

Université du Québec
Institut national de la recherche scientifique
Centre Energie Matériaux Télécommunications

**Détermination d'un patron représentatif et analyse de la complexité des
données cinématiques tridimensionnelles du genou**

Par
Imen Mechmeche

Mémoire présenté pour l'obtention du grade de
Maîtrise en Sciences, M.Sc.
en Télécommunication

Jury d'évaluation

Examineur externe	David Labbé Ecole de Technologie Supérieure
Examineur interne	Jacob Benesty Institut National de la Recherche Scientifique
Directeur de recherche	Amar Mitiche Institut National de la Recherche Scientifique
Codirectrice de recherche	Neila Mezghani la TÉLUQ

Déposé le 24/11/2016

©Imen Mechmeche, 2016

Remerciements

Arrivant à l'étape de rédaction de ce rapport, je me sentais envahie de joie et de reconnaissance. Une reconnaissance envers tous ceux qui ont rendu ce travail possible, tous ceux qui m'ont assisté dans des conditions si agréables.

Je commence certainement par mon directeur de recherche Amar Mitiche, qui m'a donné la chance de travailler sous sa direction, et qui m'a beaucoup aidé à prendre les bonnes décisions durant les différentes étapes du projet.

Je remercie vivement ma co-directrice Neila Mezghani, pour son assistance, son implication et ses conseils qui m'ont été d'une aide précieuse tout au long de ma maîtrise.

Un grand merci à Youssef Ouakrim, qui m'a assisté et veillé à ce que mon projet se déroule dans les meilleures conditions possibles, je n'oublierai jamais son sens d'humour et son dévouement qui m'aidaient à surmonter les périodes de stress.

A toute l'équipe du Laboratoire de recherche en Imagerie et Orthopédie, vous êtes une famille agréable. Merci infiniment pour l'accueil chaleureux et l'ambiance conviviale. Je remercie particulièrement Gerald Parent et Abdelmonime Aissaoui pour leur soutien et leur disponibilité.

Je fini par un chaleureux merci à la source de bonheur qui m'alimentait durant ces deux années: ma chère famille et mes merveilleuses colocatrices Chaima, Bouthaina et Cherifa. Merci d'être là, merci pour vos encouragements et surtout pour votre confiance.

Du fond de mon cœur, merci à tous!

Résumé

L'analyse des données cinématiques du genou est une nouvelle approche qui a permis une meilleure compréhension du mode de fonctionnement de cette articulation en état de mouvement. Ce type d'analyse a ouvert des issues pour effectuer un diagnostic non invasif des différentes pathologies musculosquelettiques du genou. En effet, plusieurs études ont appréhendé la classification des données cinématiques pour le développement de système d'aide à la décision de pathologies du genou. Cependant, aucune étude n'a considéré l'étude de la complexité de ces données avant leur classification.

Dans cette étude, notre premier objectif est d'évaluer pour la première fois dans la littérature la complexité des données cinématiques indépendamment du classifieur. Cette mesure de complexité permettra d'avoir une idée sur leur caractéristiques intrinsèque et pourra orienter le choix du classifieur à utiliser.

Le deuxième objectif vient suivre le premier et porte sur la détermination d'un patron représentatif des données cinématiques afin d'améliorer leur précision. Ce patron est déduit à partir d'observations pour chaque sujet qui décrivent les variations des angles de rotation du genou durant plusieurs cycles de marche. Ces courbes observées souffrent d'une part de la présence de données aberrantes et d'autre part de la variabilité en phase entre elles, ce qui fausse généralement la moyenne obtenue et qui est considérée comme le patron représentatif du sujet. Ainsi, nous proposons dans ce travail d'utiliser l'analyse fonctionnelle de données (FDA) pour un traitement des observations afin d'obtenir un patron robuste et représentatif de la cinématique du genou.

Mots-clés Données cinématiques 3D, Complexité, patron représentatif, Analyse fonctionnelle

Abstract

Gait analysis is an emerging technique providing noninvasive tools for a subject's gait evaluation. In rehabilitation medicine, sport science, orthopedics, kinesiology, and related field, several studies have tried to investigate kinematic and/or kinetic data to obtain crucial information about musculoskeletal pathologies. In fact, several studies focused on the classification of knee kinematics data for the development of decision support systems of knee pathologies. However, none of them has considered the complexity of the data before their classification.

This study has two goals. The first one is to evaluate for the first time in the literature the complexity of these kinematic data before classification. These complexity measures will give as a better insight about the data characteristics and could be helpful while choosing the convenient classifier.

The second one is to determine a representative pattern of knee kinematics to improve their precision. In fact, the difficulty with 3D knee kinematics relates to the variability of measurements. Observations appear as temporal waveforms representing joint angle measures in the sagittal, frontal and transverse planes for each stride. These observations are contaminated with outliers and phase variability, so that simple averaging of the data can significantly influence subsequent analyses. Within this study, we propose to deal with knee kinematics curves as functional data and propose a new scheme for pattern identification.

Keywords 3D knee kinematics, Complexity, representative pattern, functional data analysis

Table des matières

Remerciements	iii
Résumé	v
Abstract	vii
Table des matières	ix
Liste des figures	xi
Liste des tableaux	xiii
Liste des abréviations	xv
1 Introduction	3
2 Cadre de l'étude: Analyse cinématique des cycles de marche	5
2.1 La biomécanique et l'analyse de la marche	5
2.1.1 Le cycle de marche	5
2.1.2 L'analyse de la marche humaine	8
2.2 La cinématique du genou	9
2.2.1 Articulation du genou	9
2.2.2 Mouvement de rotation du genou	10
2.2.3 Acquisition des signaux cinématiques tridimensionnels du genou	12
2.3 Classification des données biomécaniques du genou	14
2.3.1 Description des données cinématiques 3D du genou	15
2.3.2 Classification des données cinématiques 3D du genou	16
2.4 Problématique de la recherche et objectifs	17
3 Analyse de complexité des données cinématiques 3D du genou	19
3.1 Revue de la littérature: Mesures de complexité des données	19
3.1.1 Mesures de chevauchement entre les points d'intérêts	20
3.1.2 Mesures de la séparabilité des classes	23
3.2 Évaluation de la complexité des données cinématiques	26
3.2.1 Base de donnée	26
3.2.2 Méthode de calcul des paramètres de complexité	27
3.3 Résultats expérimentaux: Évaluation des paramètres de complexité	28
4 Détermination d'un patron représentatif des données cinématiques 3D du genou	31

4.1	Revue de la littérature: Détermination classique du patron représentatif des données cinématiques du genou	32
4.1.1	Détermination du patron représentatif en moyennant les 15 cycles les plus corrélés	32
4.1.2	Détermination du patron représentatif en moyennant les courbes contenues dans l'intervalle de confiance via le bootstrap	33
4.2	Traitement des données continues: Analyse fonctionnelle des données	35
4.2.1	Principe de l'analyse fonctionnelle des données	35
4.2.2	Mesure de la profondeur fonctionnelle pour la détermination des courbes aberrantes	37
4.2.3	Alignement temporel des courbes pour la minimisation de la variabilité	41
4.3	Résultats expérimentaux: Approche fonctionnelle pour la détermination du patron représentatif	44
4.3.1	Résultats et validation du patron obtenu	47
5	Conclusion	53
	Références	55

Liste des figures

2.1	Cycle de la marche. tirée du site www.cofemer.fr	6
2.2	Les sous phases de la phase d'appui [16]	7
2.3	Les sous phases de la phase d'envol [16]	8
2.4	Articulation du genou. Tirée du site www.mon-arthrose.com	9
2.5	Description des axes de rotation au genou selon la convention de l'ISB tel que décrit par Grood and Suntay 1983 [24]. Illustration tirée de Noyes 2010 [25]	10
2.6	Amplitude des mouvements de flexion du genou selon la position de la cuisse et du bassin [16]	11
2.7	Mouvement d'abduction (a) et d'adduction (b) du genou droit [28]	12
2.8	La caméra et l'ordinateur peuvent tous deux être montés sur un chariot rendant l'ensemble du système mobile [30]	13
2.9	Fixation de l'harnais sur le genou [30]	13
2.10	Essai sur le tapis roulant [30]	14
2.11	Visualisation en temps réel des mouvements du genou en 3D [30]	14
2.12	Exemple de données cinématiques 3D du genou pour des sujets AS et OA [33]	15
2.13	Exemple de base de données (cycles de flexion)	16
2.14	Exemple d'observations de la flexion du genou pour des sujet AS et OA avant traitement	18
3.1	Illustration de différents cas de distances entre les classes [35]	20
3.2	Illustration d'un problème linéairement séparable. Tirée du site www.asi.insa-rouen.fr	24
3.3	Exemple d'un arbre couvrant minimal: Les cercles noirs appartiennent à la 1 ^{ere} classe et les carrés blancs à la 2 ^{eme} classe. Les lignes relient les voisins les plus proches [35]	25
3.4	Exemple de représentation des points d'intérêts sur les signaux cinématiques [33]	27
3.5	Exemple d'exécution du code pour le calcul des mesures de complexité	27
4.1	Procédure proposée pour la détermination du patron représentatif des données cinématiques 3D	32
4.2	Intervalle de confiance 90% pour les courbes de flexion. Tirée et modifié de [45]	34
4.3	(a) Bande définie par les deux courbes x_1 , x_2 et une 3 ^{eme} courbe x appartenant à la bande; (b) bande déterminée par trois courbes x_1 , x_2 et x_3 [54].	40
4.4	Exemple de diagramme en boîte fonctionnel [51]	41
4.5	Influence de la variabilité en phase sur la moyenne obtenue [57]	42
4.6	Illustration de la fonction de déformation temporelle (γ). γ est construite de façon à ce que les point de $f_1(t)$ coïncident avec $f_1(\gamma(t))$ [57]	43
4.7	Transformation d'un cycle de flexion en objet fonctionnel	44
4.8	Transformation d'un cycle d'abduction en objet fonctionnel	44
4.9	Transformation d'un cycle de rotation en objet fonctionnel	44

4.10	Diagramme en boite pour les courbes de flexion	45
4.11	Diagramme en boite pour les courbes d'abduction	45
4.12	Diagramme en boite pour les courbes de rotation	46
4.13	Exemple d'alignement d'une courbe de flexion	46
4.14	Exemple d'alignement d'une courbe d'abduction	46
4.15	Exemple d'alignement d'une courbe de rotation	47
4.16	Résultat d'alignement de toutes les observations de flexion pour un sujet donné . . .	47
4.17	Résultat d'alignement de toutes les observations de d'abduction pour un sujet donné	47
4.18	Résultat d'alignement de toutes les observations de rotation pour un sujet donné . .	48
4.19	Comparaison entre les 3 patrons représentatifs de flexion	48
4.20	Comparaison entre les 3 patrons représentatifs de d'abduction	48
4.21	Comparaison entre les 3 patrons représentatifs de rotation	49
4.22	Distance euclidienne	49
4.23	Distance de Fréchet	49

Liste des tableaux

- 3.1 Mesures de complexité des données cinématiques 3D du genou ainsi que pour d'autres bases de données de référence 29
- 4.1 Pourcentage des cas où le patron proposé est plus représentatif que la moyenne des données brutes et la moyenne des 15 observations les plus corrélées. 51

Liste des abréviations

AAF-TS	Articulation Fémoro-Tibiale
AS	Asymptomatique
BD	Band Depth
CM	Cycle de marche
FDA	Functional Data Analysis
IQR	Inter-Quartile Range
ISB	International Society of Biomechanics
MBD	Modified Band Depth
MVA	Multivariate analysis
OA	Gonarthrosique
SSA	Singular Spectrum Analysis

Chapitre 1

Introduction

L'analyse de la marche humaine a longtemps suscité l'intérêt des chercheurs jusqu'à l'heure actuelle. Plusieurs travaux se sont concentrés sur la mesure objective et quantitative des différents paramètres caractéristiques de la marche, afin de les appliquer à divers domaines tels que le sport [1–3], l'identification des personnes à des fins de sécurité [4, 5], et la médecine [6–8].

Dans le domaine médical particulièrement, le développement des technologies d'acquisition a permis de pousser la recherche vers le diagnostic des pathologies musculosquelettiques en utilisant les signaux cinématiques tridimensionnels du genou [9]. Ces signaux décrivant la rotation du genou en fonction du temps, permettent d'analyser son état dans le cas dynamique contrairement à l'imagerie qui permet une analyse statique de cette articulation.

De ce fait, plusieurs travaux de classification ont immergé comme outil de diagnostic se basant sur ces données. L'objectif étant de distinguer les sujets asymptomatiques des sujets arthrosiques ou encore de classifier les sujets selon le grade de l'arthrose [10, 11]. Cependant, cette classification s'avère difficile vue la nature complexe des données qui sont fournies sous la forme de vecteurs de mesures de grande dimension rendant l'analyse vulnérable à la malédiction de la dimensionnalité. D'autant plus, ces données présentent une grande variabilité intraclasse et une petite variabilité

interclasse. Ainsi, une haute précision est requise dans la détermination du patron représentatif des sujets à classifier.

C'est à ce niveau qu'intervient la contribution de ce travail qui a pour premier objectif d'évaluer pour la première fois dans la littérature la complexité des données cinématiques tridimensionnelles du genou. Les mesures de la complexité seront basées sur l'évaluation des frontières des classes de point de vue séparabilité, chevauchement et caractéristiques géométriques [12]. Nous proposons d'évaluer deux groupes de mesures: (1) Mesure de chevauchement entre les points d'intérêt qui sont les facteurs biomécaniques dans le cas de la cinématique du genou et (2) Mesure de la séparabilité des classes qui seront les sujets asymptomatiques (AS) et gonarthrosiques (OA). Le niveau de complexité sera déduit à partir d'une comparaison avec des bases de données de références.

Le deuxième objectif de cette étude est d'effectuer un prétraitement des données cinématiques afin de maximiser la précision du patron représentatif obtenu. Pour ce faire, nous proposons d'utiliser une analyse fonctionnelle de données [13]. Il s'agit d'une nouvelle approche de traitement des données continues qui a démontré sa robustesse dans divers domaines tel que les variations climatiques [14], les études médicales traitants le coeur humain [15] ou encore l'évolution de la taille des enfants au cours du temps [13].

Ce mémoire est organisé comme suit: le chapitre 2 introduit l'analyse de la marche, la cinématique du genou et particulièrement la classification de ce type de données. Les chapitres 3 et 4 traiteront séparément les deux objectifs principaux de cette étude. Enfin une conclusion générale fera l'objet du dernier chapitre.

Chapitre 2

Cadre de l'étude: Analyse cinématique des cycles de marche

Ce chapitre introductif représente une mise en contexte de cette étude. Nous allons, alors, expliquer les différentes notions de base essentielles à la compréhension des chapitres suivants détaillant les traitements réalisés sur les données cinématiques 3D du genou. Qu'est ce qu'un cycle de marche? comment décrire le mouvement du genou? comment obtenir les signaux cinématiques 3D et les classifier? les réponses à ces questions préliminaires feront l'objet de la suite de ce chapitre.

2.1 La biomécanique et l'analyse de la marche

2.1.1 Le cycle de marche

La marche humaine normale est définie comme un phénomène complexe qui implique la coordination de mouvements de rotations des segments corporels pour maintenir l'équilibre du corps pendant son déplacement vers l'avant [16] . Selon Winter [17] « La marche humaine est l'une des tâches de déplacement les plus difficiles à apprendre, mais une fois apprise, elle devient presque

inconsciente ». Ainsi ce mouvement s’acquiert à travers un processus d’apprentissage et le patron caractéristique de la marche se développe en quelques années. Une seule séquence de ce mouvement répétitif est appelée cycle de marche (CM) qui débute par le contact initial du pied et se termine lors du nouveau contact du pied au sol.

Le cycle de marche est composé d’une phase d’appui et une phase d’envol. Ce découpage sert de référentiel de discussion à l’analyse de la marche humaine. Les résultats de ces analyses sont en général normalisés par rapport à la durée du cycle de marche et les deux phases sont exprimées en pourcentage de celui-ci. Généralement il est convenu que la phase d’appui présente 60 % du cycle et la phase d’envol (appelée aussi phase oscillante) les 40 % restants. L’appui est la phase durant laquelle le pied est sur le sol et l’envol est le temps où le pied est en l’air pour l’avancement de la jambe. La figure 2.1 décrit les différentes étapes d’un cycle de marche [16].

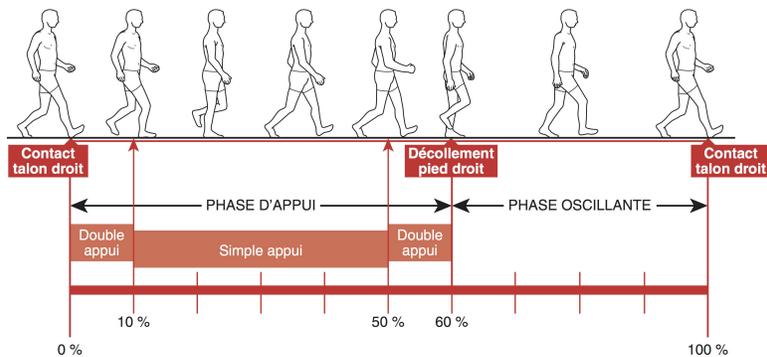


Figure 2.1 – Cycle de la marche.
tirée du site www.cofemer.fr

La phase d’appui est composée de 5 sous-phases selon l’évolution de l’état du genou et du pied en mouvement [18] :

1. Contact initial : 1 à 2 % du CM.

Le contact initial est caractérisé par l’amortissement du mouvement de la jambe avec le sol. Une fois le pied au sol, la stabilité de la jambe est assurée.

2. Phase de chargement : 1 à 10 % du CM.

Le genou continue sa progression vers l’avant et il se stabilise tout en absorbant le poids corporel. Le genou est une articulation qui assure, en même temps, la stabilité et la mobilité.

3. Milieu de la phase d'appui : 10 à 30 % du CM.

Durant cette phase, le corps pivote par-dessus le pied qui est au sol. L'articulation du genou assure principalement la stabilité des membres inférieurs.

4. Fin de la phase d'appui : 30 à 50 % du CM.

Durant cette phase, le genou connaît un mouvement de flexion complet, ce qui lui permet d'allonger la longueur du pas. Le genou assure la stabilité du corps pendant l'appui unipodal.

5. Phase de préenvol : 50 à 60 % du CM.

Le genou permet à la jambe de se mettre dans une position adéquate pour effectuer une poussée vers l'avant.

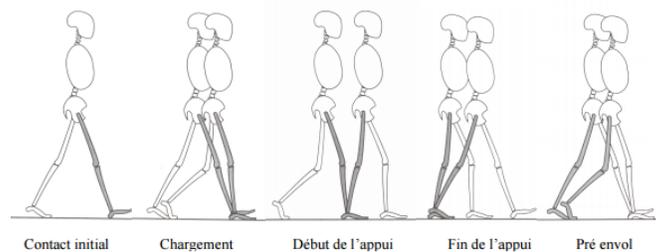


Figure 2.2 – Les sous phases de la phase d'appui [16]

La phase d'envol, quant à elle, est répartie en 3 sous-phases:

1. Début de l'oscillation 60-73 %.

Elle commence lorsque le pied est déplacé au dessus du sol. Le genou subit une flexion pour libérer le pied du sol tout en fournissant une progression de la jambe vers l'avant.

2. Oscillation intermédiaire 73-87%.

Cette période permet l'avancement du membre: le genou subit une extension passive soutenue par l'impulsion générée par la flexion de la hanche.

3. Oscillation terminale 87-100 %.

Lors de cette sous-phase finale, le membre est décéléré et préparé pour la phase suivante d'appui [18]. Le genou ralentit son extension pour préparer le contact au sol. Cette phase marque la fin du déplacement de la jambe vers l'avant.

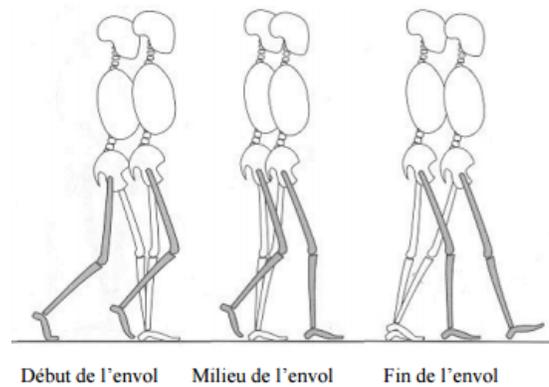


Figure 2.3 – Les sous phases de la phase d’envol [16]

2.1.2 L’analyse de la marche humaine

La recherche sur la marche humaine comprend l’évaluation qualitative et quantitative des différents éléments qui la caractérisent. Les facteurs d’intérêt varient selon le domaine de recherche. Par exemple, dans le cas de la sécurité, l’intérêt peut se concentrer sur l’identification des personnes à travers une caractérisation générale de leur silhouette et leur patron représentatif de marche [19]. Par contre, dans le domaine du sport, les études peuvent se concentrer sur l’analyse des différentes forces exercées sur chaque muscle en mesurant l’EMG (électromyogramme) [20].

Du point de vue clinique, l’importance de l’analyse de la marche humaine réside dans le fait que les troubles de la marche affectent un pourcentage élevé de la population du monde et constituent des problèmes clés dans les maladies neurodégénératives telles que les tumeurs cérébrales, les maladies neuromusculaires (myopathies), les pathologies vasculaires cérébrales, maladies cardiaques ou vieillissement physiologique. Ainsi, l’étude des caractéristiques de la marche humaine peut être utile pour des applications cliniques, et elle a été l’objet de nombreux travaux récents dont ceux de Mummolo et al. [21]. Plus spécifiquement, dans l’analyse de la marche, on trouve des études qui se sont concentrées sur l’analyse de la biomécanique du genou (signaux cinétiques et cinématiques). Ces études permettent une meilleur analyse des pathologies musculosquelettiques relatives au genou tel que l’arthrose [11, 22]. Dans cette étude, nous allons nous concentrer sur les signaux cinématiques qui ont été récemment utilisés dans plusieurs travaux prometteurs de classification.

Ainsi, notre analyse de complexité et le prétraitement proposés serviront d'un complémentaire pour l'amélioration des résultats obtenus avec ces études récentes.

2.2 La cinématique du genou

2.2.1 Articulation du genou

Le genou est l'articulation qui relie l'os de la cuisse, le fémur, au tibia et au péroné. À ce trio, s'associe un petit os, la rotule (Fig. 2.4). La partie terminale du fémur est arrondie alors que la partie haute du tibia est presque plate: il y a donc un vide à combler qui se fait grâce au cartilage qui protège les surfaces des frottements et le ménisque qui constitue le coussinet des articulations. Constitué de 90% d'eau, ce ménisque sert essentiellement à amortir les chocs [23].

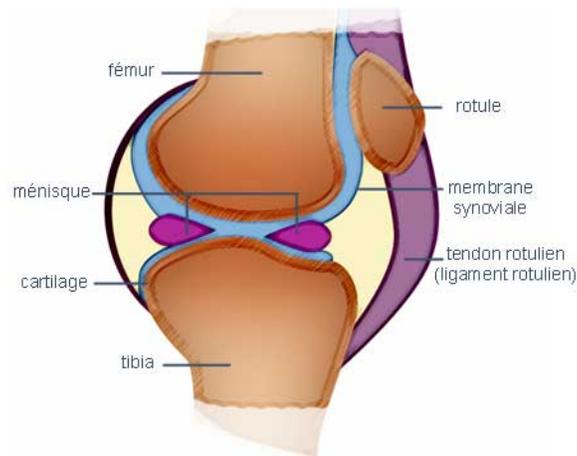


Figure 2.4 – Articulation du genou.
Tirée du site www.mon-arthrose.com

Le genou étant une articulation synoviale, il est entouré d'une capsule articulaire qui augmente légèrement la stabilité passive du genou. Cependant, le principal rôle de la capsule est d'enfermer le liquide synovial. Ce liquide visqueux permet essentiellement de lubrifier l'articulation et de fournir les nutriments nécessaires au cartilage articulaire (hyalin) du fémur et du tibia, et aide aussi légèrement à l'absorption des chocs.

2.2.2 Mouvement de rotation du genou

Le genou joue un rôle primordial durant la marche. En effet, il assure l'absorption du choc suite au contact initial durant la phase d'appui, et joue le rôle de stabilisateur en extension pour supporter le poids du corps. Durant la phase d'envol, ses mouvements rapides de flexion et extension permettent un passage optimisé de la jambe [16]. Pour parvenir à remplir ces rôles, le genou réalise une grande amplitude de mouvement dans le plan sagittal, en flexion/extension, et des mouvements plus fins dans les autres plans pour permettre l'ajustement de l'équilibre et de la stabilité [16]. La figure 2.5 illustre les axes de rotation du genou selon la convention l'ISB (*International Society of Biomechanics*), qui propose que la rotation de flexion/extension se fasse autour de l'axe fixe médiolatéral du fémur, l'abduction/adduction autour de l'axe flottant et la rotation interne/externe autour de l'axe fixe longitudinal du tibia.

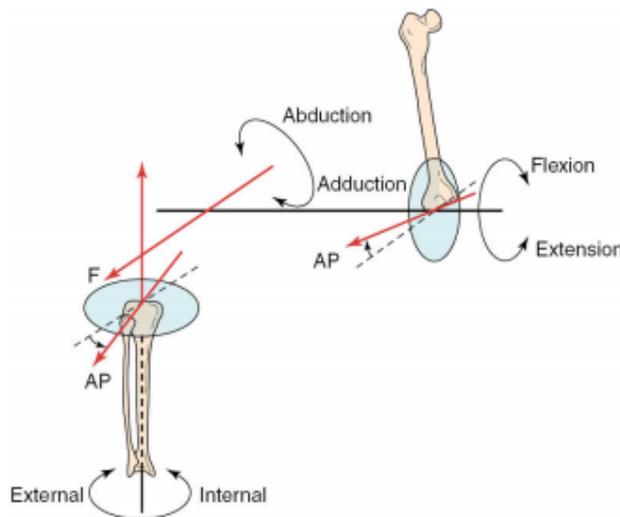


Figure 2.5 – Description des axes de rotation au genou selon la convention de l'ISB tel que décrit par Grood and Suntay 1983 [24]. Illustration tirée de Noyes 2010 [25]

La flexion du genou est le mouvement qui rapproche la face postérieure de la jambe de la face postérieure de la cuisse. Son amplitude va de 0° quand l'axe de la jambe est dans le prolongement de l'axe de la cuisse, jusqu'à 120° si la hanche est en extension. Mais elle peut atteindre 140° si la hanche est fléchie préalablement. Et même 160° en s'asseyant sur les talons, le talon touchant alors la fesse comme illustré dans la figure 2.6.

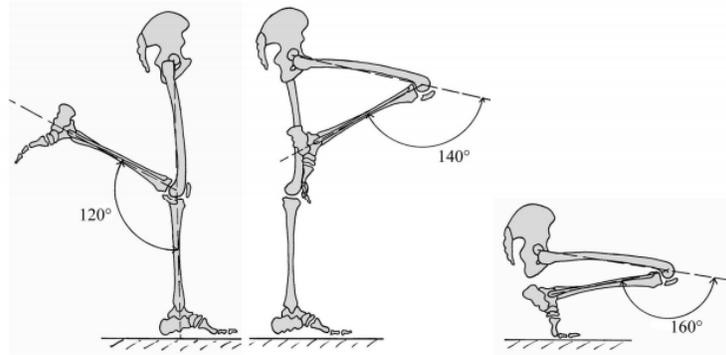


Figure 2.6 – Amplitude des mouvements de flexion du genou selon la position de la cuisse et du bassin [16]

Les mouvements de rotation dans le plan frontal et transverse de l'articulation fémoro-tibiale sont bien documentés et illustrés par Lafortune et al. [26]. Peu d'auteurs ont investigué le mouvement dans les plans nommés précédemment; la plupart des recherches se sont concentrées sur le phénomène qui se produit sur le plan sagittal. Lafortune et al. notent que les résultats de ces études démontrent de nombreuses dissimilitudes sur le mouvement d'abduction/adduction et de rotation interne/externe dans le plan frontal et transverse respectivement. Ces différences sont attribuées aux différents systèmes d'acquisition de mouvements utilisés. Lafortune et al. ont employé, pour l'acquisition des variables cinématiques dans le plan frontal et transverse pendant la marche, des tiges intra-corticales qui ont été implantées sur le tibia et le fémur des participants. Dans le plan frontal, les résultats de l'étude ont démontré que pendant la phase d'appui, l'articulation fémoro-tibiale (AF-T) exécute des mouvements de petites amplitudes d'abduction et d'adduction, en moyenne 13 de 5° . Lors du contact initial au sol du talon, l'AF-T fait un mouvement d'abduction d'environ $1,2^\circ$. Au chargement et l'appui unipodal, l'AF-T continue en abduction de presque 3° [27].

Durant la phase d'envol, l'AF-T exécute une amplitude maximale en abduction entre 6° et 8° . En général, le tibia reste en adduction par rapport au fémur tout au long du cycle de marche. Kettelkamp et al. [29] ont, de leur part, mesuré les mouvements dans le plan frontal durant la marche et ont noté une abduction maximale durant l'extension du genou lors du contact du talon au sol et au début de la phase d'appui. Par ailleurs, l'adduction maximale a été notée durant la

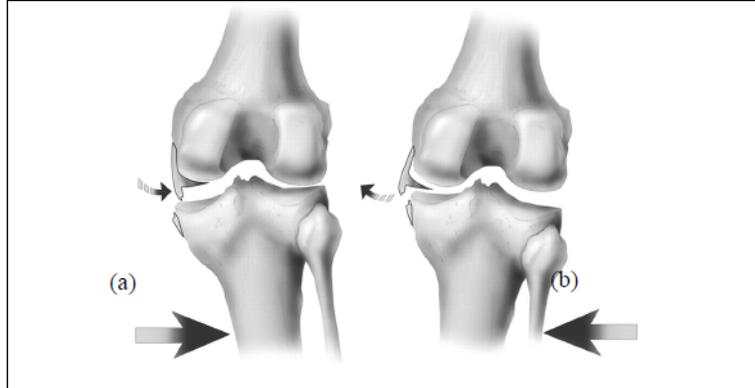


Figure 2.7 – Mouvement d’abduction (a) et d’adduction (b) du genou droit [28]

flexion du genou durant la phase d’envol. Des mouvements maximaux de 11° ont été enregistrés. Dans le plan transverse, Lafortune et al. [26] ont trouvé que, au contact initial, le fémur fait une légère rotation externe par rapport au tibia. Durant la phase de chargement, le tibia commence à accélérer en rotation interne et le fémur suit, mais à un taux légèrement plus lent. Pendant l’appui intermédiaire, lorsque l’AF-T est en extension complète dans le plan sagittal, la rotation est pratiquement nulle (0°). Lors de la poussée des orteils, le tibia effectue une rotation interne de 5° par rapport au fémur. À la fin du cycle de marche, au moment de la phase d’oscillation, le tibia commence à réaliser une rotation externe de $9,4^\circ$. Ensuite le tibia recommence à revenir en position de rotation interne pour commencer le nouveau cycle de marche.

Les signaux cinématiques sont très dépendants de la méthode d’acquisition vue la précision requise et la sensibilité des mouvements. Ces dernières années, ce problème est pratiquement résolu avec l’outil d’acquisition KneeKG que nous présenterons dans le paragraphe suivant.

2.2.3 Acquisition des signaux cinématiques tridimensionnels du genou

Le KneeKG est composé d’un harnais fixé sur le genou (KneeKG 3D-Tracker), un système de capture de mouvement infrarouge (caméra Polaris Spectra, Northern Digital Inc.), et un ordinateur équipé par le logiciel Knee3DTM (Emovi, Inc.) [30]. La caméra et l’ordinateur peuvent tous deux

être montés sur un chariot rendant l'ensemble du système mobile (Fig. 4.22). Le système mesure et analyse la position 3D et le mouvement du genou du patient.



Figure 2.8 – La caméra et l'ordinateur peuvent tous deux être montés sur un chariot rendant l'ensemble du système mobile [30]



Figure 2.9 – Fixation de l'harnais sur le genou [30]

Au cours de l'examen KneeKG, le patient doit porter un short pour que l'harnais soit en contact direct avec la peau. Après la fixation de l'harnais, une procédure d'étalonnage comme décrit par Hagemester et al. [31] est effectuée pour identifier les centres communs (hanche, genou, cheville) et pour définir un système de coordonnées sur chaque segment du corps (par exemple, fémur et tibia). Le protocole d'acquisition prend entre 15 et 20 min lorsqu'il est effectué par un technicien qualifié. Tous les mouvements sont pris à une fréquence de 60 Hz par la caméra infrarouge. Vu que les patients peuvent ne pas être habitués à une marche sur le tapis roulant, une période d'essai entre 6 et 10 min devrait être lancée avant la collecte de données pour assurer que les données cinématiques 3D soient reproductibles et minimiser les risques de données aberrantes.

Les essais sont ensuite enregistrés durant environ 45 s de marche confortable (Fig. 2.10). Lors de la collecte de données, les modèles de reconstruction osseuse 3D (tibia et fémur) sont affichés sur l'écran, ce qui permet une visualisation en temps réel des mouvements du genou 3D (Fig. 2.11). Des conditions de marche différentes peuvent être enregistrées séquentiellement et comparées,

comme la marche lente et rapide, rythmée, avec ou sans chaussures ou avec et sans prothèses. Une fois que la collecte de données est terminée, un rapport mettant en évidence les caractéristiques biomécaniques dans les trois plans de mouvement et pendant les sous-phases du cycle de marche est généré automatiquement.



Figure 2.10 – Essai sur le tapis roulant [30]

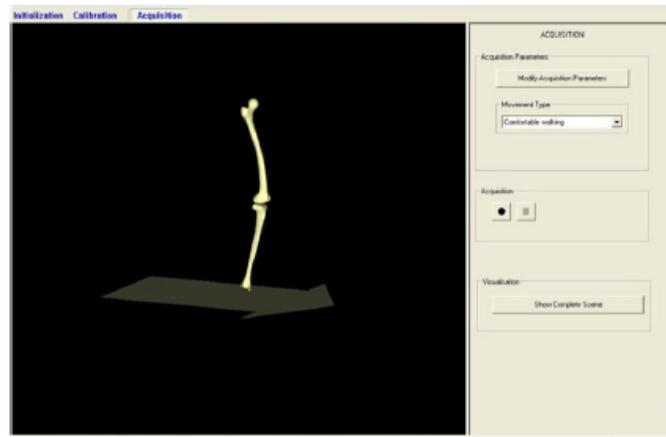


Figure 2.11 – Visualisation en temps réel des mouvements du genou en 3D [30]

2.3 Classification des données biomécaniques du genou

La collecte des données biomécaniques avec précision a ouvert de nouveaux issues pour leur utilisation dans le milieu clinique. En effet, une des façons d'utilisation de ces données pour le diagnostic des pathologies musculosquelettiques, est leur classification. Une problématique récente qui a fait l'objet de plusieurs travaux permettant de classer les sujets sains des sujets arthrosiques et distinguer les sujets selon le grade de l'arthrose. Avant d'expliquer davantage la classification qui nous a incité à développer cette étude d'analyses et traitements préliminaires, nous allons décrire d'une façon plus explicite les données à classer issues de l'acquisition avec l'outil KneeKG et qui feront la base de ce travail.

2.3.1 Description des données cinématiques 3D du genou

Comme nous l'avons expliqué, la rotation du genou est effectuée sur les trois plans sagittal, frontal et transverse. Ainsi, à l'issue de l'acquisition nous disposons de 3 courbes continues décrivant la variation angulaire du genou sur ces trois plans en fonction du temps. Pour une meilleure exploitation de ses courbes, une première étape de filtrage s'avère nécessaire. En effet, ces données sont essentiellement affectées par un bruit causé par les artefacts d'acquisition de données. Pour filtrer ce bruit, la méthode automatique Singular Spectrum Analysis (SSA) est utilisée avec une fenêtre de longueur 10 [32]. Cette méthode est une approche non paramétrique qui est appliquée pour l'analyse des séries chronologiques.

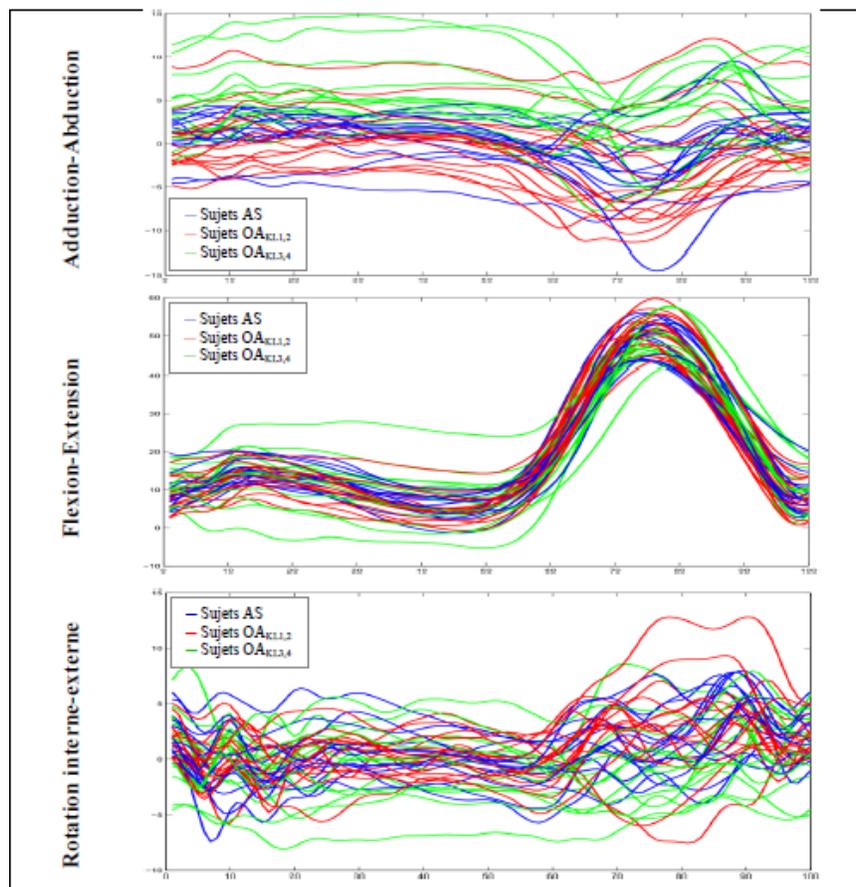


Figure 2.12 – Exemple de données cinématiques 3D du genou pour des sujets AS et OA [33]

Par la suite, chacune des 3 courbes obtenues est découpée en cycles normalisés (100 points) de façon à ce que chaque cycle de flexion, d'abduction et de rotation corresponde à un cycle de marche.

Selon la vitesse du sujet et la largeur du pas du sujet, nous obtenons généralement entre 35 et 45 courbe pour chaque plan. Finalement, pour caractériser ce sujet, un patron représentatif est déduit en moyennant les 15 courbes les plus répétitives.

La figure 2.12, présente un exemple de base de données contenant les patrons représentatifs des signaux cinématiques 3D de sujets asymptotiques (AS) et gonarthrosiques de différents grades (OA).

2.3.2 Classification des données cinématiques 3D du genou

La classification des données cinématiques 3D du genou est profondément étudiée par Mezghani et al. [10], [11] particulièrement celle des sujets asymptotiques et gonarthrosiques et la classification des différents grades de la gonarthrose. La figure suivante présente un exemple de base de données à classifier.

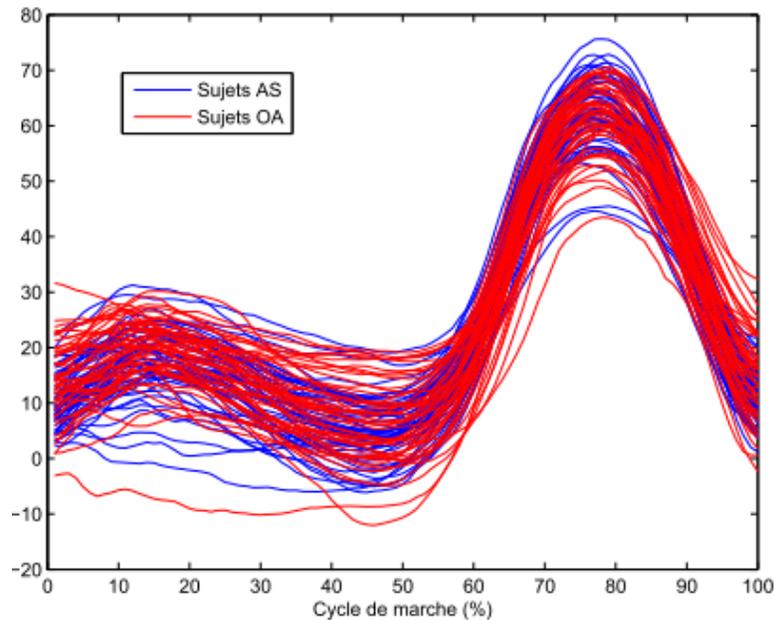


Figure 2.13 – Exemple de base de données (cycles de flexion)

On remarque la faible variabilité entre les patrons des sujets des deux classes. Cette faible variabilité interclasse nous pousse à quantifier la complexité de ce problème de classification: Une

problématique qui sera traitée dans le chapitre suivant. Cette faible variabilité nous pousse aussi à réviser la procédure avec laquelle le patron représentatif du sujet est déterminé, en effet, plus nous disposons de patrons fiables, plus les taux de classification seront meilleurs. Le 4^{ème} chapitre de cette étude présente une nouvelle approche pour la détermination d'un patron plus précis et fiable.

Notons aussi que divers facteurs interviennent dans les résultats de classification, à savoir le choix des points d'intérêt, leurs nombre, leurs positions, ainsi que le choix du classifieur lui même et ses différents paramètres. Ainsi, nous considérons ce travail comme une étude préliminaire pour les études et les tests supplémentaires dans le but d'améliorer les taux de classification de ce type de données.

2.4 Problématique de la recherche et objectifs

Après cette vue d'ensemble que nous avons présenté tout au long de ce chapitre, nous sommes en mesure de bien spécifier la problématique de la recherche abordée dans cette étude.

Comme nous avons pu le remarquer, les données cinématiques présentent une grande variabilité intraclasses et une petite variabilité interclasse, ce qui rend la classification difficile et la distinction des classes moins évidente. Généralement, pour mesurer ce degré de complexité, les chercheurs ont recours à quantifier l'erreur de classification. Cependant, la classification dépend non seulement de la nature de la base de données mais aussi du choix des points d'intérêt, et du classifieur. Ainsi, la question qui se pose est la suivante: Et si on pouvait avoir une idée sur le degré de complexité de nos données avant de les classifier? Cette question est récemment abordée dans la littérature, et nous trouvons un ensemble de mesures dont l'évaluation nous permet de quantifier la complexité des données [12,34–36]. Notre objectif est de quantifier cette complexité, ainsi, dans le prochain chapitre nous allons expliquer ces mesures et leurs interprétations, et nous allons les appliquer à un exemple de base de données de sujets AS et OA en les comparant à des bases de données de référence pour pouvoir analyser le niveau de complexité.

Suite à cette analyse, nous avons poussé encore plus l'étude pour atteindre notre deuxième objectif qui consiste à améliorer la précision du patron représentatif. La figure 2.14 illustre un ensemble de courbes d'un même participant durant plusieurs cycles de marche. L'objectif de cette étape est de les représenter par une seule courbe qui sera notre patron représentatif. On remarque la présence des courbes aberrantes et une variabilité en phase entre les observations qui influence, éventuellement la moyenne obtenue pour représenter ces données.

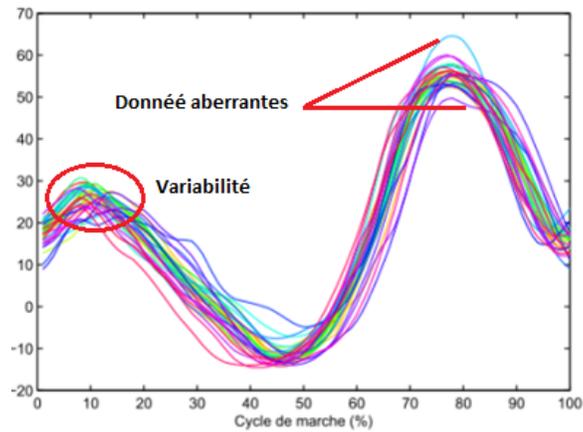


Figure 2.14 – Exemple d'observations de la flexion du genou pour des sujet AS et OA avant traitement

Ainsi le traitement que nous proposons dans ce travail vise à corriger ces deux problèmes dans nos observations en utilisant l'analyse fonctionnelle des données.

Chapitre 3

Analyse de complexité des données cinématiques 3D du genou

Comme nous l'avons décrit dans le chapitre précédent, les données cinématiques 3D du genou présentent une grande variabilité intraclasse et une petite variabilité interclasse ce qui nous pousse à croire en la difficulté de ce problème de classification. Le but de ce chapitre, c'est d'évaluer cette complexité à travers plusieurs groupes de mesures, décrivant les points d'intérêt, les frontières des classes et la topologie de notre base de données.

3.1 Revue de la littérature: Mesures de complexité des données

La complexité d'un problème est une mesure du niveau de difficulté de classification des données. Cette difficulté est essentiellement liée aux différentes caractéristiques choisies pour la description de l'ensemble de données. Les valeurs de complexité des données sont élaborées pour mesurer les différents aspects de la complexité des problèmes de classification [34]. Certaines mesures de complexité communes sont décrites dans la section 3.1.1.

3.1.1 Mesures de chevauchement entre les points d'intérêts

Le ratio maximal de Fisher: F_1

La première mesure de complexité se traduit par la distance qui sépare les centres des deux classes. Ce concept peut être représenté numériquement et calculé de différentes manières. La distance euclidienne univariée entre les centres des classes a été la première mesure utilisée par Ho et Basu [12]. Dans le cas de deux classes, la distance normalisée à partir des centres des deux groupes est évaluée en utilisant l'équation 3.1. Dans cette équation un rapport pour chaque dimension est calculé: le ratio final F , est la valeur maximale de ces rapports [35]:

$$F_1 = \max_i \left\{ \frac{(\mu_{i1} - \mu_{i2})^2}{(\sigma_{i1}^2 + \sigma_{i2}^2)} \right\} \quad (3.1)$$

où μ_{i1} , μ_{i2} , σ_{i1} , σ_{i2} sont les moyennes et variances des deux classes $i = 1, 2, \dots, p$, p étant le nombre de dimensions (points d'intérêts).

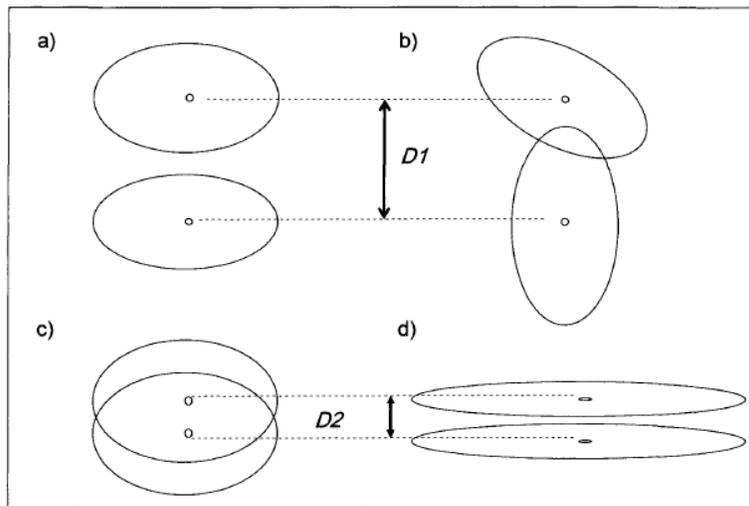


Figure 3.1 – Illustration de différents cas de distances entre les classes [35]

La mesure de Fisher est interprétée comme suit: une valeur nulle indique qu'il n'y a pas de distance entre les centres des classes, ce qui suggère que les deux moyennes sont égales, et les deux

classes se chevauchent complètement; une valeur plus élevée de cette distance indique que les classes sont plus séparées.

La figure 3.1 illustre différents cas de distances entre deux classes. Le cas (a) montre la séparabilité des deux classes suite à une large distance D_1 entre les moyennes, le cas contraire est décrit par l'illustration (c) où la distance D_2 n'est pas suffisamment large pour séparer les deux classes. Les cas (b) et (d) sont des contres exemples prouvant ainsi que cette mesure seule ne permet pas de trancher quant à la séparabilité des deux classes sans tenir compte de leur topologies et la nature des frontières. Ce qui nous amène à déduire que d'autres mesures de complexité doivent la compléter pour avoir une meilleure idée sur la séparabilité des classes.

La mesure de Fisher calculée précédemment est univariée; en effet, il était conclu dans [34] que la distance maximale entre les classes représentera le meilleur des cas vue que les points d'intérêts peuvent ne pas être tous importants dans la discrimination. Cependant, ce n'est pas toujours vrai car cela dépend de la nature des frontières et de la forme des classes, comme expliqué précédemment. Ainsi la mesure multivariée de Fisher sera plus pratique à appliquer. La mesure univariée Fisher ne prend pas en compte les corrélations entre les points d'intérêt; par contre, la version multivariée fournira une mesure de la corrélation linéaire entre les variables [37] en utilisant la covariance. Cette mesure est calculé à l'aide de l'équation suivante:

$$F_1 = (\bar{X}_1 - \bar{X}_2)^T \left(\frac{1}{n_1 - 1} \Sigma_1 + \frac{1}{n_2 - 1} \Sigma_2 \right)^{-1} (\bar{X}_1 - \bar{X}_2), \quad (3.2)$$

où n_1 et n_2 sont les tailles des instances de chaque classe, \bar{X}_1 et \bar{X}_2 sont les vecteurs estimés des moyennes de chaque classe, et Σ_1 , Σ_2 sont les matrices de covariance et $(.)^T$ est l'opérateur transposé. La valeur de F_1 dans ce cas est interprétée comme dans le cas univarié, plus elle est élevée plus les classes sont susceptibles d'être séparées, plus la valeur est faible plus le risque de chevauchement augmente.

Le chevauchement interclasses: F_2

Le paramètre F_2 mesure le chevauchement entre les distributions des deux classes. Il est défini par le ratio de l'intervalle de chevauchement contenant des attributs appartenant aux deux classes par l'intervalle entier des attributs. Ce paramètre est calculé par l'équation 3.3.

$$F_2 = \prod_i \frac{\min(\max(f_i, c_1), \max(f_i, c_2)) - \max(\min(f_i, c_1), \min(f_i, c_2))}{\max(\max(f_i, c_1), \max(f_i, c_2)) - \min(\min(f_i, c_1), \min(f_i, c_2))} / p \quad (3.3)$$

où les f_i sont les points d'intérêt (*features*), $i = 1, \dots, p$, C_j est la classe numéro $j = 1, 2$, $\max(f_i, C_j)$, $\min(f_i, C_j)$ est la valeur maximale/minimale de la variable f_i dans la classe C_j .

Contrairement à F_1 une faible valeur de ce paramètre qui est aussi appelé volume de chevauchement (*Volume of Overlap Region*) [34] permet de conclure que les attributs permettent une discrimination des données de chaque classe.

Dans la classification de données, nous sommes toujours en présence d'un grand nombre de variables ou encore de points d'intérêt quand le problème est de grande dimension comme dans notre cas avec les données cinématiques du genou. Cette situation conduit à des valeurs très faibles de cette mesure ce qui peut nous amener vers de mauvaises interprétations. Pour obtenir un résultat plus utile, une mesure modifiée de chevauchement a été introduite : le (*Overlap Minimum Range*) R .

$$R = \text{Min}_i \left[\frac{\text{MIN}(\max(f_i, c_1), \max(f_i, c_2)) - \text{MAX}(\min(f_i, c_1), \min(f_i, c_2))}{\text{MAX}(\max(f_i, c_1), \max(f_i, c_2)) - \text{MIN}(\min(f_i, c_1), \min(f_i, c_2))} \right] \quad (3.4)$$

Cette mesure représente le meilleur scénario de chevauchement entre les deux classes. Si une dimension existe sans aucun chevauchement, $R = 0$, alors nous sommes en présence d'un problème de classification peu complexe. Si le meilleur scénario révèle une valeur élevée de chevauchement,

alors nous avons un problème de complexité élevée. L'inconvénient de cette mesure est qu'elle est sensible aux données aberrantes qui pourront fausser sa valeur.

Efficacité de discrimination F_3

Le 3^{ème} paramètre F_3 , calcule l'efficacité de discrimination de chaque attribut et retourne la valeur maximale. Ainsi, pour chaque attribut on calcule le ratio du nombre d'instances non contenues dans la région de chevauchement par le nombre total des instances. La valeur maximale correspond par la suite au paramètre F_3 .

3.1.2 Mesures de la séparabilité des classes

La séparabilité linéaire(L_1, L_2)

Ces mesures illustrent à quel point deux classes sont séparables en examinant l'existence et la forme des frontières des classes. Plusieurs algorithmes existent dans la littérature pour déterminer la séparabilité linéaire, la plupart sont capable d'aboutir à des conclusions positives et réitèrent indéfiniment dans le cas échéant. Une formulation récente dans la programmation linéaire [36] a permis de traiter les deux cas séparables et non séparables et qui consiste à minimiser la fonction erreur suivante:

$$\begin{aligned} \text{Minimiser} \quad & a^T t \\ \text{Sous contrainte} \quad & Z^t w + t \geq b \\ & t \geq 0 \end{aligned}$$

où a et b sont des vecteurs arbitraires constants (généralement choisis à 1), w est le vecteur poids, t est le vecteur erreur, et Z est la matrice où chaque colonne z est un vecteur de poids d'intérêts x augmenté d'une dimension de valeur constante 1 selon la classe (c_1 ou c_2) de la façon suivante [34] :

$$\begin{cases} z = +x & \text{si } c = c_1 \\ z = -x & \text{si } c = c_2 \end{cases}$$

La valeur de cette fonction objective dans cette formulation correspond à L_1 . Cette valeur est nulle si le problème est linéairement séparable. Cette mesure peut, aussi, être facilement influencée par les données aberrantes qui peuvent ne pas se situer dans le bon côté de l'hyperplan optimal (Fig. 3.2).

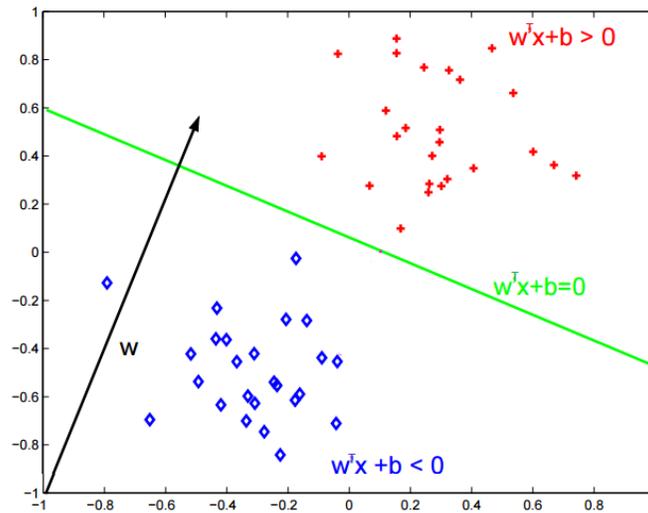


Figure 3.2 – Illustration d'un problème linéairement séparable.
Tirée du site www.asi.insa-rouen.fr

En se basant sur la description de la mesure L_1 , la mesure L_2 est définie comme étant le taux d'erreur en appliquant ce classifieur linéaire sur les données d'apprentissage. L_2 est comprise entre 0 et 1, plus sa valeur est proche de 1 plus on conclut que le problème n'est pas linéairement séparable.

Fraction des points sur la frontière de la classe N_1

Pour définir cette mesure, nous allons commencer par définir certains termes clés. Un arbre est un graphe connexe sans cycles. Un arbre couvrant (*spanning tree*) de ce graphe est un graphe qui est aussi un arbre et qui relie tous les sommets ensemble; un même graphe peut avoir de nombreux arbres couvrants. Un type particulier d'arbre est appelé un graphe de densité de bord, dans lequel

des poids sont affectés aux arêtes du graphe [38]. Un arbre de découpage minimum d'une courbe de bord pondéré est un arbre de recouvrement pour lequel la somme des poids des arêtes est un minimum [38]. L'arbre de recouvrement minimal (Minimum Spanning Tree, MST) a été créé en utilisant l'algorithme proposé par Ross (1969) [39]; d'autres algorithmes ont été développés pour améliorer le temps de calcul et profiter d'une meilleure utilisation de l'allocation de mémoire de l'ordinateur [38,40,41]. La mesure N_1 est le nombre de points de l'arbre couvrant minimal qui sont liés à la classe opposée. Dans l'exemple présenté dans la figure 3.3, cette valeur est 3. S'il n'y a pas de point où le plus proche voisin est une observation de la classe opposée, alors la valeur N_1 sera presque nulle. On peut en conclure que les classes sont complètement séparables.

$$N_1 = \frac{\text{Nombre de points connectés à la classe opposée}}{\text{Nombre total de points.}} \quad (3.5)$$

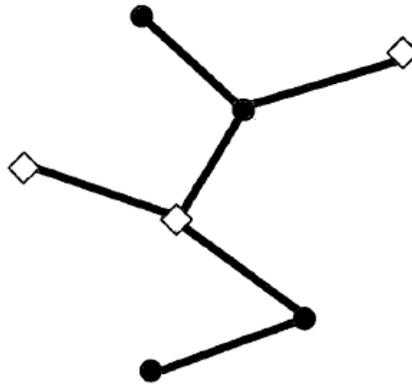


Figure 3.3 – Exemple d'un arbre couvrant minimal: Les cercles noirs appartiennent à la 1^{ère} classe et les carrés blancs à la 2^{ème} classe. Les lignes relient les voisins les plus proches [35]

Une valeur élevée de N_1 indique que la majorité des instances sont situées à la frontière de la classe ce qui rend la séparation plus difficile.

Ratio des distances moyennes des plus proches voisins intra/inter classes N_2

Le paramètre N_2 compare l'étalement des instances de la même classe par rapport à la distance des plus proches voisins appartenant à l'autre classe. Le domaine de N_2 est large, car il appartient à

l'intervalle $[0, +\infty[$. Une faible valeur de N_2 permet de conclure que les instances de la même classe sont assez proches dans l'espace des attributs. Dans le cas opposé nous pouvons conclure que les éléments de la même classe sont assez dispersés.

$$N_2 = \frac{\sum_{i=0}^n \text{intraDist}(x_i)}{\sum_{i=0}^n \text{interDist}(x_i)} \quad (3.6)$$

où x_i est une instance, n est la taille de la base de données, $\text{intraDist}(x_i)$ est la distance qui sépare x_i de ses voisins les plus proches dans la même classes et $\text{interDist}(x_i)$ est la distance qui sépare x_i de ses voisins les plus proches dans les autres classes.

3.2 Évaluation de la complexité des données cinématiques

Après avoir présenté les différentes mesures de complexité utilisées dans la littérature, nous allons voir de plus près leur utilisation dans notre cas. En effet, nous allons nous intéresser aux paramètres F_1 , F_2 , et F_3 pour les mesures de chevauchement, ainsi que N_1 et N_2 pour les mesures de séparabilité.

3.2.1 Base de donnée

La base de données que nous allons utiliser pour notre analyse est composée de 100 sujets AS et 100 OA. Les données pour chaque sujet correspondent à trois patrons représentatifs de la cinématique du genou (flexion/extension; adduction/abduction; rotation interne/externe). Pour la classification et vu la grande dimension du problème, certains points caractéristiques sont extraits de chaque courbe pour construire les points d'intérêts ou encore les facteurs biomécaniques (Fig 3.4). Ainsi notre base de données est composée de 200 sujets ($n = 200$) avec 2 classes (C_1 et C_2) et 80 attributs ($p = 80$).

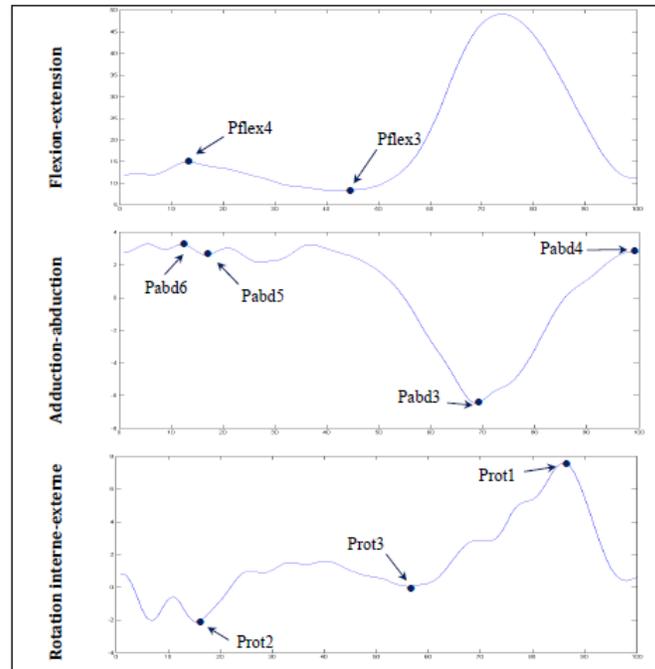


Figure 3.4 – Exemple de représentation des points d'intérêts sur les signaux cinématiques [33]

3.2.2 Méthode de calcul des paramètres de complexité

Pour l'évaluation des mesures de complexité de nos données cinématiques nous avons développé un code avec le langage C++ basé sur une librairie en ligne (DCoL) développée par Albert Orriols-Puig, Nùria Macià, et Tin Kam Ho (2010) [42]. La figure 3.5 présente un aperçu d'exécution.

```

C:\Windows\system32\cmd.exe
C:\Users\User\Dropbox\ConsoleApplication1\Release>ConsoleApplication1.exe -i cinematique.dat -o res -f
DCoL Copyright (C) 2009 Albert Orriols-Puig and Nuria Macia
The DCoL comes with ABSOLUTELY NO WARRANTY; for details type './dcol show w'.
This is free software, and you are welcome to redistribute it
under certain conditions; type './dcol show c' for details.

> PROCESSING COMPLEXITY MEASURES ON DATA SET: cinematique.dat
> Processing the maximum Fisher's discriminant ratio <F1>
> Fisher discr. ratio: 0.398379
> Processing the directional-vector maximum Fisher's discriminant ratio <F1v>
> Directional-vector maximum Fisher discr. ratio: 8.43353
> Processing the overlap of per-class bounding boxes <F2>
> Volume overlap: 8.14935e-014
> Processing the efficiency of attributes <F3/F4>
> Discriminant power of attribute 0: 0
> Discriminant power of attribute 1: 0
> Discriminant power of attribute 2: 0.04
> Discriminant power of attribute 3: 0
> Discriminant power of attribute 4: 0
> Discriminant power of attribute 5: 0
> Discriminant power of attribute 6: 0
> Discriminant power of attribute 7: 0
> Discriminant power of attribute 8: 0

```

Figure 3.5 – Exemple d'exécution du code pour le calcul des mesures de complexité

Obtenir les valeurs n'est pas encore suffisant pour conclure quand à la complexité de classification de nos données. En effet, nous avons eu recours à des bases de données de référence pour pouvoir comparé les résultats et en tirer des conclusions. Ces bases de données sont les suivantes:

1. Iris: C'est la base de données la plus connue dans la littérature pour les problèmes de classification. L'ensemble de données contient 3 classes de 50 cas chacun, chaque classe présente un type de plante d'iris. Une classe est linéairement séparable des 2 autres ; ces dernières sont pas linéairement séparables l'une de l'autre.
2. Wine: Ces données sont les résultats d'une analyse chimique des vins cultivés dans la même région en Italie, mais issus de trois cultivés différentes. L'ensemble contient 3 classes qui sont 3 types de vin définis par 13 attributs. La taille de la base de données est 178.
3. Abalone: C'est une base de données permettant la prédiction de l'âge de l'ormeau à partir de mesures physiques. L'ensemble contient 4177 instances, représentées par 8 attributs et 29 classes.

3.3 Résultats expérimentaux: Évaluation des paramètres de complexité

Le tableau 3.1 résume les résultats obtenus pour notre base de données ainsi que les 3 autres bases de données que nous avons utilisé comme référence pour évaluer le niveau de complexité de nos données:

Les mesures de complexité obtenues permettent de déduire l'aspect assez complexe du problème de classification des données cinématiques du genou. Nous remarquons que le paramètre F_1 est assez faible ce qui reflète un chevauchement important dans les instances des classes dans l'espace des attributs. Une forte valeur de N_2 prouve aussi que les instances de la même classe sont assez dispersées. Les 3 bases de données Iris, Abalone et Wine, nous ont servi comme références pour avoir

	F_1	F_2	F_3	N_1	N_2
Données cinématiques 3D du genou	0,38	$8,57 \cdot 10^{-14}$	0,065	0,5	0,94
Iris	16,82	0,0048	0,57	0,013	0,095
Abalone	14,03	0	0,99	$4,7 \cdot 10^{-4}$	0,025
Wine	4,29	$3,9 \cdot 10^{-5}$	0,76	0,06	0,49

Tableau 3.1 – Mesures de complexité des données cinématiques 3D du genou ainsi que pour d'autres bases de données de référence

une idée sur l'ordre de grandeur de ces paramètres. La base Abalone qui donne de très bons résultats de classification suit parfaitement le comportement que doivent avoir ces paramètres comme décrit précédemment. Pour conclure, nous pouvons déduire de cette étude que les données cinématique du genou sont assez complexes à classifier, ce qui est naturellement prévu vu la variabilité et la difficulté de choix des attributs. Ces mesures pourront servir à évaluer la complexité suite à chaque essai de traitement réduisant la variabilité ou encore la dimension du problème.

Chapitre 4

Détermination d'un patron représentatif des données cinématiques 3D du genou

L'évaluation de la complexité des données cinématiques a démontré que leur classification est un problème de complexité élevée. Cette évaluation se base sur les courbes représentatives de la cinématiques 3D du genou. Étant donné l'importance de ces courbes, nous abordons dans ce chapitre la détermination d'un patron représentatif des données cinématiques 3D du genou. Nous proposons ainsi une procédure basée sur l'analyse fonctionnelle de données (FDA), qui nous permet d'obtenir un patron représentatif robuste. Cette procédure est composée des deux étapes suivantes: (1) l'élimination des données aberrantes et (2) la réduction de la variabilité (Fig 4.1). Nous allons commencer alors par définir l'analyse fonctionnelle de données ainsi que les différents outils utilisés dans la procédure proposée, puis présenter l'application sur nos données cinématiques et les résultats obtenus.

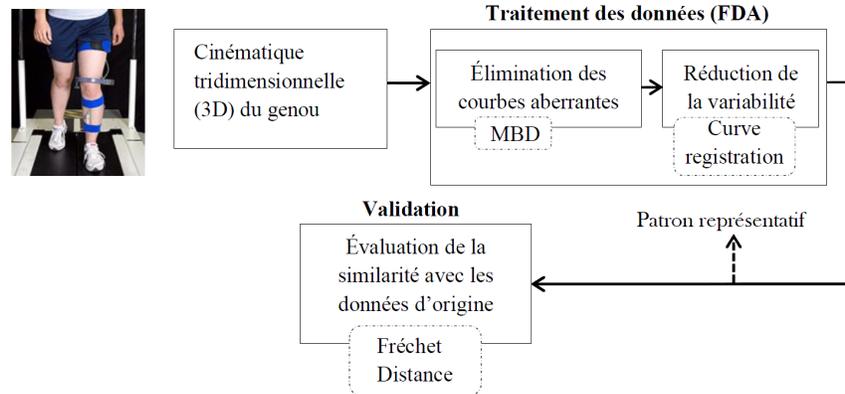


Figure 4.1 – Procédure proposée pour la détermination du patron représentatif des données cinématiques 3D

4.1 Revue de la littérature: Détermination classique du patron représentatif des données cinématiques du genou

Comme nous l’avons expliqué, le patron représentatif des données cinématiques est un élément clé dans la procédure de classification. Cependant, une revue de la littérature nous permet de conclure que ce patron est essentiellement obtenu par deux méthodes:

4.1.1 Détermination du patron représentatif en moyennant les 15 cycles les plus corrélés

Différents travaux de recherche [33], [43] ont exploité la méthodologie suivante pour l’obtention du patron représentatif de chacun des cycles de flexion/extension, abduction/adduction et la rotation interne/externe: Suite à l’acquisition des données, les étapes sont les suivantes:

1. **Filtrage:** Le filtrage sert essentiellement à supprimer le bruit qui affecte les données suite à des artefacts d’acquisition. La méthode utilisée est la méthode *automatique Singular Spectrum Analysis* (SSA) proposée dans [32]. C’est une approche non paramétrique qui est appliquée pour l’analyse des séries chronologiques.

2. **Échantillonnage:** Étant donné la variabilité entre la durée des différents cycles de la même personne, un échantillonnage permettant d'unifier le début et la fin de chaque cycle s'impose. Comme proposé dans [43], le début et la fin du cycle de marche sont déterminés à l'aide de la composante de force verticale (FV) qui correspond à l'instant où l'amplitude des forces de réactions au sol dépassent les 2 % du poids corporel des sujets. En se basant sur cette méthode, les données qui sont enregistrées durant 45 s sont découpées en cycles de marches normalisés (chaque cycle est défini par 100 points).
3. **Sélection des cycles:** A ce niveau, pour pouvoir déterminer un patron représentatif des différents cycles de chaque patient, il est important de choisir les cycles les plus représentatifs et supprimer ceux qui sont aberrants. Ces cycles sont choisis de façon à maximiser le coefficient de corrélation de l'ensemble des courbes. Selon les études de Boivin [43] une dizaine de cycles est conservée démontrant la meilleure reproductibilité inter cycles.

4.1.2 Détermination du patron représentatif en moyennant les courbes contenues dans l'intervalle de confiance via le bootstrap

La deuxième méthode pour la détermination du patron représentatif des cycles de marche diffère de la précédente au niveau de la 3^{ème} étape qui consiste à sélectionner les cycles représentatifs. En effet, la méthode utilisée est encore basée sur la corrélation entre les différentes courbes mais en définissant un intervalle de confiance qui permet de garder les courbes les plus fiables [44], [45]. Ces intervalles de confiance sont utilisés dans la recherche clinique pour déterminer l'étendue des mesures ou pour effectuer des comparaisons entre des groupes de mesures (intervalle de confiance pour la différence entre les moyennes des groupes). Des procédures statistiques appropriées pour établir ces intervalles de confiance existent et sont basées soit sur des approximations paramétriques (approximations gaussiennes ou d'autres modèles de distribution), soit sur des procédures non paramétriques comme des méthodes de bootstrap. Pour les courbes des cycles de marche, l'analogie des intervalles de confiance est les bandes de confiance: ce sont des régions délimitées par deux courbes

définies à chaque point temporel. La bande de confiance d'une population a la signification suivante: le cycle de marche d'un sujet tiré au hasard de la population est inclus dans la bande de confiance avec une probabilité prédéterminée (en général, 0,95). Pour la courbe moyenne d'une population donnée, la bande de confiance correspondante définit la région qui contient la courbe moyenne vraie avec une probabilité prédéterminée (Fig. 4.2). La fiabilité de la méthode utilisée pour construire une bande de confiance dépend de sa capacité à préserver la probabilité prédéterminée.

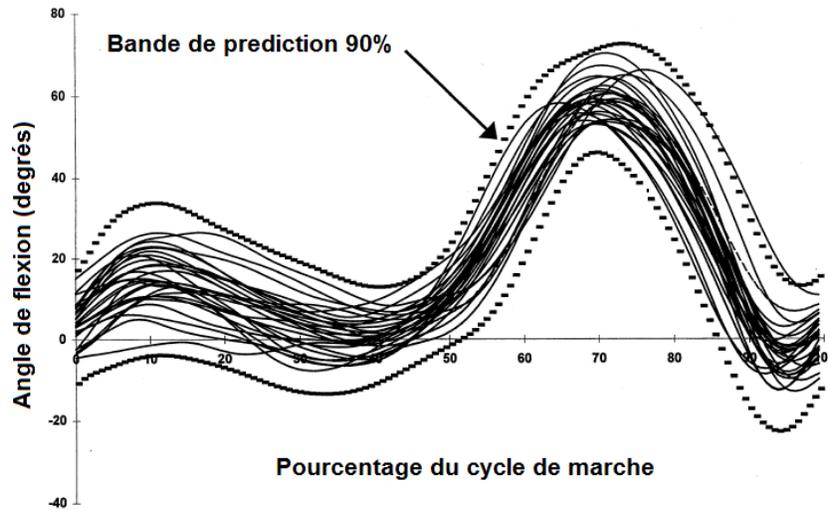


Figure 4.2 – Intervalle de confiance 90% pour les courbes de flexion. Tirée et modifié de [45]

Nous pouvons bien remarquer que la sélection des courbes représentatives par les deux méthodes que nous venons de présenter dans ce paragraphe, ne tient pas compte de l'aspect continu des courbes. D'autant plus, la variabilité existante entre les différents cycles sélectionnés risque d'influencer la fiabilité de la moyenne obtenue. Notre objectif est d'obtenir les courbes les plus homogènes et semblables possible qui nous permettrons d'obtenir un patron robuste. Nous traitons cette problématique dans ce qui suit en utilisant l'analyse fonctionnelle des données.

4.2 Traitement des données continues: Analyse fonctionnelle des données

4.2.1 Principe de l'analyse fonctionnelle des données

Bien que la notion de données fonctionnelles existe depuis longtemps, de nouvelles méthodes de traitement s'imposent vu la complexité et la dimensionnalité des données collectées dans les domaines d'analyse du comportement humain et les phénomènes naturels d'une façon plus générale. La technologie de la collecte de données a évolué au cours des deux dernières décennies, ce qui nous permet d'obtenir des observations densément échantillonnées au fil du temps, l'espace, et d'autres continuums [13]. Les outils classiques de la statistique multivariée peuvent être appliqués à ces données, mais ils ne peuvent pas exploiter des informations supplémentaires propres à ces fonctions continues. Les méthodes d'analyse de données fonctionnelles (FDA) que nous décrivons dans ce qui suit peuvent souvent extraire des informations supplémentaires contenues dans ces fonctions et leurs dérivés, ce qui n'est pas généralement possible avec les méthodes traditionnelles [46].

En effet, l'analyse fonctionnelle des données est une extension de la capacité des outils statistiques traditionnels. L'étude des propriétés statistiques d'un ensemble de données lorsque les observations sont des courbes au lieu de points exige que les outils utilisés soient adaptés à ce cadre fonctionnel pour pouvoir extraire les caractéristiques uniques de ces données fonctionnelles. Par exemple, la disponibilité des dérivés permet l'utilisation d'équations différentielles en tant que modèles, et introduit ainsi une description de la dynamique du processus, ainsi que ses caractéristiques statiques. Étant donné que les processus étudiés avec la FDA sont, par définition, une évolution en fonction du temps (ou l'espace, ou un autre continuum), la capacité de quantifier ces dynamiques est un avantage principal de ces techniques par rapport aux techniques traditionnelles statiques. Une comparaison de la FDA avec la statistique multivariée traditionnelle (MVA) est valable sur différents plans. Dans les statistiques classiques, un échantillon de points discrets est pris d'une grande population. Cet échantillon peut contenir de nombreuses sources de variations: Une

variabilité initiale des entités mesurées, l'erreur de mesure, de la variabilité due à l'échantillonnage, l'incapacité de reproduire les conditions de traitement exactement d'une entité à une autre, et la variabilité systématique qui est l'objectif de l'analyse. Dans un cadre fonctionnel, un échantillon de courbes est tiré d'un ensemble d'observations plus large. Cet échantillon fonctionnel contient les mêmes sources de variation comme pour un échantillon discret; cependant, il y a le défi supplémentaire de quantifier la variabilité au sein et entre les courbes: un problème qui ne se pose pas dans le cas discret.

Pour appliquer l'analyse fonctionnelle de données, la première étape consiste à convertir les données brutes en objets fonctionnels. Pour ce faire, les observations discrètes sont interpolées en des courbes continues. Une procédure commune dans les statistiques, les mathématiques et l'ingénierie pour représenter des données discrètes en fonction lisse est l'utilisation d'une expansion de la base comme exprimé dans l'équation [46]:

$$x_i(t) = C_{i1}\Phi_1(t) + \dots + C_{ik}\Phi_k(t) + \dots + C_{iK}\Phi_K(t), \quad (4.1)$$

où K indique le nombre de fonctions de base et $\Phi_k(t)$ est la valeur de la k^{eme} fonction de base à l'instant t . Les fonctions de base $\Phi_k(t)$ présentent un système de fonctions spécialement conçue comme des blocs de construction pour représenter une courbe lisse. Il existe différents types de systèmes de fonctions de base, dont les plus connus sont les puissances de t ou les monômes, utilisés pour construire des polynômes, et les séries de Fourier utilisées pour les courbes périodiques. Les C_{ik} sont les coefficients déterminant le poids relatif de chaque fonction de base dans la construction de la i^{eme} courbe. Ainsi, l'estimation d'une courbe exprimée comme une combinaison linéaire de fonctions de base, réduit le problème d'estimation des paramètres multivariés à l'estimation de ces coefficients. Lorsque les courbes ne sont pas périodiques et sont complexes, de sorte qu'un polynôme d'ordre inférieur ne peut pas la modéliser, les fonctions de base les plus utilisés généralement sont les fonctions B-splines [47]. Ce sont les fonctions que nous avons choisi pour modéliser nos données

cinématiques vu leurs formes complexes et non périodiques. Les B-splines sont en effet des polynômes joints bout à bout sur un ensemble de limites d'intervalles appelés les noeuds.

Les noeuds sont souvent choisis de façon à ce qu'ils soient également espacés, mais ils peuvent aussi coïncider avec les instants exactes quand les données ont été effectivement observées, ou à des moments d'intérêt prédéterminés. Les B-splines sont caractérisées par leur ordre, qui dépasse d'un degré celui des polynômes qui les construits; par exemple, les B splines d'ordre 3 comprennent des courbes quadratiques jointes aux noeuds, et les B-splines d'ordre 2 sont construites à partir de segments de droites. Suite à une étude d'un grand nombre de bases de données à modéliser, les B-splines d'ordre 4 semblent raisonnablement lisse et permettent une modélisation optimale de ces données discrètes.

Les poids des coefficients de base sont choisis de telle sorte que la courbe construite soit adaptée de façon optimale aux données avec un certain degré de lissage. Ce degré est contrôlé par le nombre de fonctions de base utilisé: Plus le nombre de fonctions de base K est important, plus l'interpolation est extrêmement lisse; sa diminution engendre des courbes moins lisses dont l'extraction des caractéristiques devient moins évidente.

4.2.2 Mesure de la profondeur fonctionnelle pour la détermination des courbes aberrantes

Dans de nombreuses études dans le domaine biomédical, les observations individuelles sont des fonctions réelles de temps, observées à des instants discrets. Chaque courbe fournit l'évolution au cours du temps d'un certain processus d'intérêt pour un individu donné. Lorsque la grille de points est assez dense, et le processus est connu pour être continu, les courbes peuvent être traitées comme des données fonctionnelles comme expliqué précédemment. Dans ce contexte, les données peuvent être considérées comme un échantillon de courbes, dans lequel il est fréquent d'observer les trajectoires atypiques. Des exemples comprennent les courbes de croissance humaine [13], les taux de fécondité et de mortalité [48] ou la concentration des polluants dans le temps [49], etc. Comme

dans le cas univarié ou multivarié, la présence d'observations atypiques peut influencer l'analyse statistique des données. Ainsi, toute analyse préliminaire devrait inclure une étape de détection des valeurs aberrantes. Le défi dans le cadre fonctionnel est que les observations atypiques peuvent être très difficiles à identifier visuellement.

En effet, une trajectoire aberrante est non seulement celle qui contient des valeurs anormalement élevées ou basses, mais aussi une trajectoire qui, même contenant des niveaux moyens dans l'ensemble de l'intervalle d'observation peut présenter une forme ou un motif différent de celui du reste des courbes de l'échantillon.

Une façon de déterminer ces courbes aberrantes dans la statistique classique est le diagramme en boîte. En effet, il est bien connu que le diagramme en boîte est une méthode graphique qui permet d'afficher cinq statistiques descriptives: la médiane, le premier et troisième quartiles, et les observations minimales et maximales non-aberrantes. Un diagramme en boîte peut également indiquer quelles observations, le cas échéant, peuvent être considérées comme des valeurs aberrantes. Les diagramme en boîte ont été initialement présentés par Tukey [50] dans l'analyse exploratoire des données, par la suite, ils ont évolué en une méthode simple mais instructive dans l'interprétation des données. La première étape pour construire un diagramme en boîte est le classement des données. Dans le cadre univarié, le classement est tout simplement de la plus petite à la plus grande observation. Cependant, le cas multivarié est beaucoup plus complexe et a suscité un intérêt considérable au fil des années [51].

Un outil naturel pour analyser ces aspects fonctionnels de données et définir un ordre de classement est l'idée de la profondeur statistique ou profondeur fonctionnelle BD (Band Depth) [52], [53]. Cette notion a été introduite pour mesurer la «centralité» ou la «non conformité» d'une observation par rapport à un ensemble de données qui sont naturellement des courbes. Ainsi, cette mesure permet d'ordonner les courbes du centre vers l'extérieur. En fait, la courbe ayant une profondeur maximale peut être définie comme une estimation du centre de la distribution fonctionnelle. Par conséquent, la profondeur et l'aberrance sont des notions inverses, et les valeurs aberrantes fonc-

tionnelles, qui sont des courbes qui devraient être loin du centre de données, vont correspondre aux courbes de faible profondeur significativement. Par conséquent, une simple façon de rechercher des données fonctionnelles aberrantes est de détecter les courbes ayant une faible profondeur significative.

Suivant cette optique, López-Pintado et Romo [54] ont introduit le concept de la profondeur fonctionnelle selon une approche basée sur les graphes. Le graphe d'une fonction $y(t)$ est la partie du plan $G(y) = \{(t, y(t)) : t \in \mathcal{I}\}$. Ou \mathcal{I} est un intervalle fermé de \mathbb{R} . La bande qui est délimitée par les courbes y_{i_1}, \dots, y_{i_k} est $B(y_{i_1}, \dots, y_{i_k}) = \{(t, x(t)) : t \in \mathcal{I}, \min_{r=1, \dots, k} y_{i_r}(t) \leq x(t) \leq \max_{r=1, \dots, k} y_{i_r}(t)\}$. Considérons J le nombre de courbes déterminant une bande, $2 \leq J \leq n$, si $Y_1(t), \dots, Y_n(t)$ sont des copies indépendantes du processus stochastique $Y(t)$ générant les observations $y_1(t), \dots, y_n(t)$ alors la version originale de la définition de la profondeur fonctionnelle d'une courbe $y(t)$ selon la mesure de probabilité P est défini par:

$$BD_J(y, P) = \sum_{J=2}^J BD^{(j)}(y, P) = \sum_{j=2}^J P\{G(y) \subset B(Y_1, \dots, Y_j)\}, \quad (4.2)$$

où $B(Y_1, \dots, Y_j)$ est une bande limitée par j courbes aléatoires [51]. Une version plus simple et facile à implémenter est définie par López-Pintado et Romo(2009) [55]. Elle consiste à calculer la proportion du temps durant laquelle la courbe est contenue dans la bande. La profondeur fonctionnelle modifiée (MBD) est définie par:

$$MBD_{\mathcal{X}}(x) = \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=i+1}^n \frac{\lambda(\{t \in \mathcal{I} | \alpha_{ij} \leq x(t) \leq \beta_{ij}\})}{\lambda(\mathcal{I})}, \quad (4.3)$$

où $\alpha_{ij} = \min(x_i(t), x_j(t))$; $\beta_{ij} = \max(x_i(t), x_j(t))$, et $\lambda(\cdot)$ est la mesure de Lebesgue sur \mathcal{R} .

Cette profondeur fonctionnelle nous permet maintenant de définir le diagramme en boîte fonctionnel (*functional boxplot*) pour l'identification des données aberrantes. Généralement, dans le cas de la statistique classique, la boîte du diagramme en boîte contient 50% des données les plus cen-

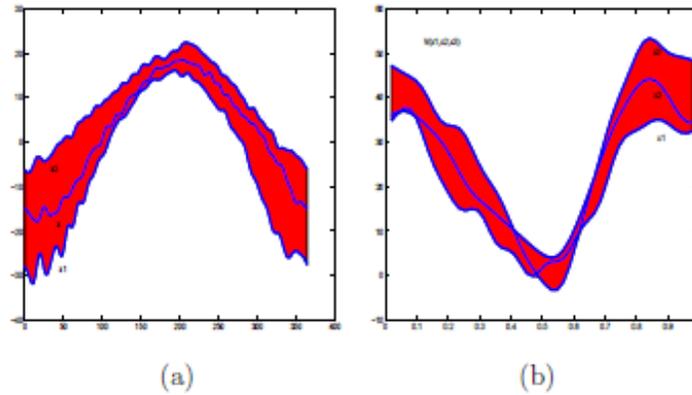


Figure 4.3 – (a) Bande définie par les deux courbes x_1 , x_2 et une 3^{eme} courbe x appartenant à la bande; (b) bande déterminée par trois courbes x_1 , x_2 et x_3 [54].

trales, qui correspond dans notre cas à la région centrale introduite par Liu, Parelius et Singh [56].

Cette région centrale est définie par:

$$C_{0.5} = \{(t, y(t)) : \min_{r=1, \dots, [n/2]} y_r(t) \leq y(t) \leq \max_{r=1, \dots, [n/2]} y_r(t)\} \quad (4.4)$$

où $[n/2]$ est la partie entière de $n/2$ plus un. Les limites de cette région définissent les extrémités de la boîte du diagramme dans la statistique classique. Cette région est encore définie comme étant la région inter-quartile (IQR) est nous donne une idée sur l'étalement des données les plus centrales parmi nos observations. Cette région est très importante dans les analyses et interprétations statistiques vue qu'elle n'est pas affectée par les données aberrantes. Un autre élément important du diagramme en boîte est la donnée ou la courbe la plus centrale qui représente la médiane. Comme nous l'avons expliqué, cette courbe possède la profondeur la plus élevée.

Les extrémités du diagramme classique indique l'enveloppe maximale des données non aberrantes. Ces données sont définies comme étant les données qui dépassent le 1.5 de la région inter-quartile. Ainsi dans le cas fonctionnel nous allons opter pour la même définition, toute courbe se situant à l'extérieur de la région centrale étendue de 1.5 de sa longueur sont de potentielles courbes

aberrantes. Dans la figure suivante (Fig 4.4) nous présentons un exemple de diagramme en boîte fonctionnel.

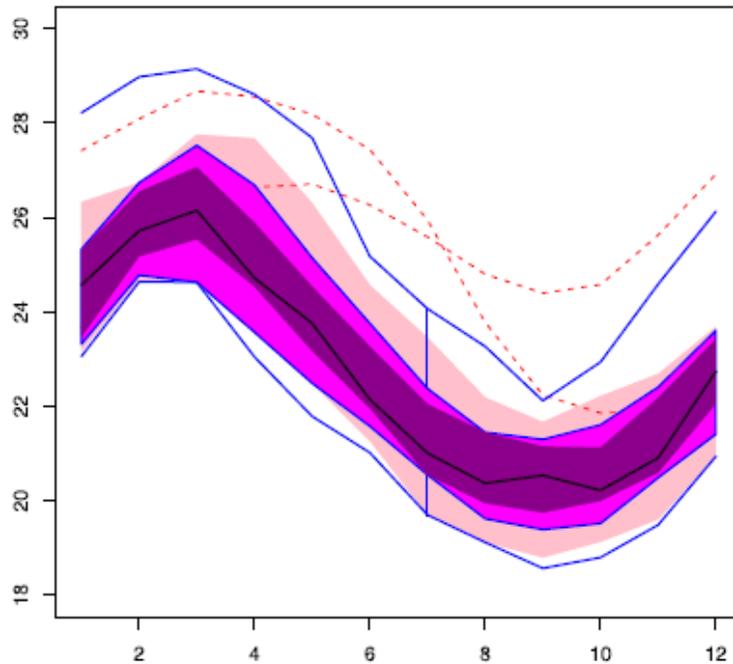


Figure 4.4 – Exemple de diagramme en boîte fonctionnel [51]

Il est important de noter que la boîte du diagramme, ses extrémités et la médiane sont une source supplémentaire d'information en exploitant leurs position, taille, longueur et même leur forme. Cela nous permet d'avoir une idée sur l'étalement des données et leur consistance.

Une fois les données aberrantes sont supprimées, l'étape suivante consiste à minimiser la variabilité entre les courbes pour avoir une moyenne plus significative et qui conserve les caractéristiques des données d'origine.

4.2.3 Alignement temporel des courbes pour la minimisation de la variabilité

Commençons par expliquer avec une simple figure l'utilité de cette étape dans l'extraction du patron représentatif (Fig 4.5).

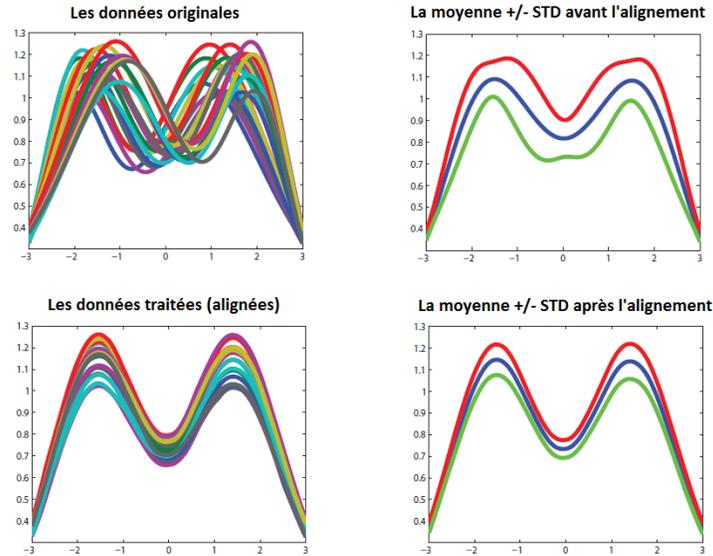


Figure 4.5 – Influence de la variabilité en phase sur la moyenne obtenue [57]

Comme nous pouvons le remarquer, la moyenne obtenue des données originales est affectée par cette variabilité en phase. L'amplitude des pics est bien réduite, et la moyenne ne peut pas représenter les données d'origine avec fiabilité. La différence est bien claire après le traitement, et nous remarquons que la moyenne obtenue est beaucoup plus représentative. Ainsi nous estimons que cette étape est nécessaire pour l'obtention d'un patron qui soit le mieux représentatif possible des données cinématiques de chaque sujet.

Expliquons un peu plus en détail cette procédure d'alignement [58]. En effet, tout se fait à l'aide des fonctions de déformation temporelle (*time warping functions*) qui consistent à effectuer des transformations non linéaires de l'échelle temporelle de chaque courbe de façon à ce que les extremums et les passages par zéro se produisent au même temps [59]. Appelons ces fonctions $h(t)$. Lorsque le processus représenté par la courbe est plus rapide que le temps normalisé, $h(t) > t$; s'il est plus lent, $h(t) < t$. Si $x(t)$ désigne la courbe d'origine, alors la courbe enregistrée serait $x[h(t)]$. Plusieurs procédures mathématiques ont été développées pour estimer les fonctions de déformation temporelle pour un ensemble de données, ces fonctions doivent en effet répondre aux conditions suivantes: Tout d'abord, ces fonctions doivent être construites de façon à ce que les débuts et les

fins des courbes restent intactes. Autrement dit, $h(t_0) = t_0$ et $h(t_{final}) = t_{final}$. Deuxièmement, une hypothèse de monotonie est imposée pour que les événements dans les courbes alignées se produisent dans le même ordre que ceux des courbes non alignées. Mathématiquement parlant, il nous faudrait alors que $h(t_2) > h(t_1)$ si et seulement si $t_2 > t_1$, ce qui implique que la fonction de déformation temporelle est monotone croissante. Enfin, nous avons pour but que nos courbes alignées possèdent toutes la même forme que les courbes originales sans affecter les amplitudes, c'est à dire que les pics et les vallées se produisent en même temps. Cela signifie que les courbes alignées dans l'ensemble de données doivent être proportionnelles entre elles. Mathématiquement, on dit que si les courbes x_1 et x_2 sont proportionnelles alors $x_1(h(t)) = ax_2(t)$ où a est une constante positive. Dans la figure 4.6 on présente un cas illustratif de ces fonctions de déformation temporelle.

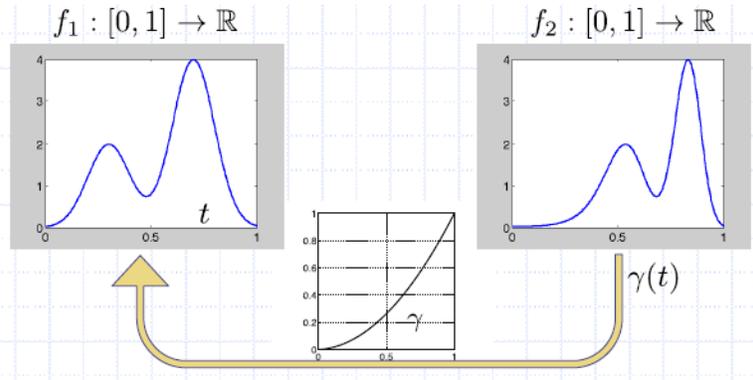


Figure 4.6 – Illustration de la fonction de déformation temporelle (γ). γ est construite de façon à ce que les points de $f_1(t)$ coïncident avec $f_1(\gamma(t))$ [57]

Pour résumer ce que nous avons expliqué dans ce chapitre, le traitement de nos données cinématiques consiste à réitérer les étapes suivantes jusqu'à convergence [60]:

1. Calculer la courbe moyenne des observations $\hat{\mu}[t]$
2. Pour chaque observation calculer la fonction de déformation temporelle $h_i(t)$ qui minimise la fonction objective suivante $J = \sum_{i=1}^N \int_T (Y_i[h_i(t)] - \hat{\mu}[t])^2 dt$, où N est le nombre de courbes retenues et Y_i la i^{eme} courbe. En d'autres termes, nous essayons d'aligner chaque courbe avec la moyenne des observations

Dans la suite de ce chapitre nous allons présenter l'application de ce traitement sur nos données et les résultats obtenus.

4.3 Résultats expérimentaux: Approche fonctionnelle pour la détermination du patron représentatif

La première étape comme nous l'avons expliqué consiste à transformer nos données en des objets fonctionnels en les exprimant en fonction des B-splines Figs 4.7, 4.8 et 4.9.

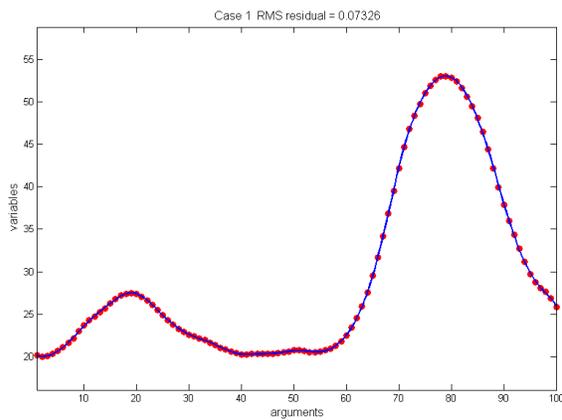


Figure 4.7 – Transformation d'un cycle de flexion en objet fonctionnel

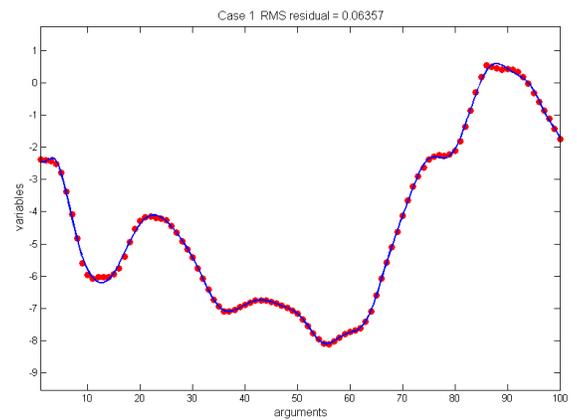


Figure 4.8 – Transformation d'un cycle d'abduction en objet fonctionnel

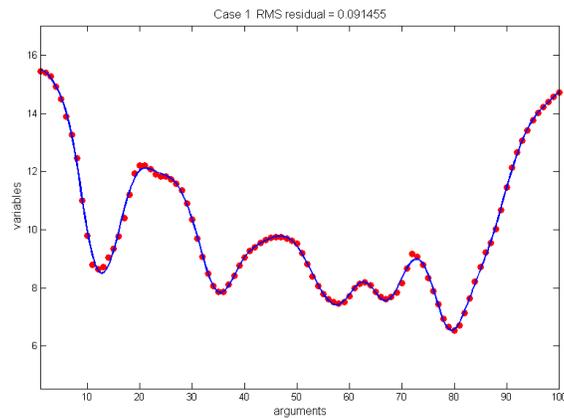


Figure 4.9 – Transformation d'un cycle de rotation en objet fonctionnel

Comme nous l'avons précisé, nous avons choisi des B-spline d'ordre 4 que nous estimons suffisamment suffisants pour obtenir des courbes lisse sans augmenter la complexité de calcul ou aboutir à un surajustement des données.

Une fois notre base de données est transformée en objets fonctionnels, nous commençons la première étape de notre approche de traitement proposée à savoir l'élimination des données aberrantes. Notons bien que la méthode communément utilisée dans la littérature est de conserver les 15 courbes les plus répétitives [44]. Cependant, cette méthode abouti à une perte importante de donnée vu que plus que la moitié des observation est abandonnée. Les figures suivantes (Fig. 4.10, 4.11, 4.12) illustrent les diagramme en boîte utilisés pour la détection des données aberrantes en se basant sur la mesure de la profondeur fonctionnelle.

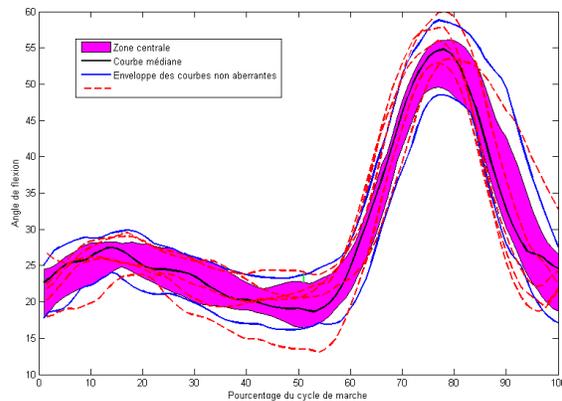


Figure 4.10 – Diagramme en boîte pour les courbes de flexion

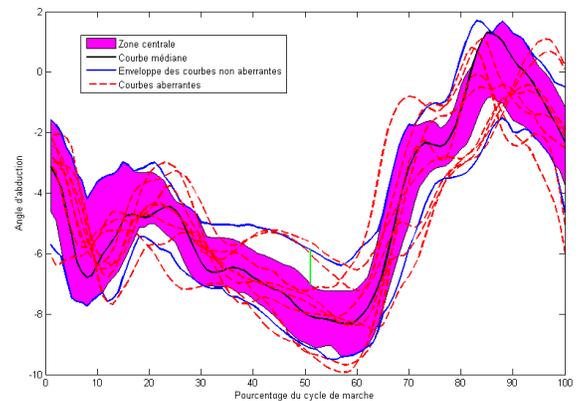


Figure 4.11 – Diagramme en boîte pour les courbes d'abduction

Nous remarquons que la variation des angles de rotation et d'abduction est beaucoup moins régulière que celle de la flexion, ce qui nous amène à confirmer davantage l'hypothèse de la non suffisance de 15 cycles parmi toutes les observation qui peuvent dépasser la quarantaine pour une meilleur représentation fiable de ses données.

Finalement la dernière étape est d'effectuer l'enregistrement des courbes pour réduire la variabilité en phase. Les figures suivantes illustrent des cas d'alignement en représentant la fonction de déformation temporelle utilisée.

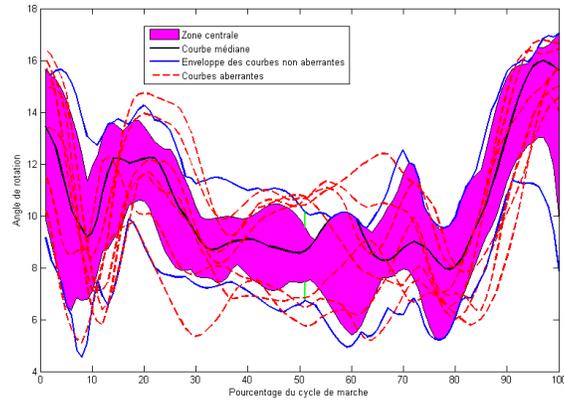


Figure 4.12 – Diagramme en boîte pour les courbes de rotation

Notons bien que le code que nous avons développé pour ce traitement est basé sur la bibliothèque open source proposée par J. O. Ramsay (2009) [13].

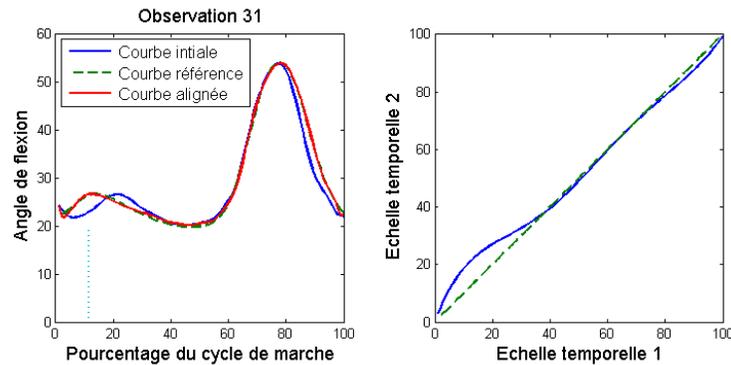


Figure 4.13 – Exemple d'alignement d'une courbe de flexion

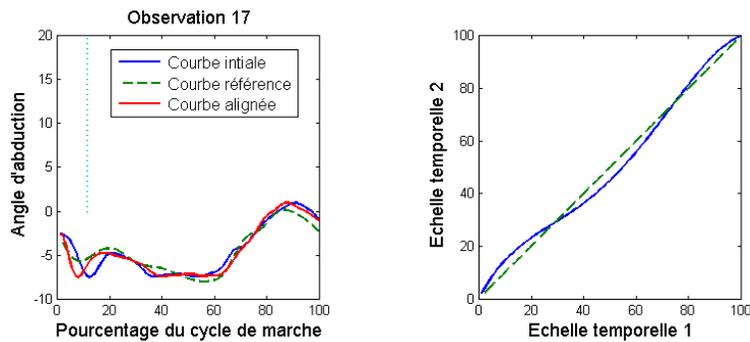


Figure 4.14 – Exemple d'alignement d'une courbe d'abduction

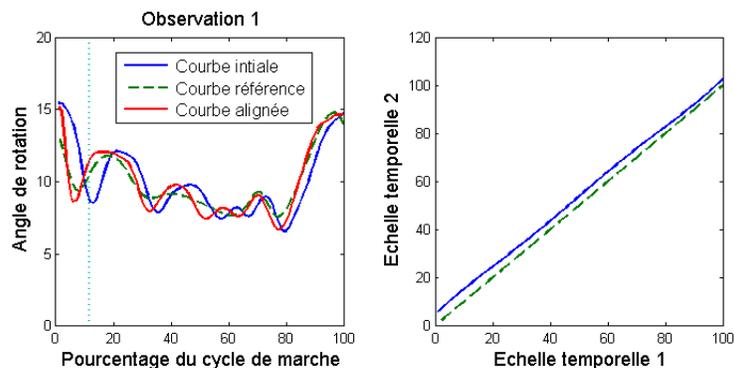


Figure 4.15 – Exemple d'alignement d'une courbe de rotation

Le résultat final appliqué sur toutes les observations d'un sujet donnée est représenté par les figures suivantes:

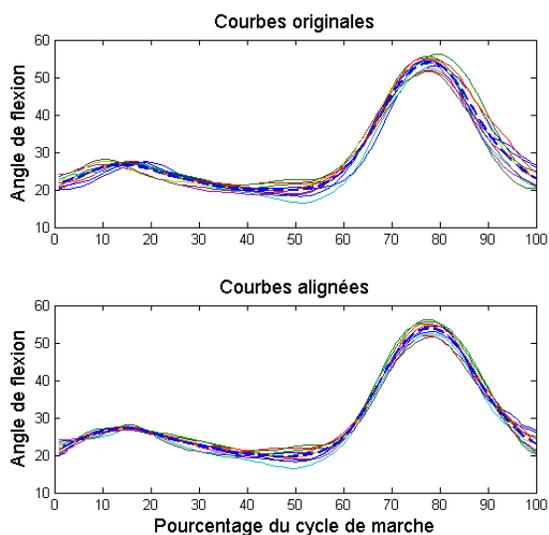


Figure 4.16 – Résultat d'alignement de toutes les observations de flexion pour un sujet donné

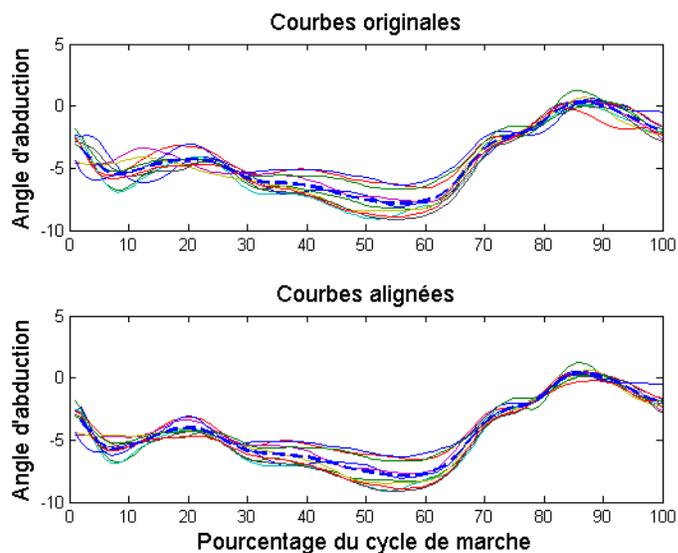


Figure 4.17 – Résultat d'alignement de toutes les observations de d'abduction pour un sujet donné

4.3.1 Résultats et validation du patron obtenu

Dans les figures 4.19, 4.20, 4.21 nous représentons les patrons représentatifs obtenus en comparaison avec les deux patrons les plus utilisés dans la littérature, à savoir, la simple moyenne des données brutes et la moyenne des 15 cycles les plus corrélés.

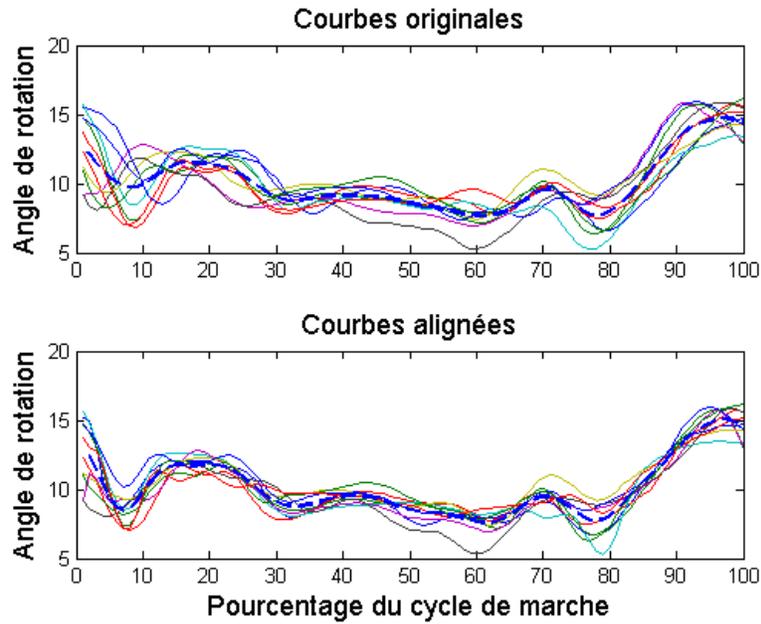


Figure 4.18 – Résultat d’alignement de toutes les observations de rotation pour un sujet donné

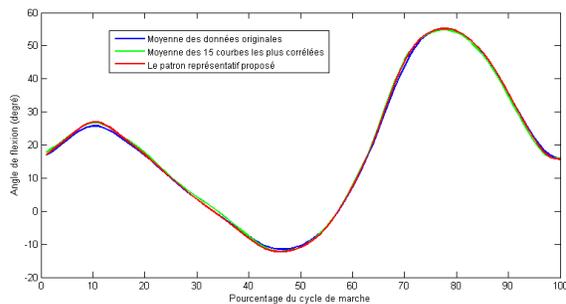


Figure 4.19 – Comparaison entre les 3 patrons représentatifs de flexion

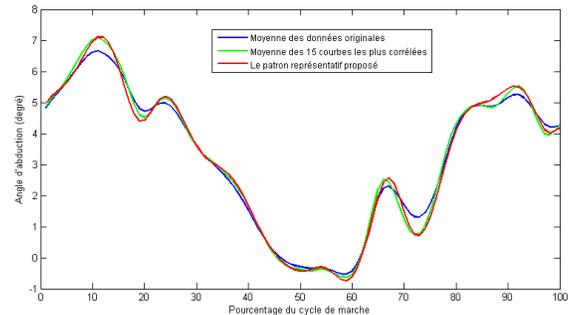


Figure 4.20 – Comparaison entre les 3 patrons représentatifs de d’abduction

Nous pouvons bien remarquer que la différence entre les 3 patrons est moins significative pour l’angle de flexion vu que c’est le signal le plus régulier et qui représente une variabilité minimale. Par contre, la différence est bien nette pour les autres angles surtout au niveau des valeurs extrêmes. Il est bien clair que les pics et les vallées sont plus importants avec notre patron.

Pour valider les résultats obtenus, nous étions amenés à répondre à la question suivante: Comment peut-on juger qu’une moyenne est meilleure qu’une autre ou plus représentative? Pour ce faire nous nous sommes basés sur l’hypothèse suivante: la courbe la plus représentative est celle qui

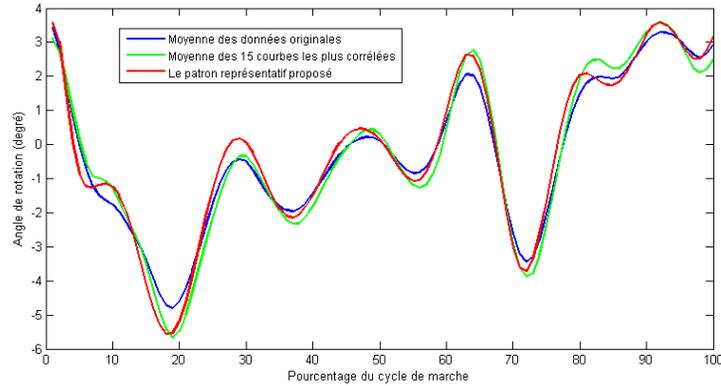


Figure 4.21 – Comparaison entre les 3 patrons représentatifs de rotation

ressemble le plus à nos données d'origine. Comme nous l'avons remarqué en expliquant l'étape de l'alignement des courbes, la variabilité en phase affecte la moyenne en atténuant les pics et modifiant des caractéristiques importantes des données originales, ce qui nous permet de conclure qu'une telle moyenne n'est pas suffisamment représentative de l'ensemble de données car elle ne les ressemble pas.

Ainsi pour mesurer cette ressemblance, nous avons opter pour la distance de Fréchet [61]. Cette distance introduite par Maurice Fréchet mesure la similarité entre deux courbes en tenant compte de la position et de l'ordre des points. Dans la figure suivante nous pouvons voir la différence entre la distance euclidienne et la distance de Fréchet:

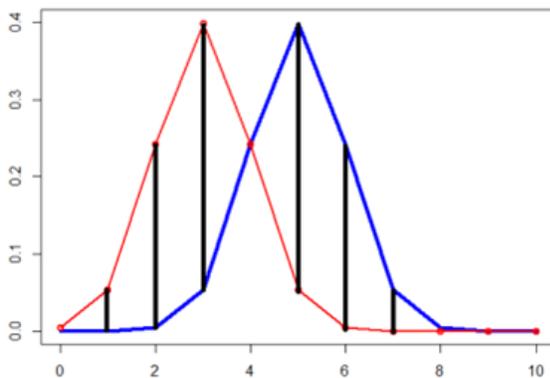


Figure 4.22 – Distance euclidienne

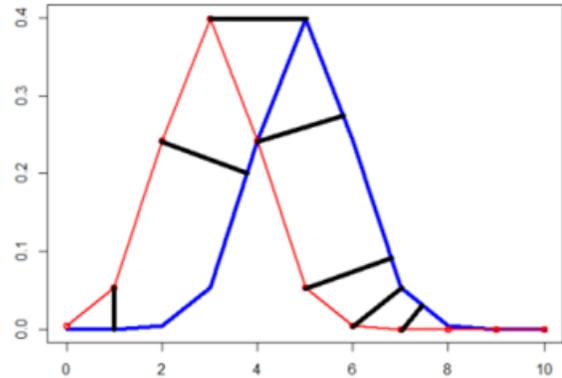


Figure 4.23 – Distance de Fréchet

Maurice Fréchet a proposé une définition intuitive de cette distance de la façon suivante: La distance de Fréchet entre deux courbes est la longueur minimale d'une laisse nécessaire pour connecter un chien et son propriétaire, contraint sur deux chemins séparés, en marchant sans retour en arrière le long de leurs courbes respectives d'un point initial à un point final. La définition est symétrique par rapport aux deux courbes. Imaginons un chien marchant le long d'une courbe et son propriétaire sur l'autre courbe, reliés par une laisse. Les deux marchent de façon continue le long de leur courbes respectives à partir du point de départ jusqu'au point final. Les deux peuvent varier leurs vitesses, et même s'arrêter à des positions arbitraires et pour des périodes arbitraires au cours du temps. Cependant, aucun ne peut revenir en arrière. La distance de Fréchet entre les deux courbes est la longueur de la laisse la plus courte qui est suffisante pour traverser les deux courbes de cette manière.

La formulation mathématique de cette distance commence par la définition d'un espace métrique \mathcal{S} muni d'une distance d . Soit A et B deux courbes de cet espace, et $\alpha : [0, 1] \rightarrow [0, 1]$ une fonction croissante surjective de paramétrisation. La distance de Fréchet entre A et B est défini comme étant le minimum absolu considérant toutes les paramétrisations α et β du maximum sur tout l'intervalle $[0, 1]$ de la distance dans l'espace \mathcal{S} entre $A(\alpha(t))$ et $B(\beta(t))$ comme l'exprime l'équation suivante:

$$F(A, B) = \inf_{\alpha, \beta} \max_{t \in [0, 1]} \{d(A(\alpha(t)), B(\beta(t)))\}. \quad (4.5)$$

En utilisant cette distance pour la mesure de similarité, nous avons comparé notre patron représentatif proposé avec les deux patrons les plus utilisés dans la littérature à savoir, la simple moyenne des données brutes et la moyenne des 15 cycles les plus répétitifs ou les encore les plus corrélés. La procédure de validation se résume dans les étapes suivantes:

1. Pour chaque sujet de la base de donnée, retenir les observations valides en éliminant les courbes aberrantes.
2. Calculer les 3 patrons représentatifs mentionnés ci-dessus.

3. Pour chaque patron, calculer la moyenne de la distance de Fréchet qui le sépare des données valides du sujet en question. En d'autres termes, mesurer la similarité entre chaque patron et les données d'origine.
4. Répéter les étapes 2 et 3 pour chaque sujet de la base de données

Le tableau suivant résume le résultat obtenu. Il représente le pourcentage de sujets de la base de données pour lesquels le patron que nous avons proposé est plus semblable aux données d'origine, ou encore la distance de Fréchet qui le sépare des données valides est la plus petite.

	Flexion	Abduction	Rotation
Moyenne des données brutes	70%	65%	78%
Moyenne des 15 observations les plus corrélées	72%	75%	76%

Tableau 4.1 – Pourcentage des cas où le patron proposé est plus représentatif que la moyenne des données brutes et la moyenne des 15 observations les plus corrélées.

On remarque que le patron représentatif que nous proposons dépasse les deux autres patrons en termes de similarité avec les données d'origine que nous voulons représenter. Nous pouvons donc conclure que la procédure de prétraitement proposée permet de conserver les caractéristiques importantes des observations à représenter ce qui sera d'une forte valeur ajoutée dans les travaux de classification ou la détection des anomalies et pathologies musculosquelettiques. Notons bien que la réduction de la variabilité nous a servis dans ce cadre pour la détermination du patron représentatif, cependant, cette variabilité en elle-même peut être l'objet de l'analyse et pourrait révéler des informations supplémentaires quant à l'état fonctionnel du genou.

Chapitre 5

Conclusion

Cette étude constitue une analyse préliminaire des données cinématiques tridimensionnelles du genou avant leur classification. L'analyse était composée de deux parties principales. La première est une étude de complexité de ces données qui a été évaluée de point de vue chevauchement entre les points d'intérêt et la séparabilité des classes. La particularité de cette étude est qu'elle est indépendante du choix du classifieur, ce qui a permis d'avoir une caractérisation intrinsèque de la base de données. Les mesures de complexité évaluées serviront d'outil d'analyse de performance, pour tout type de traitement effectué sur ces données et aidera également dans le choix du classifieur adéquat.

La deuxième partie de notre travail consistait à une amélioration du patron représentatif des ces données cinématiques qui constitue nos bases de données. La façon classique d'obtention de ce patron, ne tien pas compte de la grande variabilité existante entre les observations ce qui rend le patron moins fiable et représentatif. Ce que nous avons proposé était un traitement basé sur l'analyse fonctionnelle de données qui permet d'éliminer les observations aberrantes en premier lieu en utilisant la profondeur fonctionnelle: une approche qui tien compte de l'aspect continu de nos données contrairement à la statistique classique, puis de réduire la variabilité entre les courbes avec la technique d'alignement qui se base sur les fonctions de déformation temporelle.

La majeure difficulté que nous avons rencontrée dans cette étude est la validation des résultats obtenus vu qu'il n'existe pas un patron référence dans la littérature. Notre objectif était de maximiser la fiabilité du patron pour une meilleure exploitation et analyse dans les travaux de classification. Notre validation était donc basée sur l'idée de départ, un patron représentatif est un patron qui ressemble le plus à chacune des courbes individuellement. Cette ressemblance était effectivement améliorée en suivant la méthodologie présentée dans cette étude, ce que nous estimons un résultat prometteur qui pourra améliorer les travaux de classification et d'analyse de ces données.

Références

- [1] D. Gouwanda and S. Senanayake, “Emerging trends of body-mounted sensors in sports and human gait analysis,” in *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*, pp. 715–718, Springer, 2008.
- [2] S. L. Di Stasi, D. Logerstedt, E. S. Gardinier, and L. Snyder-Mackler, “Gait patterns differ between acl-reconstructed athletes who pass return-to-sport criteria and those who fail,” *The American journal of sports medicine*, p. 0363546513482718, 2013.
- [3] H. Lee, S. J. Sullivan, and A. G. Schneiders, “The use of the dual-task paradigm in detecting gait performance deficits following a sports-related concussion: a systematic review and meta-analysis,” *Journal of Science and Medicine in Sport*, vol. 16, no. 1, pp. 2–7, 2013.
- [4] S. S. S. Fathima and R. W. Banu, “Human gait recognition based on motion analysis including ankle to foot angle measurement,” in *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on*, pp. 1133–1136, IEEE, 2012.
- [5] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette analysis-based gait recognition for human identification,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.
- [6] D. H. Sutherland, “The evolution of clinical gait analysis part I: kinesiological emg,” *Gait & posture*, vol. 14, no. 1, pp. 61–70, 2001.
- [7] D. H. Sutherland, “The evolution of clinical gait analysis: Part II kinematics,” *Gait & posture*, vol. 16, no. 2, pp. 159–179, 2002.
- [8] D. H. Sutherland, “The evolution of clinical gait analysis part III—kinetics and energy assessment,” *Gait & Posture*, vol. 21, no. 4, pp. 447–461, 2005.
- [9] V. Medved, *Measurement of Human Locomotion*. CRC Press, 2000.
- [10] N. Mezghani, S. Husse, K. Boivin, K. Turcot, R. Aissaoui, N. Hagemeister, J. De Guise, *et al.*, “Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier,” *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 3, pp. 1230–1232, 2008.
- [11] N. Mezghani, Y. Ouakrim, A. Fuentes, N. Hagemeister, R. Aissaoui, M. Pelletier, and J. de Guise, “Knee osteoarthritis severity assessment using knee kinematic data classification,” *Osteoarthritis and Cartilage*, vol. 20, p. S97, 2012.
- [12] T. K. Ho and M. Basu, “Measuring the complexity of classification problems,” in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, pp. 43–47, IEEE, 2000.
- [13] J. O. Ramsay, *Functional data analysis*. Wiley Online Library, 2006.

- [14] P. C. Besse, H. Cardot, and D. B. Stephenson, "Autoregressive forecasting of some functional climatic variations," *Scandinavian Journal of Statistics*, pp. 673–687, 2000.
- [15] S. J. Ratcliffe, G. Z. Heller, and L. R. Leader, "Functional data analysis with application to periodically stimulated foetal heart rate data. ii: Functional logistic regression," *Statistics in medicine*, vol. 21, no. 8, pp. 1115–1127, 2002.
- [16] J. Perry and J. M. Burnfield, "Gait analysis: normal and pathological function," 1992.
- [17] D. A. Winter, *Biomechanics and motor control of human gait: normal, elderly and pathological*. 1991.
- [18] A. Fuentes-Dupré, "Apport d'une évaluation biomécanique 3d du genou dans la prise en charge orthopédique de patients ayant une rupture du ligament croisé antérieur," 2011.
- [19] A. N. M. Gomatam and S. Sasi, "Multimodal gait recognition based on stereo vision and 3d template matching," in *CISST*, pp. 405–410, 2004.
- [20] S. C. White and D. A. Winter, "Predicting muscle forces in gait from emg signals and musculotendon kinematics," *Journal of Electromyography and Kinesiology*, vol. 2, no. 4, pp. 217–231, 1992.
- [21] C. Mummolo, L. Mangialardi, and J. H. Kim, "Quantifying dynamic characteristics of human walking for comprehensive gait cycle," *Journal of biomechanical engineering*, vol. 135, no. 9, p. 091006, 2013.
- [22] I. McCarthy, D. Hodgins, A. Mor, A. Elbaz, and G. Segal, "Analysis of knee flexion characteristics and how they alter with the onset of knee osteoarthritis: a case control study," *BMC musculoskeletal disorders*, vol. 14, no. 1, p. 169, 2013.
- [23] L. Hébert and D. Camirand, "Le corps et ses mouvements: Physiologie articulaire et repérage de structures anatomiques," *Montréal: Éditions Saint-Martin*, 1996.
- [24] E. S. Grood and W. J. Suntay, "A joint coordinate system for the clinical description of three-dimensional motions: application to the knee," *Journal of biomechanical engineering*, vol. 105, no. 2, pp. 136–144, 1983.
- [25] R. A. Teitge *et al.*, "Noyes\' knee disorders: Surgery, rehabilitation, clinical outcomes," 2009.
- [26] M. Lafortune, P. R. Cavanagh, H. Sommer, and A. Kalenak, "Three-dimensional kinematics of the human knee during walking," *Journal of biomechanics*, vol. 25, no. 4, pp. 347–357, 1992.
- [27] E. Villalobos, "Analyse cinématique de la marche de travailleurs exposés à une surcharge mécanique sur l'articulation fémoro-tibiale (af-t)," 2012.
- [28] W. F. Sims and K. E. Jacobson, "The posteromedial corner of the knee medial-sided injury patterns revisited," *The American journal of sports medicine*, vol. 32, no. 2, pp. 337–345, 2004.
- [29] D. B. Kettelkamp, R. J. JOHNSON, G. L. SMIDT, E. Y. CHAO, and M. Walker, "An electrogoniometric study of knee motion in normal gait," *J Bone Joint Surg Am*, vol. 52, no. 4, pp. 775–790, 1970.
- [30] S. Lustig, R. A. Magnussen, L. Cheze, and P. Neyret, "The kneekg system: a review of the literature," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 20, no. 4, pp. 633–638, 2012.
- [31] N. Hagemeister, G. Parent, M. Van de Putte, N. St-Onge, N. Duval, and J. de Guise, "A reproducible method for studying three-dimensional knee kinematics," *Journal of biomechanics*, vol. 38, no. 9, pp. 1926–1931, 2005.
- [32] R. Aissaoui, S. Husse, H. Mecheri, G. Parent, and J. De Guise, "Automatic filtering techniques for three-dimensional kinematics data using 3d motion capture system," in *Industrial Electronics, 2006 IEEE International Symposium on*, vol. 1, pp. 614–619, IEEE, 2006.

- [33] Y. Ouakrim, *Classification de sujets asymptomatiques et gonarthrosiques en fonction des données cinématiques comparaison de l'approche globale et de l'approche locale*. PhD thesis, École de technologie supérieure, 2011.
- [34] T. K. Ho and M. Basu, "Complexity measures of supervised classification problems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 289–300, 2002.
- [35] A. I. Landeros, *Data complexity and classifier selection*. University of Alabama, 2008.
- [36] M. Basu and T. K. Ho, "The learning behavior of single neuron classifiers on linearly separable or nonseparable input," in *Neural Networks, 1999. IJCNN'99. International Joint Conference on*, vol. 2, pp. 1259–1264, IEEE, 1999.
- [37] R. A. Johnson, D. W. Wichern, *et al.*, *Applied multivariate statistical analysis*, vol. 5. Prentice hall Upper Saddle River, NJ, 2002.
- [38] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software (TOMS)*, vol. 3, no. 3, pp. 209–226, 1977.
- [39] G. Ross, "Algorithm as 13: Minimum spanning tree," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 18, no. 1, pp. 103–104, 1969.
- [40] F. J. Rohlf, "A probabilistic minimum spanning tree algorithm," *Information Processing Letters*, vol. 7, no. 1, pp. 44–48, 1978.
- [41] S. Pettie and V. Ramachandran, "An optimal minimum spanning tree algorithm," *Journal of the ACM (JACM)*, vol. 49, no. 1, pp. 16–34, 2002.
- [42] A. Orriols-Puig, N. Macia, and T. K. Ho, "Documentation for the data complexity library in c++," *Universitat Ramon Llull, La Salle*, vol. 196, 2010.
- [43] K. Boivin, *Développement d'une approche d'évaluation clinique de la cinématique tridimensionnelle du genou durant la marche pour des patients gonarthrosiques*. PhD thesis, École Polytechnique de Montréal, 2010.
- [44] A. Duhamel, J. Bourriez, P. Devos, P. Krystkowiak, A. Destee, P. Derambure, and L. Defebvre, "Statistical tools for clinical gait analysis," *Gait & posture*, vol. 20, no. 2, pp. 204–212, 2004.
- [45] M. W. Lenhoff, T. J. Santner, J. C. Otis, M. G. Peterson, B. J. Williams, and S. I. Backus, "Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data," *Gait & Posture*, vol. 9, no. 1, pp. 10–17, 1999.
- [46] D. J. Levitin, R. L. Nuzzo, B. W. Vines, and J. Ramsay, "Introduction to functional data analysis.," *Canadian Psychology/Psychologie canadienne*, vol. 48, no. 3, p. 135, 2007.
- [47] C. De Boor, "Calculation of the smoothing spline with weighted roughness measure," *Mathematical Models and Methods in Applied Sciences*, vol. 11, no. 01, pp. 33–41, 2001.
- [48] R. J. Hyndman and M. S. Ullah, "Robust forecasting of mortality and fertility rates: a functional data approach," *Computational Statistics & Data Analysis*, vol. 51, no. 10, pp. 4942–4956, 2007.
- [49] M. Febrero-Bande and W. González-Manteiga, "Generalized additive models for functional data," *Test*, vol. 22, no. 2, pp. 278–292, 2013.
- [50] J. Turkey, "Exploratory data analysis (limited preliminary edition)," 1970.
- [51] Y. Sun and M. G. Genton, "Functional boxplots," *Journal of Computational and Graphical Statistics*, vol. 20, no. 2, 2011.
- [52] R. Fraiman and G. Muniz, "Trimmed means for functional data," *Test*, vol. 10, no. 2, pp. 419–440, 2001.

- [53] A. Cuevas, M. Febrero, and R. Fraiman, “On the use of the bootstrap for estimating functions with functional data,” *Computational statistics & data analysis*, vol. 51, no. 2, pp. 1063–1074, 2006.
- [54] S. López-Pintado and J. Romo, “On the concept of depth for functional data,” *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 718–734, 2009.
- [55] A. Arribas-Gil and J. Romo, “Shape outlier detection and visualization for functional data: the outliergram,” *Biostatistics*, vol. 15, no. 4, pp. 603–619, 2014.
- [56] R. Y. Liu, J. M. Parelius, K. Singh, *et al.*, “Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh),” *The annals of statistics*, vol. 27, no. 3, pp. 783–858, 1999.
- [57] A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. Marron, “Registration of functional data using fisher-rao metric,” *arXiv preprint arXiv:1103.3817*, 2011.
- [58] J. Ramsay and X. Li, “Curve registration,” *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pp. 351–363, 1998.
- [59] M. Müller, “Dynamic time warping,” *Information retrieval for music and motion*, pp. 69–84, 2007.
- [60] E. Crane, D. Childers, E. Rothman, and G. Gerstner, *Functional Data Analysis for Biomechanics*. INTECH Open Access Publisher, 2011.
- [61] H. Alt and M. Godau, “Computing the fréchet distance between two polygonal curves,” *International Journal of Computational Geometry & Applications*, vol. 5, no. 01n02, pp. 75–91, 1995.