):575 – 586, 2006.

## 1 Avant-propos

On s'intéresse à un problème de protection des fonds marins. Ce sujet a été inspiré par le travail de Adrien Brunel. Le document scientifique décrivant la problématique est disponible en accès libre. <https://hal.archives-ouvertes.fr/hal-03445922>. Ce sujet est une très libre adaptation du travail initial. Un grand merci à Adrien Brunel, Jérémie Omer et François Clautiaux pour son montage.

## 2 Problématique

Le projet a pour objectif d'étudier différentes versions de plans de protection d'espèces marines.

On s'intéresse à la protection d'un ensemble d'espèces d'animaux marins. Le fond marin est modélisé de manière très simplifiée comme un ensemble de *secteurs* qui correspondent à des cellules dans une grille en deux dimensions (voir Figure 1). Dans chaque secteur, et pour chaque espèce, on dispose d'une estimation du nombre d'individus de l'espèce dans le secteur. Certains secteurs correspondent à des *zones terrestres*, elles ne contiennent aucun individu.

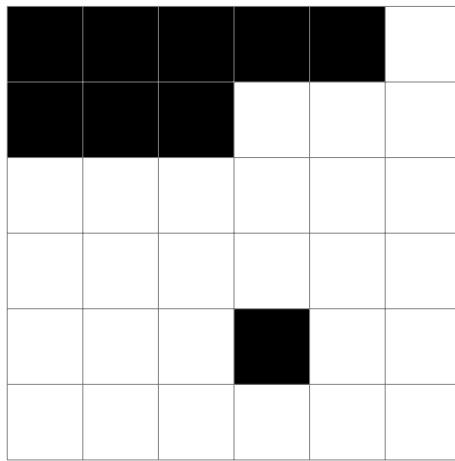


FIGURE 1 – Des secteurs représentés dans une grille  $6 \times 6$ , avec des zones terrestres. Le secteur  $(0, 0)$  est le secteur en haut à gauche de la figure. Les zones terrestres  $(0, 0), (0, 1), (0, 2), (0, 3), (0, 4), (1, 0), (1, 1), (1, 2), (4, 3)$  sont représentées en noir.

Les décisions à prendre consistent à *protéger* un certain nombre de secteurs, de manière à ce qu'un nombre suffisant d'individus de chaque espèce soit localisé dans les secteurs protégés. Lorsqu'un secteur est protégé, un certain nombre d'activités (pêche, tourisme, ...) sont interdites, de manière à préserver les espèces présentes. **Pour toutes les versions de problème considérées, on cherche à minimiser le nombre de secteurs protégés.** Pour chaque espèce, on la considérera comme suffisamment protégée si la somme du nombre d'individus de l'espèce dans les secteurs protégés est supérieure à une valeur donnée.

Les secteurs sont représentés dans une grille de taille  $W \times H$ . Ils sont numérotés de 0 à  $W * H - 1$ . Les lignes et colonnes sont respectivement numérotées de 0 à  $H - 1$  et 0 à  $W - 1$ . Le secteur d'indice  $i$  a pour coordonnées  $(\lfloor i/W \rfloor, i \bmod W)$  où la première valeur correspond au numéro de la ligne dans la grille, et la seconde au numéro de la colonne. Inversement, le secteur de coordonnées  $(x, y)$  a pour indice  $x * W + y$ .

Dans tout ce qui suit, une solution sera représentée par une liste de secteurs protégés.

## 2.1 Première hypothèse : secteurs libres

Dans la première version du problème, on peut sélectionner n'importe quel ensemble de secteurs. La seule contrainte est que l'on doit protéger suffisamment d'individus. Un exemple est donné par la Figure 2.

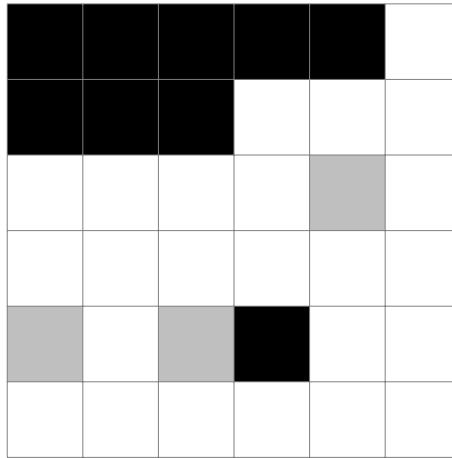


FIGURE 2 – Solution possible pour l'hypothèse 1. Les secteurs protégés sont représentés en gris. On protège trois secteurs : 16, 24 et 26.

## 2.2 Deuxième hypothèse : secteurs rectangulaires

On va maintenant chercher à se rapprocher d'une solution réaliste en introduisant une notion de connexité. On considère qu'une *zone* est un ensemble de secteurs protégés qui se touchent. Deux secteurs se touchent s'ils sont l'un au-dessus de l'autre ou l'un à côté de l'autre, mais pas en diagonal. Sous l'hypothèse 2, chaque zone doit être un rectangle, deux zones ne peuvent pas se toucher, et on ne doit pas utiliser plus de  $K$  rectangles ( $K$  un entier donné en paramètre). Une zone rectangulaire ne peut pas recouvrir un secteur terrestre (Figure 3).

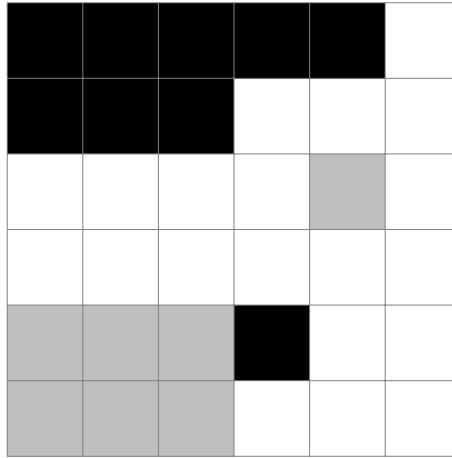


FIGURE 3 – Solution possible pour l'hypothèse 2. On protège 7 secteurs à l'aide de deux zones rectangulaires.

Cette contrainte peut augmenter le nombre de secteurs à protéger.

## 2.3 Troisième hypothèse : secteurs connexes

Dans cette hypothèse, on garde la contrainte du nombre de zones inférieur à  $K$ , mais on relâche la contrainte indiquant que les zones doivent être rectangulaires. On impose uniquement qu'elles doivent être *connexes*. Deux zones sont connexes si elles partagent une arête dans la représentation en deux dimensions (chaque secteur a donc quatre voisins, sauf les bords qui en ont trois, et les coins, qui en ont deux). Pour cette hypothèse, il n'est pas nécessaire d'empêcher deux zones de se toucher (c'est équivalent à construire une zone consistant en l'union des deux) (voir Figure 4).

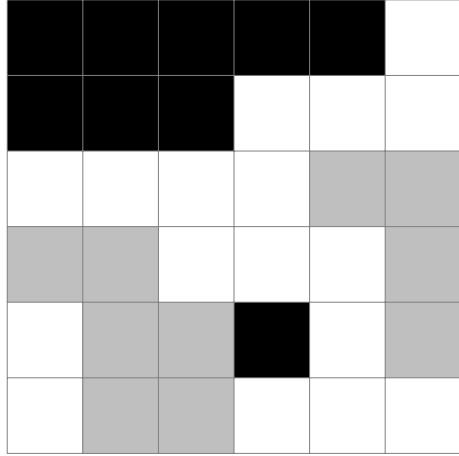


FIGURE 4 – Solution possible pour l’hypothèse 3. On protège 10 secteurs à l’aide de deux secteurs connexes.

### 3 Travail à effectuer

L’objectif du projet est de déterminer l’impact des différentes hypothèses sur le coût de la protection et de produire un rapport décrivant l’approche permettant de tirer ces conclusions, ainsi que les expérimentations. Pour les hypothèses 2 et 3, il faudra aussi étudier l’impact du nombre de zones maximum sur les résultats.

Pour cela plusieurs tâches devront être menées à termes :

- implémenter un parser permettant de lire les données des fichiers instances et de les stocker dans des structures de données ;
- proposer, pour chaque version, un modèle de programmation linéaire en nombres entiers, et résoudre ces modèles à l’aide d’un solveur ;
- implémenter des heuristiques permettant d’obtenir des solutions de bonnes qualités dans un temps raisonnable ;
- implémenter un vérificateur de solution indépendant des méthodes d’optimisation, permettant de vérifier si une solution donnée satisfait bien aux contraintes posées ;
- réaliser des expérimentations numériques sur les instances proposées, et analyser les résultats obtenus.

## A Format des fichiers

### A.1 Fichier de données

Le fichier d’entrée est organisé de la manière suivante. On indique d’abord le nombre d’espèces, puis on a une ligne par espèce, qui indique le nom et le nombre d’individus à protéger. On indique ensuite les dimensions  $W$  et  $H$  de la grille dans laquelle sont situés les secteurs. On a  $\text{largeur} \times \text{hauteur}$  lignes, une par secteur, qui indiquent leur position géographique, s’il s’agit d’une zone de terre ou une zone aquatique, et pour les zones aquatiques est indiqué le nombre d’individus pour chaque espèce, dans l’ordre. Un exemple de donnée est rapportée en Figure 5.

Le nombre de zones maximum à utiliser est un paramètre à faire varier dans les expérimentations.

### A.2 Format du fichier résultat

Le fichier résultat doit contenir sur la première ligne le nombre de secteurs protégés. La deuxième ligne doit contenir le nombre de zones utilisées. Ensuite chacune des lignes suivantes indique respectivement les secteurs dans chacune des zones. Pour l’hypothèse 1, on considère que tous les secteurs sélectionnés sont dans la même zone (non connexe). Par exemple, pour une solution de valeur 3 contenant les secteurs 15, 28 et 57 dans une seule zone, on doit produire le fichier de la figure 6. Si on utilise deux zones  $\{1, 2\}$  et  $\{12, 16\}$ , on obtient le fichier de la figure 7.

```

3
Requin blanc;40
Requin marteau;30
Requin scie;20
4;4
0 ; 0 ; 1
0 ; 1 ; 1
0 ; 2 ; 0 ; 10 ; 20 ; 30
0 ; 3 ; 0 ; 10 ; 10 ; 3
1 ; 0 ; 0 ; 10 ; 0 ; 10
1 ; 1 ; 1
1 ; 2 ; 0 ; 0 ; 0 ; 0
1 ; 3 ; 0 ; 0 ; 0 ; 0
2 ; 0 ; 0 ; 10 ; 10 ; 10
2 ; 1 ; 0 ; 10 ; 10 ; 5
2 ; 2 ; 0 ; 10 ; 10 ; 5
2 ; 3 ; 0 ; 0 ; 0 ; 0
3 ; 0 ; 0 ; 0 ; 0 ; 0
3 ; 1 ; 0 ; 0 ; 0 ; 0
3 ; 2 ; 0 ; 10 ; 10 ; 6
3 ; 3 ; 0 ; 10 ; 10 ; 10

```

FIGURE 5 – Exemple de fichier d'entrée pour une instance avec 3 espèces pour lesquelles on doit respectivement protéger 40, 30 et 20 individus. On utilise une grille de taille 4x4 (et donc 16 secteurs). Le premier secteur (secteur 0) a pour coordonnées (0,0). C'est un secteur "terre", il a donc un 1 en troisième position, et pas de valeurs pour les espèces. Le troisième secteur (secteur 2), a pour coordonnées (0,2). C'est un secteur aquatique (0 en troisième position), il a donc trois valeurs indiquées, une par espèce. On trouve donc 10 requins blancs, 20 requins marteau, et 30 requins scie.

3
1
15;28;57

FIGURE 6 – Exemple de fichier de sortie, si on utilise les trois secteurs 15, 28 et 57 dans une seule zone.

4
2
1;2
12;16

FIGURE 7 – Exemple de fichier de sortie, si on utilise les quatre secteurs {1, 2, 12, 16} dans deux zones.

# Réduction de dimensions en apprentissage non supervisé : comparaison entre Analyse en Composantes Principales et Autoencodeurs

---

Vincent Couallier

[Vincent.Couallier@u-bordeaux.fr](mailto:Vincent.Couallier@u-bordeaux.fr)

L'ACP telle qu'elle a été vue en cours est une méthode de réduction de dimension qui fournit une représentation comprimée des données d'un tableau rectangulaire de type *individusxvariables*, le plus souvent avec  $n$  individus et  $p$  variables numériques,  $n > p$ , en cherchant à reconstruire au mieux les données, avec de fortes connexions avec la décomposition en valeurs singulières d'un tableau rectangulaire. Surtout, quand elle est appliquée dans un objectif d'analyse statistique (et non comme méthode de pré-processing pour préparer les données à un autre traitement), on cherche à interpréter les résultats, pour comprendre les corrélations entre variables, ou les proximités entre les individus.

Les autoencodeurs forment une partie des modèles de réseaux de neurones qui peuvent répondre à la même problématique, en nécessitant des méthodes numériques différentes, des calculs parfois couteux, en proposant une généralisation du problème d'optimisation au cas non linéaire. Par contre, l'aspect réduction de dimensions est souvent plus important que l'aspect "explicatif".

Ce projet vise à explorer les différences entre ces deux méthodes, à les comparer mathématiquement et par simulation, pour finalement les appliquer sur un jeu de données réelles fourni par l'enseignant.

On propose de décomposer le travail selon les tâches suivantes :

1. Rappeler les fondements mathématiques de l'ACP, en insistant sur le problème d'optimisation et la décomposition aux valeurs propres d'une matrice à définir. Rappeler la méthode de construction des composantes principales et montrer pourquoi le problème est linéaire.
2. Analyser le lien entre ACP décrite typiquement par la diagonalisation d'une matrice carrée à définir, et la décomposition en valeurs singulières.
3. Avec les mêmes notations, présenter le principe de l'autoencodeur avec fonction d'activation linéaire, comparer le problème d'optimisation avec celui de l'ACP et montrer l'équivalence exacte des résultats de projection sur un sous-espace de dimension réduite (disons pour illustrer  $d = 2$ ). Quelles différences existe-t-il dans la construction des axes entre les deux méthodes ? Analyser le problème en terme de variance projetée, d'erreurs des moindres carrés, sur les axes et sur le plan.
4. Programmer les deux méthodes (en Python ou R). On s'appuie sur les méthodes existantes de diagonalisation, mais on programme l'algorithme d'optimisation du réseau de neurones.
5. Etendre le cas de l'autoencodeur avec deux couches non linéaires cachées symétriques et analyser les différences dans des simulations. Montrer le gain (potentiel) dans la diminution des erreurs, au prix d'une perte (potentiel) d'autres critères.

Dans les tâches proposées, on utilisera comme fil conducteur pour illustrer les méthodes une application à un jeu de données de mesures météorologiques et de pollution sur la ville de Périgueux ( 8000 observations horaires d'un vecteur de mesures de température, pression atmosphérique, vent, direction, ..., et de pollution NO<sub>2</sub>, NO, ozone....).

# Calcul précis de la période des solutions d'un modèle d'électrocardiologie

4TTV606U : Groupe de travail applicatif

2025-2026

---

Yves Coudière

[Yves.Coudière@u-bordeaux.fr](mailto:Yves.Coudière@u-bordeaux.fr)

## Contexte

Le nœud sinusal du cœur est un ensemble de cellules qui imposent un le rythme cardiaque. Un modèle mathématique simplifié de ces cellules est donné par le système d'équations différentielles donné ci-dessous [1] :

$$\begin{aligned} v'(t) &= \frac{h(v+a)(v+a-\lambda)(v-1)}{\tau_{\text{in}}} + \frac{v}{\tau_{\text{out}}}, \\ h'(t) &= \frac{h_\infty(v) - h}{\tau(v)}, \quad \text{avec } \frac{1}{\tau(v)} = \frac{1}{\tau_{\text{open}}} h_\infty(v) + \frac{1}{\tau_{\text{close}}} (1 - h_\infty(v)), \end{aligned}$$

où  $\tau_{\text{in}}$ ,  $\tau_{\text{out}}$ ,  $\tau_{\text{open}}$ ,  $\tau_{\text{close}}$ ,  $a$  et  $\lambda$  sont des paramètres et  $h_\infty(v) = \frac{1}{2} \left(1 - \tanh\left(\frac{v-v_g}{\epsilon}\right)\right)$  où  $v_g$  et  $\epsilon$  sont deux autres paramètres. Des valeurs typiques (en ms) des paramètres sont données dans le tableau 1.

$\tau_{\text{in}}$	$\tau_{\text{out}}$	$\tau_{\text{open}}$	$\tau_{\text{close}}$	$v_g$	$\epsilon$	$a$	$\lambda$
0.3	6.0	120.0	150.0	0.13	1.e-3	0.02	0.0

TABLE 1 – Exemple de paramètres du modèle.

Dans ce modèle, la fonction inconnue  $v(t)$  représente la tension électrique (normalisée entre 0 et 1) à travers la membrane des cellules et la fonction  $h(t)$  est une variable d'état de la membrane cellulaire.

## Objectifs

Les objectifs du travail sont

- d'étudier ce système différentiel d'un point de vue théorique,
- de calculer des solutions approchées,
- de déterminer la ou les solutions périodiques et de calculer leurs périodes le plus précisément possible.

## Travail

Dans un premier temps, on pourra étudier les fonctions qui définissent le système différentiel d'un point de vue théorique : continuité, différentiabilité, étude de signe, et obtenir une vision qualitative du portrait de phase. On s'intéressera en particulier au changement qui apparaissent dans ce portrait lorsque le paramètre  $a$  varie.

En parallèle, on mettra en place un ou plusieurs schémas à un pas pour calculer des solutions approchée, notamment des schémas d'ordre élevé à pas de temps adaptatif (de type méthodes de Runge et Kutta emboîtées). On pourra alors tracer ces solutions dans le plan de phase pour confirmer l'intuition construite par l'étude qualitative du système.

Pour rechercher la solution périodique, si elle existe, plusieurs stratégies sont possibles : approche par point fixe, par méthode de Newton directes, par méthode de Newton après changement de variable en temps. On programmera au moins une de ces méthodes, avec comme principal objectif de calculer la période de la solution.

## Références

- [1] Karima DJABELLA, Mayer LANDAU et Michel SORINE. “A two-variable model of cardiac action potential with controlled pacemaker activity and ionic current interpretation”. In : *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007.  
DOI : [10.1109/cdc.2007.4434970](https://doi.org/10.1109/cdc.2007.4434970).