

IMPACT DE LA MÉTRIQUE POUR LE CLUSTERING DE SÉRIES TEMPORELLES DE QUATERNIONS: APPLICATION AUX PATIENTS ATTEINTS DE SCLÉROSE EN PLAQUES

Klervi Le Gall¹ & Lise Bellanger² & Aymeric Stamm³ & David Laplaud⁴

¹ *Laboratoire de Mathématique Jean Leray, UMR CNRS 6629, Nantes Université, France.
klervi.legall@univ-nantes.fr*

² *Laboratoire de Mathématique Jean Leray, UMR CNRS 6629, Nantes Université, France.
lise.bellanger@univ-nantes.fr*

³ *Laboratoire de Mathématique Jean Leray, UMR CNRS 6629, Nantes Université, France.
aymeric.stamm@cnrs.fr*

⁴ *CR2TI, INSERM U1064, CHU de Nantes, Nantes Université, France.
david.laplaud@univ-nantes.fr*

Résumé. Le choix des métriques en classification non supervisée est une étape clé pour la construction de groupes séries temporelles de quaternions unitaires aux allures similaires. Nous détaillons ici la construction de plusieurs dissimilarités, ainsi la méthode de classification non supervisée par compromis choisie, et le cadre qui nous permet de d'évaluer ces partitions. Nous appliquerons cette méthodologie à un biomarqueur appelé Signature de Marche (SdM) qui caractérise la rotation de la hanche d'un individu au cours d'un cycle de marche moyen où les rotations sont représentées par des quaternions unitaires. Ces SdM ont été mesurées chez des patients atteints de sclérose en plaques pour lesquels nous obtenons une partition de données de marche interprétable et stable. La stabilité de cette méthode est étudiée à l'aide de l'inclusion de volontaires sains ainsi qu'avec la génération de données synthétiques.

Mots-clés. Séries temporelles de quaternions, données fonctionnelles, analyse de la marche, sclérose en plaques, classification non supervisée

Abstract. The choice of metrics in clustering is a key stage in the construction of clusters of unit quaternion time series with similar appearances. In this paper, we describe the construction of several dissimilarities, the clustering by compromise method chosen, and the framework that allows us to evaluate these clusters. We will apply this methodology to a biomarker called the Individual Gait Pattern (IGP), which characterises the rotation of an individual's hip during an average gait cycle, where rotations are represented by unit quaternions. These IGPs have been measured in multiple sclerosis patients for whom we obtain an interpretable and stable gait data partition. The stability of this method is studied by including healthy volunteers and by generating synthetic data.

Keywords. Quaternion time series, functional data, gait analysis, multiple sclerosis, clustering

1 Introduction

Les séries temporelles de quaternions unitaires sont des objets mathématiques complexes, pour lesquels il est nécessaire d’adapter les outils statistiques classiques. Nous savons que le choix de la métrique est un élément clé dans tous travaux de classification non supervisée, et que de nombreuses dissimilarités ou distances peuvent être construites entre deux séries temporelles de quaternions. L’objectif de ce travail est de proposer un cadre dans lequel nous pouvons appliquer une méthode de classification non supervisée à ces différentes métriques et ainsi évaluer leur pertinence pour créer des groupes interprétables et stables.

Le dispositif eGait permet de construire un biomarqueur appelé signature de marche (SdM). La SdM caractérise la rotation de la hanche d’un individu au cours d’un cycle de marche moyen. Les rotations sont représentées par des quaternions unitaires. La SdM fournit une mesure quantitative de la démarche à un moment donné sur laquelle nous travaillons afin de comprendre les troubles de la marche dans le contexte de la sclérose en plaques.

Dans le cadre du partitionnement de ces données nous souhaitons prendre en compte des informations externes, ou secondaires, qui viennent consolider les informations contenues dans les Signatures de Marche afin d’obtenir des groupes de patients aux démarches, et degrés de pathologie similaires. Ces informations sont apportées par le score EDSS et ce partitionnement est rendu possible par une méthode de classification non supervisée par compromis.

Nous détaillons ici la méthodologie nous permettant d’évaluer la pertinence des différentes métriques définies pour la classification non supervisée par compromis des séries temporelles de quaternions unitaires.

2 Matériel et méthodes

2.1 Données de marche

La démarche des patients est enregistrée à l’aide d’un système de capteurs qui transmet le vecteur d’orientation absolue du dispositif positionné à la hanche des patients sous la forme d’un quaternion unitaire toutes les 0,01 secondes, une série temporelle de quaternions unitaires est donc recueillie pour chaque patient. Ces données brutes sont ensuite segmentées en cycles de marche et la *signature de marche* (SdM) d’un individu correspond au centre des cycles de marche détectés, ce centre est obtenu avec une méthode de k-means alignement Sangalli et al. (2010) pour lequel le nombre de cluster est égal à 1. La série temporelle est ensuite exprimée en pourcentage de la durée totale. Ces SdM sont des séries temporelles de quaternions (QTS) unitaires et décrivent les rotations ”moyennes” de la hanche pendant un cycle de marche. Les étapes de l’algorithme menant à la SdM d’un individus se trouvent dans l’article de Drouin et. al (2022a).

Les données sur lesquelles nous nous appuyons sont issues de la recherche eMSGait, une étude monocentrique non contrôlée sur la quantification des troubles de la marche chez les

patients atteints de sclérose en plaques. Cette étude vise à évaluer la possibilité de répartir les patients SEP inclus dans l'étude en groupes homogènes selon leurs caractéristiques. Entre août 2021 et août 2022, les patients âgés de 18 à 64 ans atteints de SEP, qui n'ont pas observé de poussée au cours des 5 semaines dernières semaines, capables de marcher avec canne (ou deux cannes) ou sans assistance, non opposés à l'étude et affiliés à un régime de sécurité sociale ont été inclus au CHU de Nantes.

L'étude comporte une visite unique correspondant à une consultation programmée dans le cadre de la prise en charge et du suivi habituel. Les données de marche ont été recueillies par le capteur de mouvement MetaMotionR r0.4 de MbiEntLab placé sur une ceinture au niveau de la hanche droite lors de deux tests neurologiques *Time 25 Foot Walk* (T25FW) Kieseier-Pozzilli (2012) espacés de 5 minutes. Des données cliniques décrivant la sévérité de leur pathologie ont également été mesurées par les cliniciens du Centre d'Investigation Clinique du CHU de Nantes.

Parmi les données cliniques mesurée figure le score EDSS (Expanded Disability Status Scale) décrit par Kurtzke (1983). Cet indicateur, évalué par un neurologue, permet d'attribuer un score entre 0 (examen neurologique normal) et 10 (Décès à cause de la SEP) aux patients. L'examen clinique est composé de sept ou huit fonctions neurologiques à évaluer qui peuvent impacter ou non la démarche des patients. Les patients peuvent alors être répartis en 3 groupes en fonction de la sévérité globale de leur pathologie représentée par leur score EDSS:

- Sévérité faible (SF) : $EDSS < 2$
- Sévérité modérée (SM) : $2 \leq EDSS < 4$
- Sévérité élevée (SE) : $4 \leq EDSS$

Ces groupes sont illustrés dans la figure 1.

2.2 Méthodes

2.2.1 Séries temporelles de quaternions unitaires

Un quaternion est un élément $(w,x,y,z) \in \mathbb{R}^4$, qui est une extension des nombres complexe et qui s'écrit comme suit:

$$\mathbf{q} = (w, x, y, z)^\top = w + ix + jy + kz \in \mathbb{R}^4, \quad (1)$$

où i, j et k respectent $i^2 = j^2 = k^2 = ijk = -1$, Voight (2005).

Les quaternions unitaires sont ceux pour lesquels $\|\mathbf{q}\| = w^2 + x^2 + y^2 + z^2 = 1$ de sorte que leur norme soit égale à 1. Le groupe des quaternions unitaires \mathbb{H}_u forme un groupe de Lie isomorphe au groupe unitaire spécial $SU(2)$ qui couvre deux fois le groupe des matrices de rotations en 3-Dimensions, Dijkhuizen (1994).

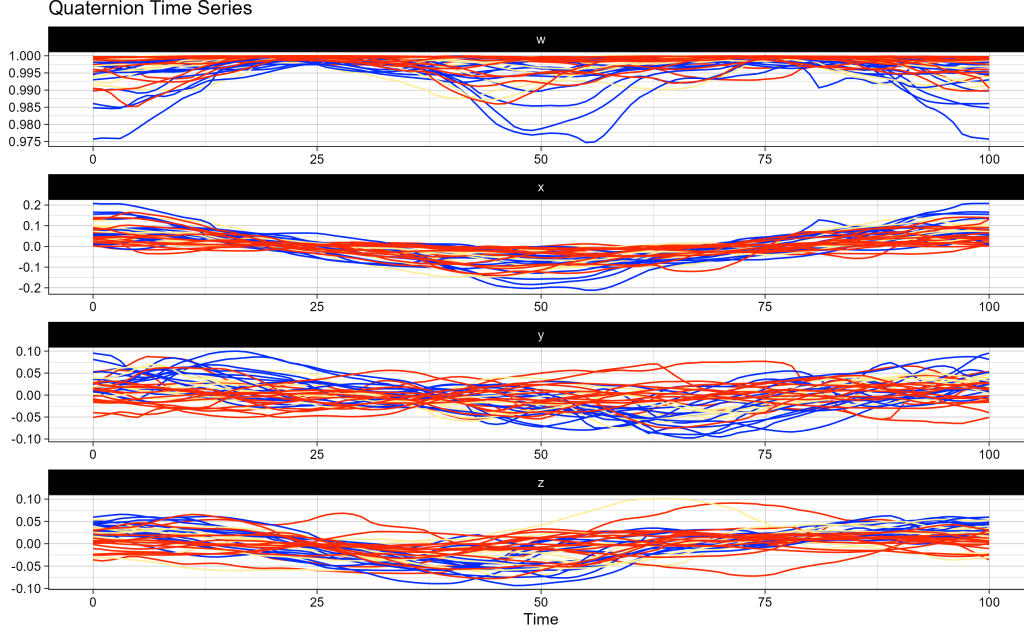


Figure 1: Signatures de Marche des 39 patients de l'étude eMSGait colorés par leur degré de sévérité pathologique, bleu: SF, jaune: SM, rouge: SE.

Les quaternions unitaires décrivent une rotation d'un angle θ autour d'un axe $\mathbf{u} = (u_x, u_y, u_z)^\top \in \mathbb{S}^2$ où \mathbb{S}^2 est la 2-sphère qui peut être exprimée comme suit:

$$\mathbf{q} = \left(\cos \frac{\theta}{2}, \mathbf{u}_x \sin \frac{\theta}{2} i, \mathbf{u}_y \sin \frac{\theta}{2} j, \mathbf{u}_z \sin \frac{\theta}{2} k \right)^\top \quad (2)$$

L'équation d'Euler pour les nombres complexes unitaires peut se généraliser généralise de la même manière aux quaternions unitaires. Il est alors possible d'écrire un quaternion unitaire sous sa *forme polaire*:

$$\mathbf{q} = \exp(\mathbf{u} \frac{\theta}{2}) = \cos \frac{\theta}{2} + \mathbf{u} \sin \frac{\theta}{2} \quad (3)$$

Série temporelle de quaternion unitaires

Une série temporelle de quaternions unitaires (QTS) est un ensemble de quaternions unitaires suivant une grille temporelle $t_{i,1}, \dots, t_{i,n}$. On note une QTS comme: $\mathbf{Q}_i = (\mathbf{q}_{i,1}, \dots, \mathbf{q}_{i,n})$. Elle représente des rotations 3D consécutives dans le temps.

Espace tangent d'une série temporelle de quaternions unitaires

Nous pouvons déduire le *logarithme* d'un quaternion unitaire depuis sa forme polaire:

$$\ln(\mathbf{q}) = \ln(\exp(\mathbf{u} \frac{\theta}{2})) = \mathbf{u} \frac{\theta}{2} \quad (4)$$

Cette transformation est une application entre l'espace des quaternions unitaires \mathbb{H}_u et l'espace tangent $\mathfrak{T}_{\mathbf{q}_0}(\mathbb{H}_u) \subseteq \mathbb{R}^3$, au point $\mathbf{q}_0 = (1, 0, 0, 0)$, Piórek (2020). La transformation inverse est possible à l'aide de l'exponentielle:

$$\exp(\mathbf{q}) = \exp(w) \left(\cos(\|\mathbf{v}\|), \frac{\mathbf{v}^\top}{\|\mathbf{v}\|} \sin(\|\mathbf{v}\|) \right)^\top, \quad (5)$$

avec \mathbf{v} la partie vectorielle de $\mathbf{q} = (x, y, z)^\top$.

2.2.2 Comment mesurer la distance entre Signatures de Marche?

La distance euclidienne dans \mathbb{R}^4 définit une métrique qui n'est pas fermée par rapport à l'algèbre des quaternions. Les quaternions unitaires vivent sur une sphère 3, donc la plus courte distance entre deux points correspond à une ligne géodésique, elle peut aussi être interprétée comme l'angle de rotation entre \mathbf{q}_1 et \mathbf{q}_2 . La **distance géodésique** entre deux quaternions peut s'écrire:

$$d(\mathbf{q}_1, \mathbf{q}_2) = 2\arccos \operatorname{Re}(\mathbf{q}_1^{-1} \mathbf{q}_2), \quad \mathbf{q}_1, \mathbf{q}_2 \in \mathbb{H}_u, \quad (6)$$

Elle a notamment été utilisée dans des travaux de Piórek-Jabłoński (2020).

Pour mesurer la distance entre deux signatures de marche, plusieurs options s'offrent à nous selon que l'on considère l'approche série temporelle ou l'approche fonctionnelle:

Il est possible d'utiliser une généralisation de l'algorithme de Dynamic Time Warping, fréquemment utilisée pour mesurer des dissimilarités entre séries temporelles, aux quaternions à l'aide de la distance géodésique comme propose Jabłoński (2012). Cette dissimilarité, s'écrit

$$QDTW(\mathbf{Q}_1, \mathbf{Q}_2) = \min_W \sum_{i=1}^K d(\mathbf{q}_{1,w_{i,1}}, \mathbf{q}_{2,w_{i,2}}), \quad (7)$$

Où $d(., .)$ est la distance géodésique présentée dans l'équation 6, W est une fonction d'alignement, et K est le nombre d'alignements nécessaires entre \mathbf{Q}_1 et \mathbf{Q}_2 .

Comme détaillé dans la section précédente, il est possible de projeter les séries temporelles de quaternions dans un espace tangent à l'hypersphère. Cette propriété nous permet d'utiliser des distances entre des fonctions, pour lesquelles on pourra choisir de faire de l'alignement fonctionnel ou non.

Soient $f_{\mathbf{Q}_1}$ et $f_{\mathbf{Q}_2}$ deux fonctions (deux séries temporelles de log-quaternions dans notre cas) correspondant aux projections de \mathbf{Q}_1 et \mathbf{Q}_2 dans l'espace tangent $\mathfrak{T}_{\mathbf{q}_0}(\mathbb{H}_u) \subseteq \mathbb{R}^3$, la **distance L_2 normalisée** s'écrit:

$$D_n(f_{\mathbf{Q}_1}, f_{\mathbf{Q}_2}) = \frac{\|f_{\mathbf{Q}_1} - f_{\mathbf{Q}_2}\|_2}{\|f_{\mathbf{Q}_1}\|_2 + \|f_{\mathbf{Q}_2}\|_2}, \quad (8)$$

Où $\|f_{\mathbf{Q}_1} - f_{\mathbf{Q}_2}\|_2 = (\int |f_{\mathbf{Q}_1} - f_{\mathbf{Q}_2}|^2)^{1/2}$

Une fonction peut $f_{\mathbf{Q}}$ également être caractérisée par sa *Square Root Slope Function* (SRSF) $g(t) \in L^2$, Tucker-Srivastava (2013)

$$g(t) = \frac{\dot{f}_{\mathbf{Q}}(t)}{\sqrt{\|\dot{f}_{\mathbf{Q}}(t)\|_2}}, \quad (9)$$

où $\dot{f}_{\mathbf{Q}}(t)$ est la dérivée en temps d'une log-série temporelle de quaternions.

La distance D_s entre deux SRSF $g_1(t)$ et $g_2(t)$ est la norme L_2 de la différence. $D_s(f_{\mathbf{Q}_1}, f_{\mathbf{Q}_2}) = \|g_1(t) - g_2(t)\|_2$.

Enfin, on peut décider d'aligner ces SRSF afin de calculer une distance entre les amplitudes des fonctions alignées. Soient $\gamma \in \Gamma(I)$ les fonctions nécessaires à l'alignement des SRSF, cette distance s'écrit:

$$D_y(f_{\mathbf{Q}_1}, f_{\mathbf{Q}_2}) = \inf_{\gamma \in \Gamma} \|g_1 - (g_2 \circ \gamma)\sqrt{\dot{\gamma}}\|_2 \quad (10)$$

2.2.3 Classification non supervisée par compromis

Prétraitements des données

Les SdM sont calculées sur les patients dans des conditions précises et répliquables, en réalisant un test du Time-25-FootWalk (T25FW). Mais il est possible de considérer que ces données, comme un certain nombre des données issues de capteurs, peuvent être bruitées. Il faut alors se poser la question de débruiter ces données, par exemple, en appliquant une Analyse en Composante Principales Fonctionnelle Multivariée (MFPCA) sur les $f_{\mathbf{Q}}$, projections des SdM dans l'espace Tangent (TPCA), Ramsay-Silverman (2005) et Happ-Greven (2018). On choisit de reconstruire les données sur les p premières composantes principales telles que l'on conserve environ 80 % de l'inertie du jeu de données.

De plus, certains patients atteints de SEP peuvent avoir une atteinte cognitive importante, qui ne se caractérise pas par une démarche particulière, mais ces patients peuvent avoir des difficultés à comprendre les consignes du test de marche et à les appliquer. Pour ne pas biaiser les résultats, on peut exclure ces patients.

Avant de calculer les matrices de distances entre les observations du jeu de données, il y a donc quatre voies de prétraitements différentes à considérer:

- Suppression des patients fortement cérébraux,
- Suppression des patients fortement cérébraux et débruitage des SdM,

- Conservation des patients fortement cérébraux,
- Conservation des patients fortement cérébraux et débruitage des SdM.

Une méthode de CAH par compromis: HclustCompro

HclustCompro est une méthode de classification ascendante hiérarchique par compromis, qui prend en compte une source d'information principale et une source d'information secondaire et permet d'appliquer une CAH à une matrice de distance considérée comme le meilleur compromis entre ces deux sources d'informations, cette méthode a déjà été appliquée aux données de marche par Drouin et al (2022b).

La première étape de cette méthode est le choix des mesures de dissimilarités, chacune des distances listée dans la section 2.2.2 est normalisée, puis sera choisie comme source d'information principale D_1 , et la distance de Gower normalisée, qui mesure les différence entre les scores EDSS (ou score spécifique à la marche) des patients est choisie comme source d'information secondaire D_0 .

On cherche alors le paramètre α qui détermine la matrice de distance qui serait la meilleure combinaison convexe de ces deux sources d'information $D_\alpha = \alpha \times D_1 + (1 - \alpha) \times D_0$.

D_α est alors calculée pour chaque α sur une grille entre 0 et 1, et une CAH, avec stratégie d'agglomération *moyenne* ou *complète* est appliquée pour déterminer le dendrogramme T_α .

La valeur de pondération telle que T_α résulte du meilleur compromis entre les deux sources de données initiales, $\hat{\alpha}$, est déterminée par la minimisation du critère suivant:

$$CorCrit_\alpha = |Cor(D_1, D_\alpha^{coph}) - Cor(D_0, D_\alpha^{coph})|, \quad (11)$$

où D_α^{coph} est la distance cophénétique entre deux observations du dendrogramme T_α obtenu par CAH avec la même stratégie d'agrégation.

Lorsque nous travaillons avec une métrique basée sur des données fonctionnelles, les courbes à l'intérieur des groupes doivent être alignées sur leur centroïde une fois le regroupement hiérarchique effectué. Donc pour chaque nombre de classe possible, les courbes sont alignées sur les centres de classes afin de déterminer une nouvelle matrice de distances (norme L_2 sans alignement) entre les observations sur laquelle on pourra calculer les indicateurs permettant le choix du nombre de classes.

Choix du nombre de classes

Après le calcul du dendrogramme par la méthode HClustCompro, il est nécessaire de déterminer le nombre de classes dans lesquelles seront réparties les observations. Les critères *Within Sum of Squares (WSS)* et *Silhouette Rousseeuw* (1987) sont calculés pour chaque nombre de classes $K \in \{2, \dots, N - 1\}$. Ainsi soit une partition $\{C_1, \dots, C_K\}$, on calcule:

$$WSS(K) = \sum_{k=1}^K \sum_{i \in V_k} D^2(\mathbf{Q}_i, \tilde{\mathbf{Q}}_k), \quad (12)$$

où D^2 correspond à la dissimilarité choisie entre les deux observations, $\tilde{\mathbf{Q}}_k$ au médoïde de la classe et $V_k := \ell \in [[1, n]] : Q_\ell \in C_k$.

Ainsi que

$$S_{Q_i}(K) = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (13)$$

où $a_i = \frac{1}{N_k} \sum_{i,j \in V_k} D(Q_i, Q_j)$ est la distance moyenne entre une observation Q_i et toutes les observations de sa classe et $b_i = \frac{1}{N_k} \sum_{i,j \in V_{k_2}} D(Q_i, Q_j)$ est la distance moyenne entre une observation Q_i et toutes les observations de la classe la plus proche à laquelle elle n'a pas été affectée, on s'intéresse aux valeurs de: $S(K) = \frac{1}{N} \sum_{i=1}^N S_{Q_i}(K)$.

Pour les distances calculées sur les projections des SdM dans R^3 , on peut facilement adapter ces indicateurs en utilisant les matrices de distances entre fonctions ou SRSF après réalignement sur centroïdes de classes.

$$WSS(K) = \sum_{k=1}^K \sum_{i \in V_k} D^2(f_{\mathbf{Q}_1}^{(aligned)}, \tilde{f}_{\mathbf{Q}_k}), \quad (14)$$

avec $V_k := \ell \in [[1, n]] : f_{\mathbf{Q}_\ell}^{(aligned)} \in C_k$ et D la norme L_2 entre les fonctions alignées. Le critère silhouette s'adapte de manière analogue.

Le nombre de classes retenu est celui qui maximise le critère silhouette sans que cela ne soit au détriment du critère WSS.

2.3 Critères d'évaluation

2.3.1 Choix de la meilleure partition

Nous appliquons donc la méthode de CAH par compromis HclustCompro avec des stratégies d'agglomération moyennes ou complètes aux matrices de distances obtenues pour chacune des quatre voies de prétraitement possibles ainsi que chaque distance présentée dans la section 2.2.2.

Une fois que le dendrogramme représentant le meilleur compromis entre ces deux sources d'information est déterminé, nous nous assurons que la *corrélation cophénétique* entre la matrice D_α et la matrice, D_α^{coph} est suffisante, une corrélation élevée signifie que le dendrogramme est une représentation fidèle de ces données. Les dendrogrammes pour lesquels cette corrélation n'est pas suffisante seront exclus.

Parmi les possibilités restantes, nous écartons également les partitions avec un WSS trop important, indicateur d'une grande variabilité intra-classes.

Par la suite, le *V de Cramer* est calculé entre la répartition des observation proposée et

les scores EDSS. Il est est défini ainsi:

$$V = \sqrt{\frac{\chi^2}{N \times \min(K - 1, J - 1)}} \in [0; 1], \quad (15)$$

Où, N est le nombre d'observations, J est le nombre de modalités de la variable étudié (EDSS), K est le nombre de classes et χ^2 est la statistique du test du Chi-deux d'indépendance. A l'issue de cette étape, nous ne conservons que des options pour laquelle il y a une bonne relation entre le score EDSS et la partition, ce qui est traduit par un V de Cramer proche de 1.

Enfin, la meilleure partition parmi celles restantes est choisie selon la taille des groupes, l'étendue du score EDSS dans les groupes, le nombre de singletons et l'avis du Professeur David Laplaud, chef du Service de Neurologie au CHU de Nantes.

2.3.2 Stabilité de la méthode de CAH par compromis

La stabilité de cette partition est vérifiée dans un premier temps en ajoutant des volontaires sains à l'effectif de patients SEP et étudiant l'impact sur les groupes.

La stabilité de la méthode de classification semi-supervisée est testée en utilisant des jeux de données synthétiques. Nous avons développé une méthode de génération de données synthétiques de marche basée sur une approche mêlant réduction de dimension et plus proches voisins qui nous permet de générer des Signatures de Marche et des EDSS synthétiques. En utilisant ces données, nous vérifions que les groupes créés partagent des caractéristiques similaires en termes d'indicateurs spatio-temporels et cliniques.

3 Résultats et perspectives

La méthodologie suivie a permis de mettre en lumière quel prétraitement de données et quelle métrique étaient les meilleurs choix pour nous amener à former des groupes de patients à la démarche similaire, et que nous pouvons expliquer cliniquement. Nous avons également montré la stabilité de cette partition.

A l'avenir, nous nous intéresserons également à la possibilité de considérer que la distance que nous utilisons soit invariante aux rotations. Le cadre mathématique nécessaire à cette démarche a été publié par Kurtek, S., Srivastava, A., Klassen, E., et al. (2012), et l'adaptation que nous envisageons nous permettra de considérer que deux séries temporelles de quaternions sont identiques lorsque que la seule différence entre les deux séries réside dans le choix du référentiel d'origine.

Bibliographie

- Dijkhuizen, M.S. (1994), The double covering of the quantum group $SO_q(3)$, *Proceedings of the Winter School "Geometry and Physics"*. *Circolo Matematico di Palermo*, 37, pp. 47-57.
- Drouin, P. et Stamm, A. et Chevreuil, L. et al (2022a), Gait impairment monitoring in multiple sclerosis using a wearable motion sensor, *Medical Case reports and Reviews*, 5, pp. 1-5.
- Drouin, P. et Stamm, A. et Chevreuil, L. et al. (2022b), Semi-supervised clustering of quaternion time series : Application to gait analysis in multiple sclerosis using motion sensor data, *Statistics in Medicine*, 42(4), pp. 433-456.
- Happ, C. et Greven, S. (2018), Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains, *Journal of the American Statistical Association*, 113:522, pp. 649-659.
- Jabłoński, B. (2012), Quaternion dynamic time warping, *IEEE Transactions on Signal Processing*, 60, pp. 1174-1183.
- Kieseier, B.C. et Pozzilli, C. (2012), Assessing walking disability in multiple sclerosis, *Multiple Sclerosis Journal* 18.7, pp. 914-924.
- Kurtek, S., Srivastava, A., Klassen, E., et al.(2012), Statistical Modeling of Curves Using Shapes and Related Features, *Journal of the American Statistical Association*, 107(499), pp. 1152-1165.
- Kurtzke, J.F. (1983), Rating neurologic impairment in multiple sclerosis : an expanded disability status scale (edss), *Neurology*, 33(11).
- Piórek, M. (2019), Analysis of Chaos for Quaternion Time Series, *Analysis of Chaotic Behavior in Non-linear Dynamical Systems : Models and Algorithms for Quaternions*, Springer International Publishing, pp. 73-88.
- Piórek, M. et Jabłoński, B. (2020), A Quaternion Clustering Framework, *International Journal of Applied Mathematics and Computer Science*, 30.1, pp. 133-147.
- Ramsay, J.O. et Silverman, B.W. (2005), Principal components analysis for functional data, *Functional Data Analysis* ,Springer, pp. 147-172.
- Rousseeuw, P.J. (1987), Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, 20, pp. 53-65.
- Sangalli, L.M. et Secchi, P. et Vantini, S. et al (2010), k-mean alignment for curve clustering, *Computational Statistics & Data Analysis*, 54, pp. 1219-1233.
- Tucker, J. D., Wu, W. et Srivastava, A. (2013), Generative models for functional data using phase and amplitude separation, *Computational Statistics and Data Analysis*, 61, pp. 50-66.
- Voight, J. (2005), Quaternion Algebras, *Springer Nature*