

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université 8 mai 1945 Guelma



Faculté des Sciences et de la Technologie
Département de Génie Electrotechnique et Automatique
OPTION : Génie Electrique

Thèse

Pour l'obtention du diplôme de :

Doctorat en Science

Présentée par

Tawfik Thelaidjia

Conception d'un système intelligent de diagnostic des défauts dans les machines tournantes

Soutenue le : 13/12/2017

Devant le jury composé de :

| | | |
|-------------------------|---|---------------|
| Pr. Ahcène Lemzadmi | Professeur, Université de 8 mai 1945 Guelma | Président |
| Pr. Abdelkrim Moussaoui | Professeur, Université de 8 mai 1945 Guelma | Rapporteur |
| Pr. Salah Chenikher | Professeur, Université de Tébessa | Co-Rapporteur |
| Pr. Tahar Bahi | Professeur, Université d'Annaba | Examineur |
| Pr. Messaoud Ramdani | Professeur, Université d'Annaba | Examineur |

Année 2017

Table des matières

| | |
|---|-------------|
| Table des figures | v |
| Liste des tableaux | viii |
| Préface | 1 |
| I Etat de l'art | 4 |
| 1 Introduction | 4 |
| 2 Généralités sur les roulements à billes | 4 |
| 2.1 Constitution des roulements à billes | 4 |
| 2.2 Modes de dégradations d'un roulement à billes | 5 |
| 2.2.1 Effets des contraintes mécaniques | 6 |
| 2.2.2 Défauts de montage | 6 |
| 2.2.3 Corrosion | 8 |
| 2.2.4 Contamination du roulement | 8 |
| 2.2.5 Défaut de lubrification | 8 |
| 2.2.6 Température de fonctionnement | 10 |
| 3 Influence des défauts sur la structure des signaux | 12 |
| 4 Fréquences caractéristiques des défauts du roulement «calcul cinématique» . | 13 |
| 5 Synthèse bibliographique | 14 |
| 6 Conclusion | 20 |

| | | |
|-----------|--|-----------|
| II | Extraction des paramètres et Réduction de dimension pour les problèmes de classification | 21 |
| 1 | Introduction | 21 |
| A) | Extraction des paramètres | 22 |
| 1 | Les ondelettes | 24 |
| 1.1 | Historique de traitement du signal : de Fourier à Gabor. | 24 |
| 1.2 | Représentation temps-fréquence et temps-échelle. | 25 |
| 1.3 | Introduction aux ondelettes | 27 |
| 1.4 | La transformée en ondelette | 27 |
| 1.4.1 | La transformée en ondelettes continue. | 28 |
| 1.4.2 | La transformée en ondelette discrète | 29 |
| 1.5 | Analyse multi-résolution | 31 |
| 1.5.1 | Présentation de l'analyse multirésolution | 31 |
| 1.5.2 | Définition d'une analyse multirésolution | 31 |
| 1.5.3 | Relations de reconstruction et de décomposition | 32 |
| 1.5.4 | Algorithmes | 34 |
| 1.5.5 | Bilan | 36 |
| 1.6 | Paquets d'ondelettes | 37 |
| 1.7 | Orthogonalité et bi-orthogonalité | 39 |
| 1.7.1 | Les ondelettes orthogonales | 39 |
| 1.7.2 | Les ondelettes bi-orthogonales | 40 |
| 1.8 | Exemples d'ondelettes | 40 |
| 1.8.1 | Ondelettes orthogonales | 41 |
| | Ondelette de Haar | 41 |
| | Ondelette de Shannon | 41 |
| | Ondelette de Meyer | 42 |
| | Ondelette de Daubechies | 42 |
| 2 | Les paramètres statistiques | 44 |
| B) | Réduction de dimension | 48 |
| 1 | Optimisation par essaim particulaire (OEP) et Optimisation par essaim particulaire binaire | 50 |
| | Algorithme général de l'OEP | 51 |

| | | |
|---|---|-----------|
| 2 | Analyse en Composantes Principales | 53 |
| 3 | Technique d'évaluation de distance | 55 |
| 4 | Matrices de dispersion pour la Mesure de séparabilité entre les classes | 57 |
| 5 | Conclusion | 57 |
| III Les séparateurs à vaste marge | | 59 |
| 1 | Introduction. | 59 |
| 2 | Cas des données linéairement séparables | 60 |
| 2.1 | Recherche de la solution | 61 |
| 3 | Séparateurs à vaste marge souple | 64 |
| 3.1 | Recherche de la solution | 65 |
| 4 | Elaboration des séparateurs à vaste marge nonlinéaires | 66 |
| 5 | Exemples des Fonctions noyaux | 68 |
| 5.1 | Noyaux polynômiaux | 68 |
| 5.2 | Noyaux gaussiens | 68 |
| 5.3 | Noyaux exponentiels | 69 |
| 5.4 | Noyaux sigmoïdaux | 69 |
| 6 | Multi classification par les séparateur à vaste marge | 69 |
| 6.1 | Multi classification par l'approche 'un contre tous'. | 69 |
| 6.2 | Multi classification par l'approche 'un contre un' | 70 |
| 7 | Conclusion | 70 |
| IV Détection des défauts des roulements par analyse vibratoire | | 72 |
| 1 | Introduction | 72 |
| 2 | Les données expérimentales | 72 |
| 2.1 | Données temporelles | 73 |
| 2.2 | La première approche | 73 |
| 2.2.1 | Analyse des résultats expérimentaux | 77 |
| 2.3 | La seconde approche | 80 |
| 2.3.1 | Analyse des résultats expérimentaux | 82 |
| 2.4 | Résultat de simulation | 90 |
| 3 | Conclusion | 91 |
| Conclusion générale | | 94 |
| Références bibliographiques | | 96 |

| | |
|--|------------|
| Annexe A : Représentation vectorielle de signaux | 103 |
| 1 Espace vectoriel de fonctions | 103 |
| 1.1 Espace vectoriel | 103 |
| 1.2 Espace de Hilbert | 104 |
| 1.3 Espace L^2 des signaux à énergie finie | 104 |
| 2 Fonctions orthogonales | 105 |
| 2.1 Produit scalaire de signaux dans $L^2(t_1, t_2)$ | 105 |
| Annexe B : Noyaux reproduisants | 106 |
| 1 Premières définitions | 106 |
| 1.1 Définition 1 | 106 |
| 1.2 Définition 2 | 106 |
| 1.3 Définition 3 | 107 |
| 2 Noyaux reproduisants et condition de Mercer | 107 |
| Annexe C : Outils pour la classification | 108 |
| 1 Les réseaux de neurones, concepts et définitions | 108 |
| 1.1 Le Perceptron Multi Couches (MLP) | 109 |
| 2 Arbres de décisions et forêts aléatoires | 111 |
| 2.1 Arbres de décisions | 111 |

Table des figures

| | | |
|------|---|----|
| I.1 | Vue éclatée des éléments constitutifs d'un roulement à une rangée de billes . . . | 5 |
| I.2 | Aspect de la bague interne d'un roulement surchargé | 6 |
| I.3 | Ecaillage d'une piste de roulement | 7 |
| I.4 | Ecaillage d'une piste | 7 |
| I.5 | Cassure d'une bague | 8 |
| I.6 | Criques de corrosion sur une piste de roulement | 9 |
| I.7 | Marques d'indentation d'un chemin de roulement dues à la présence de contaminant | 9 |
| I.8 | Mécanismes généraux d'usure en présence de contaminant | 10 |
| I.9 | Blocage de roulement | 11 |
| I.10 | Ecaillage d'un chemin de roulement | 11 |
| I.11 | Eclatement de billes | 12 |
| I.12 | Représentation schématique de la géométrie interne d'un roulement à une rangée de billes | 13 |
| II.1 | Pavages temps fréquences associés aux représentations de Dirac et de Fourier . . . | 26 |
| II.2 | Pavages du plan temps fréquence pour la représentation temps-fréquence et la représentation temps-échelle | 27 |
| II.3 | La transformée en ondelette | 28 |
| II.4 | La nature hiérarchique de l'analyse multirésolution | 32 |
| II.5 | Décomposition en cascade. | 35 |
| II.6 | Algorithme de synthèse de Mallat. | 37 |
| II.7 | Analyse en paquets d'ondelettes. | 38 |

| | | |
|-------|---|----|
| II.8 | Fonction d'échelle et d'ondelette de type Haar | 41 |
| II.9 | Fonction d'échelle et ondelette de Meyer. | 42 |
| II.10 | Fonction d'échelle et ondelette de Daubechies pour $N = 8, 10, 20$ et 44 | 43 |
| II.11 | Fonction d'échelle et ondelette Symlets pour $N = 8, N = 10$ et $N=20$ | 45 |
| II.12 | Fonction d'échelle et d'ondelette de type Coifflets pour $N = 1, N =2$ et $N=5$ | 46 |
| III.1 | Principe des SVM dans le cas où les deux classes sont linéairement séparables. Les vecteurs supports, désignent les échantillons les plus proches de l'hyperplan séparateur. | 62 |
| III.2 | Séparateurs à vaste marge souple | 65 |
| III.3 | Séparateurs à vaste marge non lineaire | 67 |
| III.4 | Transformation de l'observation par une fonction noyau reproduisante | 68 |
| IV.1 | Banc d'essai | 73 |
| IV.2 | Représentation temporelle des signaux vibratoire d'un roulement : (a) avec défaut de bille, (b) avec défaut de bague intérieure, (c) avec défaut bague extérieure, (d) sain | 74 |
| IV.3 | Procédure de la première méthode proposée pour le diagnostic des défauts dans les roulements à bille. | 75 |
| IV.4 | Energie d'Ondelettes Maximale Relative pour diffèrent conditions de roule- ment | 76 |
| IV.5 | Paquet d'ondelette d'un roulement sain (N). | 77 |
| IV.6 | Paquet d'ondelette d'un roulement avec un défaut bague extérieure (OF1). | 78 |
| IV.7 | Trois composantes (PCs) obtenues après l'application du PCA. Données ap- prentissage. | 79 |
| IV.8 | Trois composantes (PCs) obtenues après l'application du PCA. Données Test. | 79 |
| IV.9 | Procédure de la seconde méthode proposée pour le diagnostic des défauts dans les roulements à bille. | 81 |
| IV.10 | Energie d'Ondelettes Maximale Relative pour diffèrent conditions de roule- ment | 83 |
| IV.11 | Application d'un prétraitement en ondelette optimal. Roulement avec un dé- faut bille de 0.1778mm (BF1). | 84 |
| IV.12 | Application d'un prétraitement en ondelette optimal. Roulement avec un dé- faut bague intérieure de 0.5334mm (IF2). | 85 |

| | | |
|-------|--|-----|
| IV.13 | Facteur DET de différents niveaux pour les sept paramètres statistiques. (a) le minimum, (b) la valeur crête à crête, (c) La valeur efficace, (d) la déviation standard, (e) le facteur de crête, (f) le facteur d'impulsion, (g) le kurtosis(Kur), (h) Tous les paramètres. | 86 |
| IV.14 | Variation du taux de test en fonction du nombre d'attributs | 87 |
| IV.15 | Convergence de la fonction objective en fonction du nombre d'itération | 88 |
| IV.16 | PPV de la deuxième approximation et SD du signal original (Données d'apprentissage). | 89 |
| IV.17 | PPV de la deuxième approximation et SD du signal original (Données de test). | 90 |
| A.1 | Perceptron Multi Couches. | 110 |

Liste des tableaux

| | | |
|------|--|----|
| I.1 | Causes de défauts dans les roulements. | 12 |
| IV.1 | Résultats de classification pour différentes PCs. | 77 |
| IV.2 | Vecteur d'entrée et résultat de classification. | 78 |
| IV.3 | Résultats de classification pour différents vecteurs d'entrés. | 80 |
| IV.4 | Résultats de simulation | 90 |
| IV.5 | Performance de classification en utilisant : SVM, BPNN and RF.. . . . | 91 |
| IV.6 | Comparaison de la méthode proposée avec d'autres méthodes récentes.. . . . | 92 |

Remerciement

Mes premiers remerciements sont adressés à mon directeur de thèse, Abdelkrim Moussaoui, Professeur à l'université de Guelma, qui m'a ouvert les portes de la recherche scientifiques, pour sa disponibilité, sa patience et surtout sa gentillesse. Il a su guider cette étude avec sagesse, tout en me laissant une grande liberté. Je le remercie sincèrement.

Je remercie mon co-directeur de thèse, Salah Chenikher, Professeur à l'université de Tébessa, qui ma donné les premières pistes d'initiation à la recherche. Je lui fais part de ma gratitude pour la très grande compétence avec laquelle il a suivi ce travail, mais également pour l'aide qu'il ma apporté à chaque étape clé. Je le remercie sincèrement pour sa disponibilité et pour ses qualités scientifiques et humaines.

Je remercie Ahcène Lemzadmi professeur au sein de l'université de Guelma, d'être intéressé à mon travail et d'avoir bien voulu présider le jury de cette thèse.

Je remercie les membres du jury de ma thèse, les professeurs : Tahar Bahi et Messaoud Ramdani pour avoir accepté d'être les rapporteurs de cette thèse, pour l'intérêt qu'ils ont porté à mon travail et pour l'avoir enrichi de leurs remarques.

Mes remerciements s'adressent aussi à Rouainia Mohamed Elhadi.

Enfin, mes sincères remerciements à toute ma famille qui m'a apporté le soutien moral pour accomplir ce travail.

Introduction generale

En raison de la complexité des systèmes de production moderne, leurs coûts d'exploitation, essentiellement liés à la maintenance, se sont vus augmentés.

Ces dernières années, la surveillance et le diagnostic des processus industriels, occupent une place de plus en plus prépondérante. Ils sont devenus un outil efficace pour réduire l'indisponibilité du système de production et minimiser les pertes économiques. Dans le contexte du diagnostic des systèmes industriels, les machines tournantes sont très souvent considérées comme étant le coeur de ces processus industriels. On peut citer par exemple, les machines tournantes comme celles existantes dans les stations de production d'électricité, les cimenteries et les complexes pétrochimiques, etc. Les machines tournantes ne peuvent être exemptes des pannes et de défaillances de plusieurs types, telles que : les défauts électriques, de balourd, d'accouplement, et des roulements. Plusieurs études statistiques ont montré que les pannes dues aux roulements sont la source majeure de la plupart des défauts mécaniques dans les machines tournantes. Selon la littérature, les pannes liées aux roulements demeurent le noyau principal des pannes totales avec plus d'un tiers soit 40% [1],[2]. Il est donc, d'un point de vue industriel et scientifique, nécessaire de focaliser les efforts de recherche sur le diagnostic des défauts dans les roulements. Récemment, plusieurs techniques ont été développées et mises en oeuvre pour surveiller l'état des roulements, on peut citer la thermographie infrarouge, la détection ultrasonore et l'analyse des vibrations mécaniques [3],[4]. L'analyse vibratoire est aujourd'hui largement utilisée, elle est devenue une piste prometteuse pour élaborer un système de diagnostic [5], [6], [7], [8].

La taxonomie des approches de diagnostic permet une classification en deux grandes catégories : les méthodes internes utilisant des modèles mathématiques et les méthodes externes.

Les travaux de cette thèse sont focalisés, sur le diagnostic des défauts dans les roulements en

se basant sur l'analyse des signaux vibratoires. Sans connaissance a priori sur le système étudié, nous nous plaçons dans une approche "sans modèle" utilisant des méthodes de reconnaissance de formes (RdF).

Cette thèse s'articule autour de quatre chapitres :

- Le premier chapitre est consacré à la recherche bibliographique sur les différentes approches de diagnostic appliquées pour le suivi de l'état du roulement. On commence le chapitre par des notions générales sur les roulements à billes ainsi que les différents types de défauts qui peuvent être survenir sur cet élément mécanique. On présente par la suite une synthèse bibliographique de plusieurs travaux de recherches dans le domaine du diagnostic des défauts dans les roulement à billes.
- Le deuxième chapitre, est consacré à la présentation des différentes approches utilisées pour la réduction de dimension en classification supervisée, qui se divise généralement en deux étapes : l'extraction des caractéristiques et la sélection des attributs pertinents. La première transforme l'information contenue dans les attributs initiales pour produire un espace de dimension plus réduit, tandis que la seconde peut être décortiquée en deux parties : un algorithme de recherche et un critère d'évaluation. Nous présenterons dans un premier temps l'analyse en ondelette ainsi que ses variantes. Dans le second temps nous parlerons des différentes techniques qui seront employées pour la réduction de la dimension à savoir : les paramètres statistiques, l'analyse en composantes principales, la technique d'évaluation de distance, l'optimisation par essaim des particules ainsi que ses variantes, et les matrices de dispersion qui seront employées comme étant une fonction objective pour la mesure de séparabilité entre les classes.
- Nous présenterons dans le troisième chapitre, la phase finale du processus de diagnostic : la classification, nous verrons notamment la classification supervisée où nous allons aborder les séparateurs à vaste marge .
- Dans le dernier chapitre, nous discuterons l'application expérimentale des deux systèmes de diagnostic proposés, sur une base de données reconnue, issue d'un banc d'essai expérimental implanté à l'université de Case Western Reserve (USA). Dans ce chapitre, nous exposerons les différentes étapes intervenantes dans la conception des deux systèmes proposés pour le diagnostic automatique de l'état des roulements. Dans un premier temps nous présenterons la première technique qui consiste à utiliser une combinaison entre l'algorithme de paquet d'ondelette et les paramètres statistiques, suivi par l'analyse en composante principale pour formuler le vecteur d'entrée. Par la suite, les séparateurs à vaste marge (SVM) optimisés par OEPCP seront appliqués pour discriminer les différentes classes.

Dans un second temps, nous décrirons la seconde approche de réduction de dimension, où une analyse multirésolution optimale est utilisée pour filtrer les signaux originaux. Ensuite, des paramètres statistiques seront extraits à partir des différentes bandes de fréquence. Par la suite, nous adopterons pour l'extraction des paramètres pertinents, une combinaison entre la technique d'évaluation des distances et l'algorithme des essaims des particules binaires en utilisant, pour la mesure de la séparabilité entre les classes, un critère basée sur les matrices de dispersion. Nous présenterons ensuite, les résultats d'application de trois techniques de classification à savoir : les réseaux de neurones (MLP), les forêts aléatoires et les séparateurs à vaste marge (SVM). Finalement, nous donnerons une comparaison de leurs performances et les résultats obtenus seront interprétés et discutés.

Chapitre I

Etat de l'art

1 Introduction

Dans ce chapitre nous allons tout d'abord présenter les caractéristiques mécaniques des roulements à billes, puis nous nous intéresserons à certaines causes de leurs défaillances. Nous verrons ensuite les caractéristiques propres aux roulements comportant une seule rangée de billes. Ces caractéristiques seront ensuite mises en relation avec leur comportement vibratoire. Nous effectuerons par la suite un panorama de différentes techniques de diagnostic, utilisées pour le diagnostic des défauts, en discutant leurs avantages et leurs inconvénients.

2 Généralités sur les roulements à billes

2.1 Constitution des roulements à billes

Les roulements sont des organes mécaniques, qui assurent le guidage et le soutien des arbres des machines tournantes. Ils représentent la source majeure de la plupart des défauts mécaniques, dans les machines tournantes. Ils sont généralement composés de quatre composants principaux à savoir : une bague intérieure, une bague extérieure, des éléments roulants et d'une cage. Selon l'application envisagée, les éléments roulants peuvent être des billes, des rouleaux ou des aiguilles. Chaque type de roulement possède des caractéristiques spécifiques en termes de charges radiales et axiales supportées, de vitesse limite ou encore de défaut d'alignement admissible.

Dans notre cas, nous nous sommes intéressés à un modèle de roulements largement utilisés, à savoir les roulements métalliques, graissés, à une rangée de billes, à contact radial, soit le type le plus utilisé dans les machines tournantes de moyenne gamme de puissance. Les roulements considérés sont donc composés d'une bague externe, d'une bague interne, d'une rangée de

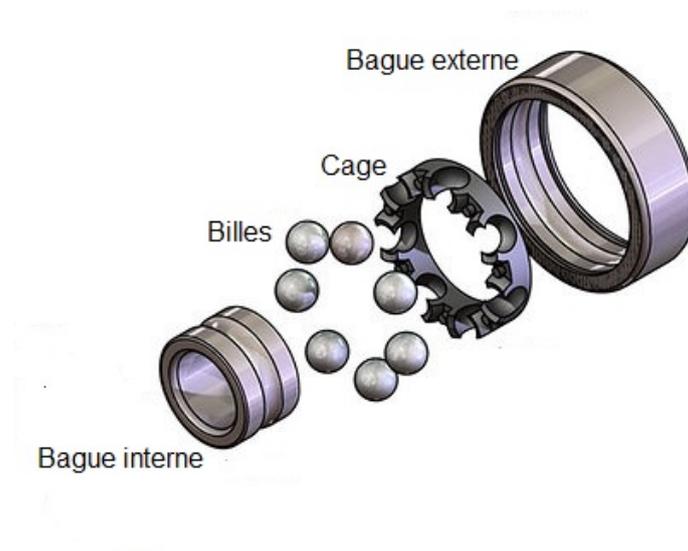


Figure I.1 – Vue éclatée des éléments constitutifs d'un roulement à une rangée de billes

billes et d'une cage, ainsi que de graisse pour la lubrification. La figure (I.1) illustre une vue éclatée de ce modèle de roulement.

2.2 Modes de dégradations d'un roulement à billes

La durée de vie d'un roulement est définie par le nombre de tours ou d'heures qu'il peut effectuer à une vitesse constante donnée, et en étant soumis à une charge spécifique, avant de montrer les premiers signes des défauts. En règle générale, une courbe de fiabilité des roulements est obtenue analytiquement à partir de la durée de fonctionnement atteinte par 90% des roulements d'un même type [9]. Cependant, ce critère de calcul de la durée de vie des roulements ne prend en considération que les phénomènes liés aux matériaux qui le compose. D'une façon générale, la dégradation d'un roulement ne peut être facilement modélisé ou décrite théoriquement, et ce car le roulement obéit à différents facteurs complexe et interdépendant. En effet, un facteur de dégradation n'intervient jamais indépendamment, mais un ensemble de facteur provoque la dégradation d'un roulement, et par conséquent la diminution de sa durée de vie. Plusieurs facteurs externes aux roulements, agissant sur leurs durées de vie, sont répertoriés. Nous pouvons citer les principaux parmi eux, à savoir : l'exercice d'une charge mécanique trop importante, les défauts liés au montage incorrect, la corrosion, la présence de particules étrangères au sein du roulement et la qualité de la lubrification ou encore la diminution de la dureté à cause d'une température trop importante.

Nous allons détailler, par la suite, ces facteurs ainsi que leurs effets.



Figure I.2 – Aspect de la bague interne d'un roulement surchargé

2.2.1 Effets des contraintes mécaniques

Après un certain temps de fonctionnement, l'état d'un roulement soumis à des efforts mécaniques se dégrade en raison de la fatigue du matériau. Si un roulement est également surchargé de manière destructrice, après un temps de fonctionnement plus court, il s'arrêtera également de fonctionner. La fissure se propage vers les bagues ou les billes tant que la surcharge se poursuit, jusqu'à ce qu'un morceau de métal se casse en laissant une fosse dans la zone de contact. La fatigue du roulement se développe plus rapidement si ce dernier est toujours surchargé, ou fonctionne avec une vitesse excessive. Ces conditions réduisent considérablement la durée de vie des roulements. Les figures (I.2;I.3) montrent les dégradations sévères que peuvent provoquer les surcharges mécaniques [10].

2.2.2 Défauts de montage

Lorsqu'un roulement est placé sur l'arbre de la machine d'un côté, et dans son logement au niveau des paliers statoriques dans l'autre côté, il obéit à des contraintes fortes sur les bagues à cause d'une mauvaise manipulation lors du montage (désalignement). Les mêmes effets que ceux observés précédemment se reproduisent, voir figure (I.4). Lorsque les contraintes exercées sont très localisées, des défauts graves allant jusqu'à la cassure des bagues seront constatés, voir figure (I.5) [10].



Figure I.3 – Ecaillage d'une piste de roulement

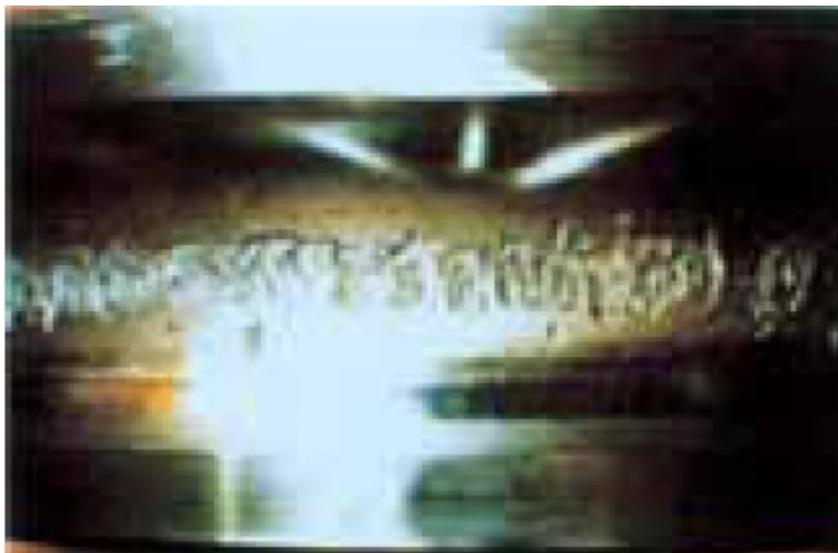


Figure I.4 – Ecaillage d'une piste



Figure I.5 – Cassure d'une bague

2.2.3 Corrosion

Les trous de rouille causés par la corrosion sur un élément roulant entraînent un bruit excessif pendant le fonctionnement. La rouille se génère lorsque le roulement est exposé à des agents corrosifs tels que l'eau et l'acide, ou lorsque les conditions de stockage sont incorrectes. Des criques seront observés sur les bagues et les billes, comme le montre la figure (I.6) [10].

2.2.4 Contamination du roulement

La contamination du roulement est une cause fréquente de dommages. La contamination du roulement provient principalement de la poussière et des particules étrangères entrant dans le roulement à travers un joint d'étanchéité inadéquat, ou une lubrification contaminée, ce qui entraîne une augmentation du frottement entre les contacts métalliques, et les changements de la géométrie des bagues interne et externe. Le roulement exposé aux dommages causés par l'usure, se détériore progressivement, entraînant une perte de dimensions, des fissures, voire un écaillage des constituants du roulement, voir figure(I.7)[10].

La figure (I.8) résume le mécanisme de la contamination du roulement.

2.2.5 Défaut de lubrification

Le lubrifiant est un organe essentiel des roulements, utilisé pour éviter un contact métallique direct entre éléments roulants, chemins de roulement et cage, permettant d'empêcher l'usure et de protéger les surfaces de la corrosion. Deux types de défauts de lubrification peuvent



Figure I.6 – Criques de corrosion sur une piste de roulement



Figure I.7 – Marques d'indentation d'un chemin de roulement dues à la présence de contaminant

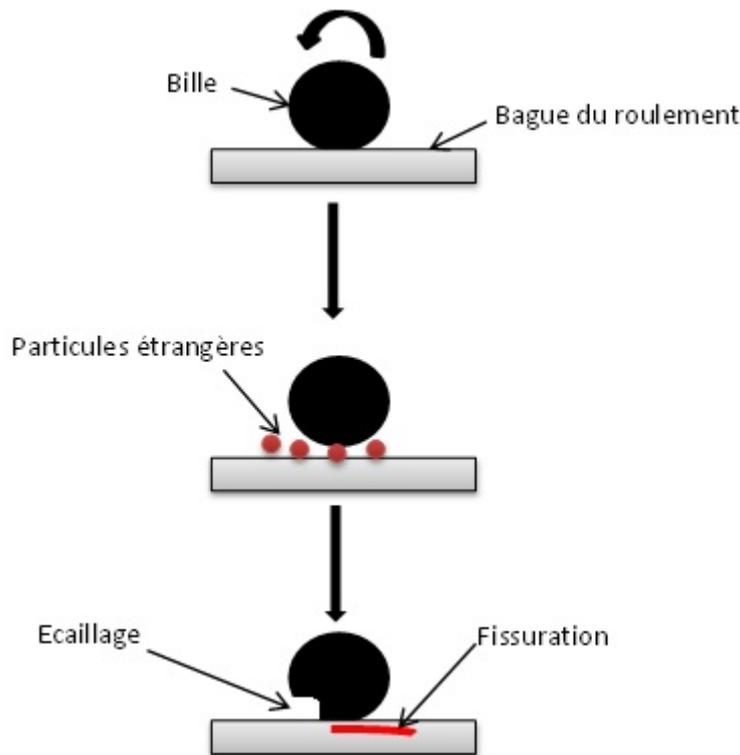


Figure I.8 – Mécanismes généraux d'usure en présence de contaminant

survenir. Le premier est une diminution du pouvoir lubrifiant d'une graisse ou d'une huile dans le temps, ou bien une absence totale de lubrifiant. En effet, l'absence de lubrifiant permet aux différents éléments métalliques d'être en contact direct, ce qui entraîne des écaillages des pistes, des éclatements des billes, un blocage et donc une destruction du roulement, comme le montre les figures (I.10), (I.11) et I.9 [10]. Le second défaut survient sur les roulements dans le cas d'un excès de lubrifiant qui entraîne une augmentation rapide de la température de fonctionnement. Dans ce cas, la diminution de l'espace libre dans le palier empêche la rotation libre des billes. Par conséquent les surfaces de contact seront rapidement dégradées.

2.2.6 Température de fonctionnement

Généralement, un roulement doit fonctionner dans une plage de température dont les valeurs varient entre $-20^{\circ}C$, $+120^{\circ}C$. Les dimensions d'un roulement changent sous l'influence de sa température de fonctionnement. En effet, les éléments métalliques se déforment en fonction de la température, par conséquent, le jeu naturel du roulement sera diminué et l'insertion du lubrifiant au niveau des points de contacts bille/bague devient très difficile, ce qui conduit à l'apparition des mêmes défauts précédemment cités.. Par ailleurs, la graisse est composée d'une



Figure I.9 – Blocage de roulement

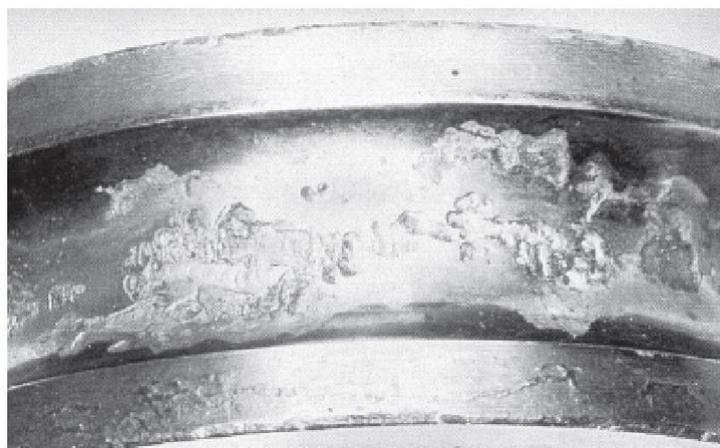


Figure I.10 – Ecaillage d'un chemin de roulement

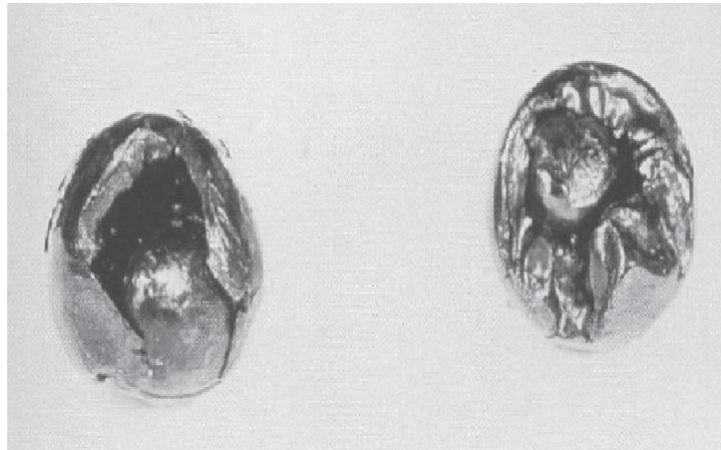


Figure I.11 – Eclatement de billes

huile lubrifiante et d'un savon liant [11]. Lorsque la température de fonctionnement augmente, le savon se solidifie, la graisse perd ses propriétés lubrifiantes et la rotation de la cage est perturbée, les billes glissent alors dans les chemins de roulement provoquant une usure rapide et prononcée.

D'une manière générale, ces causes sont récapitulées dans le tableau (I.1) .

| Causes de défauts dans les roulements | |
|---|-----|
| Cause du Défaut | % |
| Lubrification inadaptée (choix des lubrifiants, quantité, fréquence de Lubrification) | 70% |
| Pollution (poussière, détergents, liquides, etc.) | 18% |
| Montage incorrect (brutal, jeu, désalignement, échauffement excessif, etc.) | 10% |
| Fatigue (charge, contraintes) | 2 % |

Tableau I.1 – Causes de défauts dans les roulements.

3 Influence des défauts sur la structure des signaux

Lorsqu'un défaut localisé apparaît en un point dans l'un des éléments du roulement, cela entraîne une vibration à la fréquence caractéristique associée à l'élément en défaut. En effet, si un défaut se situe sur le chemin de roulement d'une des bagues ou sur une bille, il se produit un choc mécanique, à chaque fois que le défaut est confondu avec un point de contact bille/bague. Ces chocs répétés se traduisent alors par des vibrations. De manière simplifiée,

I.4 Fréquences caractéristiques des défauts du roulement «calcul cinématique»

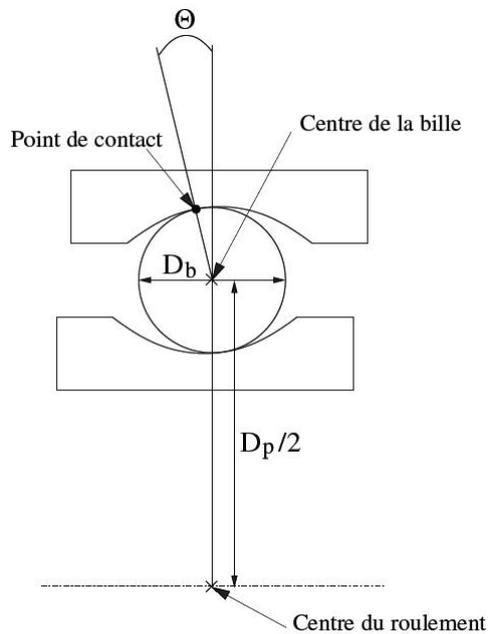


Figure I.12 – Représentation schématique de la géométrie interne d'un roulement à une rangée de billes

cette vibration peut être modélisée par un signal périodique, dont la période fait intervenir la période caractéristique de l'élément en défaut.

4 Fréquences caractéristiques des défauts du roulement «calcul cinématique»

Le roulement est l'un des éléments essentiels d'une machine tournante. En effet, chargé cycliquement, les pistes de roulement se dégradent par fatigue, conduisant aux écaillages puis à la ruine du matériau. Lors de la mise en rotation, un train d'impulsion est généré par le défaut, à une fréquence bien définie que l'on appelle «fréquence caractéristique» de défaut du roulement. Pour chaque type de roulement et en fonction de ses cotes de fabrication, On peut considérer quatre fréquences caractéristiques :

- La fréquence de passage d'un élément roulant sur un défaut de bague extérieure (Ball Passind Frequency Outer Race [**BPFO**]) :

$$BPFO = \frac{n \times N}{2} \left(1 - \frac{D_b}{D_p} \cos\theta \right) \quad (I.1)$$

- La fréquence de passage d'un élément roulant sur un défaut de bague intérieure (Ball Passind Frequency Inner Race [**BPMI**])

$$BPMI = \frac{n \times N}{2} \left(1 + \frac{D_b}{D_p} \cos\theta \right) \quad (I.2)$$

- La fréquence de passage d'un défaut de cage (Cage Frequency [**CF**]) :

$$CF = \frac{N}{2} \left(1 - \frac{D_b}{D_p} \cos\theta \right) \quad (I.3)$$

- La fréquence de passage d'un défaut de bille (Ball Frequency [**BF**]) :

$$BF = \frac{N D_p}{2 D_b} \left(1 - \left(\frac{D_b}{D_p} \cos\theta \right)^2 \right) \quad (I.4)$$

avec :

N : La fréquence de rotation en [HZ] ;

n : Le nombre de billes ;

D_p : Le diamètre moyen en [mm] ; D_b : Le diamètre de la bille en [mm] ;

θ : L'angle de contact en [degré]. (Voir figure (I.12)).

5 Synthèse bibliographique

Afin de mieux situer notre travail, il a été nécessaire de rechercher qu'elles sont les méthodes employées pour détecter et diagnostiquer une anomalie survenue sur les roulements à billes. Dans cette partie, nous allons décrire les techniques et les méthodes les plus récentes, pour le diagnostic des défauts dans les roulements à billes en précisant leurs avantages et leurs inconvénients.

Le domaine du diagnostic est très vaste, l'objectif n'est pas de faire une synthèse exhaustive mais de discuter les principales possibilités existantes pour concevoir un système de diagnostic. Plusieurs approches sont suggérées dans la littérature pour surveiller l'état de fonctionnement d'un roulement. Elles peuvent être classées comme suit :

1. **les approches basées sur des modèles analytiques ou mathématiques qui peuvent**

être issues principalement du domaine de traitement du signal.

Plusieurs travaux ont été réalisés pour le développement de diverses approches de diagnostic de défaut des roulements à bille par analyse vibratoire. La méthode de l'analyse d'enveloppe est une technique qui repose sur l'étude de l'information obtenu à travers un filtrage autour d'une ou plusieurs fréquence de résonances suivi par la transformée de Hilbert et ce afin de ramener le signal en basses fréquences [12].

Le problème de l'analyse d'enveloppe réside dans la nécessité de la connaissance préalable des fréquences de résonances ainsi que dans le choix du type du filtre. Afin de surmonter ces lacunes, une nouvelle méthode basée sur le traitement en ondelette a été développée [13]. Cette approche utilise l'analyse multirésolution en ondelette comme étant un outil puissant de filtrage, ensuite la bande de fréquence la plus appropriée est sélectionnée, tout en se basant sur le kurtosis. Cette méthode s'est avérée plus robuste, mais elle est très dépendante du bon choix de l'ondelette mère.

2. les méthodes basées sur la reconnaissance des formes

Les approches de diagnostic par reconnaissance des formes sont basées sur deux étapes principales : l'extraction d'un vecteur des attributs, appelé vecteur de forme, et le choix des règles de détection qui permettent la classification des observations.

Pour la première phase de nombreux travaux de recherches ont été réalisés afin d'extraire le vecteur forme le plus approprié.

S. Fang et W. Zijie ont présenté une nouvelle méthode basé sur l'analyse en paquet d'ondelette et les réseaux de neurone de type RBF pour le diagnostic des défauts dans les roulements à billes [14], dans cette méthode, les résultats d'application de deux types d'ondelette mère ont été comparés à savoir l'ondelette db8 et celle de sym8, les résultats démontrent que l'ondelette sym8 donne des meilleures performances. Malheureusement dans ce dernier travail seulement deux ondelettes mères ont été considérées, alors que d'autres types d'ondelettes peuvent offrir de meilleurs résultats. Le niveau de décomposition optimal n'est pas aussi justifié.

S. Fu, K. Liu, Y. Xu, et Y. Liu faisant une étude pour la surveillance de l'état du roulement basée sur les paramètres statistiques et ce afin de détecter des défauts naissants [15]. Les résultats obtenus montrent que cette méthode offre de bonnes performances. Néanmoins les paramètres statistiques sont très sensibles au bruit dû au système d'acquisition.

Plusieurs études ont montré que l'utilisation d'un vecteur forme combiné peut aider à une meilleure représentation. Malheureusement l'utilisation de cette approche conduit à une augmentation de la dimension du vecteur d'entrée et par conséquent la phase de classifi-

cation devient plus complexe.

Pour surmonter ces lacunes, certains attributs qui définissent clairement l'état du roulement devront être sélectionnés. En général, les méthodes de sélection des paramètres sont classées en deux approches principales : les méthodes de filtrage, dans lesquelles les attributs pertinents sont obtenus en fonction des critères d'évaluation indépendamment du système de classification, et des méthodes d'emballage (wrapper), dans lesquelles les paramètres pertinents sont sélectionnés à l'aide d'un retour du classifieur.

Plusieurs méthodes de filtrage ont été utilisées, telles que : l'analyse en composantes indépendantes, l'analyse en composante principale et la technique d'évaluation de distance (DET).

M. Saimurugan et K.I. Ramachandran ont fait une étude pour le diagnostic des défauts dans les machines tournantes [16], cette étude basée sur l'extraction de douze (12) paramètres statistiques, la réduction de dimension est ensuite effectuée par l'analyse en composante indépendante. Par conséquent le vecteur forme de chaque état est réduit à sept (07) composante indépendante.

T. Thelaidjia, A. Moussaoui et S.Chenikher ont proposé une nouvelle approche dans laquelle les vecteurs formes sont extraites à l'aide des paramètres statistiques combinés avec un vecteur d'énergie obtenu grâce à l'application de la transformée en ondelettes discrètes au cinquième niveau de décomposition. Après l'extraction des paramètres, ICA est utilisé pour réduire la dimension du vecteur d'entrée de treize (13) paramètre à cinq (5) composantes indépendantes [17].

L. Saidi , J. Ben Ali et F. Fnaiech , ont fait une étude expérimentale qui aide à diagnostiquer les défauts dans les roulements [18]. L'analyse bi-spectre de vibration (spectre de troisième ordre) est employée pour l'extraction du paramètre, qui sont soumis par la suite à une analyse en composante principale pour la réduction de la dimension.

T. Thelaidjia, A. Moussaoui et S.Chenikher faisant une étude pour réduire la dimension du vecteur d'entrée [19], dans cette étude un vecteur global contenant quinze paramètres (15) a été construit. Après l'application de l'analyse en composante principale un vecteur forme contenant seulement les (03) trois composantes principales est obtenu.

PCA et ICA sont souvent utilisés pour la réduction de dimension, mais si les paramètres du vecteur forme original obéissent à une relation non linéaire complexe, les méthodes de réduction avec noyaux telle que KPCA et KICA sont plus adapté [20]. Ces méthodes sont complexes et difficiles à leurs mise en oeuvre.

DET est bien établi pour la sélection des caractéristiques exceptionnelles. Il est efficace et

facile à réaliser [21]; [22]. À l'origine, il est suggéré par **Yang et al** [23]. Plus tard, il a été développé par **Lei et al** [24]. DET sélectionne un sous-espace de l'espace caractéristique en fonction du calcul des facteurs d'évaluation, de sorte que toutes les fonctionnalités sont disposées par ordre décroissant en fonction de leurs facteurs d'évaluation et les 'D' premières attributs peuvent être sélectionnés comme entrée du classificateur.

Y. Lei, Z. Hea, Y. Zia et Q. Hua ont proposé une nouvelle approche pour le diagnostic des défauts [24], cette approche est basée sur la combinaison de certaines techniques de prétraitement, comme le filtrage, la démodulation et l'EMD, pour l'extraction du vecteur forme globale. Six ensembles de fonctionnalités, y compris les caractéristiques statistiques du domaine temporel et du domaine fréquentiel ont été obtenus. Finalement, La technique d'évaluation de distance améliorée est proposée et, c'est avec elle, que six attributs parmi les six ensembles sont sélectionnés.

C. Shen, D. Wang , F. Kong et P. W. Tse ont présenté un nouveau schéma de diagnostic de défaut basé sur l'extraction de paramètres statistiques à partir des coefficients obtenus par l'application de la transformée en paquet d'ondelettes, la technique d'évaluation de distance (DET) est appliquée par la suite pour la sélection des paramètres les plus pertinents [21].

La technique d'évaluation des distances (DET) est bien établie pour la sélection des attributs pertinents. Malheureusement, DET souffre de deux inconvénients principaux : le premier inconvénient est lié au fait que son efficacité est influencée par le bruit dû au système d'acquisition de données, le deuxième inconvénient de DET est que les paramètres pertinents sont sélectionnés indépendamment du système de classification, en d'autres termes, le sous-ensemble de 'D' dimension qui contient les attributs dont les facteurs DET sont les plus grands, n'est pas nécessairement le meilleur sous-ensemble de 'D' dimension dans lequel les classes sont bien séparées.

Dans la méthode de l'emballage, plusieurs travaux ont été suggérés pour la sélection des attributs tels que les algorithmes génétiques (GA) et l'optimisation par essaims des particules (PSO) .

I. Rashedul, A. K. Sheraz et K. Jong-Myon dans leur étude ont proposé une méthode de suivi et de diagnostic des défauts dans les roulements [25]. Dans cette étude le vecteur forme original est un ensemble de paramètres statistiques hybrides calculées à partir de l'analyse temporelle, de l'analyse du domaine fréquentiel et de l'analyse du spectre d'enveloppe d'une émission acoustique. Les algorithmes génétiques sont par la suite utilisés pour sélectionner l'ensemble des attributs optimaux.

R. Ziani, A. Felkaoui et R. Zegadi ont développé une nouvel méthode basée sur

l'extraction des paramètres dans le domaine temporel, le domaine spectral et le domaine temporel-spectral. L'optimisation par essais des particules binaire est employée ensuite pour la sélection des attributs pertinents [26].

Contrairement à GA et à d'autres algorithmes évolutifs, BPSO est efficace et facile à mettre en oeuvre. Néanmoins, il présente l'inconvénient qu'il ne peut pas converger correctement. Pour surmonter cette lacune, un nouveau PSO discret binaire (NBPSO) a été développé. Il a été testé sur plusieurs problèmes d'optimisation des fonctions et a donné de meilleurs résultats que ceux obtenus par BPSO [27].

Malgré l'efficacité des méthodes de l'emballage, lorsque la taille du vecteur forme original est assez grande ces méthodes risquent d'être complexes en terme de temps de calcul et la convergence n'est pas garantie (optimum local).

Pour la phase de classification, différentes méthodes sont utilisées pour le diagnostic des défauts du roulement. Nous citons celles les plus fréquemment utilisés à savoir : les RNA basés sur l'algorithme de rétro propagation, les RBF (Radial Basis Function), ANFIS (Système d'Inférence Neuro Fuzzy Adaptive) ou (Adaptive Neural Fuzzy Inference System) et les séparateurs à vaste marge.

R. Zegadi, A. Felkaoui et M. Djouada dans leur article sur le diagnostic de défauts des roulements, ont proposé d'utiliser un réseau de neurone artificiel dont les poids sont optimisés par l'algorithme de rétro propagation et ce pour classer cinq type de défauts [28].

Malgré qu'il offre des résultats satisfaisants, le réseaux de neurones souffre de plusieurs lacunes à savoir : le choix de la structure et le choix de l'algorithme de mise à jour des poids.

Pour résoudre la première lacune, **M. Unal, M. Onat, M. Demetgul et H. Kucuk** ont suggéré d'employer les AG pour le choix de la structure en optimisant le nombre de couches cachées ainsi que le nombre de neurones dans chaque couche, mais le problème de la mise à jour des poids de réseau de neurones persiste [29].

J. Cheng, L Ai et W Xiong on développé un nouveau RNA dont les poids sont optimisés par l'application de l'algorithme de l'essaim de particule afin d'augmenter l'efficacité de la convergence du réseau [30].

Le réseau de neurone de type RBF, se comporte mieux que les RNA basés sur l'algorithme de rétro propagation dans : la capacité d'approximation, la capacité de classification et la vitesse d'apprentissage.

T. Zhou, X. Zhu, W. Peng et Y. Liu ont employé dans leur étude le réseau de neurone de type RBF (RNRBF) pour discriminer quatre états de roulements. Les résultats

obtenus montrent que le RNRBF offre des résultats satisfaisants [31].

Dans les RNRBF, la détermination du nombre de neurones dans la couche cachée est très importante, car elle affecte la complexité et la capacité de généralisation du réseau. Si le nombre de neurones dans la couche cachée est insuffisant, le RNRBF ne peut pas apprendre les données d'une manière adéquate; d'autre part, si le nombre de neurones est trop élevé, une généralisation médiocre ou une situation de sur-apprentissage peut se produire [32]. La position des centres dans la couche cachée affecte aussi considérablement les performances du réseau [33].

Pour surmonter ses lacunes, **Z. Wang, Z. Liu, X. Lan, J. Liu, S. Wang, Y. Wu et Y. Xue** ont développé une version améliorée du PSO pour l'optimisation de la structure de RNRBF en choisissant judicieusement : le nombre de noeuds cachés, le centre et la largeur correspondants à chaque noeud [34].

I. Attoui, N. Boutasseta, N. Fergani, B. Oudjani et A. Deliou ont fait une étude pour le diagnostic de différents défauts de roulement, cette étude est basée sur l'utilisation d'un système d'inférence neuro-flou Adaptative (ANFIS) pour automatiser le processus de détection et de diagnostic des pannes [35].

les principaux inconvénients de la technique ANFIS sont : sa complexité, un grand temps de calcul et le choix optimal des règles floues.

Le séparateur à vaste marge (SVM) est une nouvelle méthode d'apprentissage automatique introduite par Vapnik sur la base de la théorie de l'apprentissage statistique. Depuis le milieu des années 1990, les algorithmes utilisés pour SVM ont commencé à émerger. La principale différence entre le domaine connu du réseau neuronal artificiel (RNA) et les SVM réside dans le principe de minimisation des risques (MR). Dans le cas de SVM, le principe de minimisation des risques structurels est utilisé pour minimiser une limite supérieure du risque, alors qu'en réseau de neurone artificiel (RNA), la minimisation d'un risque empirique est utilisée pour minimiser l'erreur sur les données d'apprentissage. La différence dans le principe de minimisation du risque conduit à de meilleures performances de généralisation pour les SVM que pour les RNA. Selon la littérature, le SVM a été appliqué avec succès pour de nombreuses applications, telles que l'identification de modèle, la régression, l'approximation des fonctions [36], [37]. Les résultats donnent la preuve que la technique des SVM est non seulement très satisfaisante d'un point de vue théorique, mais qu'elle peut aussi conduire à de hautes performances dans des applications pratiques [38], [39].

G. Xian dans son étude a proposé une méthode de suivi et de diagnostic des défauts dans les roulements à billes d'une machine de moulage hydraulique par injection, et ce

en appliquant les séparateurs à vaste marge pour la classification de différent types des défauts mécaniques du roulement [40].

Plusieurs études ont montré que la capacité de généralisation des SVM dépend essentiellement du bon choix des deux paramètres, à savoir celui de la fonction noyau ainsi que celui de la régularisation [40].

Ces dernières années, plusieurs algorithmes intelligents basés sur l'étude des comportements biologiques naturels ont été proposés, comme les algorithmes génétiques (GA) et l'optimisation par essaim de particules (PSO) [41], [42]. Par rapport à d'autres algorithmes, l'algorithme du PSO est très simple, facile à mettre en oeuvre et il se caractérise par une convergence rapide.

Bien que performant et facile à mettre en oeuvre, l'algorithme du PSO souffre principalement de deux lacunes à savoir : la diminution de sa performance quand le nombre d'itération augmente et dans le cas de traitement des problèmes complexes il risque de converger vers des solutions optimales locales . Pour surmonter ces lacunes, plusieurs variantes ont été développées telles que OEP avec un Facteur de Constriction (OEPFC) et OEP avec Congrégation Passive (OEPCP) ou Particle Swarm Optimizer with Passive Congregation (PSOPC) en anglais [43], [44] .

L'OEPCP a été appliqué avec succès dans de nombreux domaines, il donne de meilleurs résultats que ceux générés par d'autres variantes de PSO [44] .

6 Conclusion

Dans ce chapitre, nous avons présenté dans un premier temps les roulements à billes, leurs processus de dégradation ainsi que les effets vibratoires associés. Ce qui nous a permis de souligner l'importance de la mise en place d'un système de surveillance pour le contrôle et le suivi de l'état du roulement.

Dans un second temps, nous avons présenté un état de l'art des différentes approches utilisées pour le diagnostic des défauts dans les roulements à billes. Après l'étude des différentes approches de diagnostic traitées dans la littérature, il s'est avéré que les méthodes de diagnostic par reconnaissance des formes offrent des résultats encourageants. Elles se composent généralement de trois étapes essentielles à savoir : l'extraction des paramètres, la sélection des attributs et la classification.

Pour la conception d'un système de diagnostic puissant et efficace, les techniques appliquées dans chaque étape doivent être judicieusement choisies.

Chapitre II

Extraction des paramètres et Réduction de dimension pour les problèmes de classification

1 Introduction

lors de l'exploration et l'analyse des données, on est confronté aux phénomènes liés à la grande dimension des données appelé **Fléau de dimensionnalité**. En effet, un grand nombre de variables affecte gravement les performances de la classification supervisée [45].

Ce problème provient du fait qu'un grand nombre d'observations exponentiellement liées à la dimension, est nécessaire pour obtenir un discriminateur avec de bonnes performances [45].

Nous traitons, dans ce chapitre, le problème d'extraction des paramètres, ainsi que celui de la réduction de dimension pour les problèmes de classification supervisée, en présentant les approches employées pour la formulation du vecteur forme.

Ce chapitre est divisé en deux parties, la première est consacrée à l'extraction des paramètres : dans cette partie, nous allons aborder, dans un premier temps, la transformée en ondelette comme étant un outil puissant de filtrage, les paramètres statistiques seront discutées dans un deuxième temps.

Dans la deuxième partie, nous discuterons les techniques de réduction de dimension telles que : l'algorithme des essaims de particules binaires, l'analyse en composantes principales et la technique d'évaluation des distances.

Finalement, le critère basé sur les matrices de dispersion sera présenté.

Première partie

Extraction des paramètres

1 Les ondelettes

1.1 Historique de traitement du signal : de Fourier à Gabor.

Le traitement en ondelettes a été mis en place il y a presque 60 ans pour surmonter une lacune entre deux méthodes de représentation d'un signal : la première consiste à représenter une fonction par son graphe classique de \mathbb{R} dans \mathbb{R} , ce qui signifie une décomposition sur la base continue des distributions de Dirac comme suit :

$$f(t) = \int_{\mathbb{R}} f(u)\delta(t - u)du \quad (\text{II.1})$$

Alors que la deuxième consiste à représenter le signal dans l'espace fréquentiel [46] :

$$f(t) = \int_{\mathbb{R}} \hat{f}(\omega)e^{j\omega t}d\omega \quad (\text{II.2})$$

avec $\hat{f}(\omega)$ représente la transformée de Fourier de $f(t)$. La première représentation offre une information temporelle précise : la quantité $f(t)$ représente la valeur du signal à l'instant t . Cependant, l'information fréquentielle est absente. En effet, on ne peut connaître aucune information sur le «contenu fréquentiel» du signal à partir de cette représentation. Par contre, la représentation de Fourier offre une information très précise sur le comportement fréquentiel, sans l'existence d'aucune information temporelle. La question qui doit être posée est la suivante : existe-t-il une représentation mixte qui permet d'obtenir une représentation fréquentielle tout en conservant la notion du temps.?, Morlet et Gabor ont introduit des nouvelles fonctions de base qui permettant l'obtention simultanée d'une bonne localisation fréquentielle et spatiale.

Une limite théorique dans cette perspective est bien connue : c'est l'inégalité d' Heisenberg [47]. Considérons une fonction de base f , que l'on suppose de norme L_2 égale à 1 :

$$\int_{\mathbb{R}} |f(t)|^2 dt = 1 \quad (\text{II.3})$$

Le centre $c(f)$ et la largeur $\Delta(f)$ d'une telle fonction sont définis comme suit :

$$c(f) = \int_{\mathbb{R}} t |f(t)|^2 dt \quad (\text{II.4})$$

$$\Delta(f) = \int_{\mathbb{R}} \sqrt{(t - c(f))^2 |f(t)|^2} dt \quad (\text{II.5})$$

Le théorème de l'inégalité d' Heisenberg est une inégalité principale qui s'écrit pour toute fonction f de norme 1 [47],[48] comme suit :

$$\Delta(f)\Delta(\hat{f}) \geq \frac{1}{2} \quad (\text{II.6})$$

Si on note : largeur fréquentielle d'une fonction f et la largeur de sa transformée de Fourier \hat{f} , cette inégalité ne permet pas donc d'avoir une fonction avec des largeurs temporelle et fréquentielle toutes deux aussi petites que l'on veut. On connaît également les fonctions optimales qui réalisent les limites imposées par cette inégalité : ce sont les fonctions proposées par Gabor, c'est-à-dire le signal gaussien translaté et modulé :

$$Ae^{-(t-t_0)^2/2\Delta t^2} e^{j\omega_0 t} \quad (\text{II.7})$$

Où A dénote un paramètre de normalisation tel que la fonction ait une norme L_2 égale à 1, qui dépend de de la résolution temporelle Δt . Pour de telles fonctions, les limites imposées par l'inégalité d' Heisenberg seront vérifiées (c'est-à-dire l'inégalité d' Heisenberg devient une égalité). Ces fonctions ont été nommée ondelettes de Gabor [47].

1.2 Représentation temps-fréquence et temps-échelle.

A une telle fonction, on associe un pavé temps-fréquence, c'est-à-dire une fenêtre rectangulaire dans le plan (t, ω) dont le centre est donné par $(c(f), c(\hat{f}))$ et dont la dimension est de $(\Delta(f) \times \Delta(\hat{f}))$.

Ce pavé est une représentation Intuitive de la couverture en temps et en fréquence d'une fonction. On associe également à une base un pavage du plan temps-fréquence, qui est un recouvrement du plan (t, ω) par des rectangles de couverture des fonctions de base. Si le centre des boites d' Heisenberg est fixé comme le point du centre spatial et fréquentiel de la fonction, leurs dimensions sont en général choisies de telle manière que les boites forment une partition du plan temps-fréquence. Cette représentation a un aspect arbitraire, d'autant qu'aucun résultat ne lie le fait qu'une famille soit une base au fait que les pavés temps-fréquence de la famille recouvrent le plan.

Le pavage temps-fréquence correspondant par exemple aux bases de fonctions de Dirac et de fonctions de Fourier sont des pavages par des rectangles infiniment fins et allongés qui sont représentées schématiquement dans la figure (II.1).

Pour représenter un signal comme combinaison de telles fonctions, il est très redondant de laisser varier indépendamment les trois paramètres t_0 , ω_0 et Δt . Deux approches différentes

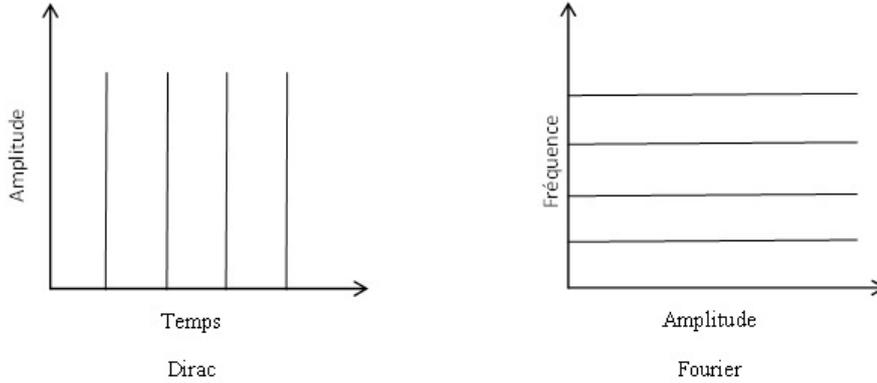


Figure II.1 – Pavages temps fréquences associés aux représentations de Dirac et de Fourier

ont prévalu :

1. La première approche est l'approche temps -fréquence, dans laquelle la largeur spatiale Δt des fonctions g est indépendante de la fréquence. Les fonctions s'écrivent sous la forme :

$$g_{t_0, \omega_0}(t) = e^{j\omega_0 t} g_0(t - t_0) \text{ où } g_0(t) = A_0 e^{-t^2/2\Delta t^2} \quad (\text{II.8})$$

Ce mode d'approximation est également appelé analyse de Fourier à fenêtre [48].

2. Une deuxième approche est l'approche temps-échelle, dans laquelle la largeur spatiale des fonctions g est inversement proportionnelle à la fréquence (le produit $\omega_0 \Delta t$ est constant et égal à c). On obtient alors, à un facteur de module 1 près, une forme relativement simple pour les différentes ondelettes :

$$g_{t_0, \Delta t}(t) = \frac{1}{\sqrt{\Delta t}} g_0\left(\frac{t-t_0}{\Delta t}\right) \quad (\text{II.9})$$

Où $g_0(t) = A_0 e^{-t^2/2\Delta t^2} e^{ict}$

Les pavages du plan temps-fréquence donnent une idée de la différence des deux approches. Dans le cas de l'analyse temps -fréquence, le pavage obtenu est un pavage par des Domaines rectangulaires qui se déduisent les uns des autres par translation dans le plan temps-fréquence. Dans le cas de l'analyse temps-échelle, les domaines sont également de surface constante, mais ont une résolution fréquentielle relative $\Delta\omega/\omega_0$ constante.

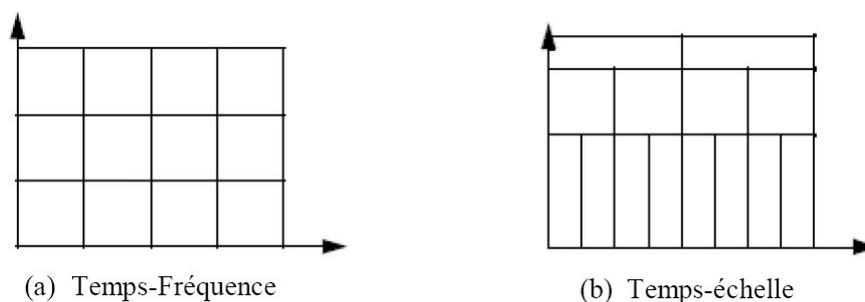


Figure II.2 – Pavages du plan temps fréquence pour la représentation temps-fréquence et la représentation temps-échelle

1.3 Introduction aux ondelettes

Dans la pratique, la plupart des signaux ne sont pas stationnaires, et c'est justement dans l'évolution de leurs caractéristiques (statistiques, fréquentielles, temporelles, spatiales) que réside l'essentiel de l'information qu'ils contiennent. Or l'analyse de Fourier propose une approche globale du signal, les intégrations sont faites de moins l'infini à plus l'infini, et toute notion de localisation temporelle disparaît dans l'espace de Fourier ; il faut donc trouver un compromis, une transformation qui renseigne sur le contenu fréquentiel tout en préservant la localisation afin d'obtenir une représentation temps/fréquence ou espace/échelle du signal.

Dans cette partie, nous présentons les fonctions ondelettes, la transformée en ondelettes et l'analyse multirésolution.

Le terme ondelette désigne une fonction qui oscille pendant un « temps donné » (si la variable est le temps) ou sur un intervalle de longueur fini (si la variable est de type spatial), au delà, la fonction décroît très vite vers zéro.

Historiquement, les premières ondelettes (introduites par Haar dans les années 1930) qui constituaient également une base de fonctions orthogonales. Les ondelettes de Haar présentent la particularité de ne pas être dérivables. Plus récemment, de nouvelles fonctions ondelettes ont été introduites [49],[50], qui constituent également une base de fonctions orthogonales, et qui, de plus, sont dérivables. Elles ont été notamment mises en oeuvre dans le cadre de l'analyse multirésolution de signaux [51].

1.4 La transformée en ondelette

D'une manière semblable à la théorie d'analyse par la transformée de Fourier fenêtrée, les ondelettes sont des fonctions élémentaires principalement employées pour la décomposition d'un signal, voir figure II.3. Le traitement d'une fonction par les ondelettes consiste à l'écrire comme

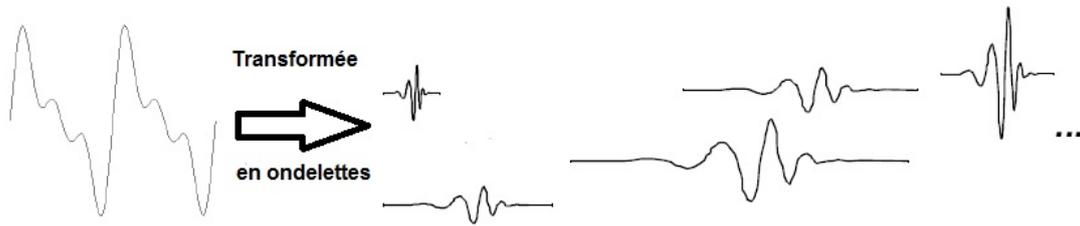


Figure II.3 – La transformée en ondelette

une somme pondérée de fonctions obtenues à partir de translations et de dilatations réalisées sur une fonction élémentaire appelée ondelette mère. On distingue deux types de transformée en ondelettes à savoir : la transformée en ondelettes continue et la transformée en ondelettes discrète. Dans la première, les facteurs de translations et de dilatations sont choisis de manière continue, alors qu'ils sont choisis d'une façon discrète dans la seconde.

1.4.1 La transformée en ondelettes continue.

La transformée en ondelettes continue permet d'analyser des signaux en utilisant des paramètres de translations et de dilatations qui peuvent varier d'une façon continue sur l'ensemble des réels \mathbb{R} (à noter que : les dilatations doivent avoir des valeurs positives).

Une ondelette mère ψ est une fonction de élémentaire qui sera translatée et dilatée pour recouvrir le plan temps-fréquences et permettre l'analyser le signal.

L'ondelette mère est une fonction de moyenne nulle qui doit vérifier la condition suivante : [52],[48],[47] :

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \quad (\text{II.10})$$

On introduit les paramètres de translation u et celui d'échelle s , on obtient une famille de fonctions définie comme suit [48],[47] :

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \quad (\text{II.11})$$

avec $u \in \mathbb{R}$ et $s \in \mathbb{R}^{*+}$.

On construit alors une famille d'ondelettes Ω engendrée à partir de l'ondelette analysante, conformément à l'équation suivante :

$$\Omega = \left\{ \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right), u \in \mathbb{R} \text{ et } s \in \mathbb{R}^{*+} \right\} \quad (\text{II.12})$$

Maintenant, nous allons discuter la transformée en ondelettes continue.

Soient f et g deux fonctions ; leur produit scalaire est définis par la relation suivante :

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t)g(t)dt \quad (\text{II.13})$$

L'existence de la transformée en ondelettes d'une fonction est soumis à la condition suivante :

$$\int_{\mathbb{R}} f^2(t)dt < \infty \quad (\text{II.14})$$

Cette condition se traduit par le fait que l'intégrale du carré de cette fonction soit fini. Dans ces conditions, la transformée en ondelette continue de la fonction f est exprimée sous la forme d'un produit scalaire de f et de ψ , [52],[48],[47] :

$$W(u, s) = \frac{1}{\sqrt{s}} \int_{\mathbb{R}} f(t) \overline{\psi\left(\frac{t-u}{s}\right)} dt \quad (\text{II.15})$$

La transformée inverse continue en ondelettes est possible si la convergence de l'intégrale suivant est acquise [47] :

$$C_{\psi} = \int_0^{\infty} \frac{|\widehat{\psi}(\omega)|^2}{\omega} d\omega \quad (\text{II.16})$$

Où $\widehat{\psi}$ représente la transformée de Fourier de ψ . Cette dernière condition est également appelée critère d'admissibilité pour une ondelette. Dans ce cas, L'inverse de la transformée en ondelettes peut être calculée à partir de la relation suivante [48] :

$$f(t) = \frac{1}{C_{\psi}} \int_{\mathbb{R}} \int_{\mathbb{R}^+} W(u, s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dud s \quad (\text{II.17})$$

La condition d'admissibilité d'une ondelette est très intéressante dans la mesure où elle offre des informations sur les caractéristiques que doit vérifier une ondelette mère pour que la reconstruction d'une fonction soit possible. En particulier, on doit avoir $\psi(0) = 0$. En remplaçant ω par 0 dans la relation de la transformée de Fourier de ψ , on remarque que cette condition devient équivalente à :

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \quad (\text{II.18})$$

1.4.2 La transformée en ondelette discrète

La transformée discrète en ondelette (DWT) est dérivée de la version continue. La différence réside dans le fait qu'elle utilise un facteur d'échelle et un facteur de translation discrétisés.

La transformée en ondelettes discrète est dite dyadique lorsque le facteur d'échelle est égal à 2^i

($a = 2^i$) . Il est clair que la DWT est utile pour le traitement de tout système numérique (PC, DSP) avec un temps de calcul acceptable. Cependant, la version continue de la transformée en ondelette nécessite un temps de calcul énorme provenant de la nature continue de ses facteurs d'échelle et de dilatation.

L'intérêt de cette partie est d'aborder la transformée en ondelettes discrète ainsi que sa relation avec la transformée en ondelettes continue.

Une transformée en ondelettes est dite discrète lorsque les facteurs de translation et de dilatation sont discrétisés (les valeurs ne sont pas nécessairement entières).

Soit ψ une ondelette mère, une famille Ω de fonctions formulée à partir de ψ peut être définie par la relation suivante :

$$\Omega(a, b, x) = \{a^{m/2}\psi(a^m x \pm nb), (m, n) \in \mathbb{Z}^2\} \quad (\text{II.19})$$

Où \mathbb{Z} représente l'ensemble des entiers relatifs. Notons que les paramètres a et b sont des réels fixes qui génèrent, avec ψ , la famille Ω ; a : représente l'échelle des dilatations et b représente le pas des translations.

Une famille d'ondelettes peut être donc complètement définie à partir de l'ondelette mère ψ et le couple (a, b) . Un membre de cette famille (c'est-à-dire une fonction) est représentée par le couple de paramètres (m, n) . Pour cette raison, on désignera dans la suite du travail, une ondelette de la famille de Ω ayant comme paramètres (m, n) par $:\psi_{m,n}(x)$.

Suivant l'équation (II.19), on obtient [47],[48] :

$$\psi_{m,n}(x) = a^{m/2}\psi(a^m x \pm nb) \quad (\text{II.20})$$

Qui peut être reformulé de la façon suivante :

$$\psi_{m,n}(x) = a^{m/2}\psi\left(\frac{x \pm na^{\pm m}b}{a^{\pm m}}\right) \quad (\text{II.21})$$

Cette relation est semblable à l'équation (II.11) qui donne l'expression d'une famille d'ondelettes dans un contexte de transformée continue, avec :

Translation : $u = na^{\pm m}b$

Dilatation : $s = a^{\pm m}$

À partir de ces relations, il s'est avéré que la translation dépend de la dilatation, alors que ces facteurs sont indépendant dans le cas de la transformée continue en ondelette.

Posons que les valeurs discrètes du facteur d'échelle 's' ainsi que celui de translation 'u' seront écrit sous la forme : $s = 2^{-j}$ et $u = k2^{-j}$ Avec ces valeurs de 's' et 'u', la relation (II.15) devient :

$$\begin{aligned} W(u, s) &= \frac{1}{\sqrt{2^{-j}}} \int_{\mathbb{R}} f(t) \psi\left(\frac{t-k2^{-j}}{2^{-j}}\right) dt \\ &= 2^{\frac{j}{2}} \int_{\mathbb{R}} f(t) \psi(2^j t - k) dt \end{aligned} \quad (\text{II.22})$$

Si la fonction $s(t)$ est discrétisée, avec une période d'échantillonnage égale à 1, pour des raisons de simplicité, la relation (II.22) devient alors :

$$W(u, s) = 2^{\frac{j}{2}} \sum_n f(n) \psi(2^j n - k) \quad (\text{II.23})$$

1.5 Analyse multi-résolution

1.5.1 Présentation de l'analyse multirésolution

Le principe général de l'analyse multirésolution a été introduit par Meyer et Mallat . L'idée de l'analyse multirésolution est de définir des opérateurs linéaires pour analyser un signal à différentes échelles. Nous pourrions dire que la construction d'une multirésolution permet d'observer un signal de "très proche" ou de "très loin". Ce Zoom est effectué en utilisant une fonction d'échelle, qui se dilate en fonction des échelles. La projection d'un signal sur cette fonction fournis une représentation de notre signal original à une échelle supérieure. La représentation obtenue (coefficients de projection) offre un zoom arrière de notre signal original, d'où le terme approximation. Pour reconstruire notre signal original, à partir des coefficients d'approximation, ce signal doit être également projeter sur un espace perpendiculaire afin de préserver toute l'information. La fonction qui génère ce deuxième espace vectoriel sera une ondelette. Nous pouvons résumer que, le signal original est projeté sur une fonction d'échelle pour obtenir une approximation et sur une ondelette pour récupérer l'information perdue lors de la première projection. Cette deuxième projection contient les détails du signal original. Les bases $\phi_{j,k}(t)$ et $\psi_{j,k}$ sont appelées fonctions d'échelles et ondelettes, respectivement.

1.5.2 Définition d'une analyse multirésolution

Une analyse multirésolution de $L^2(\mathbb{R})$ est, par définition, une famille de sous-espaces vectoriels fermés emboîtés V_j avec $j \in \mathbb{Z}$ de $L^2(\mathbb{R})$ vérifiant les propriétés suivantes [47] :

- $V_j \subset V_{j+1}$; pour tout j, V_j est emboîté dans V_{j+1}
- $s(t) \in V_j \Leftrightarrow s(2t) \in V_{j+1}$; V_{j+1} est l'image de V_j par un rapport de dilatation égale à 2.
- $s(t) \in V_j \Leftrightarrow s(t-2^{-j}k) \in V_j ; \forall k \in \mathbb{Z}$; la même fonction translatée par 2^{-j} appartient encore à V_j
- $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$; la réunion des V_j est dense dans L^2 .

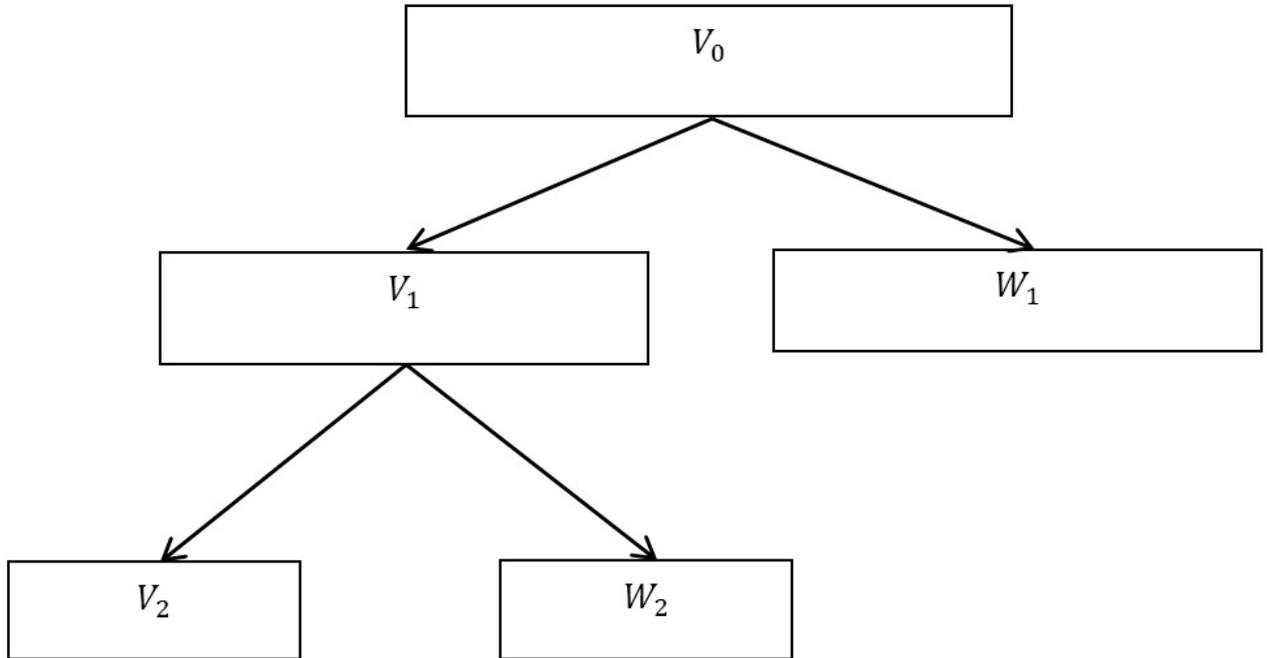


Figure II.4 – La nature hiérarchique de l’analyse multirésolution

- $\bigcap_{j \in \mathbb{Z}} V_j = 0 = \lim_{j \rightarrow \infty} V_j$; l’intersection de V_j est seulement la fonction nulle .
- De plus il doit exister une fonction d’échelle $\phi \in V_0$ tel que l’ensemble des fonctions $\phi(t - k), k \in \mathbb{Z}$ soit une base de Riesz pour V_0 .

Les sous espaces W_j sont définis comme étant les compléments orthonormés de V_j dans V_{j+1} .

$$V_j \oplus W_j = V_{j+1} \quad (\text{II.24})$$

$$V_j \cap W_j = 0 \quad (\text{II.25})$$

1.5.3 Relations de reconstruction et de décomposition

D’après la définition de l’analyse multirésolution, $\phi \in V_0 \subset V_1$ et $\psi \in W_0 \subset V_1$, il existe deux filtres discrets $\{h_1[k]\} \in L^2$ et $\{g_1[k]\} \in L^2$ tel que la fonction d’échelle ϕ et l’ondelette associée ψ , à un niveau donné, dépendent des coefficient des filtres et de la fonction d’échelle au niveau juste au dessus. Ces relations récurrentes sont appelées : relations de reconstruction.

$$\phi(t) = \sum_k h_1[k] \phi(2t - k) \quad (\text{II.26})$$

$$\psi(t) = \sum_k g_1[k] \phi(2t - k) \quad (\text{II.27})$$

En général, $\forall j \in \mathbb{Z}$, la relation reliant les sous espaces V_j, W_j et V_{j+1} est donnée par les deux relations suivantes [52] :

$$\phi(2^j t) = \sum_k h_1[k] \phi(2^{j+1} t - k) \quad (\text{II.28})$$

$$\psi(2^j t) = \sum_k g_1[k] \phi(2^{j+1} t - k) \quad (\text{II.29})$$

Par la transformée de Fourier des relations (II.26) et (II.27), on obtient :

$$\widehat{\phi}(w) = H(z) \widehat{\phi}\left(\frac{w}{2}\right) \quad (\text{II.30})$$

$$\widehat{\psi}(w) = G(z) \widehat{\phi}\left(\frac{w}{2}\right) \quad (\text{II.31})$$

Où :

$$H(z) = \frac{1}{2} \sum_k h_1[k] z^k; z = e^{-\frac{jw}{2}} \quad (\text{II.32})$$

$$G(z) = \frac{1}{2} \sum_k g_1[k] z^k; z = e^{-\frac{jw}{2}} \quad (\text{II.33})$$

En posant que $\widehat{\phi}(0) = 1$. En appliquant ce résultat dans la relation (II.30), on obtient :

$$H(1) = \frac{1}{2} \sum_k h_1[k] = 1 \Rightarrow \sum_k h_1[k] = 2 \quad (\text{II.34})$$

L'ondelette se comporte comme étant un filtre passe-bande, $\widehat{\Psi}(0) = 0$ à l'aide de la relation (II.31), on déduit :

$$G(1) = \frac{1}{2} \sum_k g_1[k] = 0 \Rightarrow \sum_k g_1[k] = 0 \quad (\text{II.35})$$

L'analyse multirésolution permet d'écrire : $V_1 = V_0 + W_0$; $\phi(2t) \in V_1$, $\phi(2t - 1) \in V_1$, il existe donc deux filtres $\{h_2[k]\} \in L^2$ et $\{g_2[k]\} \in L^2$, tel que :

$$\phi(2t) = \sum_k [h_2[-2k] \phi(t - k) + g_2[-2k] \psi(t - k)] \quad (\text{II.36})$$

$$\phi(2t - 1) = \sum_k [h_2 [1 - 2k] \phi(t - k) + g_2 [1 - 2k] \psi(t - k)] \quad (\text{II.37})$$

En se basant sur ces deux dernières équations, on déduit :

$$\phi(2t - n) = \sum_k [h_2 [n - 2k] \phi(t - k) + g_2 [n - 2k] \psi(t - k)] \quad (\text{II.38})$$

L'équation II.38 peut être généralisée comme suit :

$$\phi(2^{j+1}t - n) = \sum_k [h_2 [n - 2k] \phi(2^j t - k) + g_2 [n - 2k] \psi(2^j t - k)] \quad (\text{II.39})$$

Cette équation est appelée : relation de décomposition.

1.5.4 Algorithmes

Deux algorithmes principaux ont été mis en évidence : l'algorithme à trous et l'algorithme de Mallat . Le premier concerne des analyses multirésolution non orthogonales, le second est pratiquement le seul utilisé dans le cas des analyses multirésolution orthogonales et biorthogonales.

Algorithme de décomposition L'algorithme de décomposition (analyse) est souvent utilisé en traitement du signal, et spécialement pour le débruitage de données. L'algorithme est basé sur la relation de décomposition (II.39).

Si la fonction des $\phi_{j,k}$ et $\psi_{j,k}$ (j et k décrivant R^2), est une base orthonormée de $L^2(R)$, alors un signal $s(t) \in L^2(R)$ sera décomposé en fonction des différentes échelles 'j' en : un signal $x_j(t)$ appelé 'approximation', et un signal $y_j(t)$ nommé 'détail', cette opération se déroule de la manière suivante [52] :

$$x_{j+1}(t) \in V_{j+1} \Rightarrow x_{j+1} = \sum_k a_{j+1} [k] \phi_{j+1,k}(t) \quad (\text{II.40})$$

$$x_j(t) \in V_j \Rightarrow x_j = \sum_k a_j [k] \phi_{j,k}(t) \quad (\text{II.41})$$

$$y_j(t) \in W_j \Rightarrow y_j = \sum_k d_j [k] \psi_{j,k}(t) \quad (\text{II.42})$$

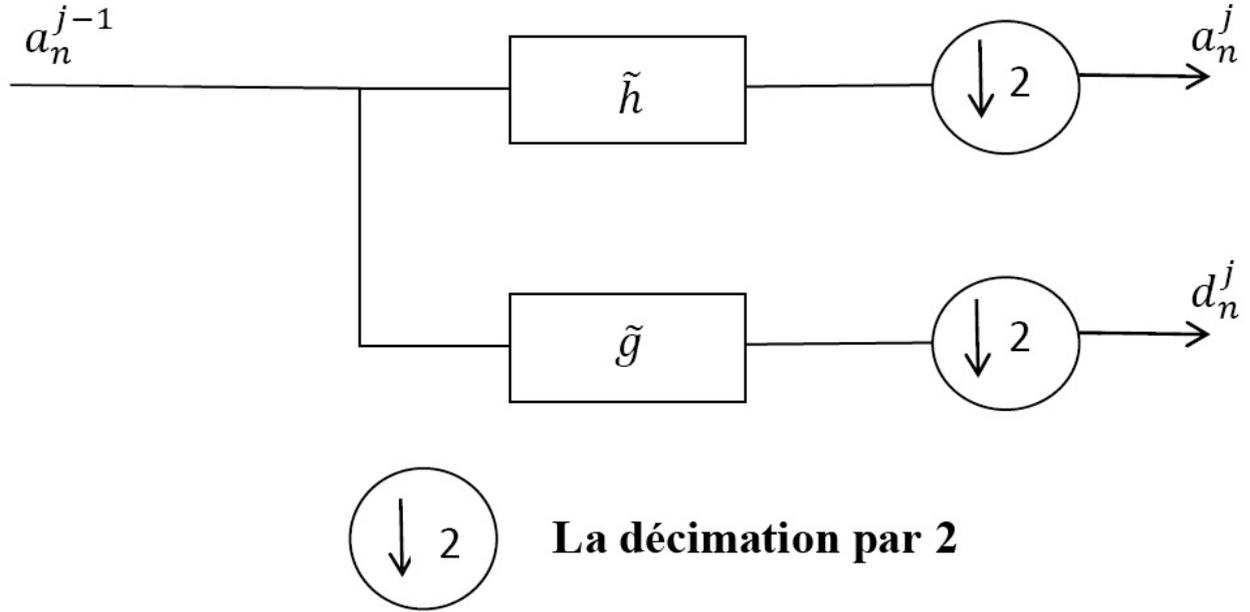


Figure II.5 – Décomposition en cascade.

Avec

$$a_j[k] = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} s(t) \overline{\phi(2^j t - k)} dt \text{ et } d_j[k] = 2^{\frac{j}{2}} \int_{-\infty}^{+\infty} s(t) \overline{\psi(2^j t - k)} dt \quad (\text{II.43})$$

L'analyse multirésolution exige que la condition suivante soit vérifiée :

$$V_{j+1} = V_j + W_j \quad (\text{II.44})$$

Donc :

$$x_{j+1}(t) = x_j(t) + y_j(t) \quad (\text{II.45})$$

$$\sum_k a_{j+1}[k] \phi(2^{j+1}t - k) + \sum_k d_{j+1}[k] \psi(2^j t - k) = \sum_k a_j[k] \phi(2^j t - k) + \sum_k d_j[k] \psi(2^j t - k) \quad (\text{II.46})$$

Substituons l'équation de décomposition (II.39) dans la relation (II.46) pour trouver une relation dont toutes les bases sont à l'échelle j . Après avoir interchangé l'ordre des sommations et comparé les coefficients de $\phi(2^j t - k)$ et $\psi(2^j t - k)$ de chaque côté de l'équation, on obtient

la décomposition suivant l'algorithme de Mallat [52] :

$$a_j [k] = \sum_n h_2 [n - 2k] a_{j+1} [k] \quad (\text{II.47})$$

$$d_j [k] = \sum_n g_2 [n - 2k] d_{j+1} [k] \quad (\text{II.48})$$

À partir des relations (II.47) et (II.48) on remarque que les coefficients de la fonction d'échelle et de l'ondelette à l'échelle j , peuvent être calculés à partir de ceux de l'échelle immédiatement supérieure.

En répétant ce processus, on obtient la décomposition en cascade illustrée dans la figure II.5.

Algorithme de reconstruction Il existe une unique transformation en ondelettes discrète inverse, qui porte aussi le nom de synthèse, tel que le signal original peut être reconstruit parfaitement à partir de ses approximations et ses détails à différentes échelles. L'algorithme de synthèse est basé sur les équations de reconstruction (II.26) et (II.27). En substituant ces deux relations dans (II.46), on obtient :

$$\begin{aligned} \sum_k a_j [k] \sum_n h_1 [n - 2k] \phi(2^{j+1}t - n) + \sum_k d_j [k] \sum_n g_1 [n - 2k] \phi(2^{j+1}t - n) \\ = \sum_n a_{j+1} [k] \phi(2^{j+1}t - n) \end{aligned} \quad (\text{II.49})$$

En comparant les coefficients de $\phi(2^{j+1}t - n)$ dans les deux côtés de la relation (II.49), on obtient la reconstruction suivant l'algorithme de Mallat :

$$a_{j+1} [k] = \sum_k \{h_1 [n - 2k] a_j [k] + g_1 [n - 2k] d_j [k]\} \quad (\text{II.50})$$

L'algorithme de reconstruction est illustré par la figure(II.6).

1.5.5 Bilan

L'analyse multirésolution consiste à l'utilisation des algorithmes de décomposition et de synthèse pour l'étude de signaux en les décomposants dans différents sous-espaces emboîtés orthonormaux. Pour calculer les paramètres de la séquence $h_1 [k], g_1 [k], h_2 [k], g_2 [k]$ dans le cas de bases orthogonales, il suffit de trouver seulement les paramètres du filtre $h_1 [k]$ puisque les trois autres filtres peuvent être calculés à partir de ce dernier :

$$g_1 [k] = (-1)^k h_1 [1 - k] \quad (\text{II.51})$$

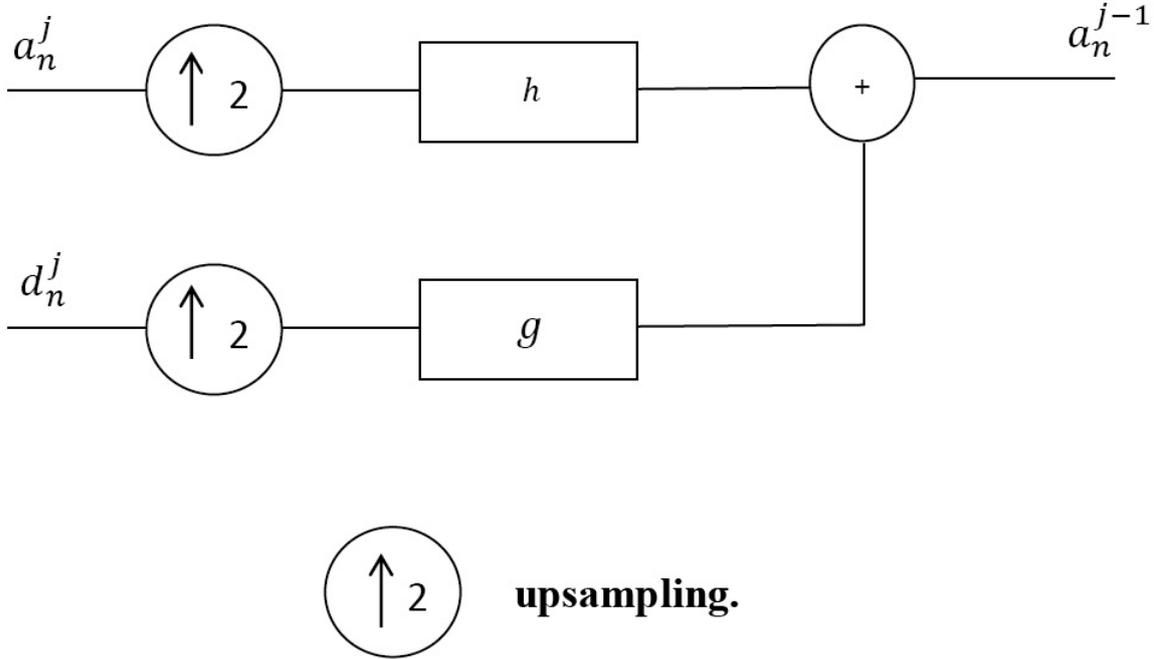


Figure II.6 – Algorithme de synthèse de Mallat.

$$h_2[k] = \frac{1}{2}h_1[-k] \quad (\text{II.52})$$

$$g_2[k] = \frac{1}{2}(-1)^k h_1[k+1] \quad (\text{II.53})$$

Les algorithmes de décomposition et de synthèse sont utilisées pour tous les types de fonctions d'échelles et d'ondelettes (orthogonales, biorthogonales, semiorthogonales).

Un signal $s(t) \in V_0$ correspondant à l'échelle J_0 (voir figure(II.4)), se décompose à l'aide de l'analyse multirésolution comme suit ($J \in \mathbb{R}^2$ représente l'échelle la plus basse) :

$$s(t) = \sum_{k \in \mathbb{Z}} a_J[k] \phi(2^J t - k) + \sum_{j=J}^{j_0} \sum_{k \in \mathbb{Z}} d_j[k] \psi(2^j t - k) \quad (\text{II.54})$$

1.6 Paquets d'ondelettes

L'idée de l'analyse multirésolution de l'espace L^2 des fonctions continues d'une variable réelle et de carré intégrable, peut être généralisée à des sous-espaces de celui-ci.

Cette nouvelle analyse peut être obtenue en utilisant les même bases de fonctions d'échelle et

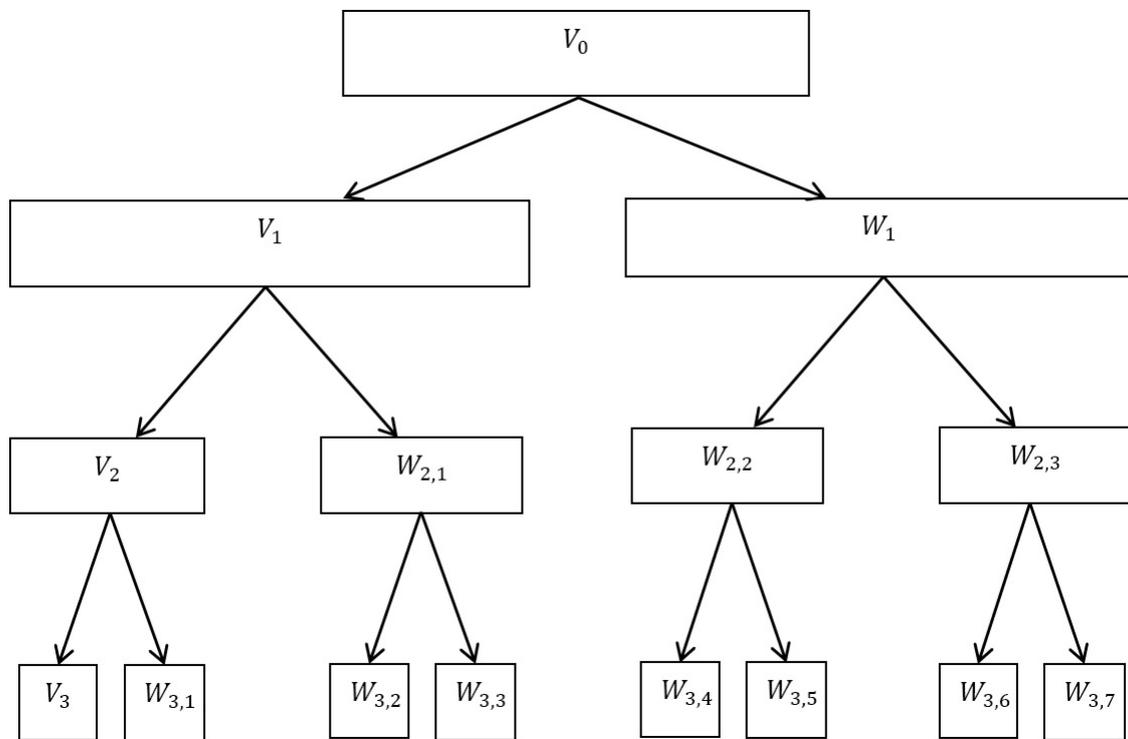


Figure II.7 – Analyse en paquets d’ondelettes.

d'ondelettes ou avec des bases différentes. On peut, en plus, changer de fonctions de base à chaque échelle. Dans ces méthodes, la synthèse parfaite est garantie par la réutilisation lors de la reconstruction de la même base que celle utilisée lors de l'analyse, et ce pour une résolution donnée.

La méthode présentée permet de plus une analyse des sous-espaces de détail du signal. Cette méthode est appelée analyse en paquets d'ondelettes. La figure (II.7) décrit le principe de l'algorithme de paquets d'ondelettes.

1.7 Orthogonalité et bi-orthogonalité

Les ondelettes orthogonales sont souvent utilisées pour éviter la redondance ainsi que pour garantir une reconstruction complète et parfaite du signal.

1.7.1 Les ondelettes orthogonales

L'analyse multirésolution orthogonale est un algorithme multirésolution dans lequel les espaces d'ondelette W_j sont définis comme le complément orthogonal de V_j dans V_{j+1} .

Par conséquent, les espaces W_j avec $j \in Z$ sont tous mutuellement orthogonaux.

Pour qu'une analyse multi-résolution soit orthogonale la condition suivante doit être satisfaite : $(W_0 \perp V_0)$, avec : $\{\phi(x-n)\}, n \in Z$ et $\{\psi(x-n)\}, n \in Z$ sont des familles des vecteurs orthogonales. Cela signifie que pour tout $n \in Z$:

$$\langle \psi, \phi(-n) \rangle = 0, n \in Z \quad (\text{II.55})$$

Reprenons la relation de la famille d'ondelettes discrètes :

$$\psi_{m,n}(x) = \frac{1}{\sqrt{|a_0^m|}} \psi\left(\frac{1}{a_0^m}x - nb_0\right) \quad (\text{II.56})$$

Y.Meyer a démontré que pour un facteur d'échelle $a_0 = 2$ et un paramètre de dilatations $b_0 = 1$, Les ondelettes sont dites orthogonales et on obtient alors :

$$\psi_{m,n}(x) = \frac{1}{\sqrt{2^m}} \psi\left(\frac{1}{2^m}x - n\right) \quad (\text{II.57})$$

Une fonction d'échelle orthogonale est une fonction ϕ telle que l'ensemble $\{\phi_{j,l}, l \in Z\}$ est une base orthonormée, où :

$$\langle \psi, \phi(-l) \rangle = 0, l \in Z \quad (\text{II.58})$$

Avant une telle fonction ϕ , la collection de fonctions $\{\phi(x-1)/l, l \in Z\}$ est une base orthonormée de ν_0 et la collection de fonction $\{\phi_{j,l}, l \in Z\}$ est une base orthonormée de V_j . L'analyse multi-résolution consiste à passer le signal à analyser à travers deux filtres distincts pour obtenir une approximation et un détail. Si les fonctions employées (fonction d'échelle ou ondelette) ne sont pas orthogonales ; L'information entre les deux signaux résultants sera corrélée, d'où la redondance. En conséquence, la reconstruction du signal d'origine à partir de ses transformées sera difficile.

1.7.2 Les ondelettes bi-orthogonales

Nous avons montré les avantages que présentait la transformée en ondelettes orthogonale pour le traitement du signal. Cependant, il n'existe pas de filtres RIF orthonormaux à phase linéaire qui assurent la reconstruction parfaite, sauf ceux correspondant à la base de Haar. Pour surmonter cette lacune, l'utilisation de bases bi-orthogonales permet la construction de filtres associés à des ondelettes possédant un haut degré de régularité. Les bases d'ondelettes bi-orthogonales sont une généralisation des bases d'ondelettes orthogonales.

Dans le cas bi-orthogonal, on utilise deux ondelettes mères $\psi(x)$ et $\tilde{\psi}(x)$ pour construire les bases d'ondelettes $\psi_{a,b}(x)$ et $\tilde{\psi}_{a,b}(x)$ et on emploie le même principe pour construire les fonctions d'échelles $\phi_{a,b}(x)$ et $\tilde{\phi}_{a,b}(x)$.

Ainsi, l'ondelette utilisée pour l'analyse $\psi_{a,b}(x)$ est orthogonale à la fonction d'échelle $\tilde{\phi}_{a,b}(x)$ utilisée pour la construction et la fonction d'échelle utilisée pour l'analyse $\phi_{a,b}(x)$ est orthogonale à l'ondelette $\tilde{\psi}_{a,b}(x)$ utilisée pour la construction, d'où le nom de bi orthogonalité. Par contre, et contrairement au cas orthogonale les ondelettes et les fonctions d'échelle associées ne sont pas orthogonales.

l'analyse multi résolution utilisant ces bases bi-orthogonales définit les espaces de détail W_m et \tilde{W}_m , et d'approximation V_m et \tilde{V}_m répondant à la propriété suivante $W_m \perp \tilde{V}_m$ et $\tilde{W}_m \perp V_m$.

1.8 Exemples d'ondelettes

Les premières ondelettes proposées par Meyer et Mallat, sont les ondelettes orthogonales. Parmi celles-ci les ondelettes de Daubechies sont couramment utilisées. Les familles d'ondelettes non orthogonales nommées aussi les ondelettes biorthogonales dont le préfixe bi rappelle que deux bases d'ondelettes sont employées, une pour la décomposition (la base duale) et l'autre pour le synthèse. Le fait que la fonction d'échelle et la fonction ondelette soient biorthogonales ne signifie pas nécessairement que l'analyse multirésolution ne soit pas orthogonale. Si les fonctions d'échelles et ondelettes biorthogonales permettent d'obtenir une analyse multirésolution orthogonale, on les appelle ondelettes semiorthogonales .

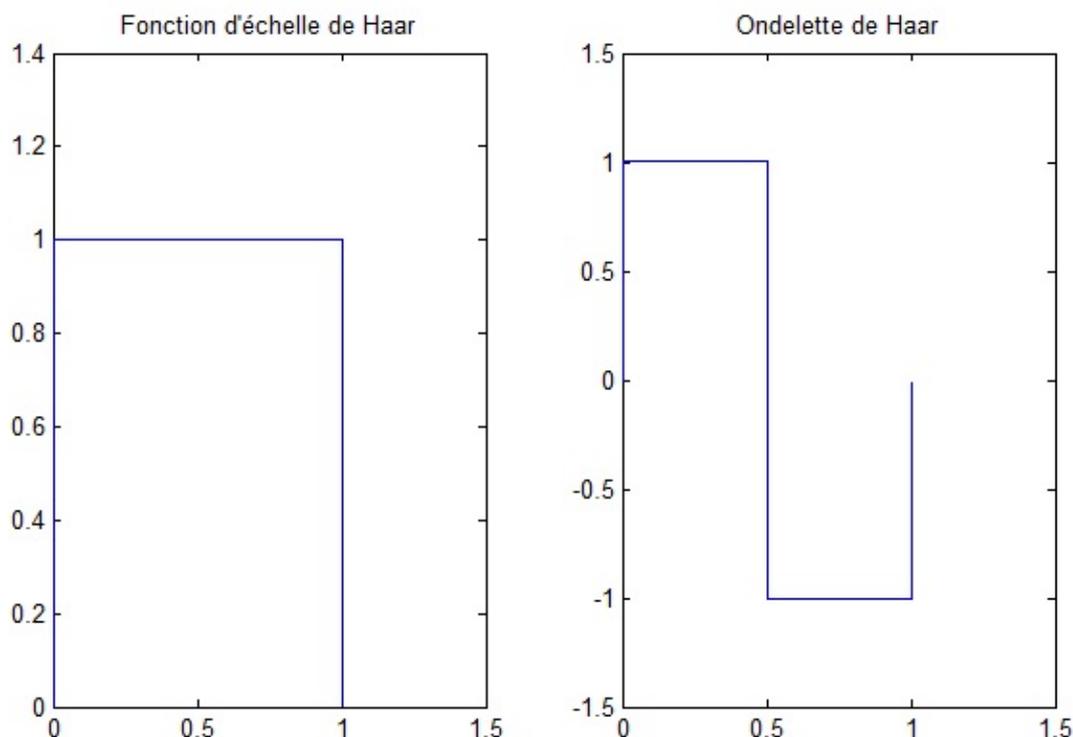


Figure II.8 – Fonction d'échelle et d'ondelette de type Haar .

1.8.1 Ondelettes orthogonales

Ondelette de Haar

Parmi la famille des ondelettes orthogonales nous parlerons maintenant de l'ondelette de Haar, certainement la plus ancienne des ondelettes. La fonction d'échelle dans ce cas est donnée par $\phi_{Haar}(t) = X_{[0,1]}(t)$ et l'ondelette associée est $\psi_{Haar}(t) = X_{[0,1/2]}(t) - X_{[1/2,1]}(t)$.

En pratique, l'ondelette de Haar n'est pas beaucoup employée car elle est discontinue..

Ondelette de Shannon

L'ondelette de Shannon est définie par :

$$\psi_{SH} = \frac{\sin 2\pi t - \sin \pi t}{\pi t} \quad (\text{II.59})$$

dont la fonction d'échelle $\phi_{SH} = \text{sinc}(\pi t)$ n'est pas beaucoup employée car elle se caractérise par une faible décroissance à l'infini. En plus de ça, ψ_{SH} est très mal localisé en temps ($\Delta\psi_{SH} = \infty$). La raison de cette mauvaise localisation en temps vient du fait que dans le domaine fréquentiel,

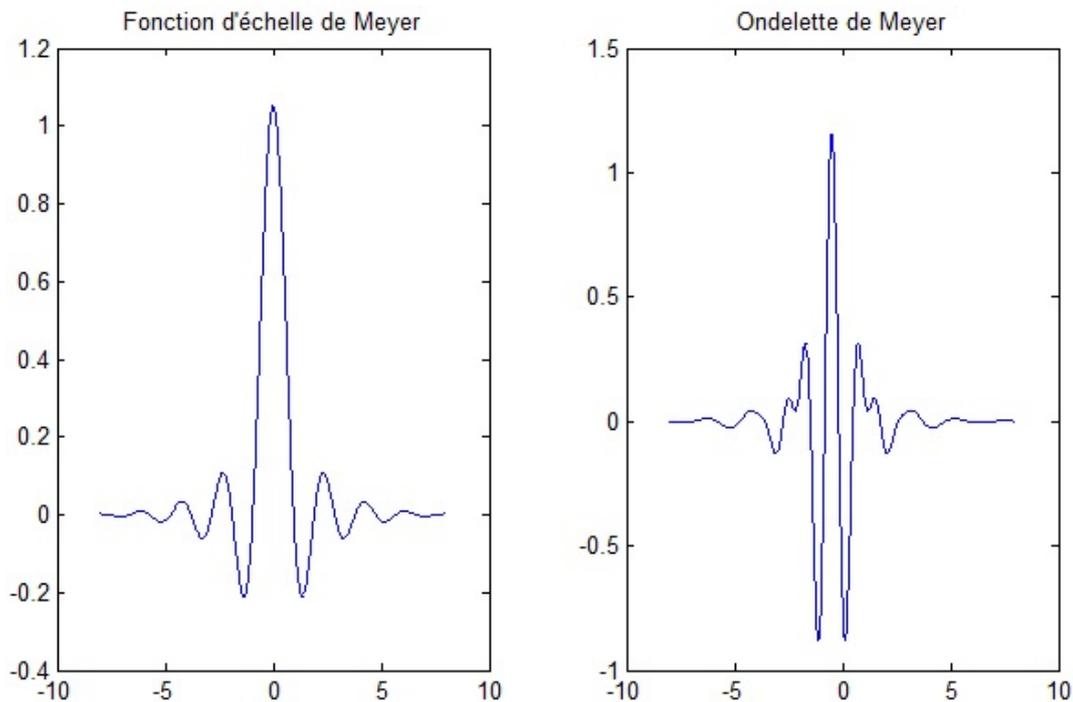


Figure II.9 – Fonction d'échelle et ondelette de Meyer.

$\hat{\psi}_{SH}(w)$ comporte deux discontinuités à $-\pi$ et à π . En conséquence, dans le domaine temporel, la fonction décroît en $1/t$ et donc $\Delta\psi_{SH} = \infty$.

Ondelette de Meyer

Meyer a trouvé une fonction d'échelle, voir figure(II.9), de telle façon que sa transformée de Fourier soit lisse aux endroits de discontinuités de $\hat{\psi}_{SH}(w)$ (voir ondelette de Shannon). Dans le domaine temporel, cela se traduit par une décroissance plus rapide à l'infini par rapport à celle de Shannon. La fonction d'échelle et l'ondelette sont symétriques respectivement par rapport à 0 et $-1/2$. Les ondelettes de Meyer sont des ondelettes indéfiniment dérivables, de support infini, leur implémentation se fait plutôt dans le domaine fréquentiel.

Ondelette de Daubechies

Cette famille d'ondelettes est probablement la plus employée parmi les ondelettes orthogonales. L'ondelette de Daubechies est la première ondelette à support compact (il existe un intervalle fermé et borné tel que la fonction est égale à zéro en dehors de cet intervalle.). Ces ondelettes seront représentées par : dbN, avec : db est le symbole donné pour Daubechies, et

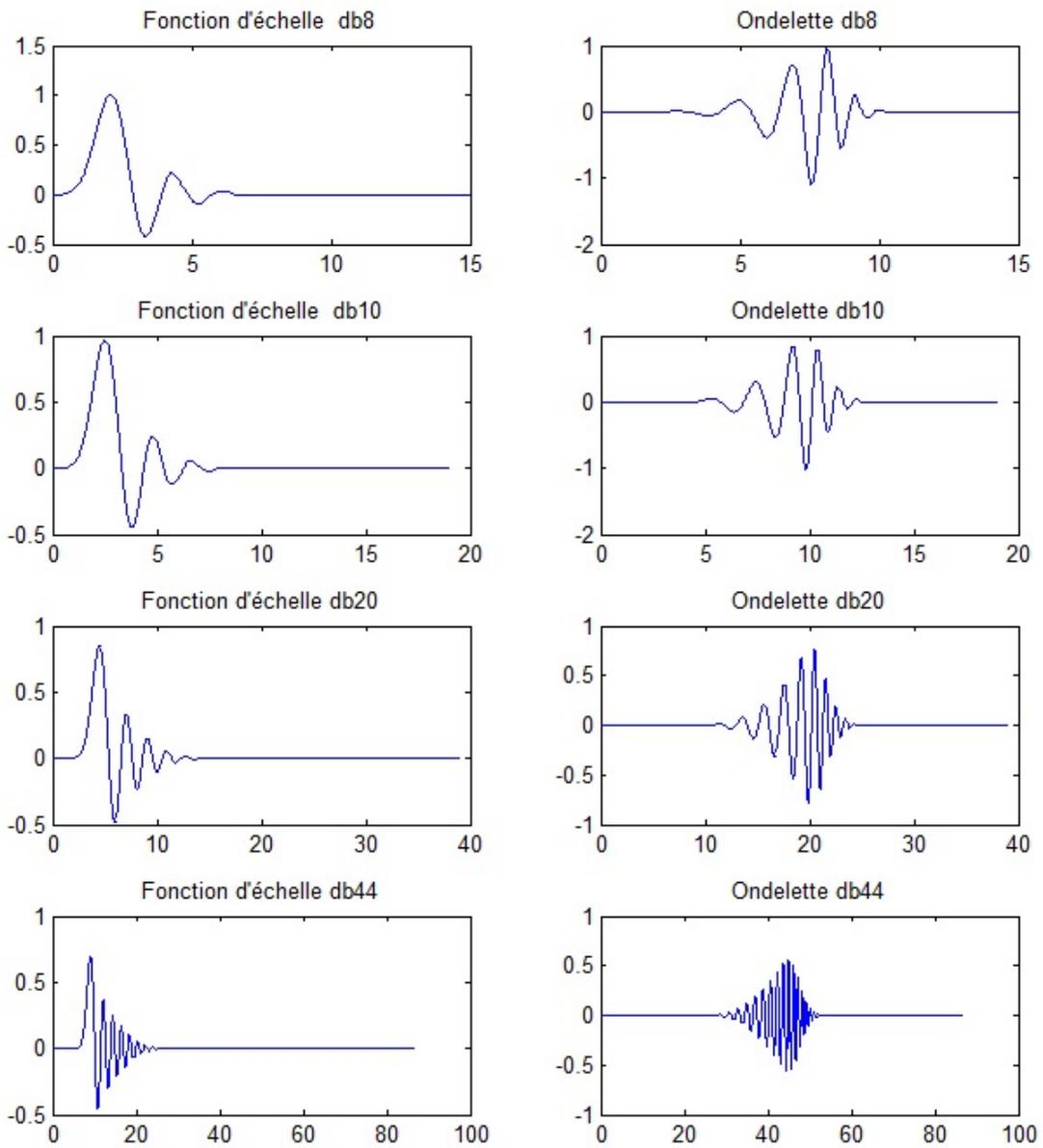


Figure II.10 – Fonction d'échelle et ondelette de Daubechies pour $N = 8, 10, 20$ et 44 .

N définit l'ordre de l'ondelette. Les ondelettes de Daubechies sont supportées sur un intervalle de longueur $2N - 1$. Un des inconvénients de ces ondelettes vient du fait qu'elles sont très asymétriques. Pour un ordre $N = 1$, l'ondelette db1 correspond à l'ondelette de Haar. La Figure (II.10) illustre les fonctions d'échelles et ondelettes pour les ordres $N = 8$, $N = 10$, $N=20$ et $N = 44$.

Symlets Cette famille d'ondelettes introduit par Daubechies, constitue une famille d'ondelettes presque symétrique, le filtre h_1 doit être symétrique ou antisymétrique par rapport au centre de son support ($\widehat{h}_1(w)$ a une phase linéaire). Daubechies a construit ces ondelettes de telle façon que la phase de ($\widehat{h}_1(w)$) soit la plus linéaire possible. Le support des symlets est $2N + 1$. La Figure (II.11), illustre la fonction d'échelle et l'ondelette Symlets pour $N = 8$, $N=10$ et $N=20$. Une meilleure symétrie par rapport à l'ondelette de Daubechies ($N = 8$) peut être constatée.

Coiflets sur la demande de Coifman, Daubechies a construit une famille d'ondelettes avec N moments nuls et un support de taille minimum ($3N - 1$ au lieu de $2N - 1$ pour une ondelette de Daubechies.), et dont la fonction d'échelle satisfait :

$$\int_{-\infty}^{+\infty} \phi(t)dt = 1, \int_{-\infty}^{+\infty} t^k \phi(t)dt = 0 \text{ pour } 1 \leq k \leq N \quad (\text{II.60})$$

La Figure (II.12), représente la fonction d'échelle et l'ondelette de type Coiflets pour $N = 1$, $N = 2$ et $N=5$.

2 Les paramètres statistiques

Dans la deuxième partie du chapitre nous allons aborder les paramètres statistiques. Dans cette approche, la caractérisation d'un signal de vibration est effectuée en utilisant plusieurs paramètres ou des indicateurs statistiques temporelles tel que : le minimum (Min), la valeur crête à crête (PPV), la valeur efficace (RMS), la déviation standard (SD), le facteur de crête (CRF), le facteur d'impulsion (IMF), le kurtosis (Kur).

1. Le minimum :

$$Min = \min(x) \quad (\text{II.61})$$

2. :La valeur crête à crête

$$PPV = \frac{1}{2} [\max(x) - \min(x)] \quad (\text{II.62})$$

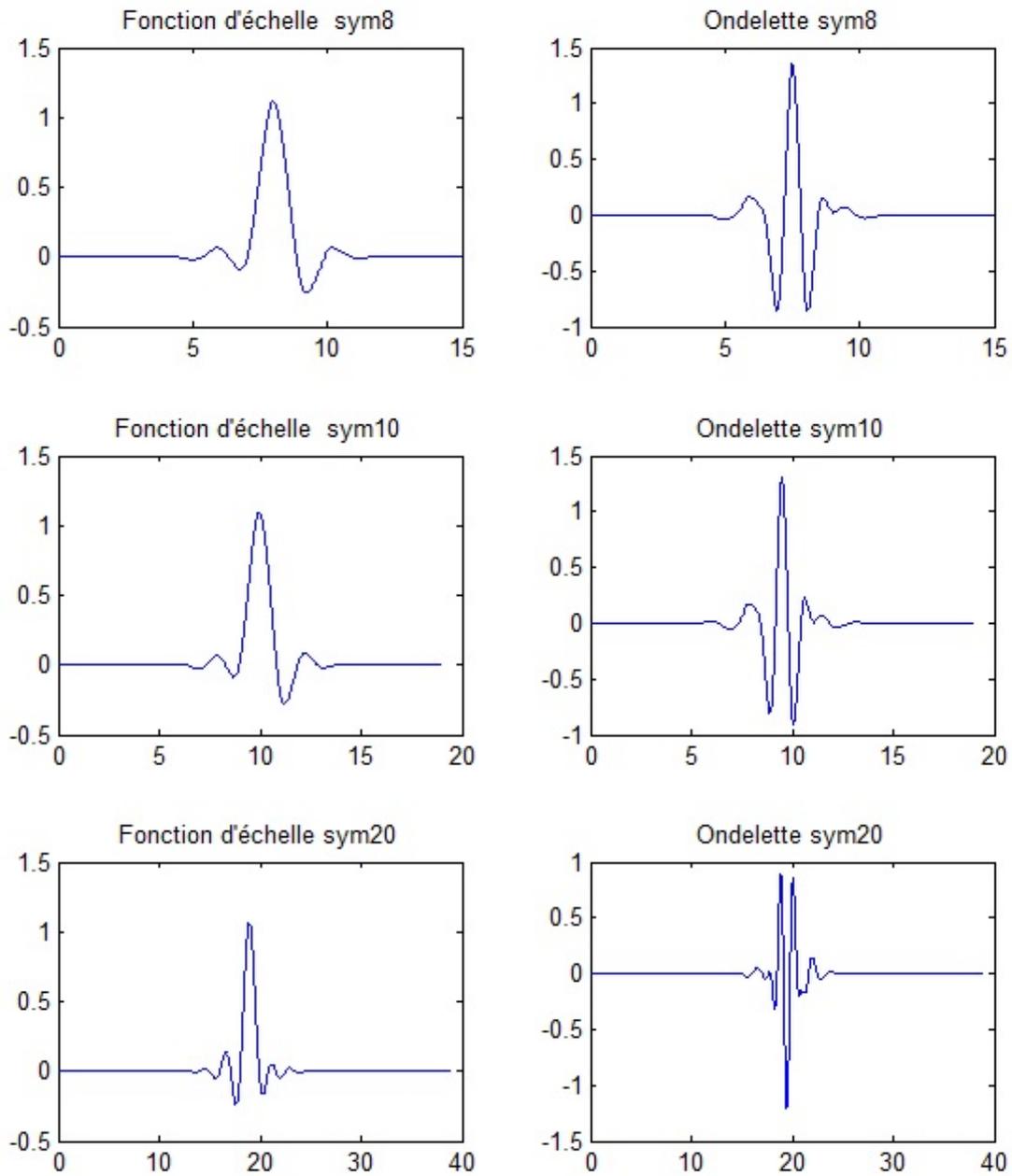


Figure II.11 – Fonction d'échelle et ondelette Symlets pour $N = 8$, $N = 10$ et $N=20$.

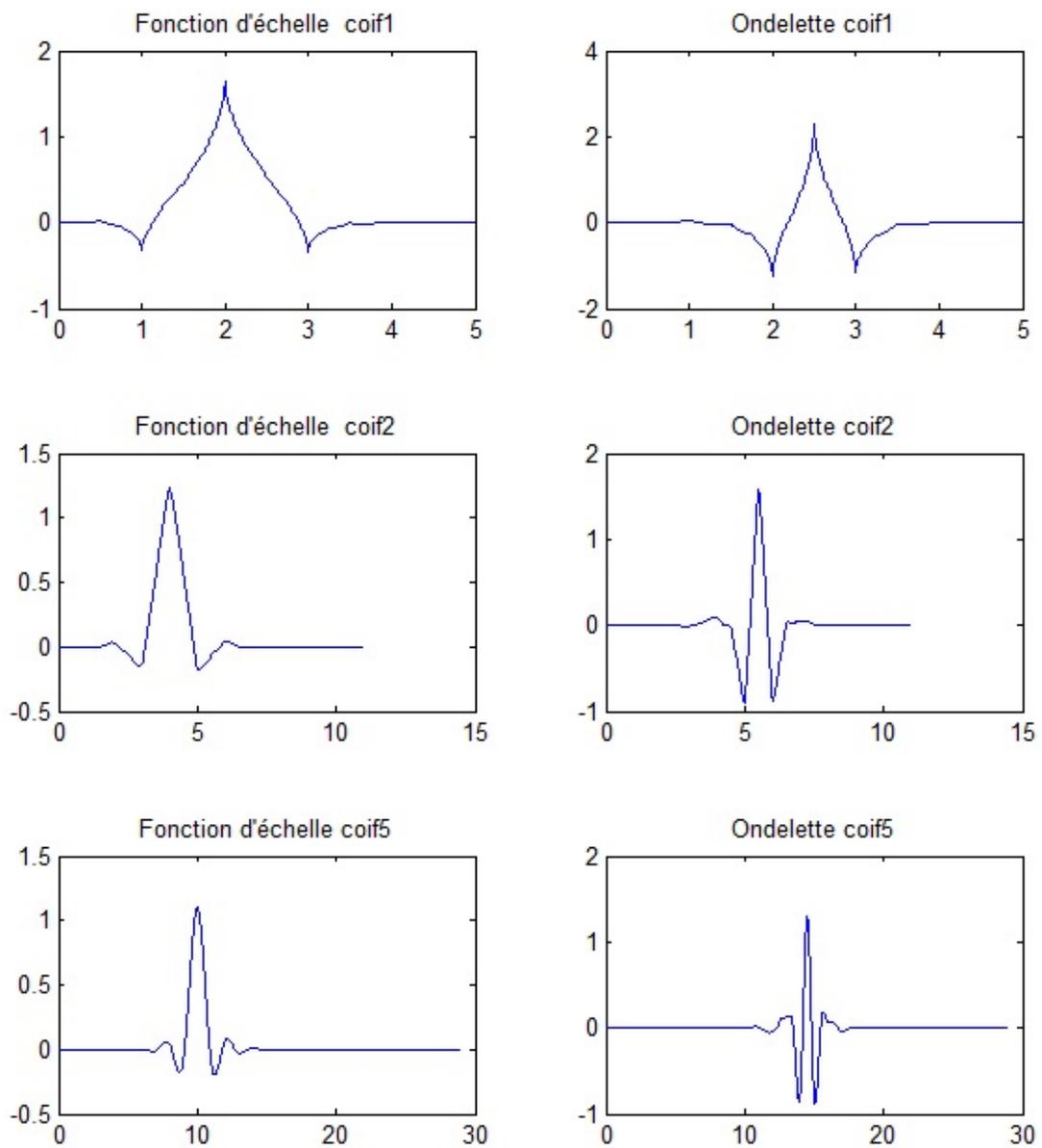


Figure II.12 – Fonction d'échelle et d'ondelette de type Coiflets pour $N = 1$, $N = 2$ et $N = 5$.

3. La valeur efficace

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (x(i))^2} \quad (\text{II.63})$$

4. La déviation standard

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x(i) - \bar{x})^2} \quad (\text{II.64})$$

5. Le facteur de crête :

$$CRF = \frac{PV}{RMS} \quad (\text{II.65})$$

6. Le facteur d'impulsion :

$$IMF = \frac{PV}{\frac{1}{N} \sum_{i=1}^N x(i)} \quad (\text{II.66})$$

7. Le kurtosis :

$$Kur = \frac{\frac{1}{N} \sum_{i=1}^N (x(i) - \bar{x})^4}{\left[\frac{1}{N} \sum_{i=1}^N (x(i) - \bar{x})^2 \right]^2} \quad (\text{II.67})$$

Deuxième partie

Réduction de dimension

1 Optimisation par essaim particulaire (OEP) et Optimisation par essaim particulaire bianire

Kennedy et Eberhart, proposent en 1995 une nouvelle méthode d'optimisation nommé Optimisation par Essaim de Particule PSO, (en anglais : Particles Swarm Optimization). PSO est une méthode d'optimisation stochastique basée sur une population de particules, un Essaim regroupe plusieurs particules (individus). Chaque particule prend sa décision en utilisant sa propre expérience et les expériences de leur voisinage.

PSO est inspiré du comportement social des nuées d'oiseaux et des bancs de poissons qui ont tendance à imiter les comportements réussis qu'ils observent dans leurs entourage, tout en y apportant leurs variations personnelles.

PSO Démarre le processus d'optimisation par une population des solutions aléatoires qui se déplacent dans l'espace de recherche. Le déplacement de chaque particule dans l'espace de recherche, est basé sur sa position actuelle et la mise à jour de sa vitesse [41], [42], [53] :

$$Z_{id}^k = Z_{id}^{k-1} + V_{id}^k \quad (\text{II.68})$$

Tel que [53] :

Z_{id}^k, Z_{id}^{k-1} : Position de la particule id à l'itération k et $k - 1$ respectivement.

V_{id}^k : Vitesse de la particule id à l'itération k .

Chaque particule dans l'essaim, change sa vitesse suivant deux informations essentielles. Une, est liée à son expérience personnelle, qui est la meilleure position trouvée par la particule durant le processus de recherche **pbest**. La deuxième information, concernant la meilleure position trouvée par les voisins (**lbest**) (ou par tout l'essaim, dans la version globale de l'algorithme **gbest**). Cette information est obtenue à partir de la connaissance de la façon dont les autres agents ont exécuté leurs recherches.

Le principe de changement de la vitesse est défini par [41], [53] :

$$V_{id}^k = \omega.V_{id}^{k-1} + c_1.r_1.(P_{id} - Z_{id}^{k-1}) + c_2.r_2.(P_{Gd} - Z_{id}^{k-1}) \quad (\text{II.69})$$

Où :

V_{id}^k : Vitesse de la particule id à l'itération k ,

ω : Fonction de pondération,

c_1, c_2 : Facteurs de pondération,

r_1, r_2 : Nombre aléatoire entre 0 et 1,

II.1 Optimisation par essaim particulaire (OEP) et Optimisation par essaim particulaire bianire

Z_{id}^{k-1} : Position actuelle de la particule id à l'itération $k - 1$,

P_{id} : Meilleure position trouvée par la particule id jusque ici,

P_{Gd} : Meilleure position trouvée par l'essaim jusque ici.

La fonction de pondération ω est donnée par l'équation suivante :

$$\omega(k) = \omega_{max} - \frac{(\omega_{max} - \omega_{min})}{k_{max}}.k \quad (\text{II.70})$$

Tel que :

ω_{max} : Poids initial,

ω_{min} : Poids final,

k_{max} : Nombre d'itérations maximum,

k : Itération courante.

La fonction de pondération " ω " joue un rôle important dans la procédure de recherche. Elle garantit un équilibre entre la recherche locale et la recherche globale, un bon choix de cette fonction augmente l'efficacité de la méthode pour avoir une solution globale.

Algorithme général de l'OEP

Les concepts de l'OEP présentés dans les sections précédentes est donné comme suit :

- **Étape 1** : Génération d'état initial de chaque agent ;
Les points de recherche initiaux, position et vitesse de chaque agent sont produits aléatoirement. Pour chaque agent le point de recherche courant est placé à $pbest$.
- **Étape 2** : Evaluation du point de recherche de chaque agent ;
Les points de recherche courants sont évalués au moyen de la fonction objective relative au système d'étude. Si la valeur d'un agent est meilleure que son $pbest$ courant, $pbest$ prend cette nouvelle valeur. Si la meilleure valeur de $pbest$ est meilleure que $gbest$ courant, $gbest$ est remplacé par cette meilleure valeur et l'agent qui correspond à cette meilleure valeur est ainsi stocké.
- **Étape 3** : Modification du point de recherche ;
La modification du point de recherche se fait en moyennant les équations (II.68),(II.69).
- **Étape 4** : Critère d'arrêt ;
La procédure de recherche est arrêtée lorsque le nombre courant d'itération devient égal à la valeur maximale d'itérations préalablement déterminée. Dans ce cas de figure, la dernière valeur $gbest$ peut être considérée comme solution. Dans l'autre cas, la recherche continue en revenant à l'étape 2.

Les avantages d'optimisation par Essaims de particules peut être présentée par les points suivants :

- La méthode OEP peut essentiellement manipuler le problème d'optimisation continu.
- L'OEP utilise plusieurs points de recherche qui convergent graduellement vers le point optimal en utilisant les valeurs *pbest* et *gbest*.
- La méthode OEP utilise La diversification et l'intensification pour une recherche efficace.
- L'OEP présente un avantage certain sur les méthodes classiques dans le sens où elle permet une exploration aléatoire de l'espace des solutions. De plus, elle est simple à mettre en oeuvre et son exécution conduit à l'obtention de très bons résultats , rapidement et efficacement avec peu des paramètres à ajuster[41], [42],[54],[55].

Par rapport à d'autres algorithmes d'optimisation heuristiques, OEP est efficace et facile à mettre en oeuvre. Néanmoins, il présente l'inconvénient que sa performance peut diminuer quand le nombre d'itération augmente. De plus, dans le cas de traitement des problèmes complexes, OEP convergera facilement vers des solutions optimales locales. Pour surmonter ces lacunes, plusieurs variantes ont été développées telles que OEP avec un Facteur de Constriction (OEPFC) et OEP avec Congrégation Passive (OEPCP) ou Particle Swarm Optimizer with Passive Congregation (PSOPC) en anglais. L'OEPCP a été appliqué avec succès dans de nombreux domaines, il donne de meilleurs résultats que ceux générés par d'autres variantes de PSO. La mise à jour de la vitesse de chaque particule, dans le cas de l'algorithme OEPCP est exprimée comme suit :

$$V_{id}^k = \omega.V_{id}^{k-1} + c_1.r_1.(P_{id} - Z_{id}^{k-1}) + c_2.r_2.(P_{Gd} - Z_{id}^{k-1}) + c_3.r_3.(P_R - Z_{id}^{k-1}) \quad (\text{II.71})$$

Avec : P_R est une particule choisie aléatoirement à partir de l'essaim.

Dans l'algorithme BPSO, la position de chaque particule est exprimée sous la forme d'un vecteur de bits binaires composé de zéros (0) et des uns (1). Chaque position est mise à jour en fonction de la valeur de sa vitesse qui sera convertie en valeurs de probabilité à l'aide de la fonction suivante [27] :

$$F(V_{id}^k) = 2 * \left| \text{sigmoid}(V_{id}^k) - 0.5 \right| \quad (\text{II.72})$$

Après avoir transformé les vitesses aux valeurs de probabilité en utilisant l'équation II.73, la nouvelle position binaire de chaque particule est déterminée comme suit [27] :

$$\text{Si } \text{randn} < F(V_{id}^k), \text{ Alors } Z_{id}^k = \text{échanger } Z_{id}^{k-1} \text{ Sinon } Z_{id}^k = Z_{id}^{k-1} \quad (\text{II.73})$$

2 Analyse en Composantes Principales

L'analyse en composantes principales (ACP) est une technique qui permet de caractériser un système complexe multidimensionnel en dimension inférieure. Donc, le but est de définir un sous-espace de dimension d' , inférieure à d , pour mieux représenter l'ensemble d'apprentissage .

Soit $X = \{x_1, x_2, \dots, x_N\}$ l'ensemble des N vecteurs (matrice des données) de dimension d . Si les composantes de ces vecteurs ne sont pas homogènes, alors les résultats de l'ACP et de la classification risquent d'être biaisés.

Afin de parer à cette éventualité les données sont normalisées de la manière suivante : Le centre de gravité de l'ensemble d'apprentissage est le vecteur défini par :

$$m = \frac{1}{N} \sum_{k=1}^N x_k$$

et les composantes de la variance sont :

$$\sigma_j^2 = \frac{1}{N} \sum_{k=1}^N (x_{kj} - m_j)^2 \text{ avec } j = 1 \dots d.$$

Ainsi, à partir des deux relations précédentes on en déduit les vecteurs formes normalisés (c'est-à-dire de moyenne nulle et d'écart type unité) : $x_{k\text{norm}} = \Delta^{-\frac{1}{2}} (x_k - m)$

Avec Δ : la matrice diagonale des variances σ_j^2 :

$$\Delta = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_d^2 \end{bmatrix}$$

Pour la suite de l'étude, les données seront considérées comme étant normalisées. La base du sous-espace est définie par les vecteurs propres de la matrice de **variance-covariance** [56] :

$$S = \frac{1}{N} \sum_{k=1}^N x_k \cdot x_k^t$$

S une matrice carrée ($d \times d$), réelle et symétrique, c'est aussi la matrice d'inertie du nuage de points considéré. Ses éléments diagonaux sont les variances de l'ensemble d'apprentissage. Elles valent 1 car les données sont centrées et réduites. Et les éléments non diagonaux sont les covariances.

Les vecteurs propres (II.74) de la matrice de variance- covariance S sont les vecteurs v_j tels que $S.v_j = \lambda_j.v_j$. Les scalaires λ_j , repris dans la matrice diagonale (II.75), sont les valeurs propres.

$$V=[v_1, v_2, \dots, v_{d'}, \dots, v_d] \quad (\text{II.74})$$

$$\lambda = \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \dots & & \\ & & & \lambda_{d'} & \\ & & & & \dots \\ 0 & & & & & \lambda_d \end{bmatrix} \quad (\text{II.75})$$

Les premiers d' vecteurs propres, rangés par ordre décroissant des valeurs propres vont constituer la base du sous-espace de dimension d' . Le vecteur x_j^p dont les N composantes sont les projections des N vecteurs x_k sur le vecteur v_j est appelé $j^{\text{ième}}$ composante principale.

Chaque composante x_{kj}^p du vecteur x_j^p aura pour expression :

$$x_{kj}^p = x_k^t \cdot v_j \quad \text{avec } k = 1, \dots, N$$

La nouvelle matrice des données s'écrit :

$$X = [x_1^p, x_2^p, \dots, x_{d'}^p, \dots, x_d^p] = X^t \cdot V$$

Les axes définis par ces vecteurs x_j^p sont dits factoriels, et sont les vecteurs propres. Ces axes sont les combinaisons linéaires des d axes initiaux.

On aboutit ainsi à une représentation optimale de la matrice des données X dans un espace de dimension inférieure. L'ACP garantit que la déformation du nuage de points est minimale au sens d'un critère j_p [57] basé sur la moyenne des carrés des distances entre les vecteurs x_k et leurs projections notées x_k^p (telles que $x_k^p = [x_{k1}^p, x_{k2}^p, \dots, x_{kd}^p]^t$) :

$$j_p = \frac{1}{N} \sum_{k=1}^N d^2(x_k, x_k^p)$$

Les résultats de l'ACP peuvent être visualisés en considérant la projection du nuage de points sur les deux ou trois premiers axes factoriels. Autrement dit, en visualisant la matrice des données dans l'espace défini par les deux ou trois premiers vecteurs propres.

Enfin la qualité de la représentation peut être quantifiée par le calcul du pourcentage d'inertie expliqué sur chaque axe factoriel ou sur le sous-espace de dimension d' obtenu. Ce pourcentage d'inertie est défini en fonction des valeurs propres :

- Pourcentage d'inertie sur l'axe i :

$$\Gamma_i = \frac{\lambda_i}{\sum_{j=1}^d \lambda_j}$$

- Qualité de la représentation dans le sous-espace de dimension d' : $\Gamma = \frac{\sum_{j=1}^{d'} \lambda_j}{\sum_{j=1}^d \lambda_j}$

Ce rapport doit tendre vers 1 pour une représentation la plus proche possible de la réalité. Mais en pratique l'utilisateur fixe un seuil de qualité autour de 80%, 90%, puis sélectionne les vecteurs propres correspondant aux valeurs propres ayant permis d'atteindre ce seuil, afin de définir la base du sous-espace.

Une fois le sous-espace $R^{d'}$ défini, il est possible d'évaluer la qualité de la représentativité de la projection x_k^p du vecteur x_k , dans ce sous-espace, en calculant le cosinus de l'angle α_k que forme ce dernier avec R^d :

$$\cos^2(\alpha_k) = \frac{\sum_{j=1}^{d'} (x_{kj}^p)^2}{\sum_{j=1}^d (x_{kj})^2}$$

Ainsi, plus $\cos^2(\alpha_k)$ est proche de l'unité plus la projection de x_k en x_k^p est représentative.

3 Technique d'évaluation de distance

La technique d'évaluation des distances (DET) est bien établie pour la sélection de attributs pertinents, elle est efficace et facile à réaliser. À l'origine, elle est suggérée par Yang, plus tard, elle a été améliorée par Lei. DET sélectionne un sous-ensemble à partir de l'espace original en calculant des facteurs d'évaluation, de tel sorte que tous les attributs sont disposés par ordre décroissant en fonction de leurs facteurs d'évaluation, puis les 'D' premiers attributs peuvent être sélectionnés comme étant une entrée du classifieur.

Supposons qu'un ensemble de paramètres de 'C' conditions est définie par :

$$\{q_{m,c,j} \quad m = 1, 2, \dots, M_c; \quad c = 1, 2, \dots, C; \quad j = 1, 2, \dots, J\}, \quad (\text{II.76})$$

Avec : $q_{m,c,j}$ représente le jn paramètre de la mn échantillon pour la condition cn, M_c : est le nombre des échantillons dans la classe c et J est le nombre de paramètre de chaque échantillon. La méthode de sélection des attributs basée sur la technique d'évaluation de distance se déroule comme suit :

- **Etape 1.** Calcul de la distance moyenne des échantillons de la même classe.

$$d_{c,j} = \frac{1}{M_c(M_c - 1)} \sum_{l,m=1}^{M_c} |q_{m,c,j} - q_{l,c,j}|, \quad l, m = 1, 2, \dots, M_c, \quad l \neq m, \quad (\text{II.77})$$

Puis on obtient la distance moyenne pour C conditions :

$$d_j^{(w)} = \frac{1}{C} \sum_{c=1}^C d_{c,j}. \quad (\text{II.78})$$

-
- **Etape 2.** Définir et calculer le facteur de variance de $d_j^{(w)}$ comme suit :

$$v_j^{(w)} = \frac{\max(d_{c,j})}{\min(d_{c,j})}. \quad (\text{II.79})$$

- **Etape 3.** Calcul de la valeur moyenne de tous les échantillons pour une même condition :

$$u_{c,j} = \frac{1}{M_c} \sum_{m=1}^{M_c} q_{m,c,j}, \quad (\text{II.80})$$

Puis on obtient la distance moyenne entre les échantillons de toutes les conditions :

$$d_j^{(b)} = \frac{1}{C(C-1)} \sum_{c,e=1}^C |u_{e,j} - u_{c,j}|, \quad c, e = 1, 2, \dots, C, \quad c \neq e. \quad (\text{II.81})$$

- **Etape 4.** Définir et calculer le facteur de variance de $d_j^{(b)}$ comme suit :

$$v_j^{(b)} = \frac{\max(|u_{e,j} - u_{c,j}|)}{\min(|u_{e,j} - u_{c,j}|)}, \quad c, e = 1, 2, \dots, C \quad \text{avec } c \neq e. \quad (\text{II.82})$$

- **Etape 5.** Définir et calculer le facteur de compensation comme suit :

$$\lambda_j = \frac{1}{\frac{v_j^{(w)}}{\max(v_j^{(w)})} + \frac{v_j^{(b)}}{\max(v_j^{(b)})}} \quad (\text{II.83})$$

- **Etape 6.** Calcul du rapport entre d_j^b et d_j^w et le multiplier par le facteur de compensation :

$$\alpha_j = \lambda_j \frac{d_j^{(b)}}{d_j^{(w)}} \quad (\text{II.84})$$

Puis on normalise α_j par sa valeur maximale et obtention des critères d'évaluation de distance

$$\bar{\alpha}_j = \frac{\alpha_j}{\max(\alpha_j)} \quad (\text{II.85})$$

De toute évidence, un grand $\bar{\alpha}_j, j = (1, 2, \dots, J)$ signifie que le paramètre correspondant est préférable pour séparer les C conditions. Ainsi, les attributs pertinents peuvent être sélectionnés à partir du $q_{m,c,j}$ selon leurs critères d'évaluation de distance α_j de grande valeur à petite valeur.

4 Matrices de dispersion pour la Mesure de séparabilité entre les classes

Dans le présent travail, nous proposons une nouvelle fonction objective qui quantifie le pouvoir discriminatoire d'un sous-ensemble de dimension D , et ce en utilisant des matrices de dispersion (MD). La fonction objective proposée, définie par l'équation (II.86) est employée pour choisir les attributs qui la maximisent.

$$J = \text{trace} \{ S_w^{-1} S_b \} \quad (\text{II.86})$$

Ici, S_w et S_b : représentent respectivement, la matrice de dispersion intra-classe et la matrice de dispersion inter-classe, elles sont donnés comme suit :

$$S_w = \sum_{i=1}^c S_i \quad (\text{II.87})$$

$$S_b = \sum_{i=1}^c N_i (m_i - m_0)(m_i - m_0)^t \quad (\text{II.88})$$

Avec $S_i = \sum_{x \in D_i} N_i (x - m_i)(x - m_i)^t$ désigne la matrice de dispersion intra-classe pour une classe $i = 1, 2, \dots, c$ et $m_0 = \sum_{i=1}^c N_i m_i$.

'C' Représente le nombre de classes, m_i et N_i désignent respectivement, la moyenne du vecteur et le nombre d'échantillons d'une classe i .

Une grande valeur de 'J' indique que les données représentées par le sous ensemble associé, ont une faible variance dans la même classe et une grande variance entre les classes. Par conséquent, une augmentation de 'J' conduit à une meilleur séparation entre les C conditions.

5 Conclusion

L'objectif de ce chapitre était de discuter des différentes méthodes utilisées pour réduire la dimension des données associées à chaque défaut du roulement et trouver ainsi une signature de taille réduite qui le caractérise, et par conséquent le différencier des autres défauts du roulement appartenant aux autres classes.

Pour cela, nous nous sommes concentrés, dans un premier temps, sur la transformée en ondelettes, en faisant un tour d'horizon sur les ondelettes, et en donnant un aperçu rapide sur la notion échelle, la caractéristique principale qui doit vérifier une onde mère, le célèbre algorithme

de MALLAT ainsi que l'algorithme de paquet d'ondelette. Dans un second temps, nous avons présenté les paramètres statistiques qui sont des outils de suivi et de diagnostic de l'état du roulement, ils sont efficaces et simples à réaliser.

L'analyse en composantes principales (PCA) est ensuite abordée, elle est largement utilisée pour l'extraction des attributs et pour obtenir des informations pertinentes à partir de jeux de données multidimensionnels.

Enfin, nous avons discuté l'algorithme d'évaluation des distances (DET), l'algorithme des essaims de particules binaires et le critère basé sur les matrices de dispersion.

Chapitre III

Les séparateurs à vaste marge

1 Introduction.

Depuis quelques années, de nouvelles techniques d'apprentissage se développent sur la base de la Théorie Statistique de l'Apprentissage de Vapnik et Chervonenkis [58]. Celles-ci visent à estimer la structure des règles de décision d pour une classe D , supposée de VC-dimension h , de probabilité d'erreur $P_e(d)$ minimale, à partir d'un ensemble d'apprentissage de taille n . Selon Vapnik et Chervonenkis, un intervalle de confiance qui lie la probabilité d'erreur P_e et la probabilité d'erreur empirique P_{emp} pour tout détecteur de D peut être formulé comme suit [59] :

$$P_e(d) \leq P_{emp}(d) + \phi(n, h, \epsilon) \quad (\text{III.1})$$

avec une probabilité au moins égale à $1 - \epsilon$, où la largeur de l'intervalle de confiance ϕ est donnée par [59] :

$$\phi(n, h, \epsilon) = \sqrt{\frac{h(\ln \frac{2n}{h} + 1) - \ln \frac{\epsilon}{4}}{n}} \quad (\text{III.2})$$

Remarquons à partir de (III.1) et (III.2) que, lorsque la quantité $\frac{n}{h}$ est importante, l'utilisation du principe de minimisation du risque empirique devient satisfaisante puisque la largeur ϕ de l'intervalle de confiance est faible. Dans le cas contraire, la largeur de l'intervalle de confiance est telle que P_{emp} ne peut donner d'information significative sur $P_e(d)$. Afin de minimiser conjointement les deux termes de la borne de droite de (III.1), on peut par exemple fixer la largeur de l'intervalle de confiance ϕ et minimiser le risque empirique, ou encore fixer la valeur du risque empirique à une valeur très petite, idéalement à 0, et minimiser ϕ . Selon Vapnik, la première méthode est appliquée dans les réseaux de neurones artificielles, où le choix de la structure du réseau permet de fixer ϕ . La deuxième méthode est mise en oeuvre pour obtenir des systèmes décisionnels appelés séparateurs à vaste marge (SVM). Le SVM est un outil de classification qui cherche à minimiser les erreurs d'apprentissage tout en maximisant la marge entre les fron-

tières de chaque classe. La maximisation de la marge est une méthode de régularisation qui réduit la complexité du classifieur. Ce processus produit un ensemble réduit des échantillons faisant partie de données d'apprentissage qu'on appelle communément vecteurs supports. Son comportement est, par ailleurs, conditionné par le type de la fonction noyau utilisé ainsi que les valeurs octroyées à ses paramètres. Le noyau d'un SVM est une fonction symétrique définie positive qui permet la redescription des données dans un espace souvent de grande dimension dans lequel s'opère plus facilement la discrimination des classes. Les valeurs des paramètres du noyau influencent directement la complexité de la fonction de décision du classifieur ainsi que le nombre de vecteurs supports. De nombreux travaux ont démontré la supériorité du SVM sur les méthodes discriminantes classiques telles que les MLP.

Les Machines à Vecteurs Supports sont des méthodes d'apprentissage supervisé qui font appel aux notions de Minimisation du Risque Empirique (MRE) et de Minimisation du Risque Structurel (MRS). Ils sont par essence des classifieurs binaire qui visent à discriminer les échantillons de chaque classe au moyen d'une règle de décision choisie de manière à garantir un maximum de distance entre la frontière de séparation et n'importe quelle observation d'apprentissage.

2 Cas des données linéairement séparables

Soit l'ensemble d'apprentissage $A_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ qui constitue un ensemble de données linéairement séparables, où x_k représente un exemple d'entrée ou une observation, et y_k l'étiquette ou la classe associée que l'on suppose appartenir à $\{-1, +1\}$, l'objectif des SVM est de trouver un hyperplan optimum d'équation :

$$\langle w; x \rangle - \lambda_0 = 0. \quad (\text{III.3})$$

Avec $\langle w; x \rangle$ est le produit scalaire des deux vecteurs w et x .

Dans le cas des séparateurs à vaste marge un hyperplan est dit optimal s'il permet de séparer sans erreur les données d'apprentissage A_n et s'il maximise la distance entre les vecteurs x_k les plus proches parmi ceux de A_n , Appelés vecteurs supports .

Un hyperplan vérifiant de telles conditions est appelé hyperplan à marge maximale.

Il existe une infinité de solutions pour un tel problème. On peut chercher une fonction discriminante dont les paramètres w et λ_0 vérifient [60] :

$$\min_{x \in A_n} |\langle w; x \rangle - \lambda_0| = 1 \quad (\text{III.4})$$

III.2 Cas des données linéairement séparables

Ainsi que nous le verrons par la suite, cette contrainte est préférée à d'autres, car elle simplifie la formulation du problème.

La distance d'un point x_k de l'espace à l'hyperplan d'équation $\langle w; x \rangle - \lambda_0 = 0$ est définie par [60] :

$$\delta_k = \frac{|\langle w; x \rangle - \lambda_0|}{\|w\|} \quad (\text{III.5})$$

Des équations (III.4) et (III.5), on peut constater que la norme $\|w\|$ de 'w' est égale à l'inverse de la distance de l'échantillon le plus proche du plan séparateur défini par $\langle w; x \rangle - \lambda_0 = 0$.

Nous déduisons à partir de l'équation (III.4) les deux inégalités suivantes :

$$\langle w; x \rangle - \lambda_0 \geq +1, \text{ si } y_k = +1 \quad (\text{III.6})$$

$$\langle w; x \rangle - \lambda_0 \leq -1, \text{ si } y_k = -1 \quad (\text{III.7})$$

Ces deux conditions peuvent être résumer par une seule inéquation :

$$y_k(\langle w; x \rangle - \lambda_0) \geq +1, \forall k \in \{1, \dots, n\} \quad (\text{III.8})$$

Des équations (III.5) et (III.8), on déduit que pour tout x_k de A_n , l'inégalité suivante est vérifiée [60],[61] :

$$\delta_k \geq \frac{1}{\|w\|} \quad (\text{III.9})$$

On peut conclure que la valeur de la marge pour laquelle la distance entre l'hyperplan et les échantillons les plus proches de chaque classe sera maximisée est définie par $\frac{1}{\|w\|}$, comme l'illustre la figure (III.1). Ainsi l'hyperplan à marge maximale recherché maximise $\frac{1}{\|w\|}$ ou encore minimise $\frac{\|w\|^2}{2}$ tout en respectant les contraintes (III.8) [59],[60],[61].

2.1 Recherche de la solution

Ce problème peut être résolu en utilisant l'expression duale. On est amené à trouver un point selle $w^*, \lambda_0^*, \alpha^*$ du lagrangien :

$$L(w, \lambda_0, \alpha) = \frac{\|w\|^2}{2} - \sum_{k=1}^n \alpha_k [y_k(\langle w; x \rangle - \lambda_0) - 1] \quad (\text{III.10})$$

où $\alpha = (\alpha_1 \dots \alpha_n)$ désigne les 'n' coefficients de Lagrange ($\alpha_i \geq 0, \forall i$). Au point selle, le lagrangien est minimum pour $w = w^*$ et $\lambda_0 = \lambda_0^*$, et maximum pour $\alpha = \alpha^*$. Le problème primaire

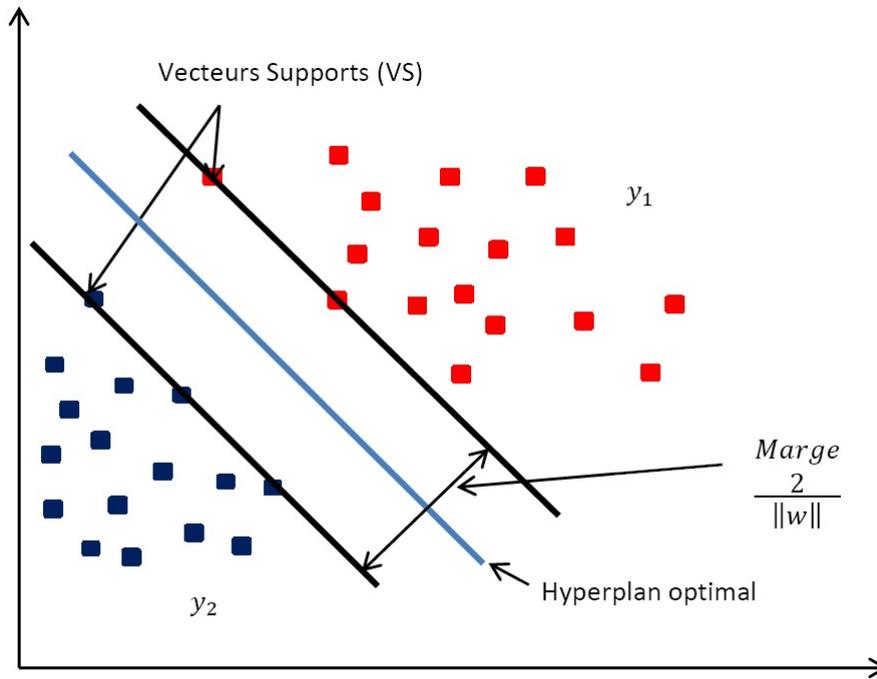


Figure III.1 – Principe des SVM dans le cas où les deux classes sont linéairement séparables. Les vecteurs supports, désignent les échantillons les plus proches de l’hyperplan séparateur.

doit satisfaire les conditions de KKT (Karush-Kuhn-Tucker) consistant à annuler les dérivées partielles du Lagrangien III.10 par rapport aux w et λ [59],[61] :

$$\begin{cases} \frac{\sigma L}{\sigma w} = 0 \Rightarrow w^* = \sum_{k=1}^n \alpha_k^* y_k x_k, \\ \frac{\sigma L}{\sigma \lambda} = 0 \Rightarrow \sum_{k=1}^n \alpha_k^* y_k = 0 \end{cases} \quad (\text{III.11})$$

nous obtenons ainsi le problème dual qui consiste à maximiser par rapport à α la forme suivante du lagrangien [59] :

$$\begin{cases} W(\alpha) = \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,k'=1}^n \alpha_k \alpha_{k'} y_k y_{k'} \langle x_k; x_{k'} \rangle, \\ \text{sous les contraintes } \sum_{k=1}^n \alpha_k y_k = 0 \text{ et } \alpha_k \geq 0. \end{cases} \quad (\text{III.12})$$

L’écriture matricielle du problème d’optimisation (III.12) peut être exprimée comme suit [59] :

$$\begin{cases} W(\alpha) = 0.5 \alpha^T H \alpha - f^T \alpha, \\ \text{sous les contraintes } y^T \alpha = 0 \text{ et } \alpha_k \geq 0, k = 1 \dots n. \end{cases} \quad (\text{III.13})$$

III.2 Cas des données linéairement séparables

avec $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$, H représente la matrice Hessien ($H_{ij} = y_i y_j x_i^T x_j$) et f est un vecteur colonne de dimension n , $f = [1, 1, \dots, 1]^T$.

La résolution du problème d'optimisation quadratique (III.13) permet d'obtenir α_k^* avec $k = 1, \dots, n$ et calculer en suite $\langle w^*; x \rangle = \sum_{k=1}^n \alpha_k^* y_k \langle x_k; x \rangle$. Les conditions de *Kuhn – Tucker* donnent [60] :

$$\alpha_k^* [y_k (\langle w^*; x_k \rangle - \lambda_0^*) - 1] = 0. \quad (\text{III.14})$$

et ce pour toutes les observations x_k de l'ensemble de l'apprentissage A_n . Ainsi, seuls les points x_k qui satisfaites :

$$y_k (\langle w^*; x_k \rangle - \lambda_0^*) = 1. \quad (\text{III.15})$$

ont des coefficients de Lagrange α_k^* non nuls. Ces points désignent les vecteurs supports (VS). Ils sont les échantillons les plus proches de l'hyperplan optimal, on a en particulier $\delta_k^* = 1$. Pour les autres observations, on a $\alpha_k^* = 0$. Sur la figure (III.1), les x_{i-}^* et x_{j+}^* représentent les vecteurs supports (VS), associés à l'une ou l'autre des hypothèses en compétition. On note qu'à partir de deux vecteurs supports x_{i-}^* et x_{j+}^* on peut facilement calculer la valeur du terme de biais optimal λ_0^* correspondant à l'hyperplan optimal, en utilisant l'équation (III.15). Ainsi on a :

$$\lambda_0^* = \frac{1}{2} [\langle w^*; x_{i-}^* \rangle + \langle w^*; x_{j+}^* \rangle]. \quad (\text{III.16})$$

Ceci peut être aussi calculé en fonction des coefficients de Lagrange en combinant les équations (III.11) et (III.16) :

$$\lambda_0^* = \frac{1}{2} \sum_{k=1}^n \alpha_k^* y_k [\langle x_k; x_{i-}^* \rangle + \langle x_k; x_{j+}^* \rangle]. \quad (\text{III.17})$$

Ici, la valeur de λ_0^* peut être calculée à partir de deux vecteurs supports seulement, mais il est parfois conseillé, pour une meilleure robustesse numérique, de la calculer en utilisant tous les vecteurs supports [61].

La fonction de décision sera finalement formulée comme suit :

$$\langle w^*; x \rangle - \lambda_0^* = \sum_{k=1}^n \alpha_k^* y_k \langle x_k; x \rangle - \lambda_0^* \quad (\text{III.18})$$

où λ_0^* est donnée par (III.17).

Pour savoir à quelle classe un nouveau point 'x' appartient, il suffit de trouver le signe de la fonction de décision associée (III.18), afin de voir de quel côté est l'échantillon par rapport à l'hyperplan et pouvoir ainsi prendre une décision.

la fonction de décision est donnée par [61] :

$$D(x) = \sum_{k \in S} \alpha_k^* y_k x_i^T x - \lambda_0^* \quad (\text{III.19})$$

avec S est l'ensemble des indices des supports vecteurs.

une entrée quelconque est alors classée dans :

$$\begin{cases} \text{classe 1} & \text{si } D(x) > 0, \\ \text{classe 2} & \text{si } D(x) < 0 \end{cases} \quad (\text{III.20})$$

3 Séparateurs à vaste marge souple

En pratique, les données peuvent être bruitées et non séparables par un hyperplan. Il convient de modifier l'approche précédente afin de surmonter ces difficultés. La figure (III.2) montre une telle situation. Une solution à ce problème consiste à reformuler les conditions (III.6) et (III.7) en introduisant des paramètres ressort ζ_k permettant de relâcher et d'assouplir les contraintes sur la marge :

$$\langle w; x \rangle - \lambda_0 \geq +1 - \zeta_k, \text{ si } y_k = +1 \quad (\text{III.21})$$

$$\langle w; x \rangle - \lambda_0 \leq -1 + \zeta_k, \text{ si } y_k = -1 \quad (\text{III.22})$$

où $\zeta_k \geq 0$. Cette approche est connue sous le nom de séparateurs à vaste marge et à marge molle (soft margin Support Vector Machine) [59],[62]. Une erreur est commise sur un échantillon x_k par la fonction de décision utilisée si ζ_k est supérieur à 1. $\sum_{k=1}^n \zeta_k$ constitue donc une fonction de perte liée aux erreurs commises sur les observations de l'ensemble d'apprentissage. Le problème d'optimisation doit être reformulé de manière à minimiser la fonction de pertes conjointement avec la maximisation de la marge. Ceci revient à minimiser le problème primal suivant :

$$\frac{1}{2} \|w\|^2 + c \sum_{k=1}^n \zeta_k, \quad (\text{III.23})$$

sous les contraintes (III.21) et (III.22), "c" est interprété comme étant un paramètre de pénalisation des erreurs, préalablement fixé, permettant de contrôler le compromis entre la largeur de la marge et la valeur de l'erreur de classification commise sur l'ensemble d'apprentissage. On parle donc d'une classification à marge souple. Notons qu'il est souvent préférable d'auto-

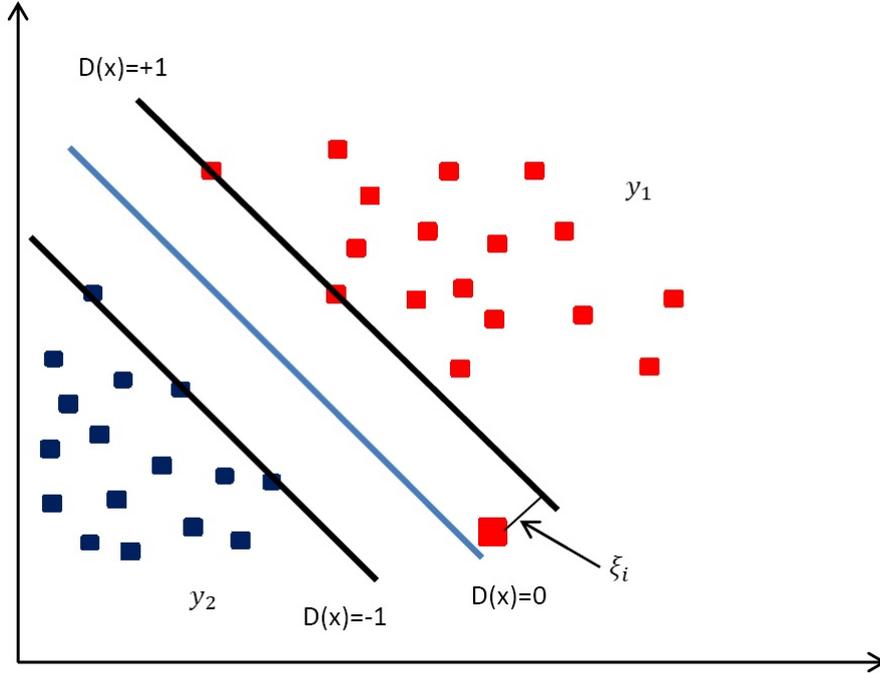


Figure III.2 – Séparateurs à vaste marge souple

riser quelques erreurs, au bénéfice d'une marge plus grande car ces erreurs peuvent être dues à des échantillons aberrants de la classe qui leur est associée. Le fait de donner une valeur importante à 'c' revient à augmenter le coût porté sur les erreurs de classification. La fonction de décision dont les paramètres minimisent (III.23) sous les contraintes (III.21) et (III.22) est aussi appelé hyperplan *optimal généralisé*.

3.1 Recherche de la solution

En introduisant les coefficients de Lagrange α_k et β_k , on est amené à trouver un point selle $w^*, \lambda_0^*, \zeta^*, \alpha^*, \beta^*$ du lagrangien :

$$L(w, \lambda_0, \alpha) = \frac{1}{2} \|w\|^2 + c \sum_{k=1}^n \zeta_k - \sum_{k=1}^n \alpha_k [y_k (\langle w; x \rangle - \lambda_0) - 1 + \zeta_k] - \sum_{k=1}^n \beta_k \zeta_k \quad (\text{III.24})$$

Avec $\alpha = (\alpha_1 \dots \alpha_n)$ et $\beta = (\beta_1 \dots \beta_n)$ désignent les coefficients de Lagrange ($\alpha_i \geq 0$ et $\beta_i \geq 0, \forall i$). Au point selle, la valeur du lagrangien est minimum pour $w = w^*, \zeta = \zeta^*, \lambda_0 = \lambda_0^*$ et maximum pour $\alpha = \alpha^*$ et $\beta = \beta^*$.

Les conditions de *Kuhn – Tucker* donnent :

$$\alpha_k [y_k (\langle w; x \rangle - \lambda_0) - 1 + \zeta_k] = 0 \text{ et } \beta_k \zeta_k = (c - \alpha_k) \zeta_k = 0 \quad (\text{III.25})$$

Chapitre III. Les séparateurs à vaste marge

Les conditions d'annulation des dérivées partielles du lagrangien par rapport aux variables primales, conduit directement aux relations vérifiées par la fonction de décision optimale :

$$w^* = \sum_{k=1}^n \alpha_k^* y_k x_k, \sum_{k=1}^n \alpha_k^* y_k = 0, \alpha_k^* + \beta_k^* = c \quad (\text{III.26})$$

afin d'obtenir α^* . On peut ainsi déduire le problème dual à résoudre, consistant à maximiser la forme suivante du lagrangien :

$$W(\alpha) = \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,k'=1}^n \alpha_k \alpha_{k'} y_k y_{k'} \langle x_k; x_{k'} \rangle \quad (\text{III.27})$$

sous les contraintes

$$\sum_{k=1}^n \alpha_k y_k = 0 \text{ et } 0 \leq \alpha_k \leq c \quad (\text{III.28})$$

Notons ici la similarité des problème duales (III.12) et (III.27). Cependant, le point selle de (III.27) doit être calculé en respectant une contrainte supplémentaire sur les multiplicateurs de Lagrange α_k ($0 \leq \alpha_k \leq c$).

Il suffit de remplacer w^* par sa valeur calculée en utilisant la relation (III.26) pour trouver l'équation de la fonction de décision optimal en fonction des α_k^* .

Comme dans le cas des données linéairement séparables, les vecteurs supports correspondent aux échantillons les plus proches de l'hyperplan optimal, mais de plus aux observations correctement classées par la fonction de décision constituée, il est toutefois possible qu'il existe des échantillons qui soient mal classées, à ces échantillons sont aussi associés des coefficients de Lagrange non-nuls.

4 Elaboration des séparateurs à vaste marge nonlinéaires

Pour permettre aux SVM de répondre à des problèmes complexes de classification, l'algorithme initial présenté ci-dessus a été généralisé pour constituer des fonctions de décision non-linéaires, voir figure (III.3). L'extension au cas non-linéaire, consiste à transformer les échantillons à l'aide d'une fonction ϕ , puis en applique un hyperplan linéaire [62],[63],[64], voir figure (III.4).

En appliquant la relation (III.27) dans l'espace engendré par ϕ , le problème dual devient :

$$W(\alpha) = \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,k'=1}^n \alpha_k \alpha_{k'} y_k y_{k'} \langle \phi_k; \phi_{k'} \rangle$$

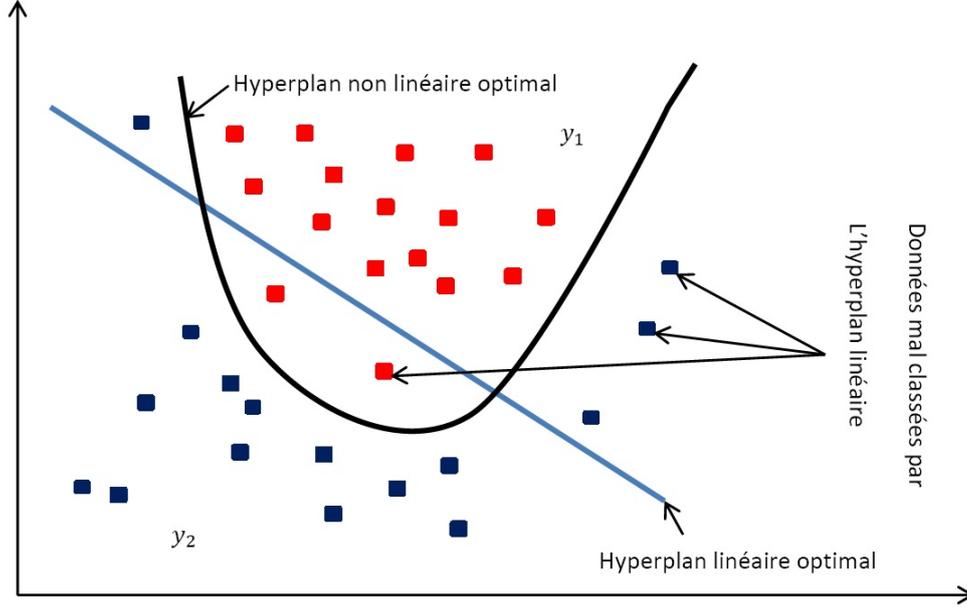


Figure III.3 – Séparateurs à vaste marge non linéaire

$$= \sum_{k=1}^n \alpha_k - \frac{1}{2} \sum_{k,k'=1}^n \alpha_k \alpha_{k'} y_k y_{k'} \kappa(x_k, x_{k'}) \quad (\text{III.29})$$

sous les contraintes $\sum_{k=1}^n \alpha_k y_k = 0$ et $0 \leq \alpha_k \leq c$ où κ désigne la fonction noyau tel que $\kappa(x_k; x_{k'}) = \langle \phi_k; \phi_{k'} \rangle$. On obtient l'équation de la fonction de décision optimal :

$$w^* \cdot \phi(x) - \lambda_0^* = \sum_{k=1}^n \alpha_k^* y_k \kappa(x_k, x) - \lambda_0^* \quad (\text{III.30})$$

Avec :

$$\lambda_0^* = \frac{1}{2} \sum_{k=1}^n \alpha_k^* y_k [\kappa(x_k; x_{i-}^*) + \kappa(x_k; x_{j+}^*)]. \quad (\text{III.31})$$

Pour décider à quelle classe appartient une observation 'x', il suffit de trouver le signe de la fonction de décision :

$$D(x) = \sum_{k \in S} \alpha_k^* y_k \kappa(x_k, x) - \lambda_0^* \quad (\text{III.32})$$

avec λ_0^* est donnée par (III.31).

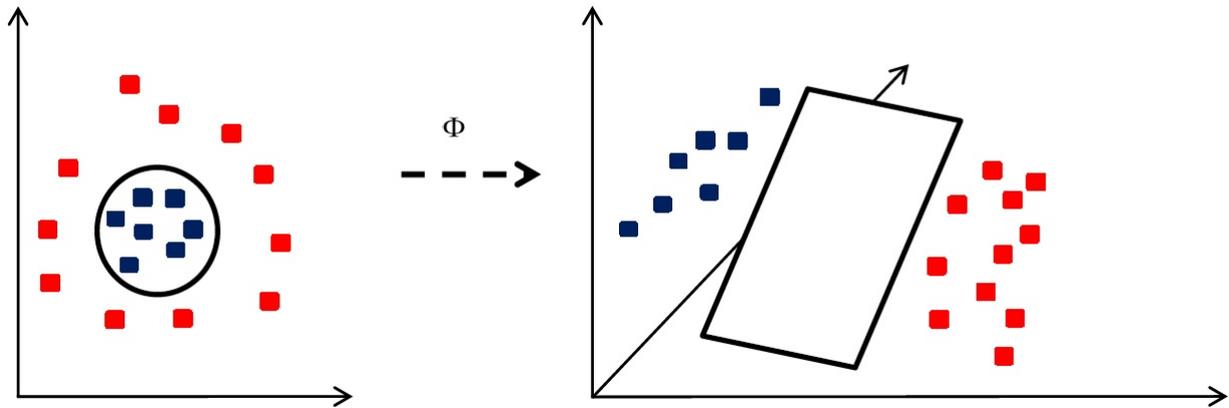


Figure III.4 – Transformation de l'observation par une fonction noyau reproduisante

5 Exemples des Fonctions noyaux

La transformation non-linéaire ϕ est implicitement définie dans l'équation (III.30) par le choix d'une fonction noyau κ , à partir duquel il est possible de redécrire le problème dans un espace de plus grande dimension, dans lequel les observations des deux classes deviennent linéairement séparable.

5.1 Noyaux polynômiaux

Le noyau polynômial est de la forme [59] :

$$\kappa(x_1, x_2) = (1 + \langle x_1; x_2 \rangle)^q \quad (\text{III.33})$$

Avec q est le degré du polynôme. Cette fonction réalise implicitement un produit scalaire des observations.

5.2 Noyaux gaussiens

Le noyau gaussien est un noyaux de type *radial*, il représente une mesure de similarité entre les observations $(x_1 - x_2)$. Ces noyaux sont définies sous la forme suivante[59] :

$$\kappa(x_1, x_2) = \exp(-\|x_1 - x_2\|^2 / \sigma^2) \quad (\text{III.34})$$

Avec σ^2 est *la variance*, c'est un paramètre de **régulation** qui doit être choisi judicieusement.

5.3 Noyaux exponentiels

Le noyau exponentiel est un autre exemple de noyau de type *radial*. Il est défini par [59] :

$$\kappa(x_1, x_2) = \exp(- \|x_1 - x_2\| / \sigma^2) \quad (\text{III.35})$$

5.4 Noyaux sigmoïdaux

Le noyau sigmoïdal est défini comme étant [59] :

$$\kappa(x_1, x_2) = \tanh(\alpha_0 \langle x_1; x_2 \rangle + \beta_0) \quad (\text{III.36})$$

Ce noyau dépend de deux paramètres α_0 et β_0 ce qui peut poser des problèmes lors de sa mise en oeuvre.

6 Multi classification par les séparateur à vaste marge

Comme il est indiqué précédemment, les séparateurs à vaste marge sont conçus pour résoudre les problèmes de deux classes, or les problèmes rencontrés en pratique, sont des problèmes de plusieurs classes. Une extension du problème au multi classification est donc nécessaire. Il existe plusieurs travaux pour construire des classificateurs multi-classes.

Nous discutons dans la suite de deux types de séparateurs à vaste marge qui traitent des problèmes multi classe :

- un contre tout support vector machines.
- un contre un support vector machines.

L'approche "un contre tous", consiste à construire, pour un problème de classification de n classes, n classifieurs binaires. Pour le $i^{\text{ème}}$ classifieur binaire, la classe i est séparé de toutes les autres classes [61],[65].

La méthode "un contre un" transforme le problème de n classes en $n(n - 1)/2$ fonctions de décisions qui couvrent tout les paires de classe [61],[65].

6.1 Multi classification par l'approche 'un contre tous'

Dans cette section on discute l'approche "un contre tous" avec une fonction de décision discrète.

Considérons un problème de n classes, pour l'approche "un contre tous" , on sera amené à chercher n fonctions de décision discrètes qui séparent une classe de toutes les autres classes.

la i^{eme} fonction de décision discrète optimale qui discrimine la classe i de toutes les autres classes est [61] :

$$D_i(x) = w_i^* \phi(x) - \lambda_{0i}^* \quad (\text{III.37})$$

en classification , si on a pour un vecteur d'entrée x :

$$D_i(x) > 0 \quad (\text{III.38})$$

est satisfaite pour une seule i , x est classé dans la classe i , car seul le signe de la fonction est utilisé , la décision est discrète [61].

6.2 Multi classification par l'approche 'un contre un'

Dans cette section on discute l'approche "un contre un".

Dans la méthode "un contre un", les fonctions de décision sont déterminées à partir des combinaisons des deux classes. Afin de déterminer une fonction de décision pour une classe paire, nous utilisons seulement les observations d'apprentissage pour les deux classes correspondantes. Ainsi, dans chaque session d'apprentissage, le nombre de données d'apprentissage est considérablement réduit par rapport aux séparateur à vaste marge de type "un contre tous", qui utilisent tous les échantillons d'apprentissage, mais le nombre de fonctions de décision est $n(n-1)/2$ comparativement avec n de "un contre tous", où n est le nombre de classes.

Soit la fonction de décision pour la classe i contre la classe j est [61] :

$$D_{ij}(x) = w_{ij}^* \phi(x) + \lambda_{0ij}^* \quad (\text{III.39})$$

Avec $D_{ij} = -D_{ji}$.

Les régions :

$$R_i = \{x \mid D_{ij}(x) > 0, j=1, \dots, n, j \neq i\} \quad (\text{III.40})$$

ne se chevauchent pas, et si $x \in R_i$, nous classons x dans la classe i [61].

7 Conclusion

Le SVM est un modèle discriminant qui tente de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes. La maximisation de la marge est une méthode de régularisation qui réduit la complexité du classifieur. Ce processus produit un

ensemble réduit de prototypes faisant partie de l'ensemble d'apprentissage qu'on appelle communément vecteurs de support. Son comportement est, par ailleurs, conditionné par le type du noyau utilisé et par les valeurs octroyées à ses paramètres. Le noyau d'un SVM est une fonction symétrique défini-positive qui permet de projeter les données dans un espace transformé de grande dimension, dans lequel s'opère plus facilement la séparation des classes.

Comme nous avons vu dans ce chapitre, le SVM procure un cadre théorique qui résout le problème du choix des paramètres du MLP, puisque celui-ci se fait automatiquement. La robustesse et l'efficacité du SVM à traiter des données de nature et de complexité arbitraire a été établie sur beaucoup d'applications de reconnaissance de formes. Les machines à noyaux en général et les SVM en particulier, apportent un certain nombre de réponses aux problèmes évoqués par les MLP.

Chapitre IV

Détection des défauts des roulements par analyse vibratoire

1 Introduction

Dans ce chapitre nous effectuerons un panorama de deux techniques mises en oeuvre pour analyser et traiter les signaux vibratoires en vue de la détection et du diagnostic des défauts de roulements. Les deux techniques seront analysées et discutées, essentiellement en ce qui concerne leurs efficacités et leurs coûts d'implantation. Nous proposerons alors deux systèmes automatiques pour le traitement des signaux vibratoires, afin de détecter les défauts de roulements. Enfin nous appliquerons ces systèmes sur une base de données reconnue.

2 Les données expérimentales

Le banc d'essai de l'université de Case Western Reserve a pour but de réaliser les expériences nécessaires à la détection des défauts de roulements. Une schématisation de l'installation du banc d'essai est illustrée en figure (IV.1), le banc d'essai est constitué d'un moteur 2 hp (à gauche), un accouplement (au centre), un dynamomètre (à droite), et l'électronique de commande (non représentée) [66].

Les défauts ont été provoqués pour des roulements rigides à bille de type 6205-2RS-EM-SKF, soit sur la bague intérieure, extérieure ou sur les billes par Électro-érosion appelée aussi EDM (Electrical Discharge Machining) avec des diamètres de défaut de 0,007 , 0,014 , 0,021 et 0,028 pouces. les Roulements défailants ont été réinstallés dans le moteur de test, et les données de vibration ont été enregistrées pour des charges de moteur de 0 à 2.2370 KW (vitesse moteur de 1730 à 1797 tour par minute). Les données de vibration ont été recueillies à l'aide d'un accéléromètre, qui a été placé dans la position de 12 heures à la fin de la carcasse du moteur.

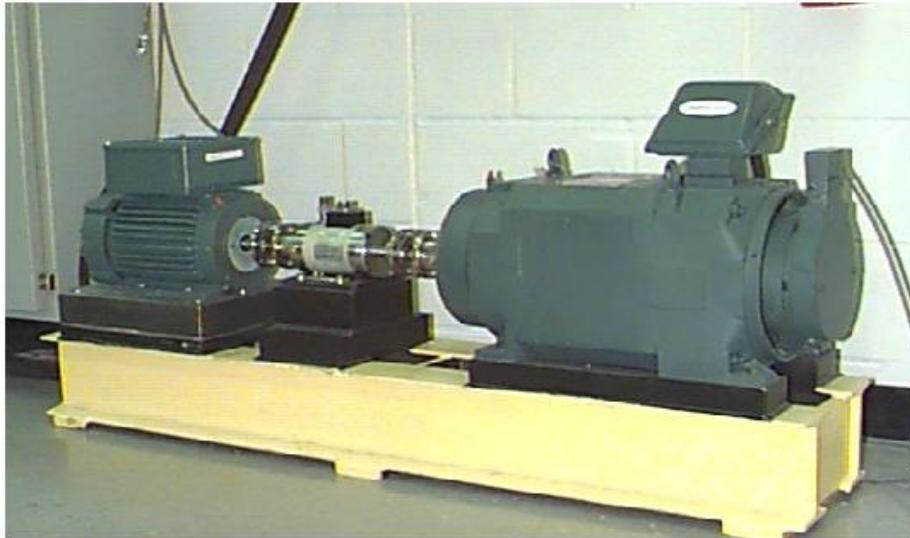


Figure IV.1 – Banc d'essai

Des signaux de vibration ont été recueillis à l'aide d'un enregistreur DAT 16 canaux, et ont été traités dans un environnement Matlab. Tous les fichiers de données sont en format Matlab (*. Mat) . Les données numériques ont été recueillis à 12.000 échantillons par seconde.

La base des données comprend 280 échantillons, couvrant sept conditions de fonctionnement du roulement à savoir : roulement sain (N), roulement avec un défaut billes de 0.1778mm (BF1), roulement avec un défaut billes de 0.5334mm (BF2), roulement avec un défaut bague interne de 0.1778mm (IF1), roulement avec un défaut bague interne de 0.5334mm (IF2), roulement avec un défaut bague externe de 0.1778mm (OF1) et roulement avec un défaut bague externe de 0.5334mm (OF2).

Pour chaque classe; Vingt-six (26) échantillons sont utilisés pour l'apprentissage du classifieur et quatorze (14) échantillons pour le test.

2.1 Données temporelles

La représentation temporelle des signaux vibratoires est illustrée par la figure (IV.2).

2.2 La première approche

La première méthode proposée pour le diagnostic des défauts dans les roulements à bille est illustrée dans la figure IV.3 . Elle contient essentiellement deux étapes : l'extraction des paramètres pertinents, et la classification.

Dans la première phase les paramètres statistiques et l'algorithme du paquet d'ondelette sont

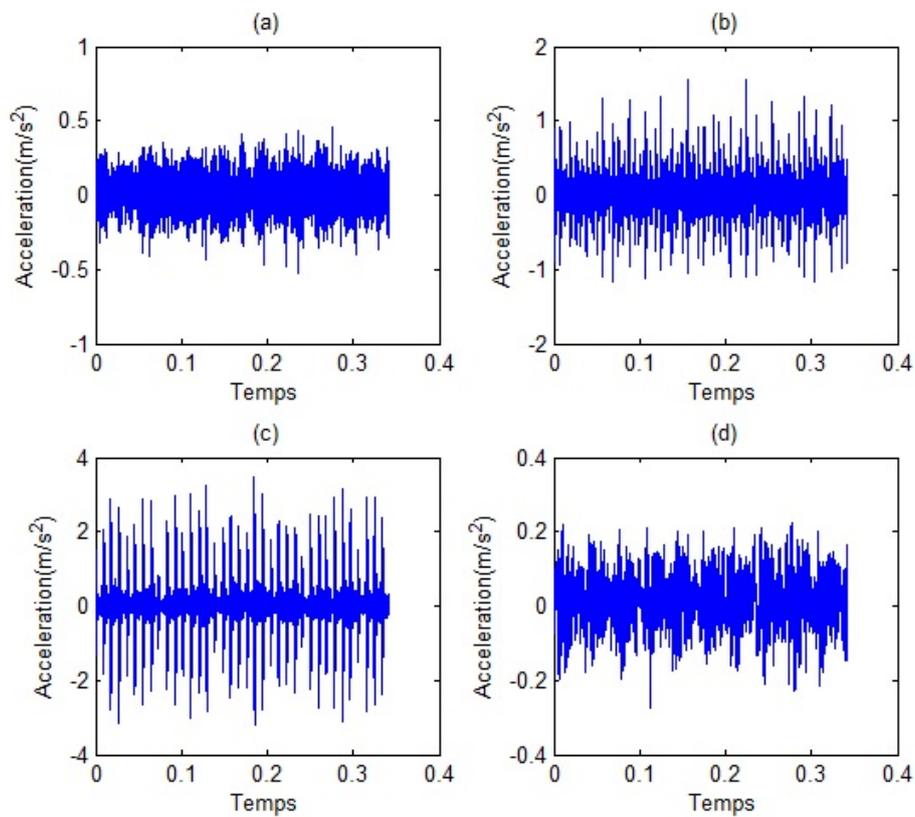


Figure IV.2 – Représentation temporelle des signaux vibratoire d'un roulement : (a) avec défaut de bille, (b) avec défaut de bague intérieure, (c) avec défaut bague extérieure, (d) sain

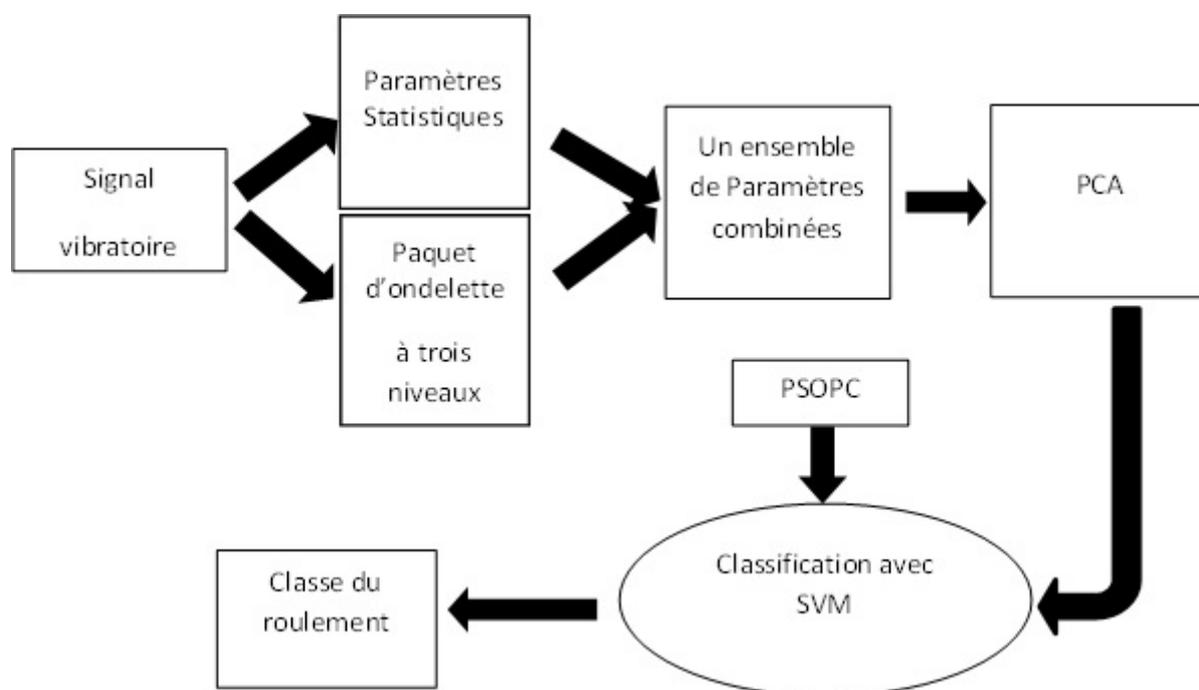


Figure IV.3 – Procédure de la première méthode proposée pour le diagnostic des défauts dans les roulements à bille.

employés pour l'extraction des paramètres. En effet, on a obtenu un vecteur de paramètre global contenant 15 attributs pour chaque état. Ensuite, l'analyse en composante principale est appliquée pour la réduction des dimensions du vecteur d'entrée.

Bien que performante sur beaucoup de problèmes, les deux principales lacunes de la transformée en ondelettes sont le choix de l'ondelette mère et le niveau de décomposition [13], [67], [68]. Pour résoudre la première lacune, cinq types d'ondelettes sont considérées à savoir (ondelette Daubechies (db5 , db8), ondelette Reverse bi orthogonal (rbio5.5), ondelette symlets (sym2) et l'ondelette coiflet (coif5). Ensuite, le critère d'Energie d'Ondelettes Maximale Relative (Maximum relative wavelet energy) (MRWE) est utilisé pour la sélection de l'ondelette la plus appropriée.

Les résultats de calcul sont représentés par la figure (IV.4).

À partir des résultats obtenus, nous pouvons conclure que l'ondelette Daubechies DB8 est celle la plus appropriée, avec une moyenne maximale de MRWE égale à 0.5503. Le niveau maximale de décomposition est égale à 3, il est calculé par :

$$n \leq \log_2\left(\frac{F_{Max}(s)}{3F_c}\right) \quad (IV.1)$$

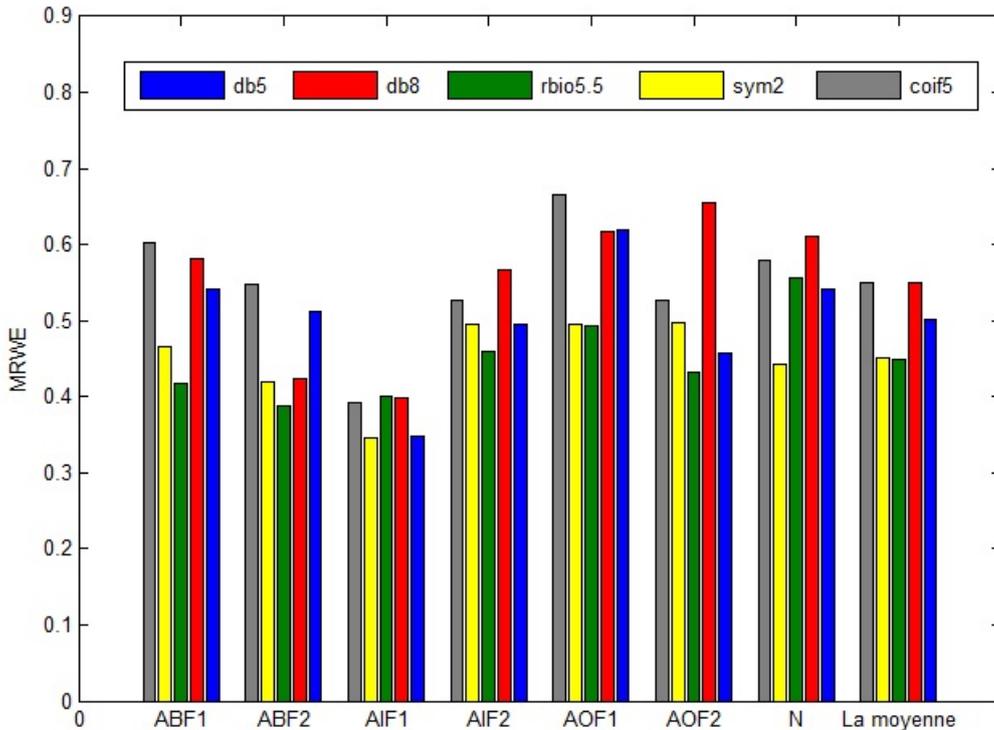


Figure IV.4 – Energie d'Ondelettes Maximale Relative pour différent conditions de roulement

Avec : n : le niveau de décomposition, F_{Max} : la fréquence maximale du signal vibratoire, et F_c la fréquence caractéristique de défaut.

L'extraction basée sur l'algorithme de paquets d'ondelettes, est présentée en quatre étapes comme suit[5] :

- Décomposition du signal vibratoire par une ondelette Daubechies (db8) en trois niveaux, et extractions des coefficients des huit bandes de fréquence à partir de la troisième couche, (figure(IV.5) et figure(IV.6))
- Le signal dans chaque bande de fréquence est extrait, et les coefficients de décomposition en paquets d'ondelettes ont été reconstruit ; D_{30} présente le signal reconstruit de d_{30} . D_{31} présente le signal reconstruit de d_{31} et ainsi de suite. Le signal original est défini comme suit :

$$D = D_{30} + D_{31} + D_{32} + D_{33} + D_{34} + D_{35} + D_{36} + D_{37}.$$

- L'énergie du signal de chaque bande de fréquence est calculée par : $P_{3j} = \sum_{k=1}^n |d_{jk}|^2 / E$.

Avec $E = \sum_{j=1}^8 \sum_{k=1}^n |d_{jk}|^2$.

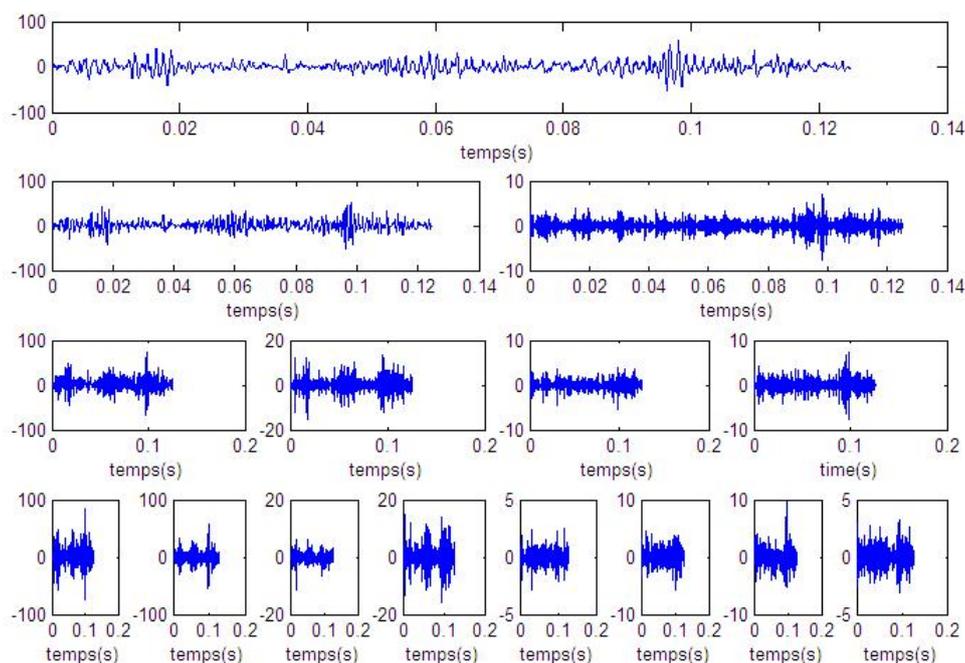


Figure IV.5 – Paquet d’ondelette d’un roulement sain (N).

– Le vecteur "x" a été construit en fonction de l’énergie de chaque bande de fréquence :

$$x = [P_{30}, P_{31}, P_{32}, P_{33}, P_{34}, P_{35}, P_{36}, P_{37}]$$

2.2.1 Analyse des résultats expérimentaux

Le tableau (IV.1) donne les résultats de classification, en utilisant un vecteur de paramètres compressés et composé des : deux premières, trois premières ou des quatre premières composantes principales. A partir du tableau (IV.1), on constate que les trois premières composantes principales sont suffisantes pour obtenir de bons résultats.

Quelques vecteurs d’entrées ainsi que leurs résultats de classification pour différentes condi-

| Taille du vecteur d’entrée | σ | c | Taux de validation% | Taux de test% |
|----------------------------|----------|----------|---------------------|---------------|
| 2 | 0.9137 | 171.3802 | 98.35 | 98.98 |
| 3 | 1.0508 | 787.9622 | 100 | 100 |
| 4 | 0.9177 | 378.1020 | 100 | 100 |

Tableau IV.1 – Résultats de classification pour différentes PCs.

tions de roulement, sont donnés dans le tableau (IV.2).

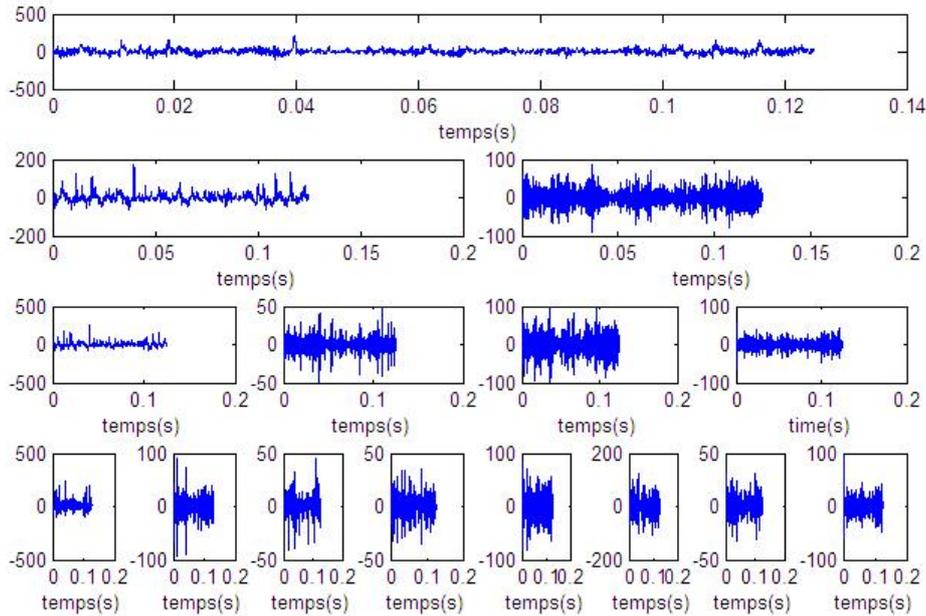


Figure IV.6 – Paquet d’ondelette d’un roulement avec un défaut bague extérieure (OF1).

Après une réduction de dimension par l’algorithme de l’Analyse en Composantes Princi-

| Entrée | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|--------|--------|--------|---------|---------|---------|---------|
| PC1 | 9.1141 | 5.9640 | 3.7421 | -1.7840 | -0.8955 | -23.722 | 10.1288 |
| PC2 | 0.4441 | 1.2340 | 0.0895 | -1.9907 | -2.8952 | 0.3143 | 0.9108 |
| PC3 | 0.0010 | 1.0766 | 0.5802 | 1.4671 | -0.8193 | -0.3941 | -0.6745 |
| Etat | N | BF1 | BF2 | IF1 | IF2 | OF1 | OF2 |
| Résultat | N | BF1 | BF2 | IF1 | IF2 | OF1 | OF2 |

Tableau IV.2 – Vecteur d’entrée et résultat de classification.

pales(ACP), La base de donnée est représentée en trois dimensions afin de visualiser la distribution des sept classes, figure(IV.7), figure(IV.8) . On remarque que les données sont bien séparées et prêtes à être classées.

Le tableau IV.3 montre les performances de classification pour ce problème de classification en utilisant :

- La méthode proposée : (PSOPC-SVM) basée sur l’extraction des paramètres pertinents par PCA).
- PSOPC-SVM en utilisant un vecteur d’entrée composé de trois paramètres choisis aléatoirement.

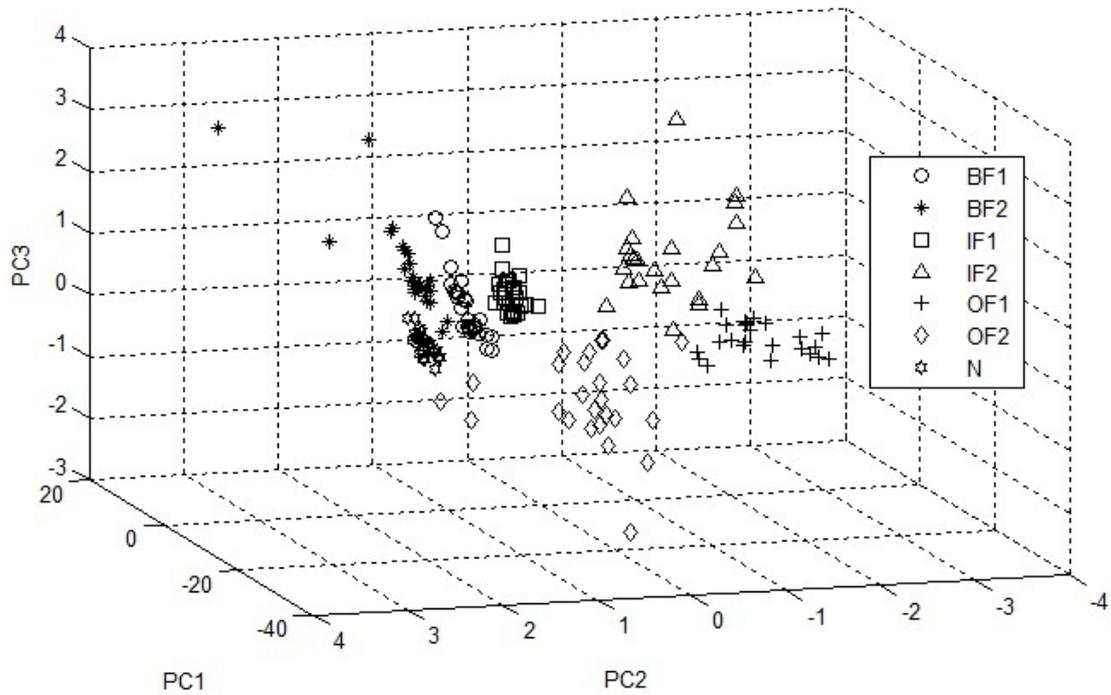


Figure IV.7 – Trois composantes (PCs) obtenues après l'application du PCA. Données apprentissage.

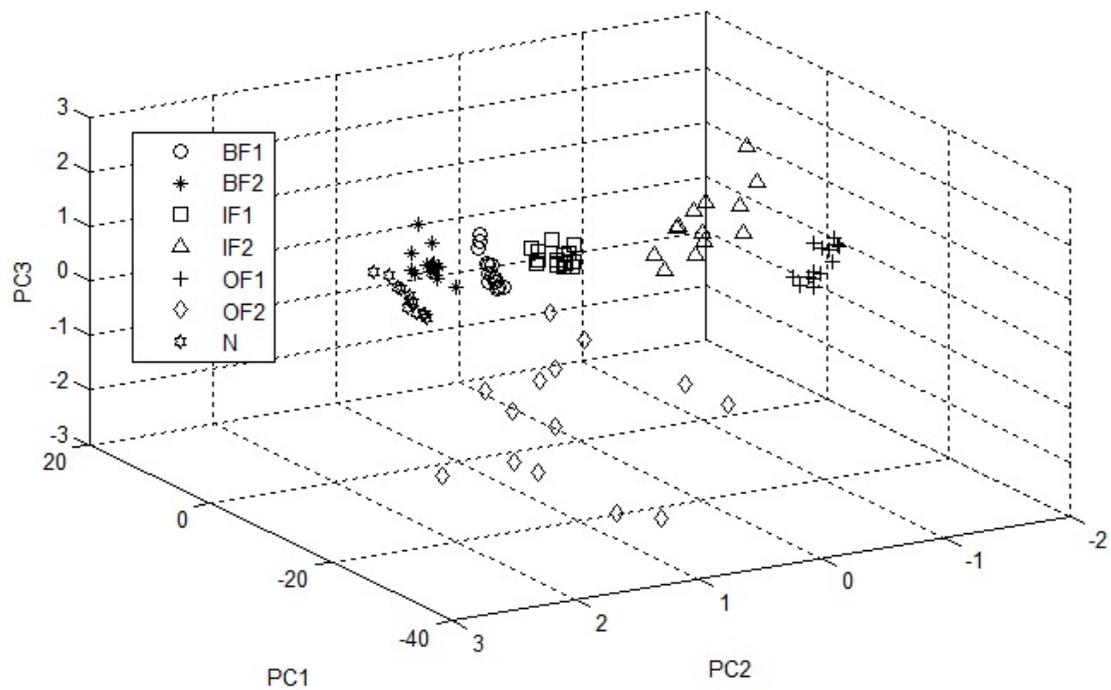


Figure IV.8 – Trois composantes (PCs) obtenues après l'application du PCA. Données Test.

- PSOPC-SVM basé sur le vecteur d’entrée globale (15 attributs).

| Méthode d’extraction des paramètres | Taux de Validation% | Taux de test% |
|--|---------------------|---------------|
| SD P_{32} IMF | 99.45 | 98.98 |
| RMS CRF Kur | 99.45 | 89.80 |
| Kur P_{31} Min | 100 | 95.92 |
| Vecteur d’entrée globale (15 paramètres) | 100 | 100 |
| La première approche proposée | 100 | 100 |

Tableau IV.3 – Résultats de classification pour différents vecteurs d’entrés.

À partir des résultats obtenus, nous pouvons conclure :

- Les meilleurs taux de validation (100%) et de test (100%) sont obtenus en utilisant l’approche proposée du SVM-PSOPC basée sur l’extraction des paramètres pertinents par PCA, où tous les cas sont correctement classés.
- Lorsque nous utilisons des vecteurs d’entrées sélectionnés aléatoirement, le taux de test diminue jusqu’à 89.80%, il est donc difficile de sélectionner le vecteur d’entrée qui donne les meilleurs résultats.
- Lorsque nous utilisons le vecteur d’entrée global (15 paramètres), un bon taux de classification est obtenu, cela s’explique par la capacité de SVM à résoudre les problèmes de classification complexes.

2.3 La seconde approche

La seconde approche proposée pour le diagnostic des défauts dans les roulements à bille, est représentée dans la figure IV.9 . Elle contient essentiellement trois étapes : l’extraction des paramètres, la sélection des attributs pertinents, et la classification.

La procédure de diagnostic basée sur OEPB se déroule comme suit :

- **Etape 1.** Prétraitement du signal vibratoire, par l’utilisation de la transformé en ondelette discrète.
- Extraction des sept paramètres statistiques du signal original et de toutes les bandes de fréquences résultantes. En conséquence, on a obtenu un ensemble de paramètres contenant 49 attributs pour chaque états.
- **Etape 2 : Sélection des attributs pertinents** : Cette phase est exécutée en deux étapes, dans la première étape l’algorithme DET est utilisé pour sélectionner les attributs

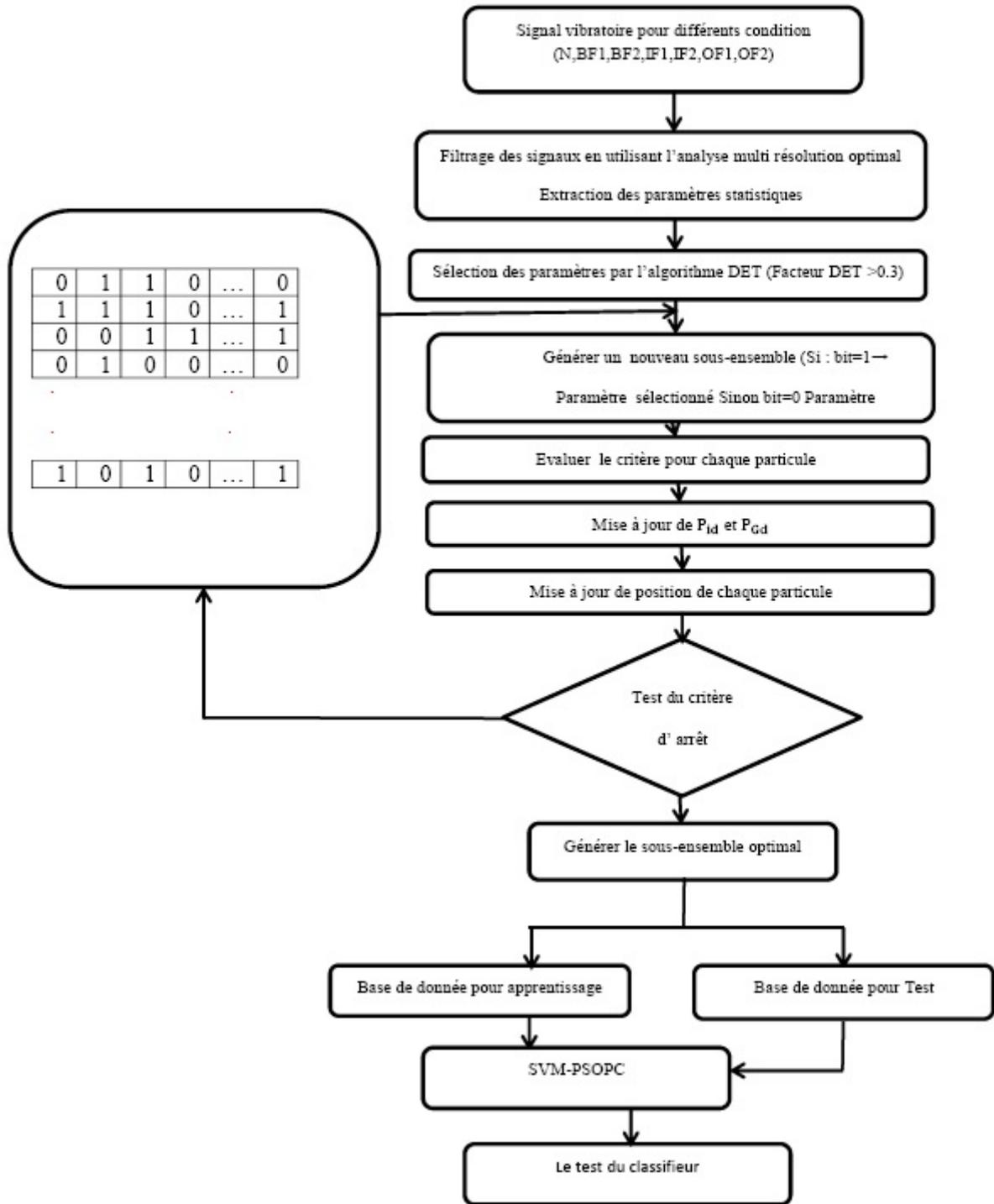


Figure IV.9 – Procédure de la seconde méthode proposée pour le diagnostic des défauts dans les roulements à bille.

dont le facteur DET est supérieur à 0,3. Dans la deuxième étape, l'algorithme OEPB est appliqué pour sélectionner le meilleur D-dimensionnel sous-ensemble, et ce en utilisant un nouveau critère pour la mesure de séparabilité des classes, basé sur les matrices de dispersion comme étant une fonction objective.

- **Étape 3.** Initialisation aléatoire de la position de chaque particule avec des valeurs binaires (0, 1), et générer un nouveau sous-ensemble composé uniquement des attributs auxquels le bit 1 a été assigné.
- **Étape 4. Évaluation de la fonction objective :** évaluer la fonction objective de chaque particule de l'essaim et calculer le critère 'J' donné par l'équation (II.86).
- **Étape 6.** Mettre à jour P_i et P_G selon le plus grand 'J'.
- **Étape 7. Manipulations de particules :** mise à jour de la position de chaque particule en utilisant les formules (II.68) et (II.71).
- **Étape 8. Test de la condition d'arrêt :** si la condition d'arrêt (nombre maximum d'itérations) est satisfaite, arrêter l'algorithme, et enregistrer le sous-ensemble optimal, sinon, aller à l'étape 4.
- **Étape 9.** Apprentissage du classifieur SVM-PSOPC en utilisant le sous-ensemble optimal comme étant une entrée. Ensuite calculer le taux de validation.

2.3.1 Analyse des résultats expérimentaux

Les paramètres du domaine temporel statistique, tels que : la variance, le kurtosis, etc., sont efficaces pour le diagnostic des défauts de roulement en raison de leur sensibilité au changement de forme du signal. Malheureusement, ils sont généralement affectés par le bruit. Pour améliorer la sensibilité des paramètres statistiques, on utilise la transformée en ondelettes, en tant qu'un outil puissant de filtrage.

Pour sélectionner l'ondelette mère la plus appropriée, cinq types d'ondelettes sont considérées à savoir (ondelette Daubechies (db44 , db8), ondelette Reverse bi orthogonal (rbio5.5), ondelette symlets (sym2) et l'ondelette coiflet (coif5). Ensuite, le critère MRWE est utilisé pour la sélection de l'ondelette optimale. Les résultats de calcul sont illustrés dans la figure (IV.10). À partir des résultats obtenus, on constate que l'ondelette Daubechies DB44 est celle la plus appropriée avec une moyenne maximale de MRWE égale à 0.424.

Le niveau maximale de décomposition est égale à 3, il est calculé en utilisant l'équation (IV.1).

Les résultats de prétraitement de quelques signaux vibratoire par l'analyse en ondelette, sont représentées par les figures (IV.11) et (IV.12).

A partir de ces figures, on constate que les signaux résultants changent d'un niveau de dé-

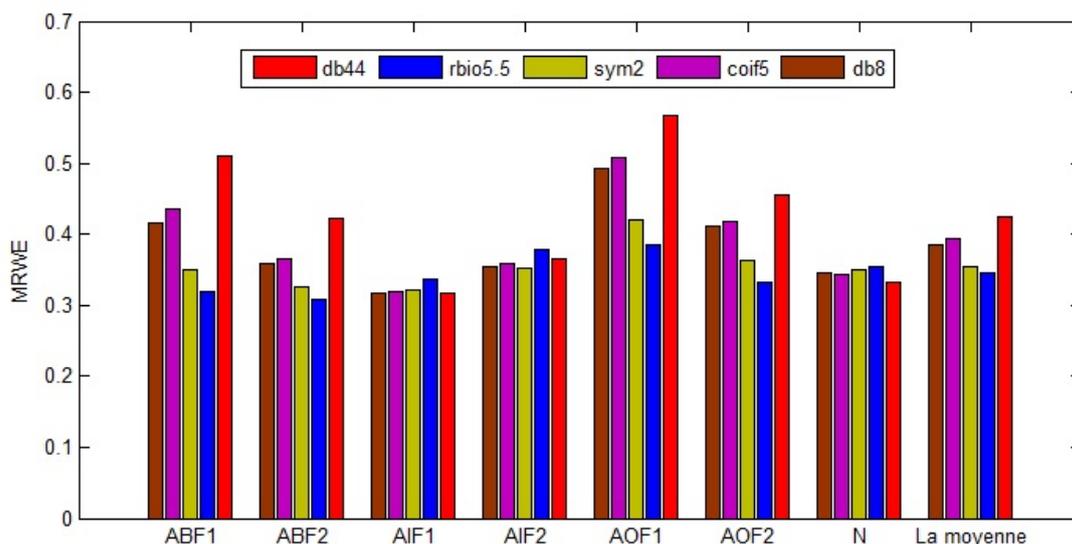


Figure IV.10 – Energie d'Ondelettes Maximale Relative pour différent conditions de roulement

composition à un autre, pour la même condition du roulement. Ils se diffèrent également d'un état à un autre.

Les paramètres statistiques ne sont pas alors seulement extraits des signaux de vibration originaux, mais aussi de toutes les bandes de fréquences obtenues grâce à une analyse DWT optimale. Cependant, si tous les attributs extraits sont utilisés, les performances du classificateur seront affectées. Pour surmonter ces lacunes, certains attributs qui définissent clairement les différentes conditions du roulement, seront sélectionnés et ce en utilisant l'algorithme DET. La figure (IV.13) illustre les résultats obtenus, après l'application de l'algorithme DET. De la figure (IV.13), nous pouvons constater que la sensibilité des paramètres statistiques est largement dépendante du rapport signal sur bruit. Elle diffère d'une bande de fréquence à une autre, prenons comme exemple :

- Le meilleur facteur DET du RMS est obtenu à partir de la deuxième approximation.
- Le meilleur facteur DET du kurtosis est obtenu à partir du troisième détail.
- Le meilleur facteur DET du facteur de crête est obtenu à partir du signal original.

Pour évaluer les performances de l'algorithme DET, les attributs se font entrer un par un au classifieur SVM-OEPCP et ce en fonction de leurs facteurs DET. Le taux de test en fonction du nombre d'attributs est représenté sur la figure (IV.14).

En observant la tendance variable du taux de test en fonction du nombre des attributs, le meilleur taux de classification (100%) est obtenu en utilisant les quatorze (14) premiers paramètres.

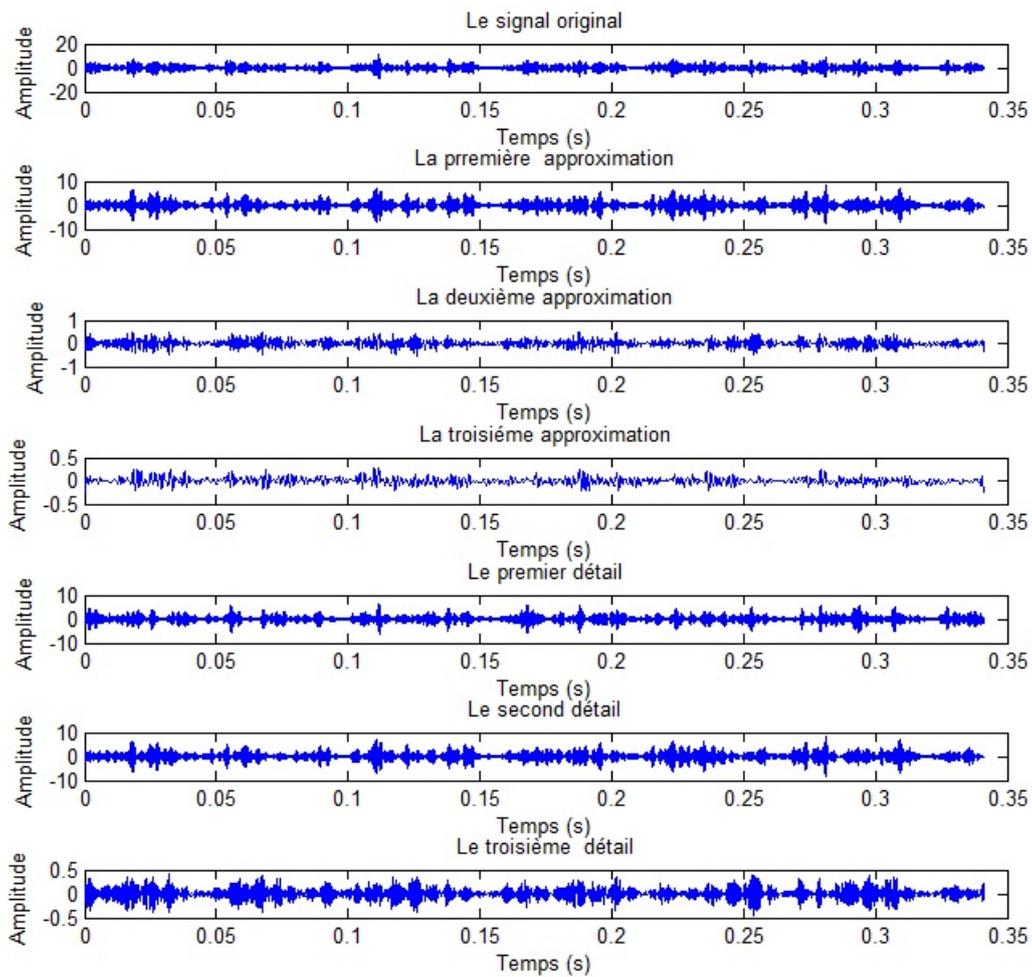


Figure IV.11 – Application d'un prétraitement en ondelette optimal. Roulement avec un défaut bille de 0.1778mm (BF1).

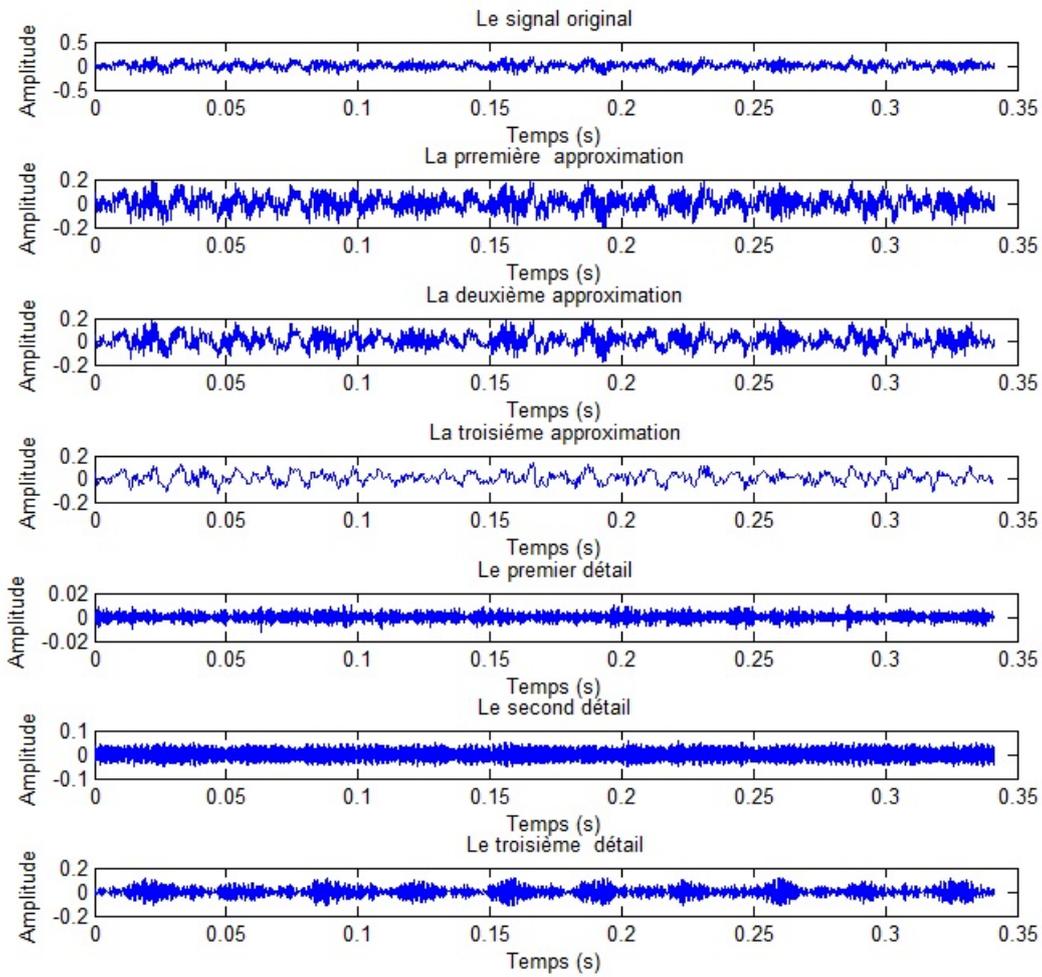


Figure IV.12 – Application d'un prétraitement en ondelette optimale. Roulement avec un défaut bague intérieure de 0.5334mm (IF2).

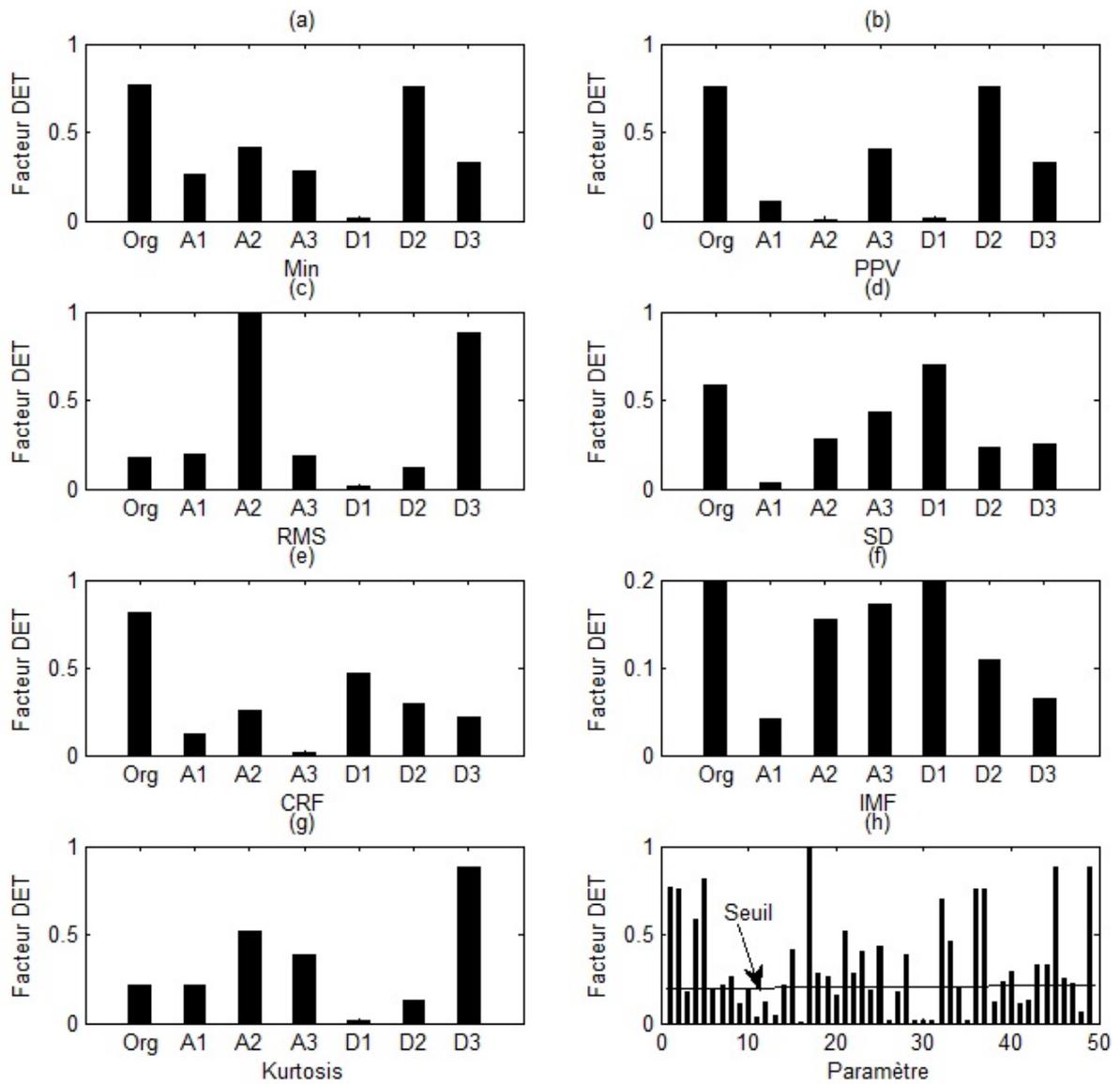


Figure IV.13 – Facteur DET de différents niveaux pour les sept paramètres statistiques. (a) le minimum, (b) la valeur crête à crête, (c) La valeur efficace, (d) la déviation standard, (e) le facteur de crête, (f) le facteur d'impulsion, (g) le kurtosis (Kur), (h) Tous les paramètres.

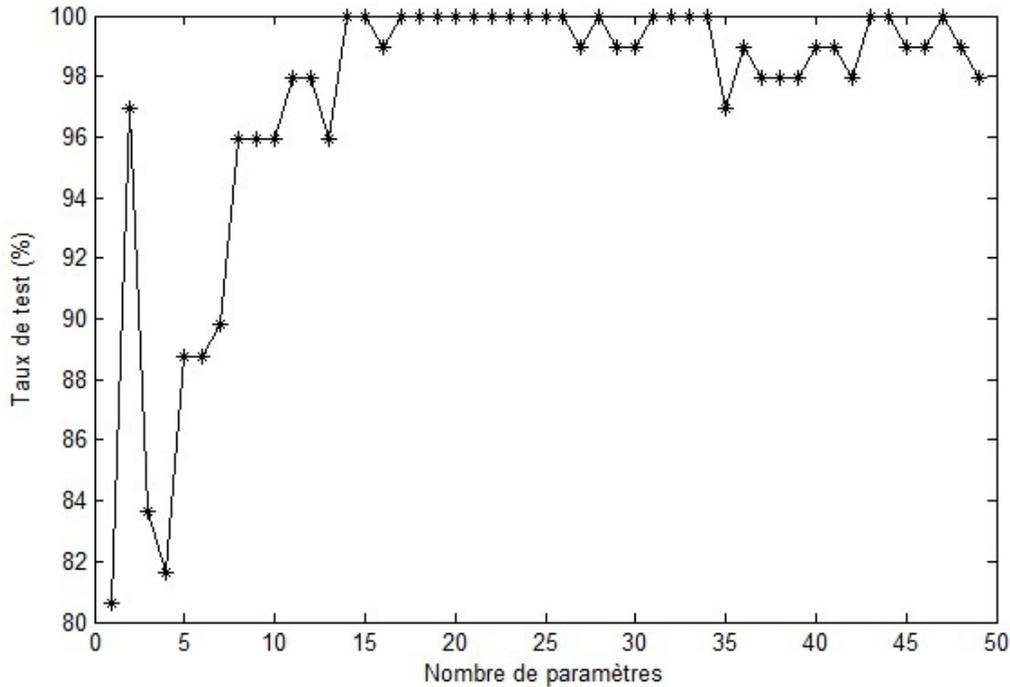


Figure IV.14 – Variation du taux de test en fonction du nombre d'attributs

Cependant, lorsque le nombre des attributs est trop grand, le taux de classification diminue, car les paramètres ajoutées entraînent une complexité du problème de classification.

La technique d'évaluation des distances (DET) est bien établie pour la sélection des attributs pertinents. Malheureusement, DET souffre de deux inconvénients principaux : le premier inconvénient est lié au fait que son efficacité est influencée par le bruit dû au système d'acquisition de données, le deuxième inconvénient de DET est que les paramètres pertinents sont sélectionnés indépendamment du système de classification, en d'autres termes, le sous-ensemble de 'D' dimension qui contient les attributs dont les facteurs DET sont les plus grands, n'est pas nécessairement le meilleur sous-ensemble de 'D' dimension dans lequel les classes sont bien séparées. C'est pourquoi, l'algorithme OEPB basé sur les matrices de dispersion comme étant une fonction objective, est introduit pour sélectionner un sous-ensemble optimal qui distingue bien les classes, et réduit le nombre d'entrées.

Le choix des paramètres de l'algorithme OEPB reste en grande partie empirique, mais suit une certaine logique. Après un grand nombre d'essais, nous avons fixé les facteurs $c_1 = 1.6$, $c_2 = 1.6$, $c_3 = 0.6$, $\omega_{max} = 0.9$ et $\omega_{min} = 0.1$. Nous avons aussi fixé la population à 30 et le nombre d'itérations à 60.

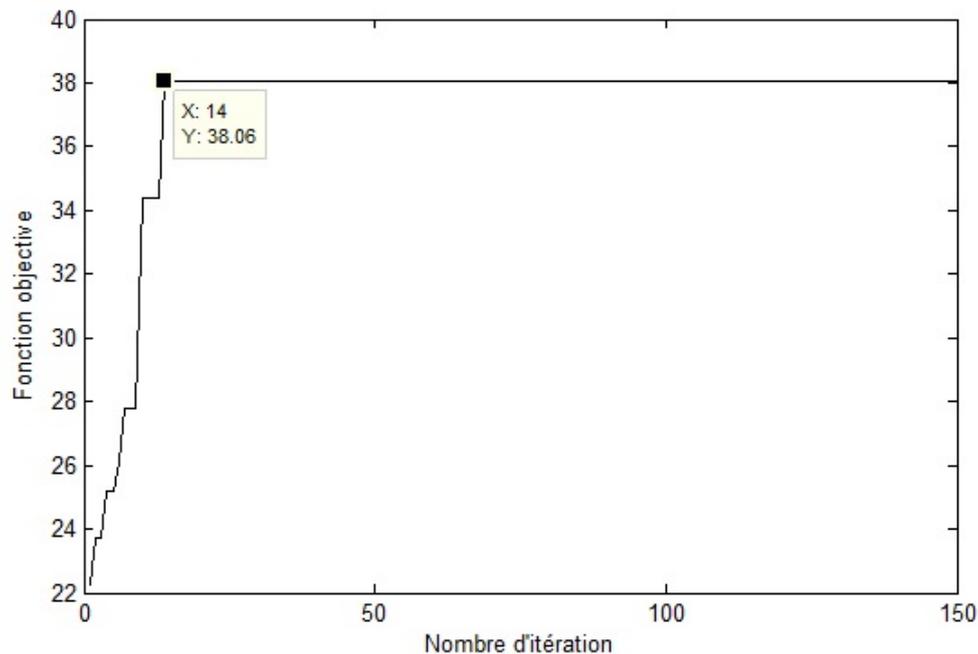


Figure IV.15 – Convergence de la fonction objective en fonction du nombre d'itération

Dans la plupart des études précédentes, la fonction objective a été déterminée en fonction de la performance du classifieur. Malheureusement, cette méthode nécessite des calculs lourds. Pour pallier à cette lacune, nous proposons un critère basé sur les matrices de dispersion comme étant une fonction objective, pour sélectionner un sous-ensemble des attributs riche en informations discriminatoires.

La figure (IV.15) illustre la convergence de l'algorithme OEPB-MD. On remarque que l'algorithme OEPB basé sur les matrices de dispersion (MD), atteint la meilleure solution après seulement 14 générations avec ' J ' égal à 38.06, et ce en utilisant uniquement deux paramètres : l'écart type (SD) du signal original et la valeur crête à crête (PPV) de la deuxième approximation.

Les figures (IV.16) et (IV.17), illustrent l'influence du sous-ensemble résultant sur la séparation des classes. Il est très clair que, dans tous les cas, les données sont bien séparées et prêtes à être classées.

Après l'application de l'approche de sélection des attributs suggérée, le sous-ensemble des paramètres résultant, est utilisée comme étant une entrée de trois classifieur à savoir : les réseaux de neurones (MLP), les forêts aléatoires et les séparateurs à vaste marge (SVM). Avec, ' n ' représente le nombre d'arbres.

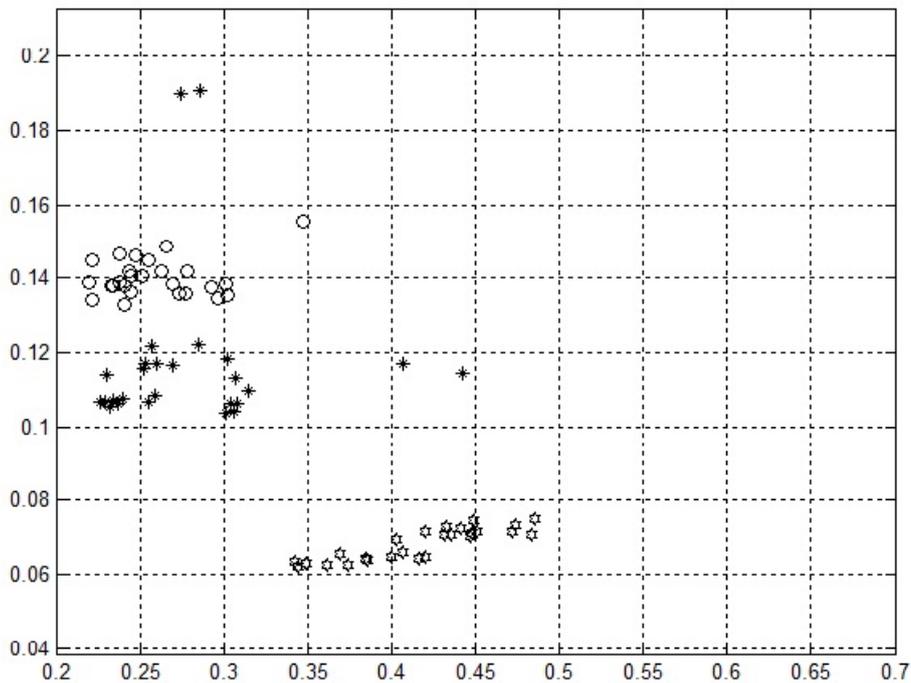
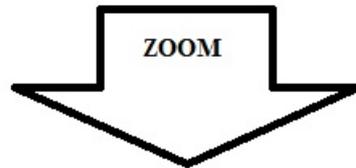
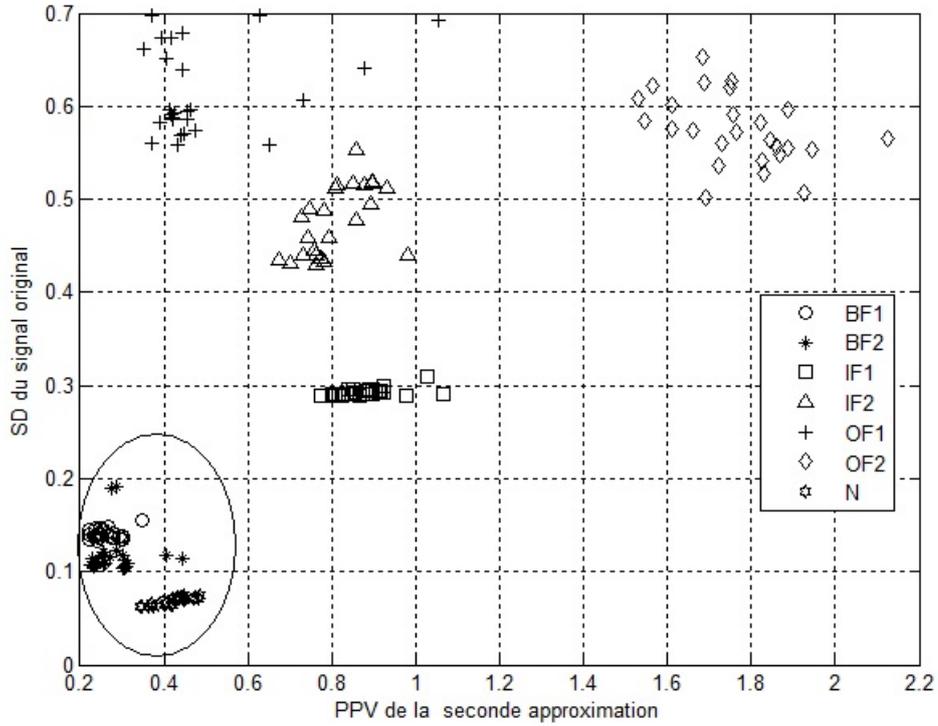


Figure IV.16 – PPV de la deuxième approximation et SD du signal original (Données d'apprentissage).

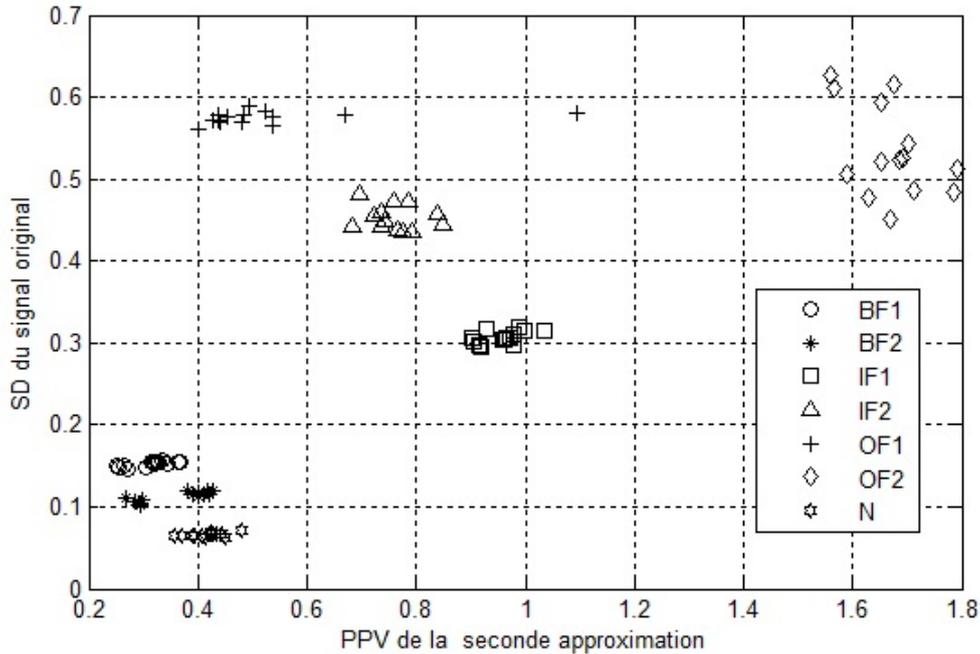


Figure IV.17 – PPV de la deuxième approximation et SD du signal original (Données de test).

2.4 Résultat de simulation

Maintenant, nous allons procéder à l'optimisation par PSOPC des paramètres des SVM, à savoir : le paramètre de régularisation "c", et celle de la fonction kernel " σ ". le tableau (IV.4) expose les paramètres des SVM obtenus par l'optimisation par PSOPC ainsi que le taux de classification..

| réduction de dimension | Fonction noyau | σ | c | Taux de Validation% | Taux de test% |
|-----------------------------------|----------------|----------|----------|---------------------|---------------|
| Paramètres statistiques +DET-OEPB | Gaussien | 0.0636 | 381.3101 | 100 | 100 |

Tableau IV.4 – Résultats de simulation

Le tableau IV.5 résume les résultats obtenus. Comme le montre le tableau , trois réseaux de neurones sont considérés :

- BPNN1 comprenant deux couches cachées, la première contient dix (10) neurones , tandis que, la seconde est composée de quinze (15) neurones,
- BPNN2 contenant cinq (05) neurones dans la première couche cachée et dix (10) neurones dans la seconde.

- BPNN3 contenant une couche cachée composée de cinq (cinq) neurones.

| Performances | Classifieur | Taux de validation | Taux de test | Temps d'apprentissage | Temps de Test |
|--------------|---------------|--------------------|--------------|-----------------------|---------------|
| | SVM | 100% | 100% | 1.080650 | 0.002775 |
| | BPNN1 10 – 15 | 96.1538% | 97.9592% | 35.221435 | 0.032019 |
| | BPNN1 5 – 10 | 98.8605% | 91.2980% | 33.342851 | 0.029779 |
| | BPNN1 10 | 98.8605% | 98.9123% | 31.001480 | 0.028055 |
| | RF (n=9) | 100% | 97.9600% | 0.009905 | 0.003832 |
| | RF (n=20) | 100% | 98.9800% | 0.011003 | 0.004473 |
| | RF (n=40) | 100% | 98.9800% | 0.012357 | 0.004030 |

Tableau IV.5 – Performance de classification en utilisant : SVM, BPNN and RF..

A partir des résultats du tableau 1, on remarque que :

- SVM offre le meilleur taux d'apprentissage et du test.
- L'erreur de généralisation des réseaux de neurones dépend fortement du bon choix de sa structure.
- La capacité de généralisation du classifieur des forêts aléatoires, dépend du nombre d'arbres dans la forêt.

Nous pouvons conclure que le classifieur SVM-OEPCP donne les meilleurs résultats de classification avec le plus petit temps de test .

Pour évaluer l'efficacité de la méthode suggérée, une comparaison entre l'approche proposée et d'autres méthodes récentes, est présentée dans le tableau (IV.6). La comparaison a été effectuée en se basant sur : la dimension du vecteur d'entrée, le classifieur utilisé, le nombre de classes considérées et l'exactitude de la classification.

Les résultats présentés dans le tableau (IV.6) montrent clairement les bonnes performances de l'approche proposée, pour le diagnostic des défauts du roulement. D'une manière générale, l'ensemble des résultats obtenus, montrent l'intérêt de l'approche développée, ils confirment ainsi que cette dernière permet de répondre aux problèmes complexes de diagnostic des défauts des roulements.

3 Conclusion

Dans ce chapitre, nous avons présenté les performances des deux méthodes testées pour le diagnostic de sévérité des défauts dans les roulements à bille.

| Référence | Dimension du vecteur d'entrée | Classifieur utilisé | Nombre de classes | Taux de classification |
|------------------|-------------------------------|------------------------|-------------------|------------------------|
| [69] | 07 | SVM et ANN | 07 | 100 % |
| [70] | 06 | SVM | 07 | 97.7143% |
| [71] | 08 | SVM | 04 | 97.5 % |
| [72] | 02 | Forêts Aléatoires (FA) | 04 | 93.54% |
| [73] | 03 | k plus proches voisins | 04 | 92.86% |
| [74] | 15 | FA,SVM,ANN | 10 | 100 % |
| Méthode proposée | 02 | SVM | 07 | 100 % |

Tableau IV.6 – Comparaison de la méthode proposée avec d'autres méthodes récentes..

Dans le premier système, nous avons adopté les paramètres statistiques, et les énergies obtenues grâce à l'application de l'algorithme de paquet d'ondelette pour l'extraction de paramètres. L'analyse en composantes principales, est ensuite utilisée pour la réduction de la dimensionnalité. Pour la phase de classification, nous avons employé les SVM basés sur OEPCP.

Les résultats obtenus démontrent que l'approche proposée produit de bonnes performances de classification.

Dans le second système, nous avons proposé pour la phase de prétraitement des données, les paramètres statistiques pour l'extraction des attributs. Il s'est avéré que ces paramètres sont affectés par le bruit. Pour résoudre ce problème, une analyse en ondelette optimale est utilisée pour le filtrage des signaux vibratoires. Ensuite, les paramètres statistiques sont extraits du signal original, ainsi que de toutes les bandes de fréquences résultantes.

La sélection des attributs pertinents est réalisée en deux étapes : Dans un premier temps l'algorithme DET est appliqué, pour sélectionner les attributs pertinents selon leurs facteurs DET. Dans un second temps l'algorithme des essaims de particules binaire basé sur les matrices de dispersion comme étant un critère pour la mesure de séparabilité des classes, est adopté pour l'extraction des paramètres les plus pertinents.

Pour effectuer l'étape de classification, les MLP, les forêts aléatoires et l'approche multi-classes "un-contre-tous" pour les SVM, ont été mise en oeuvre, les performances de trois classifieurs ont été ainsi comparées. Pour l'optimisation des SVMs, les paramètres des SVM sont obtenus par l'optimisation par essaim de particules.

Nous avons observé qu'en comparant les taux de classifications obtenus en utilisant les classifieurs présentés, qu'ils étaient systématiquement plus élevés dans le cas d'utilisation d'un classifieur SVM basé sur l'approche "un-contre-tous" optimisé par OEPCP et combiné avec

l'approche proposée pour la sélection des paramètres. Notons aussi que le classifieur SVM possède le temps de calcul le plus court.

Conclusion générale

Cette thèse s'est focalisée sur la détection et le diagnostic vibratoire des dysfonctionnements des roulements. Elle concerne la mise en oeuvre de deux nouvelles contributions, pour la détection et le diagnostic de ces défauts. Dans ce cadre, nous avons choisi d'utiliser des approches de surveillance et de diagnostic, par reconnaissance de formes à partir de l'enregistrement de signaux vibratoires sur le roulement.

Le diagnostic par reconnaissance de formes peut être décortiqué en deux étapes essentielles à savoir : la réduction de dimension et la classification. La première phase revient sur l'extraction et la sélection des paramètres pertinents, pour ce fait, on a proposé deux nouvelles contributions. Dans la première contribution, Nous avons adopté, pour l'extraction des attributs une combinaison entre les paramètres statistiques et l'algorithme de paquet d'ondelette . L'analyse des composantes principales est ensuite utilisée pour la réduction de la dimensionnalité. Dans la seconde contribution, nous nous sommes intéressés dans un premier temps aux paramètres statistiques pour l'extraction des paramètres, dont la sensibilité s'est avérée fortement dépendante du rapport signal sur bruit. Pour pallier à cette insuffisance, une analyse multirésolution optimale en ondelettes, a été appliquée pour améliorer la sensibilité des paramètres statistiques, en filtrant le bruit du système d'acquisition pour n'avoir uniquement que la signature du défaut.

Dans un second temps, nous avons admis pour l'extraction des paramètres pertinents, une nouvelle approche basée sur une combinaison entre l'algorithme DET et celui des essais des particules binaires en utilisant, pour la mesure de séparabilité entre les classes, un critère basé sur les matrices de dispersion.

Pour la phase de classification, nous nous sommes particulièrement concentrés sur les SVM, qui sont l'un des modèles de classification ayant fait ses preuves en termes de performances

en discrimination et en pouvoir de généralisation. Comme les paramètres de SVM, tel que le paramètre de régularisation et les paramètres du noyau, jouent un rôle crucial dans la construction d'un modèle de classification efficace, l'algorithme OEPCP a été appliqué avec succès pour optimiser les paramètres de SVM, il peut effectivement extraire les paramètres optimaux pour construire un bon modèle de classification.

De nombreuses expérimentations menées sur nos données ont permis de montrer que nos approches adoptées pour la réduction de dimension sont capables de fournir des variables pertinentes pour la classification des signaux vibratoires. Cette étude démontre aussi la supériorité des SVM sur les méthodes discriminantes classiques telles que le MLP et les Forêts Aléatoires. Les résultats obtenus ont révélé que les deux méthodologies proposées sont très efficaces et qu'elles peuvent diagnostiquer avec succès la condition du roulement.

Références bibliographiques

- [1] J. ZAREI, M.A. TAJEDDINI ET H.R.KARIMI. *Vibration analysis for bearing fault detection and classification using an intelligent filter*. *Mechatronics* **24**, pages 151–157. (2014). [1](#)
- [2] S. KAVATHEKAR, N. UPADHYAY ET P.K.KANKAR . *Fault classification of ball bearing by rotation forest technique*. *Procedia Technology* **23**, pages 187–192. (2016). [1](#)
- [3] S. A. MCINERNEY ET Y. DAI. *Basic vibration signal processing for bearing fault detection*. *IEEE Transactions on Education*, pages 149–156. (2003). [1](#)
- [4] W. D. LI ET C. K. MECHEFSKE. *Detection of induction motor faults : a comparison of stator current, vibration and acoustic methods*. *Journal of Vibration and Control* **12**, pages 165–188. (2006). [1](#)
- [5] L. MENG ET Z. PING. *The application of wavelet packet and svm in rolling bearing fault diagnosis*. *IEEE, Proceedings of International Conference on Mechatronics and Automation* (2008). [1](#), [76](#)
- [6] P. K. KANKAR, C. S. SATISH ET S. P. HARSHA. *Fault diagnosis of ball bearings using machine learning methods*. *Expert Systems with Applications* **38**(3), pages 1876–1886 (2011). [1](#)
- [7] L. SHUANG ET L. MENG. *Bearing fault diagnosis based on pca and svm*. *IEEE International Conference on Mechatronics and Automation August* (2007). [1](#)
- [8] T.THELALIDJIA AND S. CHENIKHER. *A new approach of preprocessing with svm optimization based on pso for bearing fault diagnosis*. *13th International Conference on Hybrid Intelligent Systems*, 04-06 December , Yasmine Hammamet, Tunisia (2013). [1](#)
- [9] T. A. HARRIS. *Rolling bearing analysis*. Wiley, New-York **3** (1991). [5](#)
- [10] R. A. GUYER. *Rolling bearings handbook and troubleshooting guide*. Chilton Book Company, Radnor, Pennsylvania (1996). [6](#), [8](#), [10](#)

- [11] M. MORET. *Roulements et butées à billes et à rouleaux*. Techniques de l'Ingénieur Mai (1990). 12
- [12] D. HOCHMANN ET E. BECHHOEFER. *Envelope bearing analysis : Theory and practice*. IEEE Aerospace Conference (2005). 15
- [13] A. DJEBALA, N. OUELAA ET N. HAMZAOUI. *Detection of rolling bearing defects using discrete wavelet analysis*. Meccanica **43**, pages 339–348 (2008). 15, 75
- [14] S. FANG ET W. ZIJIE. *Rolling bearing fault diagnosis based on wavelet packet and rbf neural network*. Proceedings of the 26th Chinese Control Conference July 26-31, , Zhangjiajie, Hunan, China (2007). 15
- [15] S. FU, K. LIU, Y. XU, ET Y. LIU. *Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy means clustering*. Shock and Vibration (2016). 15
- [16] M. SAIMURUGAN ET K.I. RAMACHANDRAN. *A comparative study of sound and vibration signals in detection of rotating machine faults using support vector machine and independent component analysis*. International Journal of Data Analysis Techniques and Strategies **6**(2), pages 188–204 (2014). 16
- [17] T.THELALDJIA, A. MOUSSAOUI ET S.CHENEIKHER. *Bearing fault diagnosis based on independent component analysis and optimized support vector machine*. IEEE 7th International Conference on Modelling, Identification and Control Sousse Tunisia, 18-20 December 2015. 16
- [18] L. SAIDI , J. BEN ALI,F. FNAIECH . *Application of higher order spectral features and support vector machines for bearing faults classification*. ISA Transactions **54**, pages 193–206 (2015). 16
- [19] T.THELALDJIA, A. MOUSSAOUI ET S.CHENEIKHER. *Feature extraction and optimized support vector machine for severity fault diagnosis in ball bearing*. Engineering Solid Mechanics **4**, pages 167–176 (2016). 16
- [20] R. SHAO , W. HU, Y. WANG ET X. QI. *The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform*. Measurement **54**, pages 118–132 (2014). 16
- [21] C. SHEN, D. WANG, F. KONG, ET P.W. TSE. *Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier*. Measurement **46**, pages 1551–1564 (2013). 17
- [22] Y. LEI, Z. HE ET Y. ZI . *Application of an intelligent classification method to mechanical fault diagnosis*. Expert Systems with Applications **36**, pages 9941–9948 (2009). 17
- [23] B.S.YANG, T. HAN ET J.L. AN. *Artkohonen neural network for fault diagnosis of rotating machinery*. Mechanical Systems and Signal Processing **18**, pages 645–657 (2004). 17

Références bibliographiques

- [24] Y. LEI, Z. HE, Y. ZI ET Q. HU . *Fault diagnosis of rotating machinery based on multiple anfis combination with ga*. Mechanical Systems and Signal Processing **21**, pages 2280–2294 (2007). [17](#)
- [25] I. RASHEDUL, A. K. SHERAZ ET K. JONG-MYON . *Maximum class separability-based discriminant feature selection using a ga for reliable fault diagnosis of induction motors*. Advanced Intelligent Computing Theories and Applications pages 526–537 (2015). [17](#)
- [26] R. ZIANI, A. FELKAOUI ET R. ZEGADI . *Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized fishers criterion*. Journal of Intelligent Manufacturing **28**(9), pages 405–417 (2017). [18](#)
- [27] H. NEZAMABADI-POUR, M. ROSTAMI-SHAHRBABAKI ET M.M. FARSANGI . *Binary particle swarm optimization : challenges and new solutions*. The Journal of Computer Society of Iran (CSI) On Computer Science and Engineering **6**, pages 21–32 (2008). [18](#), [52](#)
- [28] R. ZEGADI, A. FELKAOUI ET M. DJOUADA. *Bearing fault diagnosis using neural network and genetic algorithms with the trace criterion*. Condition Monitoring of Machinery in Non Stationary Operations pages 89–96 (2012). [18](#)
- [29] M. UNAL, M. ONAT, M. DEMETGUL ET H. KUCUK. *Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network*. Measurement **58**, pages 187–196 (2014). [18](#)
- [30] J. CHENG, L. AI ET W. XIONG. *Rolling bearing fault diagnosis based on the hybrid algorithm of particle swarm optimization with neighborhood operator*. IEEE. International Conference on Computer Science and Electronics Engineering (2012). [18](#)
- [31] T. ZHOU, X. ZHU, W. PENG ET Y. LIU. *Fault diagnosis of bearings based on wpd and rbf neural networks*. The 21st International Congress on Sound and Vibration Beijing, China, 13-17 July 2014. [19](#)
- [32] Y. LIU ,Q. ZHENG ,Z. SHI ,J. CHEN . *Training radial basis function networks with particle swarms*. Lecture Notes in Computer Science **3173**, pages 317–322 (2004). [19](#)
- [33] D. SIMON . *Training radial basis neural networks with the extended kalman filter*. Neuro-computing **48**, pages 455–475 (2002). [19](#)
- [34] Z. WANG, Z. LIU, X. LAN, J. LIU, S. WANG, Y. WU ET Y. XUE . *Research on a new hybrid intelligent fault diagnosis method and its application*. International Journal of Smart Home **10**(7), pages 133–142 (2016). [19](#)
- [35] I. ATTOUI , N. BOUTASSETA , N. FERGANI , B. OUDJANI ET A. DELIOU. *Vibration-based bearing fault diagnosis by an integrated dut-fft approach and an adaptive neuro-fuzzy inference system*. 3rd International Conference on Control, Engineering Information Technology (CEIT), 25-27 May Tlemcen, Algeria (2015). [19](#)

- [36] C. CORTES ET V. N. VAPNIK. *Support-vector networks*. Machine Learning (1995). 19
- [37] C. J. BURGES. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery **2**, pages 121–167 (1998). 19
- [38] S. R. GUNN. *Support vector machines for classification and regression*. pages 1–28 (1998). 19
- [39] J. CHRISTOPHER AND C. BURGES. *A tutorial on support vector machines for pattern recognition*. Kluwer Academic Publishers, Bell Lab, Lucent Technologies, Boston pages 1–43 (1998). 19
- [40] G. XIAN. *Mechanical failure classification for spherical roller bearing of hydraulic injection moulding machine using dwt-svm*. Expert Systems with Applications **37**, pages 6742–6747 (2010). 20
- [41] R. YUAN ET B. GUANGCHEN. *Determination of optimal svm parameters by using ga/psa*. Journal of computers **5**(8) (2010). 20, 50, 52
- [42] H. CHENG-LUNG AND D. JIAN-FAN. *A distributed pso-svm hybrid system with feature selection and parameter optimization*. Applied Soft Computing **8**(4), pages 1381–1391 (2008). 20, 50, 52
- [43] M. CLERC ET J. KENNEDY. *The particle swarm-explosion, stability, and convergence in a multidimensional complex space*. IEEE transactions on Evolutionary Computation **6**(1), pages 58–73 (2002). 20
- [44] S. HE, Q. H. WU, J. Y. WEN, J. R. SAUNDERS ET R. C. PATON. *A particle swarm optimizer with passive congregation*. Biosystems **78**(1), pages 135–147 (2004). 20
- [45] R. P. W. DUIN, A. K. JAIN ET J. MAO. *Statistical pattern recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(1), pages 4–37. (2000). 21
- [46] F. COTTET. *Traitement des signaux et acquisition de données*. Dunod, Paris (2009). 24
- [47] L. DEBNATH. *Wavelet transforms and their applications*. Birkhauser Boston (2002). 24, 25, 28, 29, 30, 31
- [48] S. MALLAT. *A wavelet tour of signal processing*. Elsevier Inc (2009). 25, 26, 28, 29, 30
- [49] Y. MEYER. *Principe d’incertitude, bases hilbertiennes et algèbres d’opérateurs*. Séminaire Bourbaki, (1985-1986). 27
- [50] Y. MEYER. *Ondelettes et opérateurs : Ondelettes*. Editions Hermann **1** (1990). 27
- [51] S. MALLAT. *A theory for multiresolution signal decomposition : The wavelet transform*. IEEE Trans. Pattern Anal. Machine Intell **11**, pages 674–693. (1989). 27

Références bibliographiques

- [52] A. N. AKANSU ET R. A. HADDAD. *Multiresolution signal decomposition : Transforms, subbands, and wavelets*. Academic Press (2001). 28, 29, 33, 34, 36
- [53] B. K. PANIGRAHI, Y. SHI ET L. M. HIOT. *Handbook of swarm intelligence concepts, principles and applications*. Springer (2011). 50
- [54] X. YUN-JIE ET X. SHU-DONG. *A new and effective method of bearing fault diagnosis using wavelet packet transform combined with support vector machine*. Journal of computers 6(11) November (2011). 52
- [55] J. LI. *A combination of pso and svm for road icing forecast*. Journal of computers 5(9) September (2010). 52
- [56] B. PINKOWSKI. *Principal component analysis of speech spectrogram images*. Pattern recognition 30, pages 777–787. (1997). 53
- [57] B. DUBUISSON. *Diagnostic et reconnaissance des formes. traité des nouvelles technologies*. Série diagnostic et Maintenance 30 (1990). 54
- [58] V.N. VAPNIK. *The nature of statistical learning theory*. Springer Verlag (1995). 59
- [59] L. WANG. *Support vector machines :theory and applications*. Springer-Verlag (2005). 59, 61, 62, 64, 68, 69
- [60] V. KECMAN. *Learning and soft computing : Support vector machines, neural networks, and fuzzy logic models*. The MIT Press Cambridge, Massachusetts London, England (2001). 60, 61, 63
- [61] A. SHIGEO. *Support vector machines for pattern classification*. Springer-Verlag (2005). 61, 62, 63, 64, 69, 70
- [62] V. N. VAPNIK. *Statistical learning theory*. New York : Wiley (1998). 64, 66
- [63] C.CORTES ET V.N.VAPNIK. *Support vector networks. machinery learning*. Machinery Learning (1995). 66
- [64] B.SCHOLKOPF, A.SMOLA, R. C.WILLIAMSON ET P. L.BARTLETT. *New support vector algorithms*. Neural Computation (2000). 66
- [65] L. HAMEL. *Knowledge discovery with support vector machines*. JohnWiley & Sons (2009). 69
- [66] K. A. LOPARO. *Bearings vibration data set.case western reserve university*. /http ://www.eecs.case.edu/laboratory/bearing/download.htm (consulté en septembre 2012). 72

- [67] A. DJEBALA, M.K BABOURI ET N. OUELAA . *Rolling bearing fault detection using a hybrid method based on empirical mode decomposition and optimized wavelet multi-resolution analysis*. The International Journal of Advanced Manufacturing Technology **79**, pages 2093–2105 (2015). [75](#)
- [68] D. H. PANDYA, S. H. UPADHYAY ET S. P. HARSHA . *Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform*. Soft Computing **18**, pages 255–266 (2014). [75](#)
- [69] A. SHARMA, M. AMARNATH ET P. K.KANKAR. *Feature extraction and fault severity classification in ball bearings*. Journal of Vibration and Control **22**, pages 176–192. (2014). [92](#)
- [70] M. VAN ET H.J.KANG. *Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection*. IET Science, Measurement Technology **9**, pages 671–680. (2015). [92](#)
- [71] P. LIU , H. LI ET P. YE. *A method for rolling bearing fault diagnosis based on sensitive feature selection and nonlinear feature fusion*. 8th International Conference on Intelligent Computation Technology and Automation, 14-15 June, IEEE. pages 30–35. (2015). [92](#)
- [72] V. VAKHARIA, V.K. GUPTA ET P.K.KANKAR . *Bearing fault diagnosis using feature ranking methods and fault identification algorithms*. Procedia Engineering **144**, pages 343–350. (2016). [92](#)
- [73] L. LIANG, F. LIU, M. LI, K. HE ET G. XU. *Feature selection for machine fault diagnosis using clustering of non-negation matrix factorization*. Measurement **94**, pages 295–305. (2016). [92](#)
- [74] A. SHARMA, M. AMARNATH ET P.K. KANKAR. *Novel ensemble techniques for classification of rolling element bearing faults*. Journal of the Brazilian Society of Mechanical Sciences and Engineering **94** (2016). [92](#)
- [75] LEO BREIMAN. *Random forests*. Machine Learning **45(1)** (2001). [112](#)

Annexe A : Représentation vectorielle de signaux

1 Espace vectoriel de fonctions

1.1 Espace vectoriel

Etant donné un corps commutatif C d'élément neutre e , on dit qu'un ensemble ζ muni d'une opération interne (notée \oplus) et d'une opération externe dont le domaine d'opérateurs est C , a une structure d'espace vectoriel sur C si :

- ζ est un groupe commutatif pour son opération interne.
- ζ est un groupe commutatif pour son opération interne, l'opération externe est telle que, pour tout $x \in \zeta$ et tout $\lambda \in C$ et $\mu \in C$, on ait : $(\lambda\mu)x = \lambda(\mu x)$ et $ex = x$
- L'opération externe est distributive par rapport à l'addition sur le corps C et par rapport à l'opération interne dans ζ :

$$(\lambda + \mu)x = \lambda x \oplus \mu x \text{ et } \lambda(x \oplus y) = \lambda x \oplus \lambda y$$

Le nombre de vecteurs de la base est appelé dimension de l'espace.

Un espace vectoriel ζ est normé si on peut définir une norme, c'est-à-dire une application de ζ dans R^+ , qui à tout vecteur \vec{x} de ζ associe $\|\vec{x}\|$. La norme possède les propriétés suivantes :

- $\forall \vec{x}, \|\vec{x}\| \geq 0$
- $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$
- $\forall(\vec{x}, \vec{y}), \|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$

Un espace vectoriel est dit métrique si à tout couple de vecteurs (\vec{x}, \vec{y}) est associé un réel $d(\vec{x}, \vec{y})$ tel que :

- $d(\vec{x}, \vec{y}) \geq 0$
- $d(\vec{x}, \vec{y}) = 0 \iff \vec{x} = \vec{y}$

Annexe A : Représentation vectorielle de signaux

$$- d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$$

La mesure $d(\vec{x}, \vec{y})$ est appelée distance. Une distance usuelle est obtenue à partir de la norme par :

$$d(\vec{x}, \vec{y}) = \|\vec{x} + \vec{y}\|$$

Une suite infinie $\{\vec{x}_n\}$ d'éléments d'un espace vectoriel métrique converge vers un vecteur \vec{x} de cet espace si :

$$\lim_{n \rightarrow \infty} d(\vec{x}_n, \vec{x}) = 0$$

Un espace vectoriel normé dans lequel toute suite est convergente est dit complet. On l'appelle aussi espace de Banach.

1.2 Espace de Hilbert

Un espace de Hilbert est un espace vectoriel normé complet (espace de Banach) dont la norme $\|\cdot\|$ découle d'un produit scalaire ou hermitien $\langle \cdot, \cdot \rangle$ par la formule $\langle x, x \rangle = \|x\|^2$.

Un espace de Hilbert est la généralisation en dimension quelconque d'un espace vectoriel euclidien ou hermitien.

1.3 Espace L^2 des signaux à énergie finie

L'ensemble des signaux à énergie finie (ou de carré sommable) sur un intervalle $[t_1, t_2]$ forme un espace noté $L^2(t_1, t_2)$ dont la norme est définie par :

$$\|x\| = \left(\int_{t_1}^{t_2} |x(t)|^2 dt \right)^{\frac{1}{2}}$$

Le carré de la norme, $\|x\|^2$, est donc égal à l'énergie du signal.

La distance entre deux signaux x et y est alors définie par la distance euclidienne :

$$d(x, y) = \|x - y\| = \left(\int_{t_1}^{t_2} |x(t) - y(t)|^2 dt \right)^{\frac{1}{2}}$$

Dans $L^2(t_1, t_2)$, on dit qu'un signal $x(t)$ converge en moyenne quadratique vers $y(t)$ si $d(x, y) \rightarrow 0$

L'exposant 2 dans $L^2(t_1, t_2)$ indique la métrique utilisée et non pas la dimension.

2 Fonctions orthogonales

2.1 Produit scalaire de signaux dans $L^2(t_1, t_2)$

Soient deux vecteurs $\vec{x} = (x_1, \dots, x_K)^T$ et $\vec{y} = (y_1, \dots, y_K)^T$ dans une base B, on peut définir le produit scalaire des vecteurs :

$$\vec{x} \cdot \vec{y} = \sum_{k=1}^K x_k y_k^*$$

où * indique la conjugaison complexe.

Par analogie, soient deux signaux $x(t)$ et $y(t)$, éléments de $L^2(t_1, t_2)$, on peut définir le produit scalaire par :

$$\langle x, y \rangle = \int_{t_1}^{t_2} x(t) y(t)^* dt$$

On remarque que $\langle x, y \rangle \neq \langle y, x \rangle$. En revanche, cette définition possède la symétrie hermitienne.

On a en effet :

$$\langle x, y \rangle = \langle y, x \rangle^*$$

Par ailleurs, en raison de la définition ci-dessus et de la linéarité de l'intégrale, le produit scalaire de sommes de signaux se développe facilement selon la règle de calcul :

$$\langle a_1 x_1 + a_2 x_2, b_1 y_1 + b_2 y_2 \rangle = a_1 b_1^* \langle x_1, y_1 \rangle + a_1 b_2^* \langle x_1, y_2 \rangle + a_2 b_1^* \langle x_2, y_1 \rangle + a_2 b_2^* \langle x_2, y_2 \rangle$$

où a_1, a_2, b_1, b_2 sont des scalaires.

Annexe B : Noyaux reproduisants

1 Premières définitions

Soit χ l'ensemble des observations, connu partiellement au travers des données x_i de l'ensemble d'apprentissage, avec $i \in 1, \dots, n$. La résolution d'un problème d'apprentissage nécessite la définition d'une mesure de similarité entre les éléments de χ . Dans le cadre des algorithmes dits à noyaux, cette tâche est accomplie par une fonction κ définie ainsi :

$$\begin{aligned} \kappa : \chi \times \chi &\longrightarrow R \\ (x_i, x_j) &\longrightarrow \kappa(x_i, x_j) \end{aligned}$$

Celle-ci est couramment appelée noyau.

Avant d'étudier des propriétés liées à la notion de noyau, il convient de présenter quelques définitions.

1.1 Définition 1

Etant donné un noyau κ et les n observations x_1, x_2, \dots, x_n d'un ensemble d'apprentissage, la matrice de dimension $n \times n$ définie par :

$$K = (\kappa(x_i, x_j))_{ij}$$

est appelée la matrice de **Gram** de κ associée à x_1, x_2, \dots, x_n

1.2 Définition 2

La matrice K de dimension $n \times n$ est dite positive si pour tous $c_1, c_2, \dots, c_n \in C$ on a :

$$\sum_{i,j} c_i \bar{c}_j K_{ij} \geq 0$$

où \bar{c}_j est le conjugué de c_j .

Dans le cas complexe, la relation précédente implique directement que : $\kappa(x_i, x_j) = \overline{\kappa(x_j, x_i)}$

Il convient de noter ici la différence entre une matrice positive et un noyau à valeurs positives.

1.3 Définition 3

Soit χ un ensemble non vide. Une fonction

$$\kappa : \chi \times \chi \rightarrow \mathbb{C}$$

qui, pour tous $x_i \in \chi$, conduit à une matrice de Gram positive est appelée noyau défini positif. Les définitions associées aux noyaux à valeur réelles sont analogues à celles données ci-dessus, mis à part que les coefficients c_i sont alors réels. Cependant, afin que le noyau soit défini positif dans ce cas, il faut exiger qu'il soit symétrique :

$$\kappa(x_i, x_j) = \kappa(x_j, x_i)$$

pour tous les couples (x_i, x_j) de $\chi \times \chi$. Notons ici que n'importe quel produit scalaire peut définir un noyau défini positif.

2 Noyaux reproduisants et condition de Mercer

D'après la théorie de Hilbert-Schmidt, toute fonction symétrique $\kappa(x_1, x_2)$ de $L_2(\chi^2)$ admet une décomposition de la forme :

$$\kappa(x_1, x_2) = \sum_i \gamma_i \phi_i(x_1) \phi_i(x_2)$$

avec $\phi_i \in L_2(\chi)$ et $\gamma_i \in \mathbb{R}$.

Les éléments ϕ_i et γ_i intervenant dans cette expression correspondent aux fonctions propres et valeurs propres de l'opérateur intégral défini par le noyau κ , soit :

$$\int \kappa(x_1, x_2) \phi_i(x_1) dx_1 = \gamma_i \phi_i(x_2)$$

D'après l'expression de décomposition fonction symétrique $\kappa(x_1, x_2)$, une condition suffisante pour que $\kappa(x_1, x_2)$ soit un produit scalaire est que les valeurs propres γ_i soient positives. Il en est ainsi, selon le théorème de Mercer, si et seulement si la condition suivante est satisfaite pour toute fonction f de $L_2(\chi)$:

$$\iint \kappa(x_1, x_2) f(x_1) f(x_2) dx_1 dx_2 \geq 0$$

Les fonctions κ vérifiant cette relation sont appelées noyaux de Mercer.

Annexe C : Outils pour la classification

1 Les réseaux de neurones, concepts et définitions

Les réseaux de neurones sont des outils d'intelligence artificielle (IA), capables d'effectuer des opérations de classification par apprentissage. Leur principal avantage par rapport aux autres outils est leur capacité d'apprentissage et de généralisation de leurs connaissances à des entrées inconnues. Une des qualités de ce type d'outils de classification est son adéquation pour la mise au point de systèmes de surveillance capables de s'adapter à un système complexe avec reconfigurations multiples. Durant une étape d'apprentissage le réseau de neurones essaye d'apprendre la relation "ensembles de données en entrée - modes de fonctionnement", ceci va lui permettre d'associer les classes représentant les modes de fonctionnement aux formes, mesurées sur le système à surveiller, en entrées. Cette étape d'apprentissage permet de déterminer ou modifier les paramètres du réseau de neurones afin d'adopter un comportement désiré. Plusieurs algorithmes d'apprentissage ont été développés depuis la première règle d'apprentissage de Hebb en 1949. Ils sont classés en deux catégories : supervisé et non supervisé.

Dans l'apprentissage supervisé, la valeur de sortie (la classe désirée) que le réseau de neurones doit associer au vecteur d'entrée x est connue. Cette valeur ou vecteur de sortie (appelé cible ou sortie désirée) est fournie par un superviseur (ou expert humain). L'apprentissage consiste à modifier les paramètres du réseau de neurones afin de minimiser l'erreur entre la sortie désirée et la sortie réelle du réseau de neurones. Par contre, dans l'apprentissage non supervisé, aucune connaissance a priori n'est fournie sur la sortie désirée, il s'agit dans ce cas de déterminer les paramètres du réseau de neurones suivant un critère à définir.

La capacité de généralisation est la capacité du réseau de neurones à élargir les connaissances acquises après l'étape d'apprentissage à des données nouvelles et d'associer donc un vecteur d'entrée x , qui n'a pas fait l'objet d'un apprentissage, à une classe donnée. En d'autres termes, c'est l'aptitude du réseau de neurones à approximer une fonction uniquement à partir d'une

partie des données. C'est l'une des raisons qui motivent l'étude et le développement des réseaux de neurones artificiels.

Il existe plusieurs types de réseaux de neurones et plusieurs algorithmes d'apprentissage. Les architectures neuronales les plus utilisées dans le domaine de la surveillance et le diagnostic industriel sont le Perceptron Multi Couche, le Réseau à Fonctions de base Radiales,...etc. Seule le Perceptron Multi Couches (MLP) est décrit dans la suite.

1.1 Le Perceptron Multi Couches (MLP)

Un MLP est un réseau dont les neurones sont disposés en plusieurs couches successives et où chaque neurone d'une couche est connecté à tous les neurones de la couche suivante. En revanche il n'y a pas de connexion entre les neurones de la même couche. Un MLP est composé d'une couche entrée, une ou plusieurs couches cachées et une couche de sortie. L'addition de cette couche cachée permet au réseau de modéliser des fonctions de décision complexes non linéaires entre n'importe quel espace d'entrée et de sortie. Dans un problème de classification, le nombre de neurones en sortie est fonction du codage adopté. Chaque sortie est dédiée à une classe donnée.

Soit le réseau à deux couches décrit par la figure (A.1) dans lequel les unités de sortie sont notées y_j , les unités cachées v_i et les unités d'entrée x_k . Les connexions des unités d'entrée aux unités cachées sont notées w_{ik} et celles des unités cachées aux unités de sortie par w_{ij} . L'entrée k a pour valeur x_k^l lorsque la donnée l est présentée au réseau.

Pour estimer les paramètres et adapter les poids du réseau de neurones une étape d'apprentissage est nécessaire. L'ajustement est obtenue en minimisant un critère par exemple la somme des erreurs quadratiques.

Pour la donnée d'entrée l , la valeur de sortie de l'unité cachée i est donnée par :

$$v_i^l = f(a_i^l) = f\left(\sum_k w_{ik}x_k^l\right) \quad (1)$$

où w_{ik} est le poids de la connexion qui relie l'unité k à l'unité i , a_i est l'activation de l'unité i , f est la fonction d'activation des unités.

Les unités de sortie ont comme valeur :

$$y_j^l = f(a_j^l) = f\left(\sum_i w_{ij}v_i^l\right) \quad (2)$$

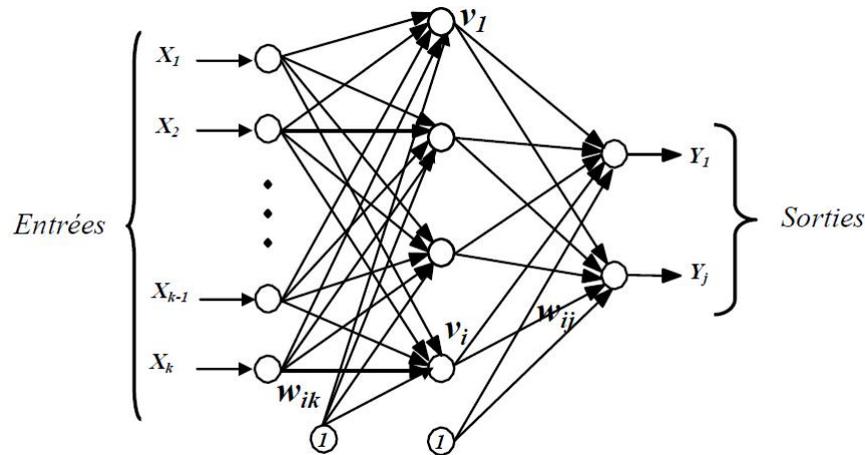


Figure A.1 – Perceptron Multi Couches.

Les fonctions d'erreurs partielles et globales sont alors définies par :

$$E^l = \frac{1}{2} \sum_j (y_j^l - S_j^l)^2 \text{ et } E = \sum_l E^l \quad (3)$$

où S_j^l est la sortie désirée et y_j^l est la sortie réelle du réseau de neurones.

La minimisation de la fonction d'erreur globale 'E' va se faire par une descente de gradient. Par conséquent, après la présentation de tous les vecteurs d'entrée de la base d'apprentissage, la valeur de chaque connexion est modifiée par la règle de delta généralisée :

$$\partial w = -\eta \frac{\partial E}{\partial w} = -\eta \sum_l \frac{\partial E^l}{\partial w} \quad (4)$$

Il existe plusieurs algorithmes d'apprentissage. L'algorithme de rétropropagation (back-propagation) est de loin le plus utilisé [28] et se divise en deux étapes :

- une phase de propagation, qui consiste à présenter les exemples à l'entrée du réseau, puis à propager cette entrée de proche en proche de la couche d'entrée à la couche de sortie en passant par les couches cachées ;
- une phase de rétropropagation, qui consiste à minimiser l'erreur commise sur la base d'apprentissage en modifiant les paramètres du réseau. Après avoir calculé la variation des poids des connexions pour tous les neurones de sortie on calcule alors la variation des poids des connexions de la couche cachée. On met ainsi à jour les poids des connexions de la couche de sortie jusqu'à la couche d'entrée : on rétropropage ainsi le signal d'erreur. C'est de là que vient le nom de cet algorithme : rétropropagation du gradient de l'erreur.

Du fait de sommer les w_{ij} pour tous les vecteurs l de la base d'apprentissage puis de remettre à jour les poids avec la variation totale ainsi calculée, l'algorithme est appelé gradient total.

L'algorithme de rétropropagation du gradient de l'erreur, utilisé pour l'apprentissage dans un MLP, a permis de dépasser les limites du Perceptron simple. Il s'avère capable de résoudre un grand nombre de problèmes de classification et de reconnaissance des formes. Il souffre néanmoins de nombreux défauts, parmi lesquels :

- le temps de calcul : l'apprentissage est très long ;
- Une grande sensibilité aux conditions initiales, c'est-à-dire l'initialisation des poids des connexions ;
- La présence éventuelle de minima locaux dans la fonction d'erreur. Ce problème est en partie résolu avec le gradient stochastique, mais il subsiste quand même ;
- Le problème de dimensionnement du réseau. La rétro-propagation apprend une base d'apprentissage sur un réseau dont la structure est fixée a priori. La structure est définie par le nombre de couches cachées, le nombre de neurones par couche et la topologie des connexions. Un mauvais choix de la structure peut dégrader considérablement les performances du réseau.

2 Arbres de décisions et forêts aléatoires

2.1 Arbres de décisions

Les arbres de décision constituent une méthode qui s'avère être très performante dans de nombreuses applications, aussi bien pour des problèmes de régression que de classification supervisée, en vue de la prédiction d'une variable qualitative à l'aide de variables de tout type (qualitatives et/ou quantitatives). Cette flexibilité constitue un avantage par rapport à certains outils de classification.

Il s'agit d'une représentation graphique d'une procédure de classification qui permet la modélisation simple d'un phénomène mesuré. En effet, la méthode consiste à découper l'espace des variables explicatives en groupes homogènes, et ce en posant une succession de questions binaires (de type oui/non) sur les attributs de chaque échantillon.

Contrairement à beaucoup d'outils de classification, les arbres de décision sont extrêmement intuitifs et fournissent une représentation graphique, parlante et facile à lire, d'un protocole de classification des individus. Cette représentation graphique est sous forme d'un arbre constitué de feuilles terminales (les classes d'individus) obtenues en suivant un chemin le long des noeuds,

Annexe C : Outils pour la classification

chaque noeud correspondant à une question binaire utilisant une variable du jeu de données. Un raffinement de la méthode consiste à faire tourner en parallèle un grand nombre d'arbres de décisions construits aléatoirement, avant de les moyenner, et ce pour obtenir une forêt aléatoire disposant d'un plus fort pouvoir discriminant [75].

RESUME

L'objectif principal de cette étude est de développer deux nouvelles approches d'extraction et de sélection des attributs dont le but est d'améliorer les performances de classification, tout en minimisant le nombre des attributs et en réduisant le temps de calcul.

La première approche, consiste à prétraiter le signal vibratoire en utilisant les paramètres statistiques et le vecteur d'énergie obtenu grâce à l'application de Db8-WPT au troisième niveau de décomposition. Après l'extraction des paramètres à partir du signal vibratoire, PCA est utilisé pour la réduction de la dimension.

Dans la seconde approche, nous avons utilisés une analyse multirésolution en ondelettes optimale pour décomposer le signal de vibration à différentes niveaux et améliorer le rapport signal/bruit. Les paramètres statistiques sont ensuite extraits à partir du signal original et de toutes les bandes de fréquences obtenues. Après la construction du vecteur global des attributs, la technique d'évaluation de la distance (DET) est combinée avec l'algorithme d'optimisation par essaims de particules binaires (BPSO) pour sélectionner le sous-ensemble optimal. Afin de maximiser la séparation des classes, un critère basé sur les matrices de dispersion est utilisé comme étant une fonction objective pour l'algorithme DET-BPSO.

Finalement, les séparateurs à vaste marge optimisé par OEPCP sont utilisés pour automatiser l'identification des différentes conditions du roulement. Les résultats de classification obtenus indiquent l'efficacité des méthodes proposées pour le diagnostic de la sévérité des défauts dans les roulements à billes.

Mots clés : Optimisation par essaims de particules ; Transformée en ondelettes discrètes ; Technique d'évaluation de distance ; Analyse en composantes principales ; Sélection des attributs ; Les matrices de dispersion ; Les paramètres statistiques.

ABSTRACT

The main objective of this study is to develop two novel feature reduction approaches with the goals of improving the classification performance, minimizing the number of features and reducing the computational time.

The first approach consists of preprocessing the bearing fault vibration signal using statistical parameters and energy obtained through the application of Db8- WPT at the third level of decomposition. After feature extraction from vibration signal, PCA is employed for dimensionality reduction.

In the second approach, an optimal discrete wavelet transform (DWT) is employed to decompose the vibration signal at different decomposition depths and enhance the signal-to noise ratio. Statistical parameters are then extracted from the raw signal and all obtained frequency bands. After the construction of the global feature set, distance evaluation technique (DET) is combined with binary particle swarm optimization (BPSO) to select the superior combining feature subset that discriminates well among classes. In order to maximize the class separability, a criterion based on scatter matrices is employed as a fitness function in the proposed DET-BPSO algorithm.

Finally, particle swarm optimization with passive congregation-based support vector machine is used to classify seven kinds of bearing faults. The classification results indicate the effectiveness of the proposed method for severity faults diagnosis in ball bearings.

Keywords : Particle Swarm Optimization ; Discrete Wavelet Transform ; Distance Evaluation Technique ; Feature Selection ; Principal Component Analysis ; Scatter Matrices ; Statistical Parameters.

الملخص

الهدف الرئيسي من هذه الدراسة هو تطوير مقاربتين جديدتين لاختيار الصفات التي تهدف إلى تحسين أداء المصنفات، مع التقليل من عدد الصفات وكذا الوقت اللازم للحساب.

في المقاربة الأولى قمنا أولاً بمعالجة إشارة الإهتزازات عن طريق استخراج كلا من المعلمات الإحصائية و شعاع الطاقة الناتج عن تطبيق خوارزمية حزمة الموجات في المستوى الثالث. بعد ذلك تم استخدام تحليل المكونات الرئيسية للتقليل من عدد الصفات.

في حين استخدمنا في المقاربة الثانية تحليلاً محسناً متعدد الوضوح للموجات من أجل تحليل إشارة الإهتزازات إلى مستويات مختلفة وتحسين نسبة الإشارة / الضوضاء. بعد ذلك تم استخراج المعلمات الإحصائية من الإشارة الأصلية ومن جميع مجالات التردد التي تم الحصول عليها، و بهذا نتحصل على الشعاع الاجمالي للسمات. فيما بعد قمنا بالجمع بين تقنية تقييم المسافات مع خوارزمية سرب الجسيمات الثنائية لتحديد المجموعة الفرعية المثلى وهذا باستخدام معياراً يستند إلى تشتت المصفوفات من أجل تحقيق أقصى قدر من الفصل بين المجموعات.

وأخيراً، تم استخدام الفواصل ذات الهوامش الواسعة المحسنة باستخدام خوارزمية سرب الجسيمات الأمثل بجماعة كامنة لأتمتة تحديد حالات التشغيل المختلفة.

وقد أظهرت النتائج المتحصل عليها فعالية الطريقة المقترحة لتشخيص شدة العيوب في المدرجات.

الكلمات المفتاحية: خوارزمية سرب الجسيمات الأمثل؛ تحليل متعدد الوضوح للموجات؛ تقنية تقييم المسافات؛ اختيار الصفات؛ تحليل المكونات الرئيسية؛ معيار تشتت المصفوفات؛ المعلمات الإحصائية