



MÈMOIRE

Pour l'obtention du diplôme de :

Magister en Electronique

FACULTE DES SCIENCES ET TECHNOLOGIE

DEPARTEMENT DE GENIE ELECTRIQUE

OPTION : Automatique et traitement du signal

Présentée par

Thelaidjia Tawfik

LA MODELISATION AUTOREGRESSIVE ET LES SEPARATEURS A VASTE MARGE POUR LE DIAGNOSTIC DES DEFAUTS DES ROULEMENTS A BILLES

Soutenue le : 11/04/2013

Devant le jury composé de :

Dr. Djalel Dib	Maître de Conférences de l'Université de Tébessa	Président
Pr. Salah Chenikher	Professeur de l'Université de Tébessa	Rapporteur
Dr. Nour Eddine Guerfi	Maître de Conférences de l'Université de Mila	Examineur
Dr. Toufik Bouden	Maître de Conférences de l'Université de Jijel	Examineur
Dr. Abderrazek Lachouri	Maître de Conférences de l'Université de Skikda	Examineur

Année 2013

Remerciement

Je tiens à remercier très vivement mon encadreur le Professeur Chenikher Salah qui a dirigé ce mémoire, et qui m'a prodigué conseils pertinents et encourageants tout au long de cette étude.

Je tiens à exprimer ma très vive gratitude au Docteur Maamri Mahmoud , pour ses conseils et son aide tout au long de cette étude.

Je remercie le Docteur Djalel Dib d'être intéressé à mon travail et d'avoir bien voulu présider le jury de ce mémoire.

Je remercie les membres de mon jury de thèse, les Docteurs : Nour Eddine Guerfi, Abderrazek Lachouri et Toufik Bouden pour avoir accepté d'être les rapporteurs de ce mémoire, 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 Nassim, Rehamnia Mohammed et Rouainia Mohamed Elhadi.

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

Table des matières

Table des figures	vii
Liste des tableaux	x
Préface	1
I Diagnostic par reconnaissance des formes	4
1 Introduction	4
2 L'architecture globale d'un système de surveillance	4
2.1 Systèmes de surveillance	4
2.2 La perception	5
2.3 Le diagnostic	6
2.4 Quelques approches pour le diagnostic	6
3 Système de diagnostic par reconnaissance des formes	8
3.1 Les approches de reconnaissance de formes	8
3.2 Principe et formulation	9
4 Construction d'un système de reconnaissance de formes	9
4.1 Phase d'analyse	10
4.1.1 Détermination de l'espace de représentation	10
4.1.2 Réduction de la dimension de l'espace de représentation	11
4.1.3 Détermination de l'espace de décision	11
4.2 Phase de choix d'une méthode de décision	12
4.3 Phase d'exploitation	13

4.3.1	Évaluation du système de diagnostic	13
5	Approches de discrimination non-paramétrique	15
5.1	L'estimation des fonctions de classement	15
6	Outils pour la classification	15
6.1	Les réseaux de neurones, concepts et définitions	15
6.1.1	Le Perceptron Multi Couches (MLP)	16
7	Conclusion	19
II	Les séparateurs a vaste marge	20
1	Introduction.	20
2	Cas des données linéairement séparables	21
2.1	Recherche de la solution	22
3	Extension au cas non-linéairement séparable.	25
3.1	Recherche de la solution	26
4	Elaboration des Supports Vector Machines nonlinéaires	27
5	Exemples de noyaux reproduisants	29
5.1	Noyaux polynômiaux	29
5.2	Noyaux gaussiens	29
5.3	Noyaux exponentiels	30
5.4	Noyaux sigmoïdaux	30
6	Multi classification par les séparateur a vaste marge	30
6.1	Multi classification par l'approche un contre tous.	31
6.2	Multi classification par l'approche un contre un	31
7	Conclusion	32
III	Réduction de la dimension	33
1	Introduction	33
2	Les ondelettes	34
2.1	Historique de traitement du signal : de Fourier à Gabor.	34
2.2	Représentation temps-fréquence et temps-échelle.	35
2.3	Introduction aux ondelettes	36
2.4	La transformée en ondelette	37
2.4.1	La transformée en ondelettes continue.	37
2.4.2	La transformée en ondelette discrète	39
2.5	Analyse multi-résolution	40

2.5.1	Présentation de l'analyse multirésolution	40
2.5.2	Définition d'une analyse multirésolution	41
2.5.3	Relations de reconstruction et de décomposition	41
2.5.4	Algorithmes	43
2.5.5	Bilan	46
2.6	Paquets d'ondelettes	46
2.7	Orthogonalité et bi-orthogonalité	47
2.7.1	Les ondelettes orthogonales	47
2.7.2	Les ondelettes bi-orthogonales	48
2.8	Exemples d'ondelettes	48
2.8.1	Ondelettes orthogonales	49
	Ondelette de Haar	49
	Ondelette de Shannon	49
	Ondelette de Meyer	49
	Ondelette de Daubechies	49
3	La modélisation autorégressive	53
3.1	Modèles rationnels linéaires	53
3.1.1	Modèle autorégressif	53
3.1.2	Relation AR-Prédiction linéaire	54
3.1.3	Méthodes d'estimation	56
4	Conclusion	59
IV Détection des défauts des roulements par analyse vibratoire		61
1	Introduction	61
2	Généralités sur les roulements à billes	61
2.1	Constitution des roulements à billes	61
2.2	Modes de dégradations d'un roulement à billes	62
2.2.1	Effets des contraintes mécaniques	63
2.2.2	Défauts de montage	65
2.2.3	Corrosion	65
2.2.4	Contamination du roulement	66
2.2.5	Défaut de lubrification	67
2.2.6	Température de fonctionnement	69
3	Influence des défauts sur la structure des signaux	70
4	Fréquences caractéristiques des défauts du roulement «calcul cinématique»	70

5	Bases des données	71
6	La première application	72
6.1	Prétraitement des signaux et calcul des paramètres	72
6.1.1	Données temporelles	72
6.1.2	Recherche de vecteur forme	74
6.2	Méthodes de la prise de décision	75
6.3	Prise de décision par les réseaux de neurones (MLP)	76
6.4	Prise de décision par les SVM	77
6.4.1	Méthode classique	77
6.4.2	Optimisation par essaim particulaire (OEP)	81
	Algorithme général de l'OEP	83
	Formulation de la fonction objective	84
6.4.3	Choix des paramètres de l'OEP	84
6.4.4	Résultats de simulation	84
7	La seconde application	85
7.1	Résultat de simulation	86
8	Conclusion	88
	Conclusion générale	90
	Références bibliographiques	92
	Annexe A : Représentation vectorielle de signaux	98
1	Espace vectoriel de fonctions	98
1.1	Espace vectoriel	98
1.2	Espace de Hilbert	99
1.3	Espace L^2 des signaux à énergie finie	99
2	Fonctions orthogonales	100
2.1	Produit scalaire de signaux dans $L^2(t_1, t_2)$	100
	Annexe B : Noyaux reproduisants	101
1	Premières définitions	101
1.1	Définition 1	101
1.2	Définition 2	101
1.3	Définition 3	102
2	Noyaux reproduisants et condition de Mercer	102

Annexe C : La maintenance	103
1 Maintenance reactive	103
2 Maintenance préventive	103
3 Pratique de la maintenance préventive conditionnelle	104
3.1 les techniques d'analyse	104
3.1.1 L'analyse vibratoire	105
3.1.2 L'analyse des lubrifiants	105
3.1.3 La thermographie infrarouge	105
3.1.4 L'analyse acoustique	106
3.1.5 Le contrôle ultrasonore	106
4 Choix d'une technique d'analyse	106
4.1 La surveillance	110
4.2 La détection	110
4.3 Le diagnostic	110
Annexe D : Analyse en Composantes Principales	111

Table des figures

I.1	L'architecture globale d'un système de surveillance. L'état (1) représente un arrêt momentané du processus avec une reprise presque immédiate. Par contre dans l'état (2), la situation est plus grave et nécessite un arrêt de fonctionnement du processus.	5
I.2	Système de reconnaissance de formes.	10
I.3	Perceptron Multi Couches.	17
II.1	Principe des SVM dans le cas où les deux classes sont linéairement séparables. Les Support Vectors, indiqués par des flèches, désignent les échantillons les plus proches de l'hyperplan séparateur.	23
II.2	Cas linéairement non séparable	26
II.3	Support vector machines non linéaire	28
II.4	Transformation de l'observation par une fonction noyau reproduisante	29
III.1	Pavages temps fréquences associés aux représentations de Dirac et de Fourier	36
III.2	Pavages du plan temps fréquence pour la représentation temps-fréquence et la représentation temps-échelle	36
III.3	La nature hiérarchique de l'analyse multirésolution	42
III.4	Décomposition en cascade.	44
III.5	Algorithme de synthèse de Mallat.	45
III.6	Analyse en paquets d'ondelettes.	47
III.7	Fonction d'échelle et ondelette de Meyer.	50
III.8	Fonction d'échelle et ondelette de Daubechies pour $N = 2, 4, 8$	51

III.9	Fonction d'échelle et ondelette Symlets pour $N = 8$	52
III.10	Fonction d'échelle et ondelette Coiflets pour $N = 5$	52
IV.1	Vue éclatée des éléments constitutifs d'un roulement à une rangée de billes . .	62
IV.2	Répartition d'une contrainte radiale au sein d'un roulement	63
IV.3	Schématisation de la zone de charge d'une bague interne et d'une bague externe	64
IV.4	Aspect de la bague interne d'un roulement surchargé	64
IV.5	Ecaillage d'une piste de roulement	65
IV.6	Ecaillage d'une piste	65
IV.7	Cassure d'une bague	66
IV.8	Criques de corrosion sur une piste de roulement	66
IV.9	Marques d'indentation d'un chemin de roulement dues à la présence de conta- minant	67
IV.10	Mécanismes généraux d'usure en présence de contaminant	68
IV.11	Ecaillage d'un chemin de roulement	68
IV.12	Eclatement de billes	69
IV.13	Représentation schématique de la géométrie interne d'un roulement à une ran- gée de billes	70
IV.14	Figure représentative du banc d'essais	72
IV.15	Représentation temporelle des signaux vibratoires d'un roulemnet :(a) sain, (b) avec défaut bague extérieure, (c) avec défaut cage avec un élément dénoué, (d) avec défaut cage avec quatre éléments dénoués	73
IV.16	Représentation des données en 3D	74
IV.17	Paquet d'ondelette d'un roulement sain	75
IV.18	Paquet d'ondelette d'un roulement avec un défaut bague extérieure	76
IV.19	Taux de test et d'apprentissage avec $\sigma = 1, c \in [50 : 100 : 2000]$	78
IV.20	Taux de test et d'apprentissage avec $c = 50, \sigma \in [1 : 1 : 10]$	78
IV.21	Taux de test et d'apprentissage avec $c = 50, \sigma = 1$, degré du modél AR \in $[11 : 1 : 21]$	79
IV.22	Taux de test et d'apprentissage avec $c = 2500, \sigma \in [1 : 1 : 10]$	80
IV.23	Taux de test et d'apprentissage avec $\sigma = 2, c \in [2000 : 100 : 4000]$	80
IV.24	Représentation temporelle des signaux vibratoire d'un roulement : (a) avec défaut de bague intérieure, (b) avec défaut de bille, (c) avec défaut bague extérieure, (d) sain	85
IV.25	Banc d'essai	86

IV.26 Représentation des données en 3D 87

Liste des tableaux

IV.1	Causes de défauts dans les roulements.	69
IV.2	Les types des défauts.	73
IV.3	Architecture de RNA utilisé	76
IV.4	Effet de l'ordre de modèle et de paramètre de noyau sur le taux de classification.	79
IV.5	Résultats de classification	81
IV.6	Résultats de simulation	84
IV.7	Base de donnée	87
IV.8	Résultats de simulation	88
IV.9	Résultats de simulation	88
A.1	Principales méthodes de détection.	107
A.2	Choix d'une technique d'analyse.	109

Introduction generale

Dans de nombreux domaines industriels, la surveillance et le diagnostic des systèmes occupent une place grandissante. En effet, en raison de la complexité accrue des systèmes électriques modernes, leurs couts d'exploitation, essentiellement liés à la maintenance, se sont vus augmentés. De manière générale, il existe trois schémas de maintenance [1] :

- la maintenance préventive ou systématique, qui répond aux exigences d'un échancier entraînant des actions d'inspection et de controle puis des réparations si nécessaire,
- la maintenance corrective, effectuée en cas de panne du système,
- la maintenance prédictive ou conditionnelle, liée à la surveillance et au diagnostic du système et n'entraînant d'action de réparation que si une panne (présente ou future) est détectée.

La maintenance systématique est largement utilisée dans le cadre des systèmes critiques pour la sécurité ou la sûreté de fonctionnement, comme par exemple en aéronautique. Ce type de maintenance est difficilement remplaçable. La maintenance corrective, quant à elle, est la plus coûteuse puisqu'elle n'intervient qu'après une panne du système et donc après une période de non fonctionnement imprévue. La maintenance prédictive vise à pallier aux coûts de la maintenance corrective, en réduisant au maximum le temps d'indisponibilité des systèmes et surtout en pouvant planifier ces indisponibilités. Cette méthode anticipative permet donc d'assurer une meilleure continuité de service et ainsi de réduire les coûts d'exploitation sur le long terme [2].

Dans le contexte du diagnostic des systèmes électriques, les machines tournantes occupent une place prépondérante. C'est pourquoi il apparait comme important d'en effectuer une surveillance, afin d'en augmenter la qualité de service par l'utilisation de la maintenance prédictive. De nombreux types de pannes peuvent être diagnostiqués comme les défauts électriques, qu'ils

soient statoriques ou rotoriques, ou les défauts d'origine mécanique par exemple.

Des études statistiques ont montré que les pannes imputables aux roulements à billes sont prépondérantes quelle que soit la gamme de puissance des machines considérées [3], [4]. Selon les différentes sources citées, les pannes dûes aux roulements représentent entre 40 et 50% des pannes totales, précédant les défauts statoriques et rotoriques. Il est donc logique, d'un point de vue industriel et scientifique, de concentrer les efforts de détection des pannes et de diagnostic des machines tournantes sur les roulements à billes.

Ce travail de mémoire s'inscrit dans le diagnostic de défauts liés aux roulements à billes. Plusieurs techniques de détection de défauts peuvent être utilisées, on peut citer la thermographie infrarouge, la détection ultrasonore et l'analyse des vibrations mécaniques [5],[6]. La maintenance par analyse vibratoire est largement utilisée pour le diagnostic des installations industrielles, notamment le diagnostic des roulements des machines tournantes[7],[8],[9],[10]. Pour notre application, la détection de défaut du roulements repose sur l'analyse des vibrations enregistrées avec des accéléromètres mis en place pour le diagnostic. Sans connaissance a priori sur le système étudié, nous nous plaçons dans une approche "sans modèle" utilisant des méthodes de reconnaissance de formes (RdF).

Ce mémoire s'articule autour de quatre chapitres :

- Le premier chapitre est consacré à la démarche du diagnostic par reconnaissance des formes (RdF). Sa définition, sa formulation et ses principes sont décrits. Les différentes phases pour construire un système par RdF sont aussi illustrées. La dernière phase du processus de diagnostic par RdF consiste en une étape de classification. Pour notre application, étant donné que les classes de défauts sont connues préalablement (classes étiquetées), nous avons adopté la classification supervisée.
- Nous consacrons le deuxième chapitre à la phase finale du processus de diagnostic, la classification. Nous nous intéressons plus particulièrement à la classification supervisée où les séparateurs a vaste marge sont traités.
- Le troisième chapitre aborde le problème de la réduction de dimension en classification supervisée. Deux approches peuvent être utilisées : La sélection de variables ou l'extraction des caractéristiques. La première comporte principalement deux étapes : un algorithme de recherche et un critère d'évaluation mesurant la pertinence des sous-ensembles de variables choisies parmi l'ensemble des variables d'origine. la seconde consiste en une projection des variables initiales dans un espace de dimension plus réduit. Nous nous intéressons plus particulièrement aux méthodes d'extraction de caractéristiques basées sur des représentations temps-échelle ou de modelisation autorégressif.
- Le dernier chapitre présente l'application expérimentale du système de diagnostic sur deux

bases de données reconnues ; il débute par la présentation des roulements à billes ainsi que de leurs processus de dégradation. Nous décrivons, par la suite, les différents modules intervenant dans la conception de notre système de diagnostic automatique de l'état des roulements. Ce système constitue un cadre d'application des différentes approches de prétraitement choisies et exposées dans les chapitres 3. Pour mieux répondre aux besoins de l'exploitant, deux techniques de classification ont été explorées : les réseaux de neurones (MLP) et les séparateurs à vaste marge (SVM). Une comparaison de leurs performances est donnée et les résultats obtenus sont largement interprétés et discutés. Sachant que les problèmes de prise de décision par SVM nécessitent la prise en compte de plusieurs facteurs et essentiellement, le paramètre de la fonction noyau et celle de régularisation, nous nous orientons pour leur résolution vers l'optimisation de ces paramètres par un algorithme récent : l'optimisation par essaim de particule (PSO). Enfin les résultats d'expérimentation de la sélection automatique de paramètres sont présentées.

Chapitre I

Diagnostic par reconnaissance des formes

1 Introduction

Ce présent chapitre est consacré à l'élaboration d'un système de diagnostic par reconnaissance de forme correspondant à la nature de nos données.

La première partie du chapitre illustre la stratégie de surveillance qui peut être mise en place pour répondre aux besoins industriels. La deuxième partie du chapitre introduit la notion de diagnostic et passe en revue les différentes approches de diagnostic. Celle qui s'applique le mieux à notre problématique repose sur l'approche "sans modèle" et sera amplement détaillée. Les différentes étapes de diagnostic sont aussi exposées.

2 L'architecture globale d'un système de surveillance

2.1 Systèmes de surveillance

La surveillance automatique s'inscrit dans un processus global de conduite et de supervision. La figure (I.1) présente l'architecture générale d'un système de surveillance. On y retrouve les trois grandes fonctions - "Voir", "Comprendre" et "Agir" - nécessaires à une bonne surveillance telle que résumée dans [11]. Les deux tâches essentielles en diagnostic sont donc l'observation des symptômes de défaillance ainsi que l'identification de la cause de défaillance à l'aide d'un raisonnement logique fondé sur des observations acquises sur le système. À partir des informations disponibles sur son état de fonctionnement, l'objectif est de détecter, localiser et identifier les défaillances qui peuvent affecter sa sûreté de fonctionnement. Dans la plupart des cas, la surveillance se base sur des systèmes de traitement d'alarmes. Des valeurs de seuil sont définies préalablement par des experts du procédé selon des critères de sécurité fixés par l'exploitant. Ce système de traitement d'alarmes est l'outil de base pour aider l'opérateur dans sa tâche de

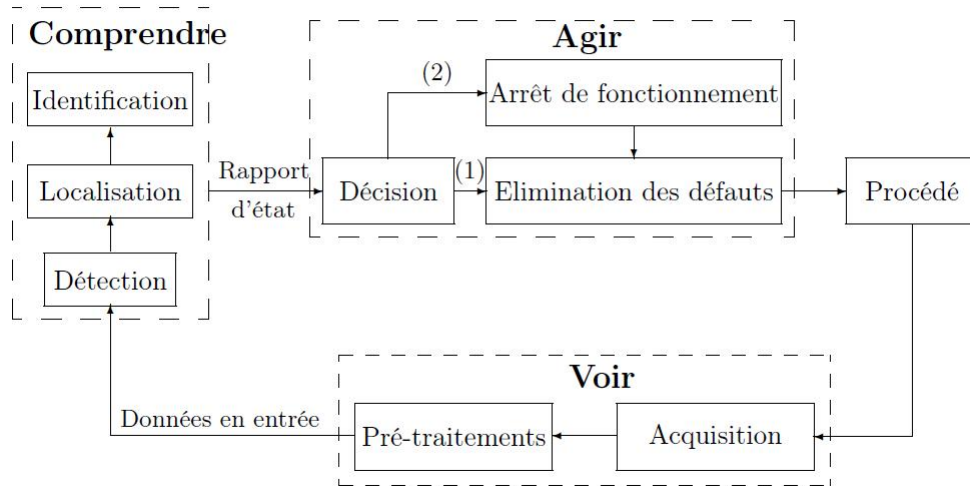


Figure I.1 – L'architecture globale d'un système de surveillance. L'état (1) représente un arrêt momentané du processus avec une reprise presque immédiate. Par contre dans l'état (2), la situation est plus grave et nécessite un arrêt de fonctionnement du processus.

surveillance. Il reste cependant à l'opérateur à analyser la situation et à prendre une décision adaptée. Son efficacité et sa performance à répondre à une alarme dépendent de son expérience. Cependant, un tel système de surveillance ne permet pas de gérer le caractère évolutif d'un état normal ou anormal du système. L'utilisation d'un système de surveillance basé sur des techniques plus évoluées permettra de pallier ces limites. Dernièrement, la surveillance et la maintenance des équipements industriels ont fait appel à des techniques de diagnostic de plus en plus sophistiquées. Il existe plusieurs procédures de diagnostic [12]. Le choix d'une approche est lié à la connaissance que l'on souhaite acquérir sur le système, mais aussi à la complexité de ce système. Dans ce travail de thèse nous traitons la surveillance des composants, les roulements, et nous nous focalisons dans la suite sur les composantes "Voir" et "Comprendre". On s'intéresse donc aux deux étapes élémentaires : la perception et le diagnostic, qui seront décrites dans les paragraphes suivants.

2.2 La perception

L'étape de perception constitue la source principale des informations sur le système. Elle se compose de deux phases. Une phase d'acquisition des données qui consiste à déterminer la configuration matérielle (le type, le nombre de capteurs à utiliser et la période d'échantillonnage, etc.) nécessaire pour le recueil des signaux sur le système étudié. Les signaux acquis doivent fournir l'information utile afin de juger de l'état de fonctionnement dans lequel se trouve le

système. Cette première phase est suivie d'une phase de conditionnement des signaux (filtrage, débruitage,...).

2.3 Le diagnostic

La définition donnée par les instances nationales et internationales de normalisation du terme diagnostic en automatique est "*l'identification de la cause probable de la (ou des) défaillance(s) à l'aide d'un raisonnement logique fondé sur un ensemble d'informations provenant d'une inspection, d'un contrôle ou d'un test*" [1].

Dans un système de surveillance, le diagnostic représente la composante principale et consiste à localiser et à identifier les éléments défectueux ou à déterminer si l'état courant du système est normal ou anormal à partir des descripteurs générés par le module de perception. Le diagnostic est constitué, dans la plupart des cas, de plusieurs étapes, figure (I.1) [13] :

- la détection consiste à identifier le mode de fonctionnement du système physique. En se basant sur une connaissance préalable des modes de fonctionnement normaux et anormaux du système, elle permet de décider s'il est en bon ou en mauvais état [14] ;
- la localisation consiste à déterminer de manière plus approfondie les composants défaillants ou la cause probable de la défaillance ;
- l'identification des défauts a pour but de déterminer le type (classe, taille, ...) d'un défaut (classification).

D'une manière générale, fournir un diagnostic sur système consiste à faire une comparaison entre l'information mesurée, lors du fonctionnement réel du système, et la connaissance a priori de ses modes de fonctionnement. En se basant sur les résultats obtenus de cette comparaison, l'utilisateur peut intervenir et mettre en place les actions correctrices nécessaires à un retour à la normale. Une bonne approche de diagnostic doit être capable de détecter les défauts de faible amplitude avant que ceux-ci ne se propagent dans tout le système et n'entraînent des pannes. Elle doit, en outre, éviter ou diminuer la fréquence des fausses alarmes qui occasionneraient des arrêts inutiles du système.

2.4 Quelques approches pour le diagnostic

Deux grandes familles de procédures sont définies dans la littérature. Les méthodes de diagnostic à base de modèles analytiques reposent sur le suivi des paramètres et des grandeurs de la machine, au moyen d'algorithmes d'observation. Le principal avantage de ces méthodes réside dans l'intégration d'une connaissance a priori du système et donc un filtrage de l'information. Les méthodes sans modèle, basées sur l'extraction d'informations, exploitent le traitement des

I.2 L'architecture globale d'un système de surveillance

signaux mesurés. Les signaux mesurables tels que la vitesse, les vibrations, les émissions sonores, etc, peuvent fournir des informations significatives sur les défauts. À partir de ces grandeurs caractéristiques du fonctionnement, la mise en oeuvre de méthodes décisionnelles permet de concevoir des systèmes de surveillance et/ou de diagnostic. Les performances de ces méthodes sont étroitement liées à la pertinence des indicateurs de défauts retenus et à la finesse d'analyse des mesures (la perception).

Dans ce qui suit, nous passons en revue quelques approches utilisées pour le diagnostic. Le domaine est très vaste, le but n'est pas de faire une synthèse exhaustive mais de montrer les différentes possibilités qui s'offrent au concepteur de système de diagnostic. Plusieurs méthodes sont proposées dans la littérature pour établir un diagnostic sur l'état de fonctionnement d'un système donné. Le choix d'une méthode par rapport à une autre dépend fortement de la connaissance a priori du système ainsi que du type de la représentation retenue pour ses modes de fonctionnement. Elles peuvent être classées comme suit :

1. les méthodes basées sur des modèles analytiques ou mathématiques qui peuvent être issus de différents domaines :
 - traitement du signal : séries temporelles, analyse spectrale, analyse temps-fréquence, etc.
 - automatique : observateurs, filtrage de Kalman, etc.
2. les méthodes symboliques : les systèmes experts,... etc.
3. les méthodes basées sur la reconnaissance des formes : approches statistiques, approches floues.

Quand les observations issues du système sont de type numérique et que l'on dispose d'un modèle mathématique du système, l'approche par modèle mathématique est privilégiée pour le diagnostic. La modélisation se base sur une bonne connaissance du comportement entrée/sortie du procédé. Cette connaissance est généralement exprimée sous forme de lois physiques, c'est ce qu'on appelle un modèle de connaissance ou modèle interne. Différentes approches de diagnostic ont été développées par la communauté de l'automatique [15],[16]. Le point commun entre toutes ces approches est la nécessité d'avoir une bonne connaissance du système pour construire un modèle et pour trouver des relations de redondance analytique. Dans ce type de modèle, un diagnostic peut être défini littéralement comme un état plausible du système physique et habituellement décomposé en deux étapes. L'étape de détection produit des symptômes. L'étape de localisation des défauts collecte les symptômes des tests de détection et fournit les diagnostics possibles. Parmi les différentes méthodes utilisant des modèles mathématiques, on trouve principalement l'espace de parité, les observateurs et l'estimation paramétrique.

Lorsque le mode de représentation de la connaissance est de type symbolique, l'approche par système expert est particulièrement adaptée [14]. Ce type d'approche est adapté aux problèmes nécessitant la manipulation d'une quantité d'information importante de données non homogènes et/ou contextuelles. Pour des applications industrielles on pourra se référer à [17]. En revanche, lorsqu'on ne peut construire un modèle mathématique correspondant aux divers modes de fonctionnement du procédé à surveiller ou si sa complexité le rend inexploitable, les méthodes basées sur la reconnaissance des formes sont les plus adaptées. Par conséquent, l'approche par reconnaissance de formes sert à la conception d'un système de diagnostic en lui procurant les moyens d'appréhender la connaissance a priori incomplète des modes de fonctionnement.

3 Système de diagnostic par reconnaissance des formes

Rappelons que sur des problèmes complexes, la modélisation du procédé est souvent difficile à mettre en oeuvre. On privilégie alors l'approche par reconnaissance de formes. Celle-ci sera présentée comme une solution alternative à l'approche de diagnostic par modèle puisque les modes de fonctionnement sont modélisés mais pas de manière analytique. On ne dispose pas de modèle issu de considérations sur les comportements du système ou de ses composantes, mais d'une base de données composée à partir d'un ensemble de mesures (individus) de ses modes de fonctionnement. En fonction de l'application considérée, un individu est un ensemble de mesures réalisées sur un système physique ou un ensemble d'informations collectées lors de l'observation d'un phénomène. Dans la suite, un individu est représenté par le terme "forme".

3.1 Les approches de reconnaissance de formes

La reconnaissance de formes se base sur la définition d'algorithmes permettant de classer des formes en les comparant à des formes-types. Elle intervient dans de nombreux domaines tels que la reconnaissance vocale, la reconnaissance de caractères, l'automatisation industrielle, le diagnostic médical, la classification de documents,...etc. De manière générale on distingue deux types de reconnaissance des formes :

- les approches syntaxiques, structurelles et la mise en correspondance des formes (Template Matching) utilisent les hypothèses sur les distributions des données à l'intérieur des classes. Les procédures de classification, dans ce cas, sont construites à l'aide d'hypothèses probabilistes (Ex. classement Bayésien). Elles se basent sur une représentation des formes à l'aide de grammaires et nécessitent des moyens de calculs importants. De ce fait, ce type d'approche n'est pas approprié aux applications en temps réel [18] ;

- la reconnaissance de formes statistique, contrairement à la précédente, s'appuie sur une représentation numérique des formes en se basant sur des méthodes paramétriques ou non paramétriques. Les méthodes non paramétriques ne posent pratiquement aucune hypothèse sur la forme des distributions (Ex K-plus proches voisins). Parmi les méthodes les plus utilisées, on cite les méthodes connexionnistes qui constituent un sous-ensemble des méthodes statistiques [19], [20]. Un état de l'art concernant les méthodes statistiques a été proposé dans [18]. C'est l'approche statistique qui sera utilisée et développée dans la suite du document.

3.2 Principe et formulation

On considère N formes, chacune définie par un ensemble de p paramètres regroupées dans un vecteur appelé vecteur forme x . L'espace R^p défini par ces paramètres est appelé l'espace de représentation. Les formes-types ou prototypes constituent des points représentatifs de l'espace de représentation. Dans le cas réel, les observations sont bruitées et sont rarement confondues avec l'un des prototypes. Cela signifie que chaque forme-type est représentée par une zone géométrique désignée par le terme "classe" en reconnaissance de formes.

Le principe est qu'on observe des formes de M classes différentes, $\Omega = \{y_i, i = 1, \dots, M\}$ L'ensemble Ω définit "l'espace de décision". L'objectif est de construire des frontières réalisant une partition de façon à affecter un nouveau vecteur forme à l'une des classes $\{y_1, \dots, y_M\}$. Cette association désigne l'opération de classement ou discrimination.

En diagnostic, l'ensemble des données issues du système de perception sert à construire le vecteur forme. La notion d'état de fonctionnement du procédé permet de construire l'ensemble d'apprentissage, celui-ci est composé de vecteurs formes associés à chaque mode de fonctionnement ou classe.

De nombreux travaux ont permis de montrer l'intérêt de ce type d'approche dans plusieurs domaines d'applications[21], [22]. Le domaine industriel est devenu l'un des premiers champs d'application pour le diagnostic par reconnaissance des formes [23], [24], [25].

4 Construction d'un système de reconnaissance de formes

Un système de reconnaissance de formes se décompose généralement en cinq étapes séquentielles [26], figure(I.2) . La qualité de chaque étape du processus dépend de la qualité des étapes précédentes. Néanmoins, pour faciliter la compréhension, nous les regroupons en trois phases principales : phase d'analyse, phase de choix de la méthode de décision et enfin la phase

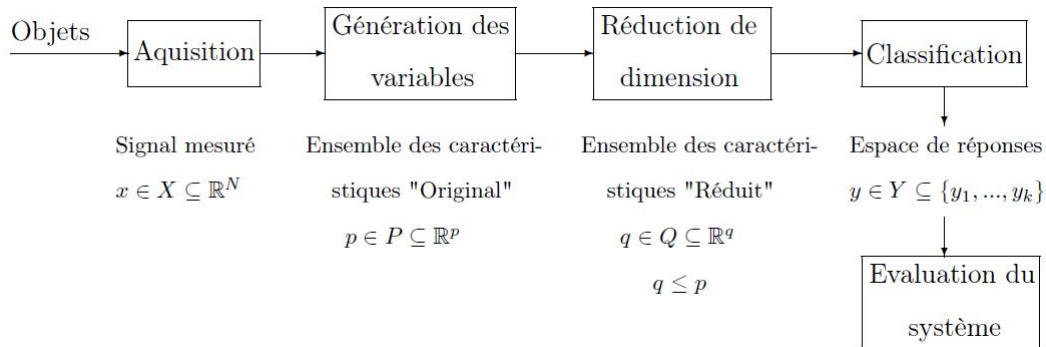


Figure I.2 – Système de reconnaissance de formes.

d'exploitation.

4.1 Phase d'analyse

Cette première phase consiste à déterminer les paramètres et les techniques de prétraitement nécessaires pour permettre une bonne classification. La collecte d'informations sur le procédé est effectuée grâce au module d'acquisition, durant l'étape de perception. Suivant le domaine d'application, ces données peuvent être de nature qualitative, quantitative ou les deux à la fois. L'exploitation de ces connaissances permet d'élaborer le module de diagnostic.

4.1.1 Détermination de l'espace de représentation

Il s'agit de construire le vecteur forme x . Les données collectées par le module de perception sont des vecteurs réalisations de variables qui correspondent à des mesures réalisées sur un système physique ou à des informations collectées lors d'une observation d'un phénomène. Ces variables ne sont pas toutes aussi informatives. Elles peuvent correspondre à du bruit, peuvent être peu significatives, corrélées ou redondantes, aberrantes ou simplement inexploitable. De ce fait, une étape de génération de variable s'impose.

L'étape de génération est définie comme la transformation initiale qui vient après la mesure du signal, produisant un ensemble original de paramètres dénoté $p \in P \subset \mathbb{R}^p$. On utilise en général des méthodes d'analyse de données ou de traitement du signal pour effectuer cette transformation. Chaque observation effectuée à un instant donné sera caractérisée par l'ensemble de p variables, $x = (x_1, x_2, \dots, x_p)$. L'objectif fondamental est d'accentuer les informations importantes du signal acquis. Cela implique une transformation du vecteur représentant le signal du domaine temporel vers un domaine où les informations contenues dans le signal seront mieux représentées. Cependant, suivant l'information recherchée plusieurs possibilités s'offrent à nous.

I.4 Construction d'un système de reconnaissance de formes

Deux techniques d'analyse du signal pouvant répondre à nos besoins sont détaillées dans le chapitre 3.

En diagnostic, l'étape de caractérisation des états de fonctionnement du système consiste à traiter les mesures recueillies lors de la phase d'acquisition pour extraire une représentation stable et concise, nécessaire à la mise en forme des caractéristiques du fonctionnement normal ou anormal du système. Elle représente le fruit de l'expérience acquise sur le processus et conditionne le succès du système de diagnostic par reconnaissance de formes.

Les données résultantes sont ensuite regroupées dans un tableau numérique de dimension $(N \times p)$ qui correspond à un ensemble de N formes pour lesquelles on connaît la valeur de p variables quantitatives. Les N formes (x_1, \dots, x_N) recueillies sur le système constituent l'ensemble d'apprentissage. La détermination de l'espace de représentation constitue l'étape clé dans la construction du module de diagnostic. La dimension p des variables disponibles à la fin de cette étape peut être très importante. Une forme peut se distinguer mieux des autres en éliminant des variables qui ne sont pas significatives ou en changeant de représentation. Ainsi, travailler dans un espace de représentation réduit, facilitera largement la tâche de la classification. En mode d'apprentissage, l'étape de réduction de variables permet de définir l'espace de représentation le plus approprié pour la classification. C'est dans cet espace de représentation que sont définis les formes à reconnaître.

4.1.2 Réduction de la dimension de l'espace de représentation

Il s'agit de réduire la dimension de l'espace de représentation afin, d'une part, de diminuer le temps de classification d'une nouvelle observation et, d'autre part, d'éviter la dégradation des performances de classification à cause d'une dimension trop élevée de l'espace de représentation. Cette opération de réduction de dimension peut être obtenue par une sélection ou une extraction de variables. La sélection de variables consiste à ne conserver, parmi l'ensemble des p variables originales, que les plus pertinentes ou les plus informatives, au sens d'un critère de séparabilité entre les classes et d'un critère de condensation des points d'une même classe, comme par exemple le rapport de Fisher. Tandis que l'approche d'extraction, contrairement à la première, utilise toute l'information des variables initiales en les projetant le mieux possible dans un espace de dimension plus petite. En combinant entre elles les variables initiales, on définit de nouvelles variables par transformations linéaires ou non linéaires.

4.1.3 Détermination de l'espace de décision

Le diagnostic par reconnaissance de formes est réalisé en utilisant une méthode de classification qui associe une nouvelle observation à une classe correspondant à un mode de fonction-

nement. Les méthodes de classification se divisent généralement en deux familles : les méthodes supervisées, et les méthodes non supervisées [26].

L'espace de décision est défini par l'ensemble des classes possibles. Si l'on dispose d'un ensemble de formes étiquetées dont la classe d'appartenance est connue, la classification est dite supervisée. L'étiquetage permet de définir (ou d'apprendre) un ensemble fini de M classes d'objets en général mutuellement exclusives. L'objectif est alors de définir des règles d'affectation pour une forme inconnue x à l'une des M classes apprises. Ainsi, on parle de prédiction en statistique et de classement en reconnaissance de formes. Nous utiliserons, par la suite, les termes discriminateur ou classifieur pour désigner une règle d'affectation.

Dans le cas contraire, si la classe d'origine de chaque observation de l'ensemble d'apprentissage - l'espace de décision - n'est pas parfaitement connue, l'apprentissage des classes se fait en mode non supervisé.

Étant donné que les deux modes de fonctionnement, sans défaut et avec défaut, ainsi que l'ensemble des classes de défauts sont connus préalablement dans notre application, nous ne nous intéressons, dans le cadre de notre travail, qu'aux méthodes de classification supervisée.

4.2 Phase de choix d'une méthode de décision

Une fois l'espace de décision défini, il s'agit de construire des frontières de décision entre les classes. Les méthodes de décision utilisées définissent une règle de décision pour la classification des nouvelles observations aux différentes classes de l'ensemble d'apprentissage. La performance de ces méthodes est estimée en utilisant un indice de performance. En général on choisit la probabilité de mauvaise classification. La règle de décision peut être construite en utilisant une approche statistique ou une approche analytique [27].

Dans une approche statistique, on associe à chaque vecteur forme x une densité de probabilité conditionnelle $f(x|y_i)$ par rapport à chaque classe de y . Les concepts issus de la théorie statistique de la décision et de l'analyse discriminante [28] peuvent être utilisés pour établir les frontières de décision entre les classes. En fonction de la connaissance, complète ou incomplète, dont on dispose sur la distribution de $f(x|y_i)$ on trouve différentes approches.

- lorsque la connaissance est complète, modèle et paramètres de $f(x|y_i)$ étant supposés connus, la règle de décision optimale est basée sur la théorie Bayésienne de la décision,
- lorsque la connaissance est incomplète c'est-à-dire dans la plupart des cas, le modèle de $f(x|y_i)$ inconnu, différentes stratégies peuvent être utilisées :
 - si la forme de $f(x|y_i)$ est supposée connue, on estime les paramètres de $f(x|y_i)$,
 - si la forme de $f(x|y_i)$ est inconnue, on utilise des méthodes non paramétriques pour

l'estimation de $f(x|y_i)$

Que ce soit pour les méthodes paramétriques ou non paramétriques, l'estimation de la densité conditionnelle $f(x|y_i)$ se fait sur un ensemble d'apprentissage.

Dans l'approche analytique, la fonction discriminante est déterminée en estimant les paramètres d'une fonction mathématique de manière à séparer au mieux les classes c'est-à-dire à minimiser la probabilité de mauvaise classification. Le choix du modèle de la fonction mathématique dépend de la complexité de la frontière de décision. En pratique, on commence par des fonctions linéaires puis si les performances de classification ne sont pas satisfaisantes, on passe à des modèles plus complexes, quadratiques, réseaux neuromimétiques, etc.

C'est à l'issue de ces deux premières phases qu'est élaboré le module de diagnostic qui décidera à quelle classe y_i sera affectée une nouvelle observation x . Nous nous sommes intéressés aux méthodes de classification non paramétrique, plus adaptées aux cas industriels, pour lesquelles la densité de probabilité est complètement inconnue.

4.3 Phase d'exploitation

4.3.1 Évaluation du système de diagnostic

Le système de reconnaissance de formes permet de classer chaque nouvelle observation recueillie sur le système dans une des classes connues, en appliquant la règle de décision élaborée dans la phase précédente. La détermination de cette classe permet de connaître le mode de fonctionnement actuel du système en élaborant un diagnostic.

Une fois l'apprentissage effectué, il est indispensable d'évaluer la qualité de la règle de classement induite. L'idée est d'appliquer une phase de test pour estimer l'erreur réelle de classement (taux de mauvais classement). Dans la pratique, l'échantillon S de taille N est divisé en deux ensembles : un ensemble d'apprentissage et un ensemble test. Selon la répartition entre les deux ensembles, il existe de nombreuses méthodes pour estimer l'erreur de classement :

- **la re-substitution** : les deux ensembles d'apprentissage et de test sont identiques et correspondent à l'ensemble de l'échantillon. Si l'on note n_e le nombre d'erreurs commises l'hors de la phase de test, on aura $Erreur = n_e/N$;
- **la méthode Holdout** : on partitionne l'échantillon S en un ensemble d'apprentissage et un ensemble test. La répartition entre les deux ensembles doit être aléatoire ; en général dans des proportions $1/2, 1/2$ pour chacun des deux ensembles ou $2/3$ pour l'ensemble d'apprentissage et $1/3$ pour l'ensemble test. Si m est la taille de l'ensemble test, on aura $Erreur = n_e/m$;

- **la D-validation croisée** [29] : l'échantillon est partitionné en D parties de tailles (approximativement) égales. L'apprentissage se fait alors sur les $(D - 1)$ sousensembles et le test sur le D^{me} sous-ensemble restant. On réalise alors D apprentissages en laissant à chaque fois une des parties de côté pour le test. Si l'on note $n_e(d)$ le nombre d'erreurs de classement commises sur le $d^{\text{ème}}$ sous-ensemble, l'estimation de l'erreur est la moyenne arithmétique des $n_e(d)$, $Erreur = \frac{1}{D} \sum_{d=1}^D n_e(d)$;
- **la méthode du Leave-one-out** : cette technique est un cas particulier de la validation croisée pour lequel $D = N$. La procédure de classement est répétée N fois sur $(N - 1)$ instances ce qui la rend très coûteuse en temps de calcul, mais recommandée si l'ensemble E est de petite taille. En effet, elle conduit à un minimum de biais dans l'estimation de l'erreur ;
- **la méthode de ré-échantillonnage (Bootstrap)** [30],[31] : étant donné un échantillon E de taille N , on tire avec remise un ensemble d'apprentissage de même taille N (un élément de N peut ne pas appartenir à l'ensemble d'apprentissage, ou y figurer plusieurs fois), l'ensemble test est N . L'erreur est alors la moyenne des erreurs obtenues pour un certain nombre d'itérations de l'algorithme d'apprentissage.

Dans le cas supervisé, la méthode de re-substitution est connue pour être facile et rapide à mettre en oeuvre. Cependant, l'ensemble de test étant identique à l'ensemble d'apprentissage, l'estimation risque de souffrir d'un biais optimiste quasi systématique. On ne teste alors pas la capacité en généralisation mais plutôt la faculté d'apprendre éventuellement par coeur. La méthode "Holdout" est connue pour être pessimiste et biaisée surtout lorsque la taille des ensembles d'apprentissage *et/ou* de test est faible. Un partitionnement différent des ensembles d'apprentissage et test donne un résultat d'estimation différent, ce qui se traduit par une variance importante. A l'inverse, la technique du "Leave-One-Out (LOO)" est sans biais mais souffre de l'influence du choix de la base d'apprentissage. Les méthodes de "Bootstrap", quant à elles, fournissent un bon estimateur de l'erreur de variance moins importante que la méthode du "Leave-one-out", mais sont encore plus coûteuses en temps de calcul. Elles sont très utiles pour les échantillons de petite taille [30]. La technique de validation croisée réalise un bon compromis entre le biais et la variance [29].

5 Approches de discrimination non-paramétrique

Les approches non-paramétriques conviennent particulièrement au cas où les distributions des données sont inconnues ou multimodales. Trois étapes sont communément suivies :

- l'estimation des densités conditionnelles $P(x|y_i)$;
- l'estimation des probabilités a posteriori $P(y_i|x)$;
- l'estimation des fonctions de classement ou la construction directe des frontières de décision.

Dans la dernière approche, nous trouvons par exemple la méthode des moindres carrés, la méthode de Fisher, les réseaux de neurones et les machines à vecteurs support (SVM) [28]. Ce sont les méthodes appartenant à cette dernière approche qui seront utilisés dans l'étape de classification.

5.1 L'estimation des fonctions de classement

Les techniques de surveillance par l'estimation des fonctions de classement sont fondées sur l'existence d'une base de données d'apprentissage constituée d'un ensemble de mesures de différents modes de fonctionnement normal ou anormal du système à surveillé. Le principe est de trouver une relation entre les variables d'entrée et les variables de sortie. Les variables d'entrée peuvent être qualitatives ou quantitatives. Pour ce dernier cas, la sortie du réseau de neurones représente l'état du fonctionnement du système considéré. Le problème de la surveillance est donc considéré comme un problème de reconnaissance des formes. La forme à reconnaître est caractérisée par l'ensemble des données (quantifiables et/ou qualifiables) et les classes d'appartenance représentant les différents modes de fonctionnements. Ce type de méthode peut ainsi nous renseigner sur l'état de fonctionnement tout en assurant à la fois la fonction de détection (fonctionnement normal ou pas) et la fonction de diagnostic, puisqu'il précise le mode de défaillance.

6 Outils pour la classification

6.1 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.

6.1.1 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

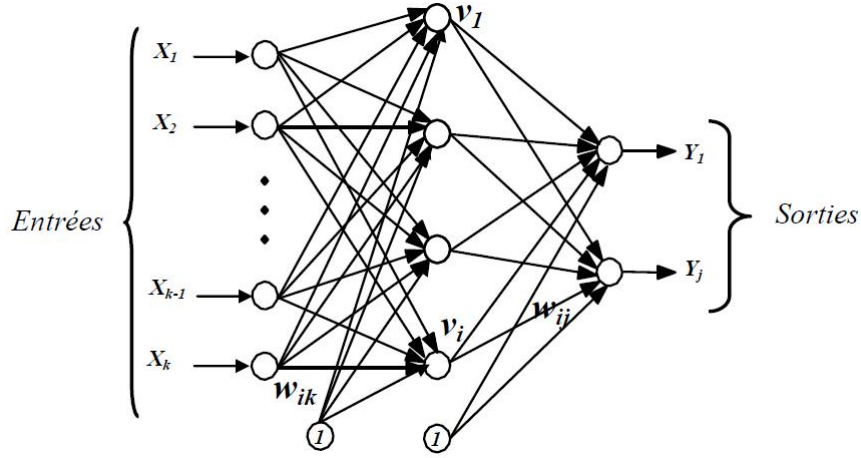


Figure I.3 – Perceptron Multi Couches.

classe donnée.

Soit le réseau à deux couches décrit par la figure (I.3) 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 :

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

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 :

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

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

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

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 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 :

$$\Delta w = -\eta \frac{\sigma E}{\sigma w} = -\eta \sum_l \frac{\sigma E^l}{\sigma w} \quad (\text{I.4})$$

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 perfor-

mances du réseau.

7 Conclusion

Nous avons consacré ce présent chapitre à la démarche de diagnostic par reconnaissance de formes (RdF). Après avoir introduit les concepts de base de l'approche de reconnaissance de formes, nous avons présenté les principales étapes de construction d'un système de diagnostic par RdF. Le module d'acquisition collecte l'ensemble des informations sur le procédé. Une fois l'étape de perception effectuée, le processus de diagnostic se déclenche en se basant sur l'ensemble des signaux vibratoires enregistrés. Seule l'approche de diagnostic par reconnaissance des formes (RdF) est prise en compte et utilisée dans nos travaux.

Dans la dernière partie du chapitre, nous avons rappelé les principes de la théorie de la décision et nous avons analysé une des principale méthode discriminante utilisée pour la reconnaissance des formes : le MLP.

Bien que performante sur beaucoup de problèmes, la principale lacune du MLP est la difficulté d'adapter correctement l'architecture du classifieur (réseau de poids) au problème de classification. L'autre maillon faible de cette méthode est son sensibilité à la dimensionalité des données. Comme nous allons le voir dans le chapitre suivant, le SVM procure un cadre théorique qui résout tous ces lacunes.

Chapitre II

Les séparateurs a vaste marge

1 Introduction.

Depuis quelques années, de nouvelles méthodes d'apprentissage se développent sur la base de la théorie de l'apprentissage statistique de Vapnik et Chervonenkis [32]. Celles-ci visent à obtenir la structure d'une classe D de détecteurs, supposée de VC-dimension h , de probabilité d'erreur $P_e(d)$ minimale, à partir d'une base d'apprentissage de taille n . Suivant Vapnik et Chervonenkis, un intervalle de confiance liant la probabilité d'erreur P_e et la probabilité d'erreur empirique P_{emp} pour tout détecteur de D peut être caractérisé la relation [33] :

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

avec une probabilité au moins égale à $1 - \epsilon$ où la largeur de l'intervalle de confiance ϕ est donnée par [33] :

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

Remarquons à partir de (II.1) et (II.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 (II.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 conventionnels, où le choix d'une architecture de réseau permet de fixer ϕ . La deuxième méthode est mise en oeuvre pour obtenir des systèmes décisionnels appelés Support Vector Machine (SVM). Le SVM est un modèle discriminant qui tente de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les don-

nées 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. Les valeurs des paramètres de noyaux affectent directement la complexité de la frontière de décision du classifieur. Aussi, ses valeurs influencent le nombre de vecteurs de support du classifieur. De nombreux travaux ont démontré la supériorité du SVM sur les méthodes discriminantes classiques telles que les MLP.

Les Machines à Vecteurs Supports sont de puissants classifieurs qui font appel aux notions de Minimisation du Risque Empirique (MRE) et de Minimisation du Risque Structurel (MRS). Ils sont par essence des classifieurs bi-classes qui visent à séparer les exemples de chaque classe au moyen d'un hyperplan choisi de manière à garder un maximum de marge de séparation entre celui-ci et n'importe quels exemples d'apprentissage.

2 Cas des données linéairement séparables

Etant donnée une base d'exemples $A_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ qui constitue un ensemble de données linéairement séparables, où x_k représente une observation et y_k la décision associée que l'on suppose appartenir à $\{-1, +1\}$, le but des SVM est de rechercher un hyperplan optimum d'équation :

$$\langle w; x \rangle - \lambda_0 = 0. \tag{II.3}$$

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

Dans le cas des Supports Vector Machines un hyperplan est dit optimal s'il sépare sans erreur les exemples de la base d'apprentissage A_n et s' il est situé à la distance maximale des vecteurs x_k les plus proches parmi ceux de A_n .

Un hyperplan possédant de telles caractéristiques est appelé hyperplan à marge maximale. Il existe une infinité de solutions pour un tel problème. On peut chercher un hyperplan canonique tel que les paramètres w et λ_0 vérifient [34] :

$$\min_{x \in A_n} |\langle w; x \rangle| - \lambda_0 = 1 \tag{II.4}$$

Ainsi que nous le verrons par la suite, cette contrainte est préféré à d'autres car elle simplifie la formulation du problème. On remarque que la distance d'un point x_k de l'espace à l'hyperplan

d'équation $\langle w; x \rangle - \lambda_0 = 0$ est donnée par [34] :

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

Des équations (II.4) et (II.5) , on peut déduire que la norme $\|w\|$ de w doit être égale à l'inverse de la distance de l'observation la plus proche à l'hyperplan défini par $\langle w; x \rangle - \lambda_0 = 0$.

D'autre part, l'équation (II.4) implique les deux inégalités suivantes :

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

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

ce qui peut se résumer par une description unifiée de l'ensemble des conditions :

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

En utilisant les deux équations (II.5) et (II.8) , on déduit que pour tout x_k de A_n , l'inégalité suivante est vérifiée [34],[35] :

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

On peut en plus déduire de (II.4) que la valeur de la marge, définie comme étant la distance des points les plus proches de l'hyperplan à ce dernier est $\frac{1}{\|w\|}$, comme l'illustre la figure (II.1). Ainsi l'hyperplan à marge maximale recherché maximise $\frac{1}{\|w\|}$ ou encore minimise $\frac{\|w\|^2}{2}$ sous les contraintes (II.8) [33],[34],[35].

2.1 Recherche de la solution

On est amené à rechercher un point selle $w^*, \lambda_0^*, \alpha^*$ du lagrangien :

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

où $\alpha = (\alpha_1 \dots \alpha_n)$ désigne les n multiplicateurs de Lagrange ($\alpha_i \geq 0, \forall i$). Au point selle, la valeur de L est minimum pour $w=w^*$ et $\lambda_0 = \lambda_0^*$, et maximum pour $\alpha = \alpha^*$. Les conditions d'annulation des dérivées partielles du lagrangien mènent directement aux relations vérifiées par l'hyperplan

II.2 Cas des données linéairement séparables

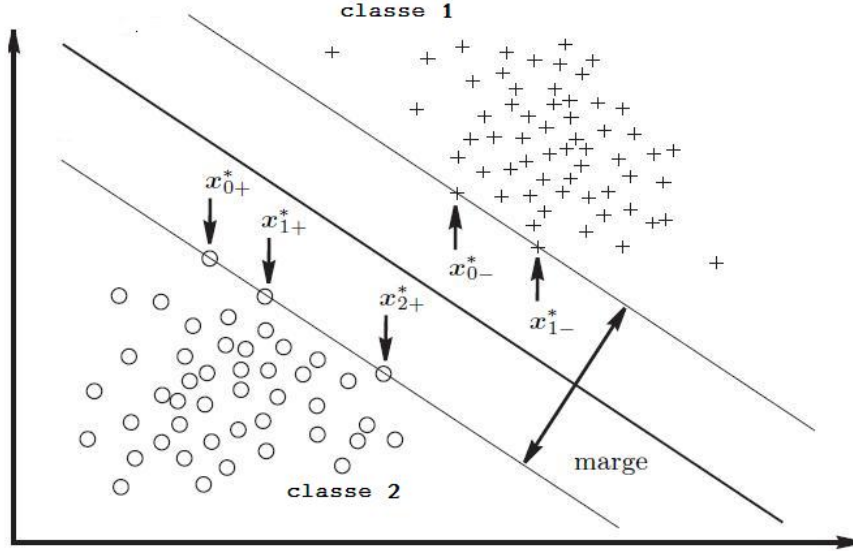


Figure II.1 – Principe des SVM dans le cas où les deux classes sont linéairement séparables. Les Support Vectors, indiqués par des flèches, désignent les échantillons les plus proches de l'hyperplan séparateur.

optimal [33],[35] :

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

On peut ainsi transformer le problème initial en son dual consistant à maximiser par rapport à α la forme suivante du lagrangien [33] :

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

le problème d'optimisation (II.12) peut être exprimé sous une forme matricielle comme suit [33] :

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

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 de dimension $(n,1)$, $f = [1, 1, \dots, 1]^T$.

afin d'obtenir α^* , il suffit alors de résoudre le problème d'optimisation quadratique (II.13) pour déterminer $\alpha_k, k = 1, \dots, n$, et calculer en suite $\langle w^*; x \rangle = \sum_{k=1}^n \alpha_k^* y_k \langle x_k; x \rangle$. Les conditions de

Chapitre II. Les séparateurs a vaste marge

Kuhn – Tucker donnent [34] :

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

pour tous les points x_k de A_n . Ainsi, seuls les points x_k qui vérifient :

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

ont des multiplicateurs de Lagrange α_k^* non nuls. Ces points sont nommés Support Vectors (SV) ou vecteurs supports. Ils sont les points 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 (II.1) , les x_{k-}^* et x_{k+}^* désignent les Supports Vectors associés à l'une ou l'autre des hypothèses en compétition. On note qu'à partir de deux Support Vectors x_{i-}^* et x_{j+}^* on peut facilement calculer le seuil λ_0^* correspondant à l'hyperplan optimal en utilisant l'équation (II.15). Ainsi on a :

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

Ceci peut être aussi exprimé en fonction des multiplicateurs de Lagrange en combinant les équations (II.11) et (II.16) :

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

Ici le seuil λ_0^* est calculé à partir de deux Support Vectors seulement, mais il est parfois conseillé, pour une meilleure robustesse numérique, de le calculer en utilisant tous les Support Vectors définissant la marge [35].

L'équation de l'hyperplan optimal peut être écrite finalement sous la forme :

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

où λ_0^* est donné par (II.17).

Pour savoir à quelle classe une nouvelle observation x appartient , il suffit de chercher le signe de l'expression (II.18) afin de voir de quel côté est l'observation par rapport à l'hyperplan et pouvoir ainsi prendre une décision.

II.3 Extension au cas non-linéairement séparable.

la fonction de décision est donnée par [35] :

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

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

une entrée quelconque est classée dans :

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

3 Extension au cas non-linéairement séparable.

Dans la plupart des problèmes rencontrés en pratique, les données ne sont pas séparables par un hyperplan .voir figure (II.2). Il convient de modifier l'approche précédente afin de prendre en compte la possibilité d'observations mal classées. Une solution à ce problème consiste à reformuler les conditions (II.6) et (II.7) en incorporant des paramètres ζ_k permettant d'assouplir les contraintes :

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

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

où $\zeta_k \geq 0$.Cette méthode est connue sous le nom de *soft margin* Support Vector Machine [33],[36]. Une erreur est commise sur une observation x_k par la structure de décision utilisée si ζ_k est supérieur à 1. La fonction $\sum_{k=1}^n \zeta_k$ constitue donc un coût caractérisant l'erreur commise sur les éléments de l'ensemble d'apprentissage. Le problème d'optimisation doit être reformulé de manière à minimiser l'erreur de classification de la structure de décision conjointement avec la maximisation de la marge. Ceci revient à minimiser :

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

sous les contraintes (II.21) et (II.22), "c" étant un réel positif préalablement fixé permettant de contrôler le compromis entre le fait de maximiser la marge et minimiser les erreurs de classification commises sur l'ensemble d'apprentissage. On parle alors de classifieur à marge souple. Notons qu'il est souvent préférable de tolérer certaines erreurs, au bénéfice d'une marge

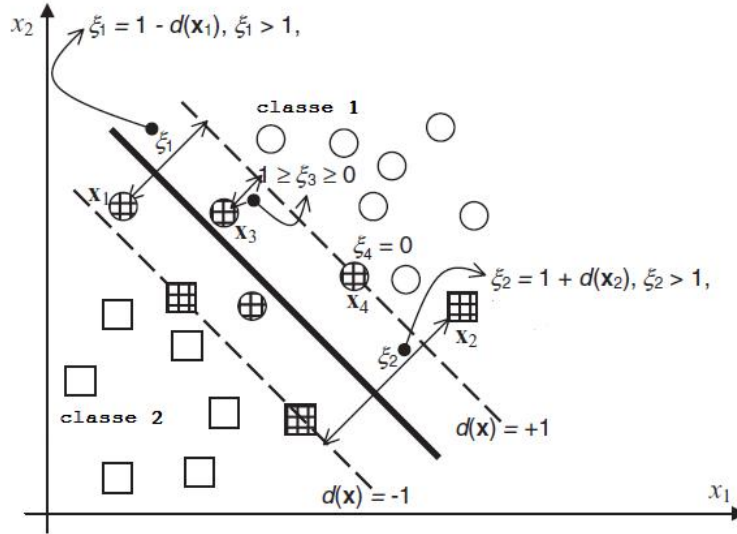


Figure II.2 – Cas linéairement non séparable

plus grande car ces erreurs peuvent être dues à des outliers, observations aberrantes, non-significatives de la classe qui leur est associée.. Le fait de donner une valeur importante à c revient à attribuer une forte pénalité aux erreurs. L'hyperplan dont les paramètres minimisent (II.23) sous les conditions (II.21) et (II.22) est parfois appelé hyperplan *optimal généralisé*.

3.1 Recherche de la solution

En appliquant les conditions de *Kuhn – Tucker* , on est amené à rechercher un point selle $w^*, \lambda_0^*, \zeta^*, \alpha^*, \beta^*$ du lagrangien :

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

où $\alpha = (\alpha_1 \dots \alpha_n)$ et $\beta = (\beta_1 \dots \beta_n)$ désignent les multiplicateurs de Lagrange ($\alpha_i \geq 0$ et $\beta_i \geq 0, \forall i$). Au point selle, la valeur de L 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 :

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

II.4 Elaboration des Supports Vector Machines nonlinéaires

Les conditions d'annulation des dérivées partielles du lagrangien mènent directement aux relations vérifiées par l'hyperplan optimal :

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

On peut ainsi transformer le problème initial en son dual consistant à maximiser la forme suivante du lagrangien :

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

sous les contraintes

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

afin d'obtenir α^* .

Notons ici l'identité des formes duales (II.12) et (II.27). Cependant, le point selle de (II.27) doit être trouvé en respectant une borne supérieure sur les multiplicateurs de Lagrange $\alpha_k (0 \leq \alpha_k \leq c)$.

Il suffit de remplacer w^* par sa valeur proposée par la première relation de (II.26) pour trouver l'équation de l'hyperplan optimal en fonction des α_k^* .

Comme dans le cas des données linéairement séparables, les Support Vectors correspondent aux observations les plus proches de l'hyperplan optimal, mais de plus correctement classées par la structure de décision constituée. Il est toutefois possible qu'il existe des observations qui soient mal classées. A ces observations sont aussi associés des multiplicateurs de Lagrange non-nuls.

4 Elaboration des Supports Vector Machines nonlinéaires

Pour permettre aux SVM de répondre à des problèmes complexes de classification , l'algorithme initial présenté précédemment a été transformé pour élaborer des structures de détection non-linéaires. voir figure (II.3). L'extension au cas non-linéaire peut être effectuée en transformant les observations à l'aide d'une fonction ϕ puis en appliquant un détecteur linéaire [28],[37],[38]. voir figure (II.4). En appliquant l'équation (II.27) dans l'espace engendré par ϕ , on obtient :

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

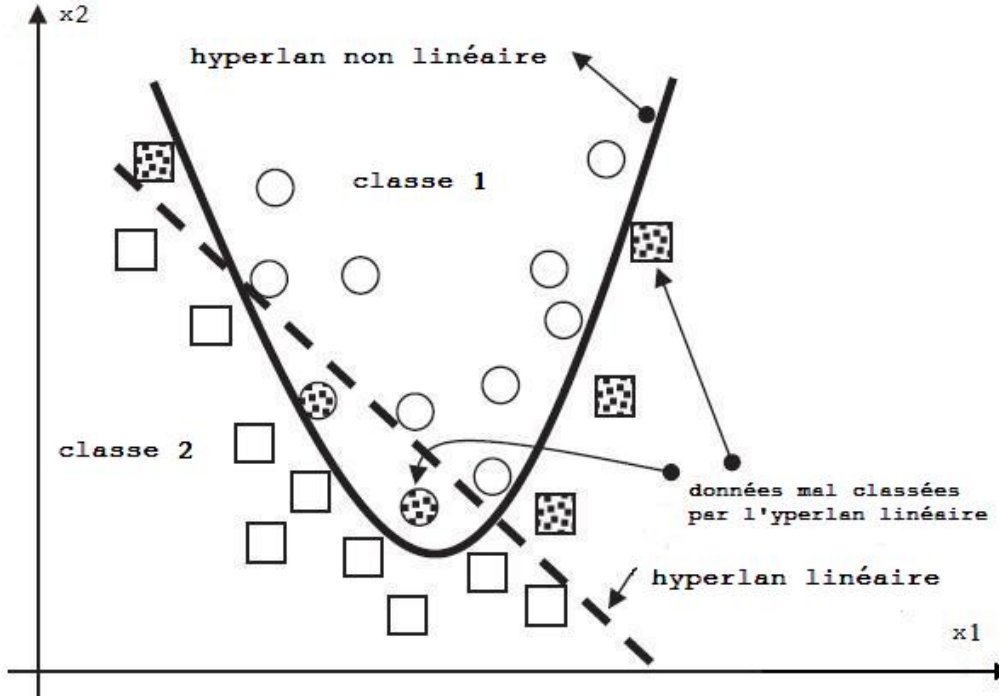


Figure II.3 – Support vector machines non lineaire

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

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

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

Avec :

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

La fonction de décision est définie par le "Sign" de

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

avec λ_0^* est donnée par (II.31).

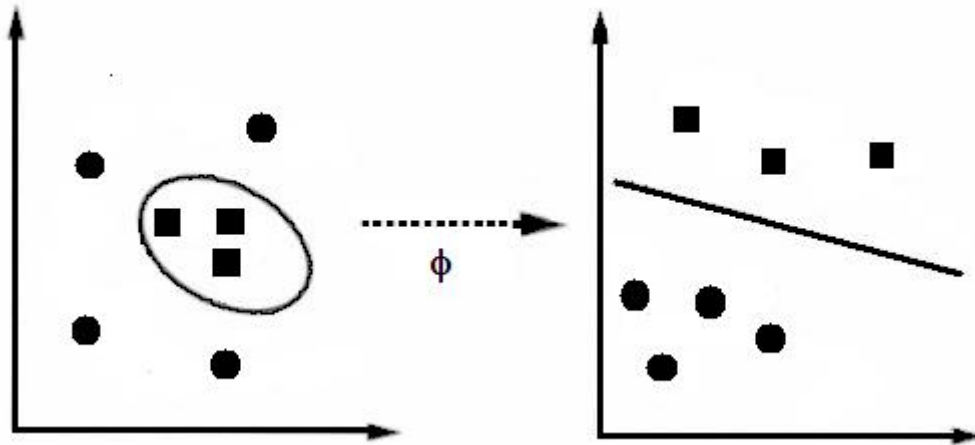


Figure II.4 – Transformation de l'observation par une fonction noyau reproduisante

5 Exemples de noyaux reproduisants

La fonction non-linéaire ϕ est implicitement définie dans l'expression (II.30) par le choix d'un noyau reproduisant κ , à partir duquel il est possible de couvrir une large classe de non-linéarités. Le choix du noyau et de ses paramètres se fait généralement d'une manière heuristique lors de tentatives du type essai-erreur.

La recherche de méthodes reposant sur des arguments théoriques solides n'en est pas moins un des défis à relever dans l'avenir des méthodes à noyaux.

5.1 Noyaux polynômiaux

Les noyaux polynômiaux sont de la forme [33] :

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

Dans ce cas, on peut montrer que la règle de décision repose sur une statistique polynômiale de degré q . Parce qu'ils sont fonction du produit scalaire des observations, de tels noyaux sont dits *projectifs*.

5.2 Noyaux gaussiens

Les noyaux gaussiens sont des noyaux de type *radial*, indiquant qu'ils dépendent de la distance $\|x_1 - x_2\|$ entre les observations. Ces noyaux sont définies par [33] :

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

Où σ^2 est appelé *largeur de bande*, il est appelé aussi le paramètre de la **fonction noyau**.

5.3 Noyaux exponentiels

Le noyau exponentiel est un autre exemple de noyau de type *radial*. Il est défini par [33] :

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

Ce noyau fournit une surface de décision linéaire par morceaux dans l'espace des observations.

5.4 Noyaux sigmoïdaux

Le noyau sigmoïdal est défini comme étant [33] :

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

L'utilisation d'un tel noyau est équivalente à celle d'un réseau de neurones à une couche cachée. 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 a vaste marge

Comme il est indiqué précédemment ,les support vector machines sont formulées pour les problèmes de deux classe. Mais parce que les support vector machines emploient des fonctions de décision directe , une extension du problème au multiclassification n'est pas simple. on discute dans la suite deux types de support vector machines qui traitent des problèmes multiclassé :

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

Dans la méthode "un contre tous", un problème de classification de n classes est convertis en n classifieurs binaire et pour le $i^{\text{ème}}$ classifieur binaire [35],[39],la classe i est séparé des autres classes.

L'approche "un contre un" transforme le problème de n classes en $n(n-1)/2$ classifieurs binaire qui couvrent tout les paire de classe [35],[39].

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 détermine n fonctions de décision discrètes qui séparent une classe des autres classes.

la i^{eme} fonction de décision discrète avec une marge maximale qui sépare la classe i des autres classes est [35] :

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

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

$$D_i(x) > 0 \quad (II.38)$$

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 [35].

6.2 Multi classification par l'approche un contre un

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

Dans l'approche "un contre un" , les fonctions de décision sont déterminées a partir des combinaisons des deux classes. Pour déterminer une fonction de décision pour une classe paire, nous utilisons les données d'apprentissage pour les deux classes correspondant. Ainsi, dans chaque session d'apprentissage, le nombre de données d'apprentissage est considérablement réduit par rapport de "un contre tous" supports vector machines , qui utilisent tous les données d'apprentissage. Mais le nombre de fonctions de décision est $n(n-1)/2$ comparativement avec n de "un contre tous" supports vector machines, où n est le nombre de classes.

Soit la fonction de décision pour la classe i contre la classe j est [35] :

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

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

Les régions :

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

ne se chevauchent pas, et si $x \in R_i$, nous classons x dans la classe i [35].

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 III

Réduction de la dimension

1 Introduction

Le besoin de réduire la dimension de l'espace, dans lequel une forme est décrite, est tout d'abord une nécessité physique. Des problèmes de stockage, de transmission, de représentation et même de perception se posent en dimension élevée. Malheureusement, l'augmentation du nombre de variables n'améliore pas systématiquement la qualité de l'apprentissage, au contraire elle fait croître le nombre de paramètres à estimer.

En effet, on est confronté au phénomène appelé "the curse of dimensionality" [18].

Ce problème provient du fait qu'un concept avec une quantité N de données en dimension p , nécessaire pour apprendre, croît exponentiellement en fonction de p [40]. La relation exacte entre la probabilité d'erreur de classification, le nombre de variables p , la taille n_i de l'ensemble d'apprentissage et la complexité de la règle de décision est très difficile à établir [41]. Dans le cas général, il est conseillé de disposer par classe d'un nombre n_i d'exemples d'apprentissage dix fois plus grand que le nombre de variables ($n_i/p > 10$) pour obtenir un discriminateur avec de bonnes performances. Plus le modèle est complexe, plus le rapport n_i/p doit être grand, pour éviter le problème de la malédiction de dimensionnalité [25].

Nous traitons, dans ce chapitre, le problème de réduction de dimension dans le cadre de la reconnaissance des formes statistiques et plus particulièrement dans le cadre de la classification supervisée. Nous présentons dans les sections suivantes quelques techniques d'extraction de variables pour la réduction de la dimension.

Dans un premier temps nous allons aborder le problème de la réduction de dimension à l'aide de l'algorithme de package d'ondelette. La modélisation autorégressive de données est discutée dans un deuxième temps.

2 Les ondelettes

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

Les ondelettes ont été introduites il y a presque 50 ans pour combler une lacune entre deux modes extrêmes de représentation du signal : la représentation d'une fonction par son graphe classique de \mathbb{R} dans \mathbb{R} , ce qui correspond à une décomposition sur la base continue des distributions de Dirac :

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

et la représentation dans la base de Fourier [42] :

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

avec $\hat{f}(\omega)$ est la transformé de Fourier de $f(t)$. La première représentation donne une information précise en temps : la valeur $f(t)$ indique l'intensité du signal au temps t . En revanche l'information fréquentielle est nulle. La valeur en un point de f ne donne aucune information sur le «contenu fréquentiel» du signal f . A l'inverse, la représentation de Fourier donne une information très précise en fréquence, mais ne donne aucune information temporelle. La question qui vient naturellement est la suivante : existe-t-il une représentation dans laquelle on puisse lire une information mixte ?, Morlet et Gabor ont cherché à concevoir des fonctions de base qui se situent à mi-chemin entre ces extrêmes, c'est-à-dire qui ont à la fois une bonne localisation fréquentielle et une bonne localisation spatiale.

Une limite théorique dans cette perspective est bien connue : c'est l'inégalité de Heisenberg [43]. Soit une fonction de base f , qu'on suppose de norme L_2 égale à 1 :

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

On définit le centre $c(f)$ et la largeur $\Delta(f)$ d'une telle fonction par :

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

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

L'inégalité de Heisenberg est une inégalité fondamentale qui s'écrit pour toute fonction f de norme 1 [43],[44] :

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

Si on appelle largeur fréquentielle d'une fonction f la largeur de sa transformée de Fourier \hat{f} , cette inégalité interdit donc d'avoir une fonction avec des largeurs temporelle et fréquentielle toutes deux aussi petites que l'on veut. On connaît également les fonctions qui réalisent le minimum de cette limite théorique : ce sont les fonctions gaussiennes translatées et modulées :

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

Où A est un coefficient de normalisation tel que la fonction ait une norme L_2 égale à 1, qui dépend de Δt . Pour de telles fonctions, et seulement pour ces fonctions-là, l'inégalité de Heisenberg devient une égalité. Ces fonctions ont été appelées ondelettes de Gabor [43].

2.2 Représentation temps-fréquence et temps-échelle.

A une telle fonction, on associe un pavé temps-fréquence, c'est-à-dire un rectangle dans le plan (t, ω) centre en $(c(f), c(\hat{f}))$ et de dimensions $(\Delta(f) \times \Delta(\hat{f}))$.

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 boîtes de 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 boîtes 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 rectangle infiniment fins et allongés qui sont représentées schématiquement dans la figure (III.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 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 :

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

Ce mode d'approximation est également appelé analyse de Fourier à fenêtre [44].

2. Une deuxième approche est l'approche temps-échelle, dans laquelle la largeur spatiale des

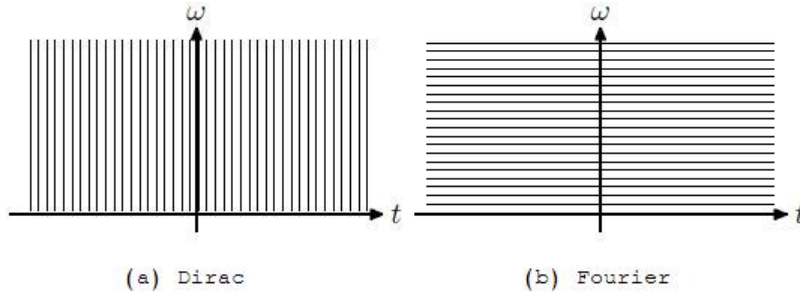


Figure III.1 – Pavages temps fréquences associés aux représentations de Dirac et de Fourier

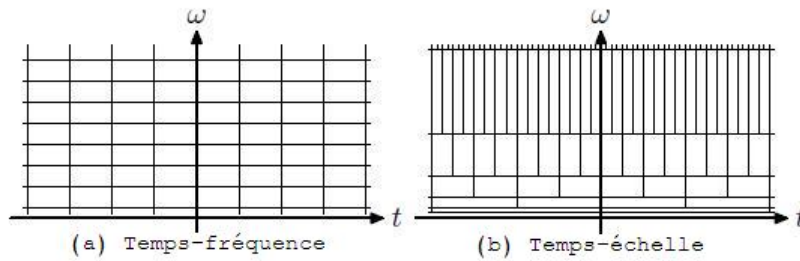


Figure III.2 – Pavages du plan temps fréquence pour la représentation temps-fréquence et la représentation temps-échelle

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 :

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

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.

2.3 Introduction aux ondelettes

La plupart des signaux du monde réel ne sont pas stationnaires, et c'est justement dans l'évolution de leurs caractéristiques (statistiques, fréquentielles, temporelles, spatiales) que ré-

side 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 ce chapitre, 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 finie (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 [45],[46], 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 multi résolution de signaux [47].

2.4 La transformée en ondelette

De manière analogue à la théorie des séries de Fourier, les ondelettes sont principalement utilisées pour la décomposition de fonctions. La décomposition d'une fonction en ondelettes consiste à l'écrire comme une somme pondérée de fonctions obtenues à partir d'opérations simples effectuées sur une fonction principale appelée ondelette mère. Ces opérations consistent en des translations et dilatations de la variable. Selon que ces translations et dilatations sont choisies de manière continue ou discrète, on parlera d'une transformée en ondelettes continue ou discrète.

2.4.1 La transformée en ondelettes continue.

Une transformée en ondelettes est dite continue lorsque les paramètres structurels des fonctions utilisées (c'est-à-dire les translations et les dilatations) peuvent prendre n'importe quelle valeur de l'ensemble des réels \mathbb{R} (les dilatations doivent néanmoins être positives).

Une ondelette mère ψ est une fonction de base que l'on va traduire et dilater pour recouvrir le plan temps-fréquences et analyser le signal.

L'ondelette doit être une fonction de moyenne nulle ; en d'autres termes, ψ doit être une onde

Chapitre III. Réduction de la dimension

, ce qui s'écrit mathématiquement [48],[44],[43] :

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

On introduit alors les facteurs de translation u Et d'échelle s [44],[43] :

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

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

On constitue ainsi une famille d'ondelettes engendrée à partir de l'ondelette mère. On la note Ω . On a alors la définition suivante :

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

Maintenant, nous allons présenter la transformée en ondelettes continue.

Soient f et g deux fonctions ; on définit leur produit scalaire par l'intégrale suivante :

$$\langle f, g \rangle = \int_R f(t)g(t)dt \quad (\text{III.13})$$

Pour que la transformée en ondelettes d'une fonction existe, il faut que cette fonction appartienne à l'ensemble des fonctions de carré sommable que l'on note par $L^2(R)$. Autrement dit, il faut que son carré soit fini. Cette condition se traduit par :

$$\int_R f^2(t)dt < \infty \quad (\text{III.14})$$

Dans ces conditions, la transformée en ondelette continue de la fonction f est définie comme le produit scalaire de f et de ψ [48],[44],[43] :

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

La reconstruction de la fonction f à partir de sa transformée est possible dans le cas où l'intégrale suivante est convergente [43] :

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

Où $\widehat{\psi}$ est la transformée de Fourier de ψ . Cette dernière condition est également appelée critère d'admissibilité pour une ondelette. Dans ce cas, f peut être reconstruit à partir de la relation

suivante [44] :

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

La condition donnée par l'équation(III.16) est très intéressante dans la mesure où elle donne des informations sur les propriétés que doit vérifier une ondelette mère (si l'on souhaite que la reconstruction de la fonction transformée soit possible). En particulier, on doit avoir $\psi(0) = 0$. En remplaçant ω par 0 dans la définition de la transformée de Fourier de ψ , on voit que cette condition est équivalente à :

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

Donc, une ondelette est une fonction à support de longueur finie et d'intégrale nulle. Ainsi, les gaussiennes radiales ne peuvent pas être considérées comme des ondelettes.

2.4.2 La transformée en ondelette discrète

La transformée discrète en ondelette (DWT) est issue de la version continue à la différence de cette dernière ; la DWT utilise un facteur d'échelle et translation discrétisés.

On appelle transformée en ondelettes discrète dyadique toute base ondelettes travaillant avec un facteur d'échelle $a = 2^i$. Il est clair que la DWT est en pratique implémentable sur tout système numérique (PC, DSP). Il est à noter que la transformée en ondelette continue est aussi implantable sur les systèmes digitaux avec un lourd calcul provenant de la nature continue de facteur d'échelle et de dilatation. L'intérêt de ce paragraphe est de présenter la transformée en ondelettes discrète pour montrer la relation avec la transformée en ondelettes continue. Une transformée en ondelettes est dite discrète lorsque les valeurs des (translations et des dilatations) sont à valeurs discrètes (pas nécessairement entières).

Soit ψ une ondelette mère (qui peut être la même que celle utilisée pour la transformée continue). Une famille Ω de fonctions obtenue à partir de ψ peut être exprimée de la manière suivante :

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

Où Z est l'ensemble des entiers relatifs. Notons que a et b sont des paramètres réels fixes qui définissent, avec ψ , la famille Ω ; a : détermine l'échelle des dilatations et b détermine le pas des translations.

Une famille d'ondelettes est donc entièrement connue par la donnée du triplet (ψ, a, b) . Un membre de cette famille (c'est-à-dire une fonction) est désignée par le Couple(m,n). Pour cette raison, on désignera dorénavant une ondelette de la famille de Ω ayant comme paramètres (m, n) par $:\psi_{m,n}(x)$.

Chapitre III. Réduction de la dimension

Suivant la relation (III.19), nous avons [43],[44] :

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

Qui peut être réécrite de la manière suivante :

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

Cette relation est la même que la relation (III.11), qui donne l'expression d'une ondelette dans un contexte de transformée continue, avec :

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

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

Ces relations montrent que la translation dépend de la dilatation, alors que ces quantités sont indépendantes dans le cas de la transformée continue. Les valeurs discrètes du facteur d'échelle s et du paramètre de translation u seront considérées sous la forme : $s = 2^{-j}$ et $u = k2^{-j}$. Avec ces valeurs de s et u , l'équation (III.15) devient :

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

Si la fonction $s(t)$ est discrétisée, en supposant une période d'échantillonnage égale à 1, pour des raisons de simplicité, l'équation (III.22) s'écrit alors :

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

2.5 Analyse multi-résolution

2.5.1 Présentation de l'analyse multirésolution

L'idée de l'analyse multirésolution a été développée par Meyer et Mallat . L'analyse multi-résolution est une théorie définissant des opérateurs linéaires permettant d'analyser un signal à différentes échelles. Nous pourrions dire que la construction d'une multirésolution permet de regarder un signal de "très près" ou de "très loin". Ce Zoom est effectué à l'aide d'une fonction d'échelle, qui se dilate à travers les échelles. Le signal projeté sur cette fonction donne une représentation de notre signal d'origine à l'échelle supérieure. Cette représentation (coefficients de projection) provoque un zoom arrière de notre signal d'origine, d'où le terme approximation. Afin de reconstruire notre signal, à partir des coefficients d'approximation, nous devons

également projeter notre signal originel sur un espace perpendiculaire (conservation de toute l'information). La fonction générant ce deuxième espace vectoriel sera une ondelette. Finalement, le signal est projeté sur une fonction d'échelle créant une approximation du signal et sur une ondelette pour récupérer l'information perdue lors de la première projection. Cette deuxième projection contient les détails du signal d'origine. Les bases $\phi_{j,k}(t)$ et $\psi_{j,k}$ sont respectivement nommées fonctions d'échelles et ondelettes.

2.5.2 Définition d'une analyse multirésolution

Une analyse multirésolution de $L^2(\mathbb{R})$ est, par définition, une suite croissante V_j avec $j \in \mathbb{Z}$ de sous espaces vectoriels fermés de $L^2(\mathbb{R})$ ayant les propriétés suivantes [43] :

- $V_j \subset V_{j+1}$; pour tout j, V_j est un sous espace de 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 une dilatation d'un facteur 2.
- $s(t) \in V_j \Leftrightarrow s(t-2^{-j}k) \in V_j ; \forall k \in \mathbb{Z}$; V_j est invariant par translation de 2^{-j}
- $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$; la réunion des V_j est dense dans L^2 .
- $\bigcap_{j \in \mathbb{Z}} V_j = 0 = \lim_{j \rightarrow \infty} V_j$; l'intersection de V_j est réduite à 0 dans L^2 .
- Il existe une fonction d'échelle $\phi \in V_0$ telle que la suite $\phi(t-k), k \in \mathbb{Z}$ soit une base de Riesz pour V_0 .

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

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

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

2.5.3 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 séquences $\{h_1[k]\} \in L^2$ et $\{g_1[k]\} \in L^2$ telles que la fonction d'échelle ϕ et l'ondelette associée ψ , à un niveau donné, peuvent être construites à partir de la fonction d'échelle au niveau juste au dessus. Ces équations seront nommées : relations de reconstruction.

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

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

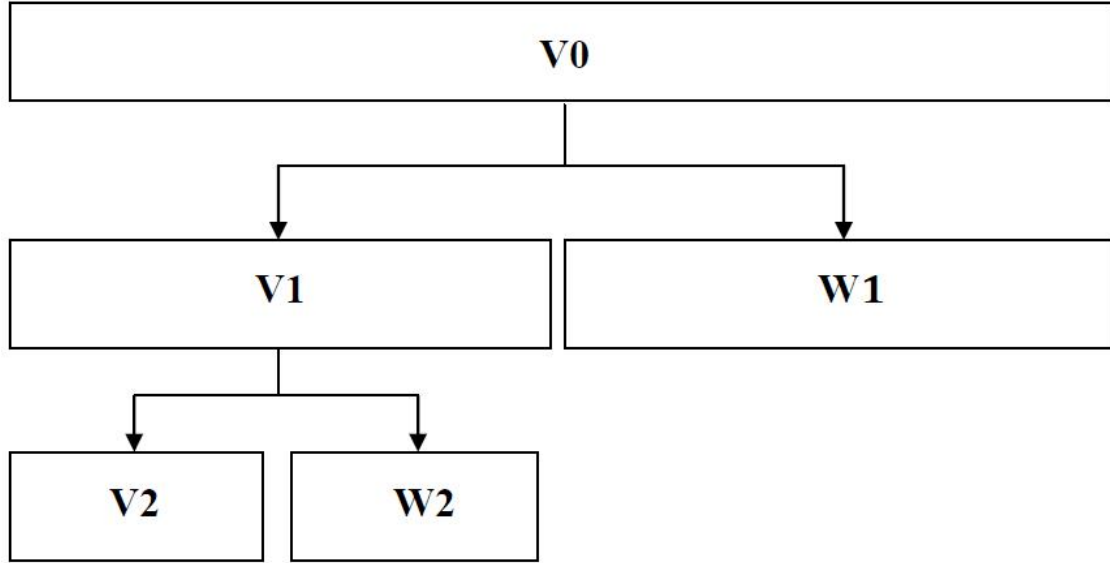


Figure III.3 – La nature hiérarchique de l’analyse multirésolution

En général, $\forall j \in \mathbb{Z}$, la relation entre les sous espaces V_j, W_j et V_{j+1} est engendrée par les deux équations suivantes [48] :

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

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

En prenant la transformée de Fourier des équations (III.26) et (III.27), on obtient :

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

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

Où :

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

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

On admettra que $\widehat{\phi}(0) = 1$. En appliquant ce résultat à l'équation (III.30), on déduit :

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

L'ondelette se comportant comme un filtre passe-bande, $\widehat{\Psi}(0) = 0$ à l'aide de l'équation (III.31), on obtient :

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

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 séquences $\{h_2[k]\} \in L^2$ et $\{g_2[k]\} \in L^2$ telles que :

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

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

En combinant ces deux relations, on obtient :

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

Dans le cas général :

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

Cette équation est nommée : relation de décomposition.

2.5.4 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 très employé en traitement du signal. Il est aussi bien utilisé pour la compression de données .L'algorithme est basé sur la relation de décomposition (III.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

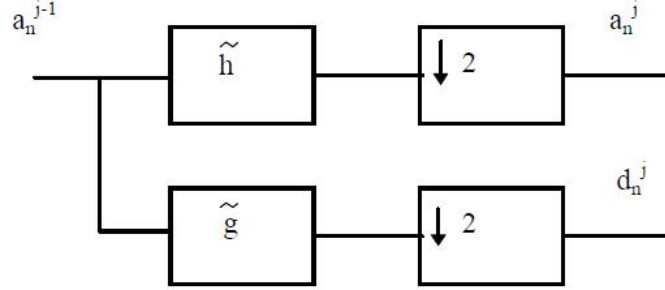


Figure III.4 – Décomposition en cascade.

signal $s(t) \in L^2(R)$ se décompose suivant les différentes échelles j en une fonction d'approximation $x_j(t)$ et une fonction de détail $y_j(t)$ de la façon suivante [48] :

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

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

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

Avec

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

L'analyse multirésolution exige que :

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

Donc :

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

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

On substitue la relation de décomposition (III.39) Dans l'équation (III.46) pour obtenir une

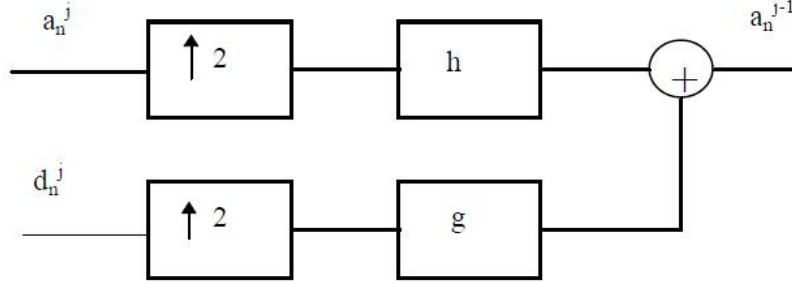


Figure III.5 – Algorithme de synthèse de Mallat.

équation où 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, la décomposition suivant l'algorithme de Mallat est obtenue [48] :

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

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

Les équations (III.47) et (III.48) montrent que les coefficients de la fonction d'échelle et de l'ondelette à l'échelle j , se calculent à partir de ceux de l'échelle immédiatement supérieure. En répétant ce processus, on obtient la décomposition en cascade présentée dans la figure III.4.

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

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

En comparant les coefficients de $\phi(2^{j+1}t - n)$ dans les deux membres de l'équation (III.49), la reconstruction suivant l'algorithme de Mallat est obtenue :

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

L'algorithme de reconstruction peut être schématisé dans la figure(III.5).

2.5.5 Bilan

L'analyse multirésolution a permis d'introduire et de mieux comprendre les algorithmes rapides de décomposition et de reconstruction. Les relations de décomposition et de reconstruction sont essentielles pour le développement de ces algorithmes. Pour connaître les filtres $h_1[k], g_1[k], h_2[k], g_2[k]$ dans le cas de bases orthogonales, il suffit de connaître seulement le filtre $h_1[k]$ puisque les trois autres filtres se calculent à partir de ce dernier :

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

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

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

Les algorithmes de décomposition et de reconstruction s'appliquent à 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(III.3)), s'écrira à l'aide de l'analyse multirésolution ($J \in \mathbb{R}^2$ étant l'échelle la plus basse) :

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

2.6 Paquets d'ondelettes

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

Cette analyse peut être menée avec les mêmes bases de fonctions d'échelle et d'ondelettes ou avec des bases différentes. On peut, de même, changer de fonctions de base à chaque échelle. Dans ces algorithmes, la reconstruction parfaite est assurée par la réutilisation lors de la synthèse, pour une résolution donnée, de la même base que lors de l'analyse pour cette résolution.

La méthode présentée qui consiste à analyser les sous-espaces de détail du signal est baptisée analyse en paquets d'ondelettes. L'algorithme correspondant peut être schématisé par la figure III.6).

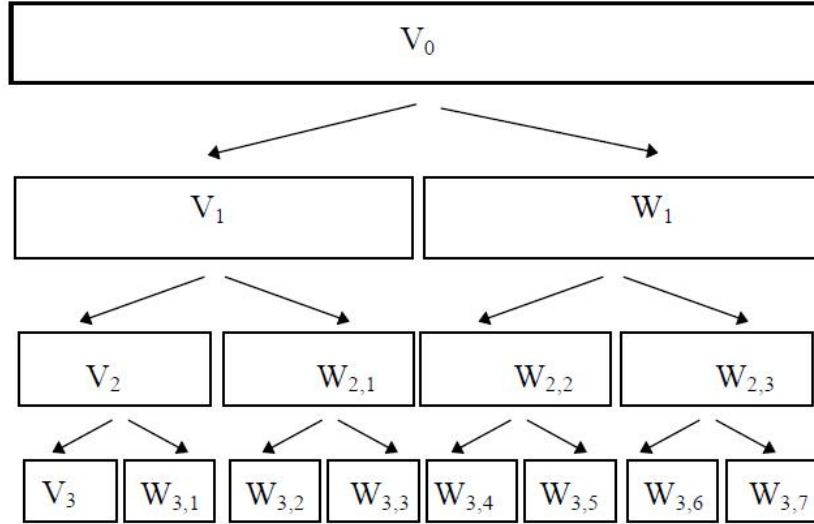


Figure III.6 – Analyse en paquets d'ondelettes.

2.7 Orthogonalité et bi-orthogonalité

Les ondelettes orthogonales sont couramment employées pour supprimer la redondance ainsi que pour assurer la reconstruction parfaite du signal.

2.7.1 Les ondelettes orthogonales

L'analyse multi-résolution orthogonale est une analyse multi résolution dans laquelle les espaces d'ondelette W_j sont définis comme le complément orthogonal de V_j dans V_{j+1} .

En conséquence, les espaces W_j avec $j \in \mathbb{Z}$ sont tous mutuellement orthogonaux.

Une condition suffisante pour qu'une analyse multi-résolution soit orthogonale ($W_0 \perp V_0$) est : $\{\phi(x-n)\}, n \in \mathbb{Z}$ et $\{\psi(x-n)\}, n \in \mathbb{Z}$ sont des familles des vecteurs orthogonaux. Cela signifie que pour tout $n \in \mathbb{Z}$:

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

Reprenons l'expression de la famille d'ondelettes discrètes :

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

Y.Meyer a montré que pour $a_0 = 2$ et $b_0 = 1$, Les ondelettes sont dites orthogonales et s'écrivent alors :

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

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

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

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 en la projection du signal à analyser sur deux espaces distincts : un signal d'approximation, et un espace détail. Si les fonctions utilisées (fonction d'échelle ou ondelette) ne sont pas orthogonales ; L'information ne sera pas décorrélée entre les vecteurs, d'où la redondance. On ne pourra plus retrouver facilement le signal d'origine à partir des transformées.

2.7.2 Les ondelettes bi-orthogonales

Nous avons vu 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 exacte à l'exception de ceux correspondant à la base de haar. Cependant, 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 utilisera 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 synthèse 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 synthèse, d'où le nom de bi orthogonalité. Par contre, et contrairement au cas orthogonale les ondelettes et les fonctions d'échelle associée 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 $\tilde{W}_m \perp V_m$.

2.8 Exemples d'ondelettes

Les premières ondelettes qui sont nées des travaux de Meyer et Mallat, sont les ondelettes orthogonales. Ils existent un certain nombre de familles d'ondelettes orthogonales couramment utilisées. Les plus connues sont sans doute les ondelettes de Daubechies. Les familles d'ondelettes non orthogonales ont reçu l'appellation d'ondelettes biorthogonales. En effet, le préfixe bi

rappelle que deux bases d'ondelettes sont utilisées, une pour la décomposition (la base duale) et une pour la reconstruction. 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 engendrent une analyse multirésolution orthogonale, on les appelle semiorthogonales.

2.8.1 Ondelettes orthogonales

Ondelette de Haar

Un premier exemple d'ondelette orthogonale est l'ondelette de Haar. La fonction d'échelle dans ce cas est $\phi_{Haar}(t) = X_{[0,1]}(t)$ et l'ondelette correspondante $\psi_{Haar}(t) = X_{[0,1/2]}(t) - X_{[1/2,1]}(t)$.

Dans la pratique, l'ondelette de Haar n'est pas beaucoup utilisée en raison de ses discontinuités.

Ondelette de Shannon

L'ondelette de Shannon :

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

dont la fonction d'échelle est $\phi_{SH} = \text{sinc}(\pi t)$ n'est pas beaucoup utilisée en raison de sa très faible décroissance à l'infini. En effet, 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, $\hat{\psi}_{SH}(w)$ comporte deux discontinuités à $-\pi$ et π . En conséquence, dans le domaine temporel, la fonction décroît en $1/t$ et donc $\Delta\psi_{SH} = \infty$.

Ondelette de Meyer

Meyer a construit une fonction d'échelle (III.7) de telle sorte que sa transformée de Fourier soit lisse aux endroits de discontinuités de $\hat{\psi}_{SH}(w)$ (voir ondelette de Shannon). En temps, cela se traduit par une décroissance plus rapide à l'infini par rapport à 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

Les ondelettes de Daubechies sont probablement les plus utilisées en ce qui concerne les ondelettes orthogonales. Elles sont à support compact (les filtres H et G ont une réponse impulsionnelle finie donc les filtres h_1 et g_1 sont finis). Ces ondelettes seront notées dbN, où db

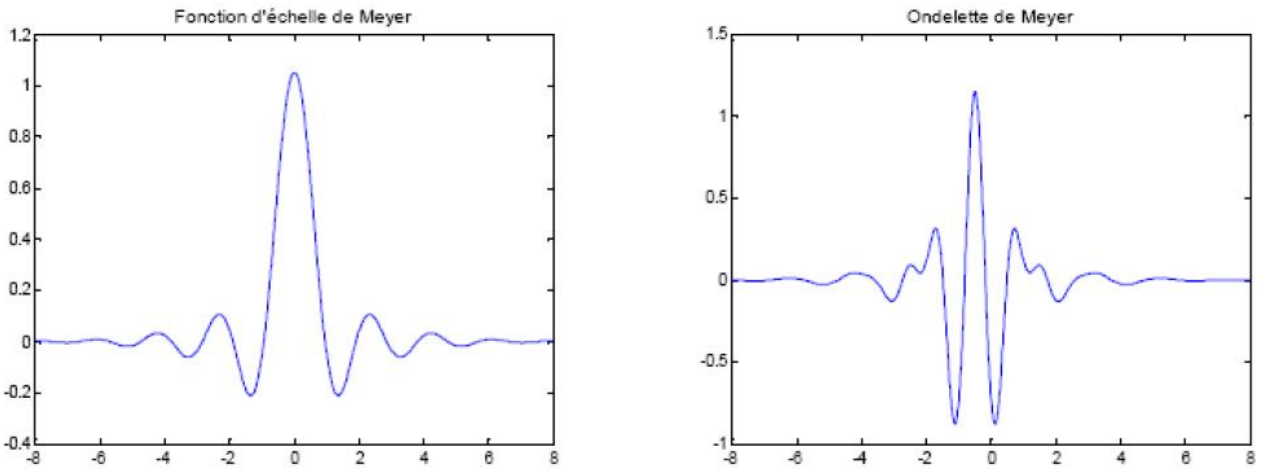


Figure III.7 – Fonction d'échelle et ondelette de Meyer.

est le symbole donné pour Daubechies, et N est le nombre de moments nuls de l'ondelette. Les ondelettes de Daubechies sont supportées sur un intervalle de longueur $2N - 1$. Ces ondelettes présentent l'inconvénient de ne pas être symétriques ou antisymétriques, excepté quand $N = 1$ ce qui correspond à l'ondelette de Haar. La Figure (III.8) représente les fonctions d'échelles et ondelettes pour $N = 2$, $N = 4$, et $N = 8$.

Symlets Pour obtenir une ondelette symétrique ou antisymétrique, le filtre h_1 doit être symétrique ou antisymétrique par rapport au centre de son support ($\hat{h}_1(w)$ a une phase linéaire). Les symlets sont des ondelettes de Daubechies construites de telle sorte que la phase de ($\hat{h}_1(w)$) soit la plus linéaire possible. Le support des symlets est $2N + 1$. La Figure (III.9), représente la fonction d'échelle et l'ondelette pour $N = 8$. Une meilleure symétrie par rapport à l'ondelette de Daubechies ($N = 8$) peut être remarquée.

Coiflets Pour une application en analyse numérique Coifman a demandé à Daubechies de construire une famille d'ondelettes avec N moments nuls et un support de taille minimum, et dont la fonction d'échelle vérifie :

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

Le résultat est l'ondelette coiflets dont la taille du support est $3N - 1$ au lieu de $2N - 1$ pour une ondelette de Daubechies. La Figure (III.10), représente la fonction d'échelle et l'ondelette pour $N = 5$.

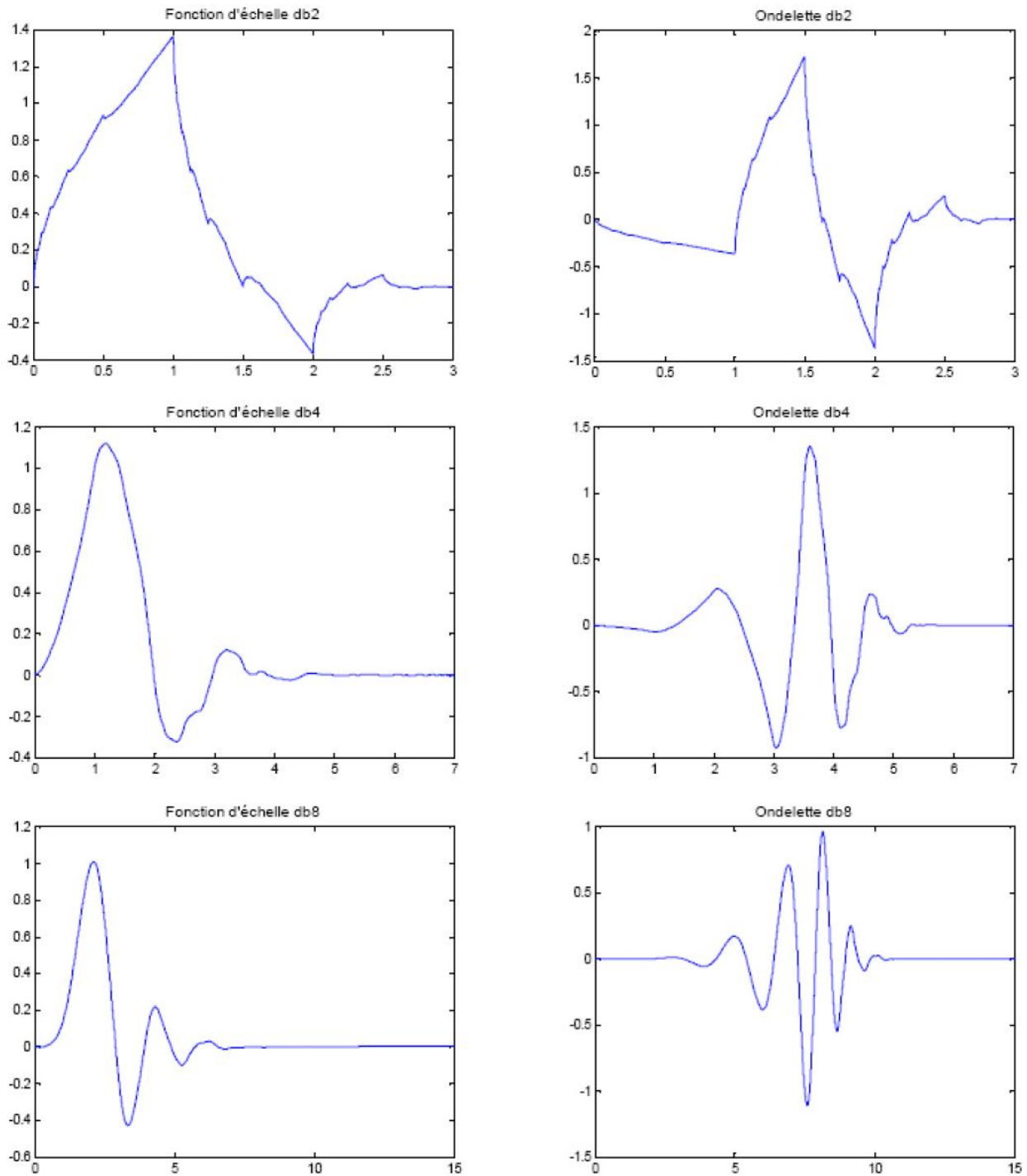


Figure III.8 – Fonction d'échelle et ondelette de Daubechies pour $N = 2, 4, 8$.

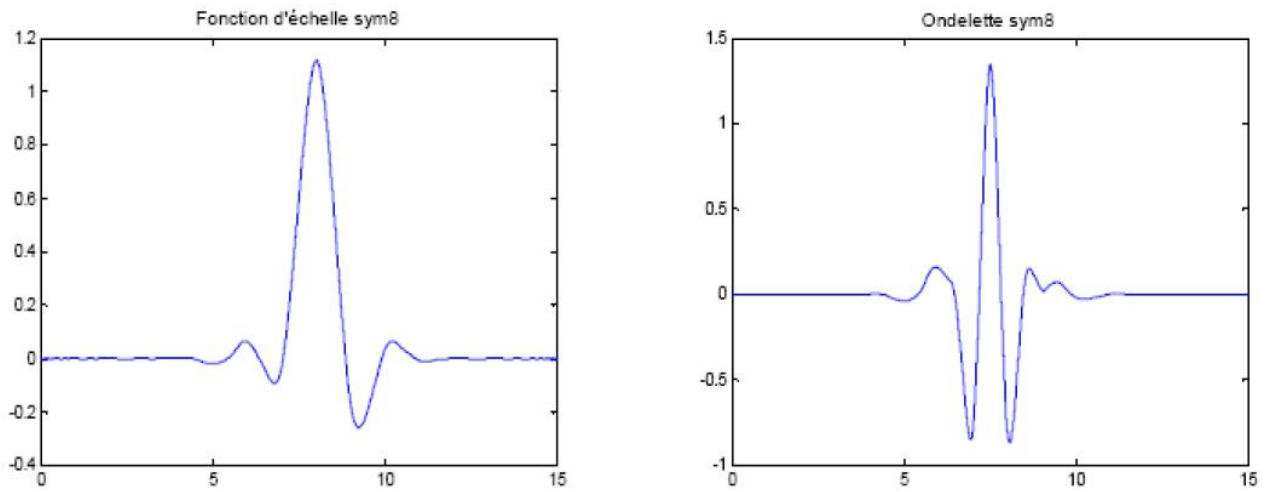


Figure III.9 – Fonction d'échelle et ondelette Symlets pour $N = 8$.

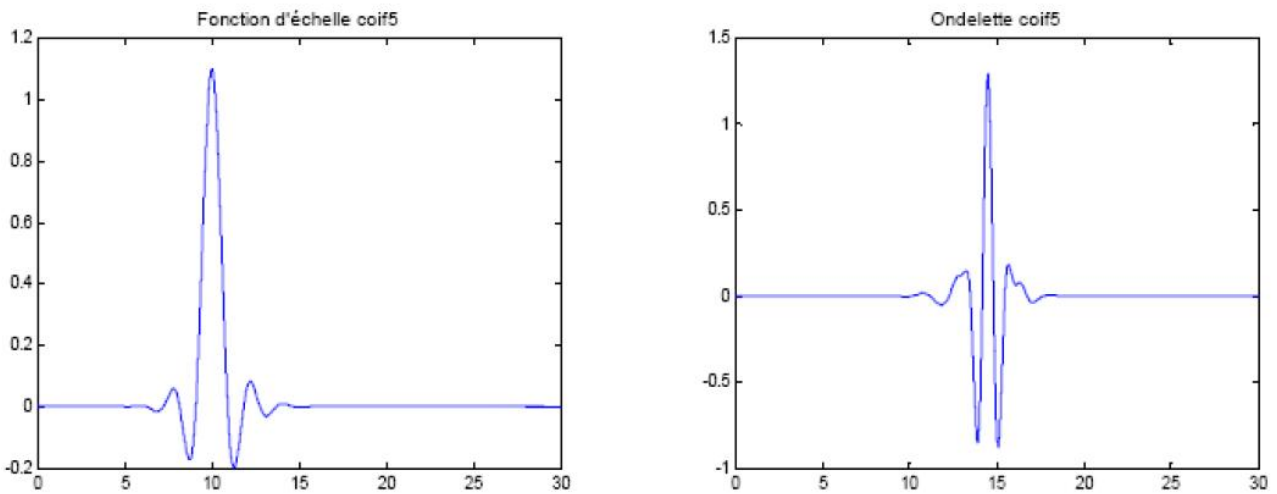


Figure III.10 – Fonction d'échelle et ondelette Coiflets pour $N = 5$.

3 La modélisation autorégressive

3.1 Modèles rationnels linéaires

La plupart des processus sont bien approximés par un modèle rationnel linéaire. Un modèle général est le modèle ARMA(p,q), (Auto Regressive Moving Average) défini par l'équation de récurrence [49] :

$$x(n) = - \sum_{k=1}^p a_k x(n-k) + \sum_{k=0}^q b_k u(n-k) \quad (\text{III.61})$$

Il peut être vu comme la sortie d'un filtre linéaire, dont l'entrée est $e(n)$ et la fonction de transfert $H(z)$ telle que :

$$H(z) = \frac{B(z)}{A(z)} = \sum_{k=0}^{\infty} h_k z^{-k} \quad (\text{III.62})$$

avec :

$$B(z) = 1 + \sum_{k=1}^q b_k z^{-k} \quad (\text{III.63})$$

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (\text{III.64})$$

On considère que l'entrée $e(n)$ du filtre est un bruit blanc de valeur moyenne nulle et de variance " σ^2 " et on suppose également en général que $e(n)$ est un processus gaussien.

3.1.1 Modèle autorégressif

La modélisation autorégressive (AR) est une méthode très populaire en analyse spectrale paramétrique [49],[50]. Le modèle AR exprime le signal à l'instant n comme étant une combinaison linéaire des p échantillons précédant l'instant n . Le modèle AR d'ordre p s'écrit :

$$x(n) = - \sum_{k=1}^p \alpha_k x(n-k) + e(n) \quad (\text{III.65})$$

Où les α_k sont les paramètres AR et $e(n)$ un bruit blanc centré . Lorsqu'un signal $x(n)$ est approché par ce modèle , le bruit $e(n)$ correspond à l'erreur de modélisation. Le signal $x(n)$ correspond au filtrage linéaire du bruit blanc $e(n)$ par un filtre stable de fonction de transfert en z :

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^p \alpha_k z^{-k}} \quad (\text{III.66})$$

Le choix d'un filtre stable contraint les racines du polynôme $A(z)$ à être de module strictement inférieur à 1, c'est-à-dire à l'intérieur du cercle unité .

Chapitre III. Réduction de la dimension

Les paramètres d'un modèle AR(p) vérifient les équations de Yule-Walker (YW) définies par :

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdot & \cdot & r_{xx}(-p) \\ r_{xx}(1) & r_{xx}(0) & \cdot & \cdot & r_{xx}(-p+1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{xx}(p) & r_{xx}(p-1) & \cdot & \cdot & r_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (\text{III.67})$$

Ce résultat montre que la fonction d'autocorrélation de $x(n)$ satisfait la même récursion que le signal.

3.1.2 Relation AR-Prédiction linéaire

Supposons que $x(n)$ soit un processus AR(p). On cherche à trouver le meilleur prédicteur d'ordre p de $x(n)$:

$$\hat{x}(n) = - \sum_{k=1}^p \alpha_k x(n-k) \quad (\text{III.68})$$

Pour cela, on cherche les coefficients α_k qui minimisent la puissance de l'erreur de prédiction :

$$\sigma_p^2 = E \{ |e(n)|^2 \} = E \{ |\hat{x}(n) - x(n)|^2 \} \quad (\text{III.69})$$

Or :

$$\begin{aligned} \sigma_p^2 &= \rho_p = E \left\{ \left[x(n) + \sum_{k=1}^p \alpha_k x(n-k) \right] \left[x^*(n) + \sum_{k=1}^p \alpha_k^* x^*(n-k) \right] \right\} \\ &= r_{xx}(0) + \sum_{k=1}^p \alpha_k r_{xx}(-k) + \sum_{k=1}^p \alpha_k^* r_{xx}(k) + \sum_{k=1}^p \sum_{m=1}^p \alpha_k \alpha_k^* r_{xx}(m-k) \\ &= r_{xx}(0) + \alpha^* r + r^* \alpha + \alpha^* R \alpha \quad (\text{III.70}) \end{aligned}$$

avec :

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)^T, r = (r_{xx}(1), r_{xx}(2), \dots, r_{xx}(p))^T \quad (\text{III.71})$$

$$R = \begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdot & \cdot & r_{xx}(-p+1) \\ r_{xx}(1) & r_{xx}(0) & \cdot & \cdot & r_{xx}(-p+2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & \cdot & \cdot & r_{xx}(0) \end{pmatrix} \quad (\text{III.72})$$

D'où, la minimisation revient à :

$$\frac{\sigma \rho_p}{\sigma \alpha} = 2r + 2R\alpha = 0 \iff R\alpha = -r \quad (\text{III.73})$$

On retrouve les équations de Yule-Walker .

Le meilleur prédicteur d'ordre p d'un processus AR(p) est donné par les paramètres AR.

On a :

$$\rho_{min} = \sum_{k=0}^p \alpha_k r_{xx}(-k) \quad (\text{III.74})$$

Le meilleur prédicteur est celui qui rend l'erreur orthogonale aux données. En effet, pour minimiser l'erreur entre $x(n)$ et une combinaison linéaire des $\{x(n-k)\}_{k=1}^p$, il faut projeter orthogonalement $x(n)$ sur l'espace engendré par $\{x(n-k)\}_{k=1}^p$.

De fait, l'erreur de prédiction devient orthogonale aux $\{x(n-k)\}_{k=1}^p$. Ceci peut se retrouver en écrivant que, $\forall k \in [1, p]$

$$E \{e(n)x^*(n-k)\} = E \left\{ \sum_{l=0}^p \alpha_l x(n-l)x^*(n-k) \right\} = \sum_{l=0}^p \alpha_l r_{xx}(k-l) = 0 \quad (\text{III.75})$$

au vu de (III.73). Les coefficients optimaux α_k sont donc ceux qui rendent l'erreur de prédiction $e(n)$ orthogonale (i.e. non corrélée) aux données $\{x(n-1), \dots, x(n-p)\}$. L'innovation $e(n)$ peut être vue comme l'information qui est contenue dans $x(n)$ et qui n'était pas dans $\{x(n-1), \dots, x(n-p)\}$.

A cause de la propriété précédente, on estime généralement les paramètres AR par minimisation de l'erreur de prédiction linéaire. Il existe une autre approche que celle de la prédiction linéaire qui a été proposée par Burg et appelée méthode du maximum d'entropie. Comme on l'a vu, si l'on suppose connus $\{r_{xx}(k)\}_{k=0,2,\dots,p}$, alors la fonction d'autocorrélation est implicitement extrapolée (via les équations de Yule-Walker) pour un processus AR. Burg s'est alors posé la question suivante pour un processus quelconque : si l'on connaît $\{r_{xx}(k)\}_{k=1,2,\dots,p}$, comment est-il possible d'extrapoler $r_{xx}(p+1), r_{xx}(p+1), \dots$ de manière à garantir le caractère semi-défini positif de la fonction de corrélation ? . Il a alors montré que parmi toutes les extrapolations possibles, celle qui donne l'entropie maximale correspond à un processus AR. En d'autres termes, le processus AR est "le processus le plus aléatoire" qui permette d'extrapoler la fonction d'autocorrélation à partir de la connaissance de $\{r_{xx}(k)\}_{k=0,2,\dots,p}$. Burg considère cette propriété importante dans la mesure où elle permet de faire le moins d'hypothèses possible (telle que corrélation nulle) sur la fonction d'autocorrélation en dehors de $\{r_{xx}(k)\}_{k=0,2,\dots,p}$. Du fait de l'analogie que nous venons de présenter, la méthode du maximum d'entropie est souvent assi-

milée à la modélisation AR. Dans la suite, elle sera plutôt associée à la méthode proposée par Burg pour estimer les paramètres AR.

Dans la suite, nous présentons différentes méthodes d'estimation de ces paramètres AR, fondées soit sur une minimisation de l'erreur de prédiction, soit sur l'algorithme de Burg.

3.1.3 Méthodes d'estimation

Algorithme de Levinson C'est un algorithme qui permet de résoudre tout système du type $Ax = b$ avec A Toeplitz, donc en particulier les équations normales de Yule-Walker :

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdot & \cdot & r_{xx}(-p) \\ r_{xx}(1) & r_{xx}(0) & \cdot & \cdot & r_{xx}(-p+1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{xx}(p) & r_{xx}(p-1) & \cdot & \cdot & r_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \rho_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (\text{III.76})$$

Principe : Trouver de façon récursive les paramètres d'ordre k en fonction des paramètres d'ordre $k - 1$.

– Initialisation :

$$a_1 [1] = -\frac{r_{xx}(1)}{r_{xx}(0)}$$

$$\rho_1 = \left(1 - |a_1 [1]|^2\right) r_{xx}(0) \quad (\text{III.77})$$

– Récursion :

pour $k=2, \dots, p$; faire :

$$a_k [k] = -\frac{r_{xx}(k) + \sum_{i=1}^{k-1} a_{k-1} [i] r_{xx}(k-i)}{\rho_{k-1}} = k_k$$

$$a_k [i] = a_{k-1} [i] + a_k [k] a_{k-1}^* [k-i] \quad i = 1, \dots, k-1$$

$$\rho_k = \left[1 - |a_k [k]|^2\right] \rho_{k-1} \quad (\text{III.78})$$

En pratique, on utilise un estimateur de la fonction de corrélation. Pour assurer un système à minimum de phase (pôles à l'intérieur du cercle unité), on prend l'estimateur biaisé. Les coefficients $k_{k=1, \dots, p}$ sont dits coefficients de réflexion. Trois représentations équivalentes sont possibles pour un processus AR :

$$\{\rho_p, a_p [1], \dots, a_p [p]\} \Leftrightarrow \{r_{xx}(0), k_1, \dots, k_p\} \Leftrightarrow \{r_{xx}(0), \dots, r_{xx}(p)\}.$$

Algorithme de Burg On a vu que Burg avait montré que dans le cas où $x(n)$ est Gaussien, on a :

$$\left\{ \begin{array}{l} \{r_{xx}(k)\}_{k=0\dots p} \text{ connu} \\ \text{entropie maximale} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \{r_{xx}(k)\}_{k=0\dots p} \text{ connu} \\ x(n) \text{ spectre AR}(p) \end{array} \right. \quad (\text{III.79})$$

Principe : c'est un algorithme récursif en ordre qui minimise par rapport au coefficient de réflexion la puissance des erreurs de prédiction directe et rétrograde. On a :

$$e_k^f(n) = x(n) + \sum_{m=1}^k a_k[m] x(n-m) = e_{k-1}^f(n) + k_k e_{k-1}^b(n-1) \quad (\text{III.80})$$

$$e_k^b(n) = x(n-k) + \sum_{m=1}^k a_k^*[m] x(n-k+m) = e_{k-1}^b(n-1) + k_k^* e_{k-1}^f(n) \quad (\text{III.81})$$

On cherche à minimiser :

$$\rho_k = \frac{1}{2}(\rho_k^f + \rho_k^b) = \frac{1}{2} \sum_{n=k}^{N-1} |e_k^f(n)|^2 + |e_k^b(n)|^2 \quad (\text{III.82})$$

Si l'on suppose connus k_1, k_2, \dots, k_{k-1} , alors ρ_k ne dépend que de k_k . On a donc :

$$k_k = \frac{-2 \sum_{n=k}^{N-1} e_{k-1}^f(n) e_{k-1}^{b*}(n-1)}{\sum_{n=k}^{N-1} |e_{k-1}^f(n)|^2 + |e_{k-1}^b(n-1)|^2} \quad (\text{III.83})$$

On en déduit :

$$a_k[k] = k_k \quad (\text{III.84})$$

$$a_k[i] = a_{k-1}[i] + a_k[k] a_{k-1}^*[k-i] \quad i = 1 \dots k-1 \quad (\text{III.85})$$

$$\rho_k = (1 - |a_k[k]|^2) \rho_{k-1} \quad (\text{III.86})$$

L'algorithme de Burg garantit un système à minimum de phase.

Algorithmes des moindres carrés En général, on utilise des méthodes qui consistent à minimiser, au sens des moindres carrés, la puissance de l'erreur de prédiction linéaire.

Soit $\hat{x}(n) = -\sum_{k=1}^p a_k x(n-k)$ le prédicteur d'ordre p de $\{x(n)\}_{n=1}^N$ et $e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^p a_k x(n-k)$ l'erreur de prédiction.

Chapitre III. Réduction de la dimension

On peut écrire sous forme matricielle :

$$\begin{pmatrix} e_1 \\ \cdot \\ e_p \\ e_{p+1} \\ \cdot \\ e_N \\ e_{N+1} \\ \cdot \\ e_{N+p} \end{pmatrix} = \begin{pmatrix} x_1 & 0 & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_p & & & x_1 & 0 \\ x_{p+1} & & & & x_1 \\ \vdots & \ddots & & & \\ x_N & & \ddots & & x_{N-p} \\ 0 & x_N & & & \\ \vdots & \ddots & & & \\ \vdots & & \ddots & & x_N \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{pmatrix} \quad (\text{III.87})$$

On cherche alors à minimiser :

$$\rho = \sum_{n=n_0}^{n_1} |e(n)|^2 = \sum_{n=n_0}^{n_1} \left| x(n) + \sum_{k=1}^p a_k x(n-k) \right|^2 = \|Xa + h\|^2 \quad (\text{III.88})$$

Selon l'horizon $[n_0, n_1]$ sur lequel on effectue la minimisation, on a des algorithmes différents :

- $[1, N + p]$ avec fenêtrage (on suppose le signal nul en dehors de $[1, N]$) → méthode des corrélations.
- $[1, N]$: préfenêtrage.
- $[p + 1, N + p]$: postfenêtrage.
- $[p + 1, N]$: sans fenêtrage (on n'utilise que la portion disponible du signal) → méthode des covariances.

Quelle que soit la méthode, le vecteur qui minimise (III.88) est donné par :

$$a = -(X^*X)^{-1} X^*h = -R^{-1}r \quad (\text{III.89})$$

avec :

$$R(i, j) = \sum_{n=n_0}^{n_1} x^*(n-i)x(n-j)$$

$$r(j) = \sum_{n=n_0}^{n_1} x(n)x^*(n-j) \quad (\text{III.90})$$

1^{er} cas : méthode des corrélations :

On a :

$$R(i, j) = [X^* X](i, j) = \sum_{n=1}^{N-i+j} x^*(n)x(n+i-j) \quad (\text{III.91})$$

La matrice R, qui est de Toeplitz, n'est autre que la matrice des corrélations des équations de Yule-Walker dans laquelle l'estimateur biaisé de la corrélation est utilisé : on peut donc utiliser l'algorithme de Levinson-Durbin pour résoudre cette équation.

2^{ème} cas : méthode des covariances :

Ici :

$$R(i, j) = [X^* X](i, j) = \sum_{n=p+1}^N x^*(n-i)x(n-j) \quad (\text{III.92})$$

Dans ce cas, la matrice R, bien qu'hermitienne n'est plus de Toeplitz : on ne peut plus appliquer l'algorithme de Levinson-Durbin. Néanmoins, Morf et al ont développé un algorithme nécessitant $O(p^2)$ opérations pour résoudre de tels systèmes.

3^{ème} cas : méthode des covariances modifiée :

Marple a proposé de minimiser, non plus la puissance de l'erreur de prédiction avant mais la demi somme des erreurs de prédiction arrière et avant :

$$\rho = \frac{1}{2}(\rho^f + \rho^b) = \frac{1}{2} \sum_{n=p+1}^N |e^f(n)|^2 + |e^b(n)|^2 \quad (\text{III.93})$$

où :

$$e^f(n) = x(n) + \sum_{k=1}^p a_k x(n-k) \quad (\text{III.94})$$

$$e^b(n) = x(n-p) + \sum_{k=1}^p a_k^* x(n-p+k) \quad (\text{III.95})$$

Dans ce cas, la matrice R est donnée par :

$$R(i, j) = \sum_{n=p+1}^N x^*(n-i)x(n-j) + x(n+i-p)x^*(n+j-p) \quad (\text{III.96})$$

4 Conclusion

L'objectif de ce chapitre était de trouver la meilleure façon pour réduire la dimension des données associées à chaque défaut de roulement. Ainsi trouver une signature de taille réduite qui le caractérise, par conséquent le différencie des autres défaut de roulement appartenant aux

Chapitre III. Réduction de la dimension

autres classes.

Pour cela nous avons présenté une technique de réduction de dimension : l'approche d'extraction de caractéristiques. Nous nous sommes concentrés, dans un premier temps, sur les paquets d'ondelettes. Dans un second temps, nous avons présenté la modélisation autorégressive.

La première partie du chapitre a été consacrée essentiellement à faire un tour d'horizon sur les ondelettes. Il donne un aperçu rapide sur la notion échelle, la caractéristique principale qui doit vérifier une onde mère ainsi que le célèbre algorithme de MALLAT. La résolution en fréquences de la transformée utilise des fonctions bien localisées dans le plan temps-fréquence lui donne beaucoup d'avantages :

- La résolution en fréquences de la transformée dépend du facteur de dilatation s par le principe d'Heisenberg, on peut donc choisir arbitrairement celle-ci suivant ce que l'on désire analyser.
- La localisation en temps est précieuse pour nombre d'applications.
- La transformée en ondelettes peut représenter complètement et efficacement un signal quelconque en peu de coefficients.
- L'algorithme de paquets d'ondelettes conduit à une décomposition en sous-bandes de fréquence du signal ; cette décomposition est ajustable par le choix des composantes. Ce type d'analyse offre une grande souplesse pour l'utilisateur et lui permet de s'adapter au signal à analyser.

Dans la deuxième partie du chapitre nous avons abordé la modélisation paramétrique autorégressive qui consiste à associer à un signal un modèle, représenté par un vecteur paramètre $\theta = [\theta_1, \theta_2, \dots, \theta_p]^T$ censé représenter au mieux le signal considéré. Etant donné que l'on choisit a priori un modèle pour le signal, cela signifie qu'en général on possède des informations a priori sur le signal lui-même qui permettent de sélectionner tel ou tel modèle. Il s'ensuit que le choix d'un modèle plutôt que d'un autre requiert au préalable une analyse du signal.

L'intérêt d'une modélisation paramétrique est de réduire l'espace de représentation. On peut ainsi représenter un ensemble de N échantillons par un vecteur de dimension $p \ll N$.

Chapitre IV

Détection des défauts des roulements par analyse vibratoire

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 des deux techniques mises en oeuvre pour analyser et traiter les signaux vibratoires en vue de la détection et du diagnostic des défauts de roulements. Ces méthodes seront analysées et discutées, essentiellement en ce qui concerne leurs efficacités et leurs coûts d'implantation. Nous proposerons alors un système automatique pour le traitement des signaux vibratoires afin de détecter des défauts de roulements. Enfin nous appliquerons ce système sur deux bases de données reconnues.

2 Généralités sur les roulements à billes

2.1 Constitution des roulements à billes

De manière générale, les roulements sont constitués de quatre composants essentiels. Un organe externe, un organe interne, des organes roulants et un organe de lubrification. En fonction de l'application envisagée, les organes de roulement 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.

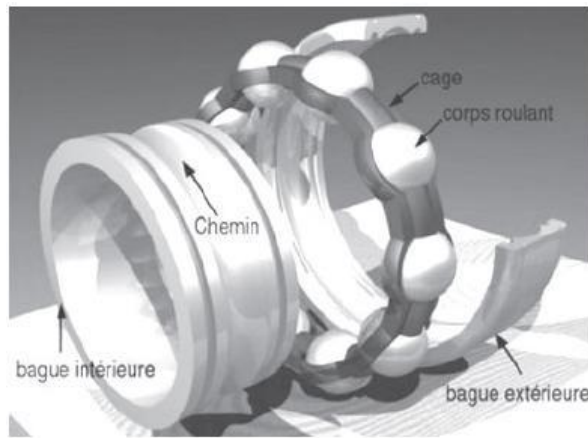


Figure IV.1 – Vue éclatée des éléments constitutifs d'un roulement à une rangée de billes

Dans ce travail, nous nous concentrons sur un type de roulements couramment utilisés, à savoir les roulements métalliques, graissés, à une rangée de billes, à contact radial, soit le modèle le plus courant dans les machines électriques de moyenne gamme de puissance. Les roulements concernés sont donc constitués d'une bague externe, d'une bague interne et d'une rangée de billes maintenues par une cage, ainsi que de graisse pour la lubrification. Un éclaté de ce type de roulement est présenté en figure (IV.1). Par ailleurs, les billes sont en contact, d'une part avec la piste de roulement interne et d'autre part avec la piste de roulement externe. On parlera d'ailleurs sans distinction de piste de roulement, de bande de roulement ou bien de chemin de roulement.

2.2 Modes de dégradations d'un roulement à billes

Par définition, la durée de vie d'un roulement est une donnée statistique caractérisant le nombre de cycles, c'est-à-dire de tours qu'il peut effectuer en étant soumis à une contrainte spécifique, avant de présenter des défauts d'écaillage des surfaces de contact. 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 [51]. Cependant, cette évaluation de la durée de vie des roulements ne prend en compte que les phénomènes de fatigue des matériaux. De manière générale, l'usure d'un roulement peut difficilement être décrite théoriquement du fait de la complexité des mécanismes mis en jeu et de l'interdépendance de ceux-ci. En effet, un mécanisme d'usure n'intervient jamais seul, mais un ensemble de mécanismes conduit à la dégradation d'un roulement et à la diminution de sa durée de vie. De nombreux facteurs externes aux roulements, influençant leur durée de vie, sont répertoriés. Nous pouvons citer par exemple,

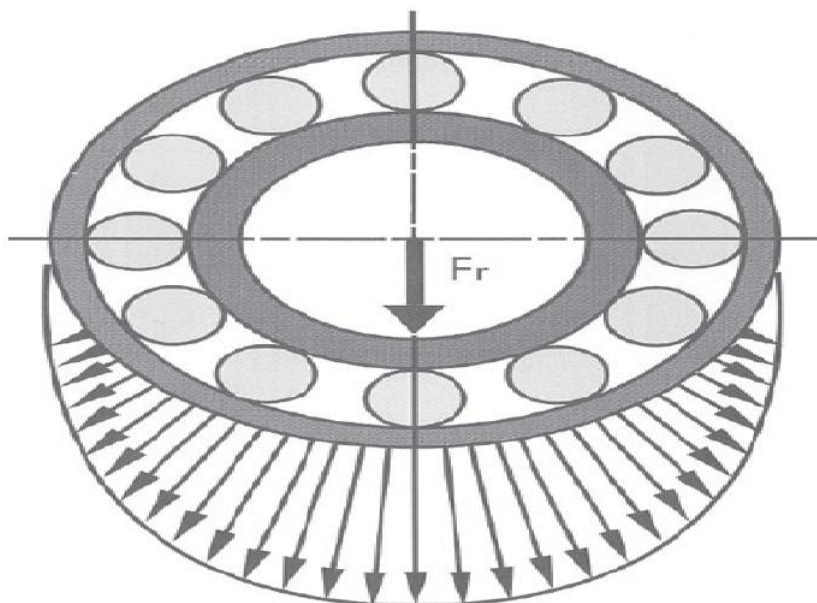


Figure IV.2 – Répartition d'une contrainte radiale au sein d'un roulement

et de manière non exhaustive, la présence de charges mécaniques excessives, les défauts de montage, la circulation de courants électriques, la corrosion, la présence de contaminants au sein du roulement, la qualité de la lubrification ou encore la température de fonctionnement. Nous allons détailler ces mécanismes ainsi que leurs effets.

2.2.1 Effets des contraintes mécaniques

Lorsqu'un roulement est placé dans une machine, il subit des contraintes mécaniques radiales et axiales. Les contraintes axiales apparaissent plus particulièrement dans des machines entraînant un ventilateur ou lorsqu'il existe une précharge axiale, c'est-à-dire un dispositif appliquant une poussée ou une traction dans l'axe de l'arbre de rotation. Les contraintes radiales, quant à elles, existent nécessairement du fait de la masse de l'arbre de rotation de la machine. Ces dernières s'exercent donc sur la bague interne du roulement qui les transmet à la bague externe par l'intermédiaire des billes. Ainsi, les contraintes radiales ne s'exercent pas en un seul point de la bague externe mais se répartissent dans la moitié inférieure de celle-ci, comme le schématise la figure (IV.2). Dans cet exemple, la contrainte radiale est unique et est symbolisée par la force F_r . Considérons le cas où la bague interne est rotative alors que la bague externe est fixe. La zone de charge couvre alors un arc de la piste de roulement externe. Par ailleurs, de par la rotation de la bague interne, la zone de charge couvre la longueur totale de la piste de roulement interne, figure (IV.3) [52].

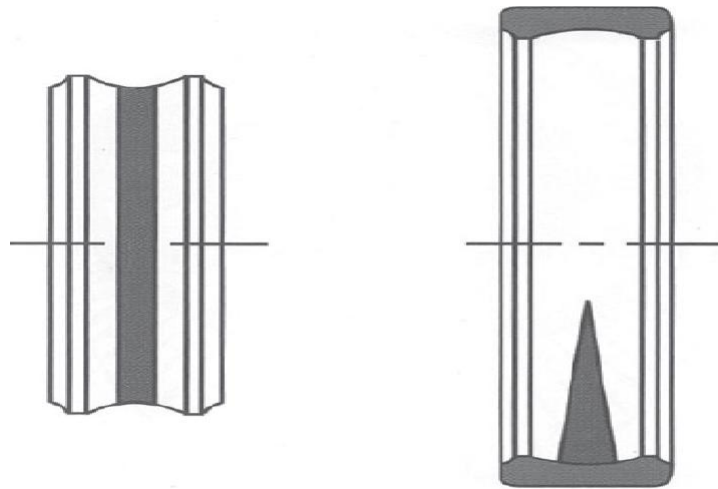


Figure IV.3 – Schématisation de la zone de charge d'une bague interne et d'une bague externe



Figure IV.4 – Aspect de la bague interne d'un roulement surchargé

Par ailleurs, les efforts exercés au niveau des points de contact se répartissent au sein des matériaux en une contrainte de compression et une contrainte de cisaillement. La contrainte de compression est maximale au point de contact bille-bague et décroît avec la profondeur dans le matériau. La contrainte en cisaillement est très faible au point de contact, augmente, présente un maximum puis décroît avec la profondeur [53].

Lorsque les efforts exercés sont trop importants, il y a surcharge du roulement. La résistance des matériaux des bagues ou des billes est mise à mal et il se produit des criques et fissures internes, voire un écaillage des pistes de roulement ou des billes si les efforts sont cycliques. La dégradation du roulement est alors considérée comme mortelle.

Les figures (IV.4;IV.5) montrent les dégradations sévères que peuvent provoquer des surcharges mécaniques [52].

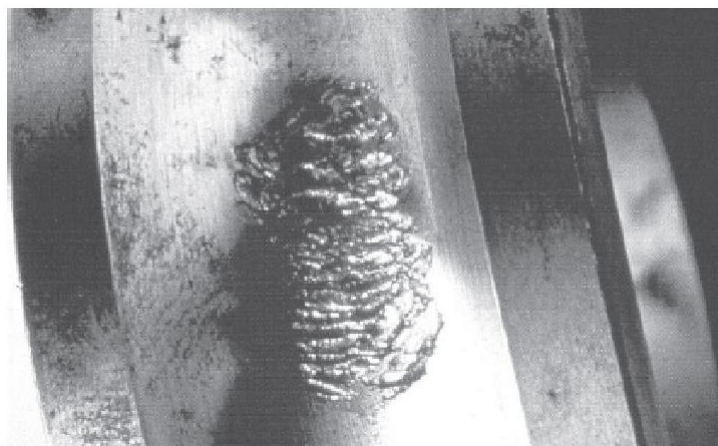


Figure IV.5 – Ecaillage d'une piste de roulement



Figure IV.6 – Ecaillage d'une piste

2.2.2 Défauts de montage

La mise en place d'un roulement sur l'arbre de la machine d'une part et dans son logement au niveau des paliers statiques d'autre part, nécessite d'être effectuée avec soin. Si les alignements ne sont pas respectés, des contraintes fortes sont exercées sur les bagues. Les mêmes effets que précédemment sont observés, figure (IV.6). Les contraintes exercées peuvent également être très localisées, allant jusqu'à la rupture complète des bagues, figure (IV.7) [52].

2.2.3 Corrosion

Des agents corrosifs, tels que des acides, sont susceptibles de s'introduire dans les roulements et de dégrader les matériaux métalliques les constituant. Des criques apparaissent alors sur les bagues et les billes, comme le montre la figure (IV.8) [52]. Sous l'effet de la corrosion

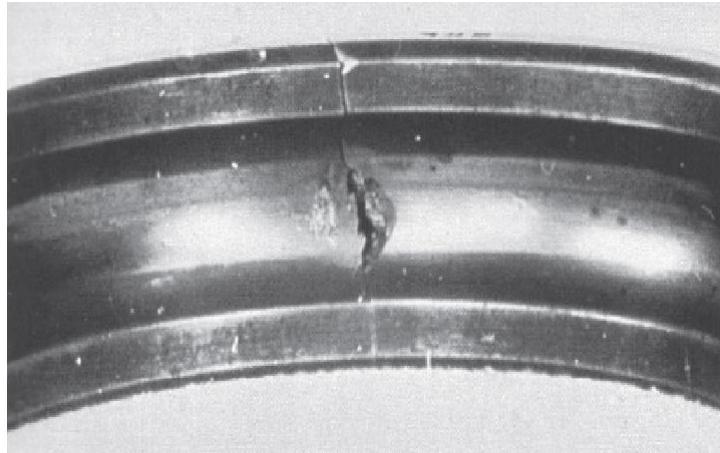


Figure IV.7 – Cassure d'une bague

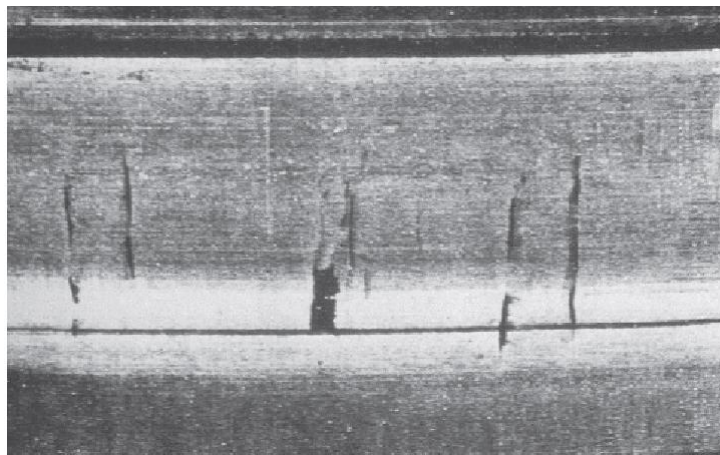


Figure IV.8 – Criques de corrosion sur une piste de roulement

et/ou d'autres facteurs, les criques évoluent en fissures entraînant des dégats mortels pour le roulement.

2.2.4 Contamination du roulement

Les machines peuvent être installées dans des environnements poussiéreux ou en extérieur, là où les vents peuvent amener des sables ou des poussières. Pour les applications en milieux dits contaminés, les roulements sont de préférence étanches et le lubrifiant choisi est une huile circulant en circuit fermé. Ces systèmes de lubrification sont complexes et chers et ne sont donc pas toujours mis en oeuvre.

Les roulements non étanches, utilisés en milieux contaminés, peuvent donc se trouver eux-mêmes contaminés, c'est-à-dire que des impuretés entrent dans le roulement et dégradent alors

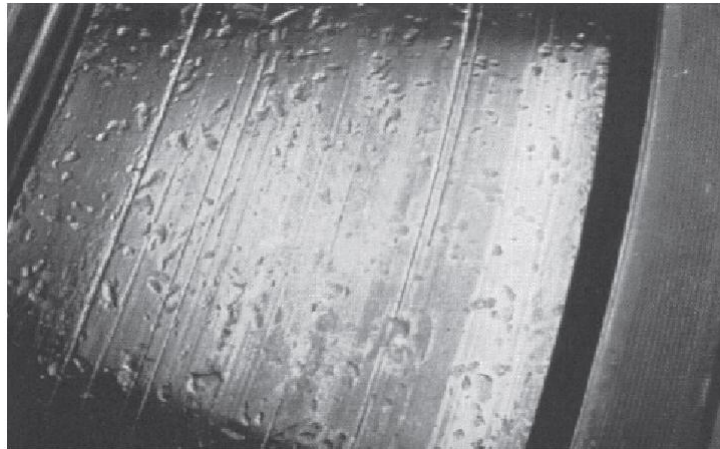


Figure IV.9 – Marques d'indentation d'un chemin de roulement dues à la présence de contaminant

les surfaces de contact. Plusieurs mécanismes peuvent être mis en jeu. Lorsque les particules de contaminant s'insèrent au niveau du contact bille/bague, il peut se produire un phénomène d'abrasion à trois corps et/ou une indentation des particules dans les billes ou les pistes, figure(IV.9)[52]. Ceci ne se produit que si les particules sont constituées d'un matériau plus dur que celui des organes internes du roulement. Les dégradations peuvent alors s'intensifier et créer des fissures, voire un écaillage des constituants du roulement. Ces mécanismes sont résumés par la figure (IV.10).

2.2.5 Défaut de lubrification

Le lubrifiant est un organe essentiel des roulements dans la mesure où il crée un film d'huile protecteur au niveau des contacts bille/bague permettant de protéger les surfaces. Deux types de défauts de lubrification principaux peuvent se présenter. Le premier est une dégradation des qualités lubrifiantes de la graisse ou de l'huile ou bien une absence de lubrifiant. En effet, en l'absence de lubrifiant, les surfaces métalliques se trouvent en contact direct, frictionnent et s'érodent rapidement, jusqu'à entraîner des écaillages des pistes, des éclatement des billes, un blocage et donc une destruction du roulement, comme le montre les figures (IV.11),(IV.12) [52]. Le second type de défaut de lubrification intervient dans les roulements par l'intermédiaire d'un sur-graissage (la graisse occupe tout le volume libre interne au roulement).

Dans ce cas, la viscosité de la graisse empêche la libre rotation de la cage et par conséquent des billes, ce qui entraîne un glissement de ces dernières au niveau des points de contact. Les surfaces de contact se dégradent alors rapidement.

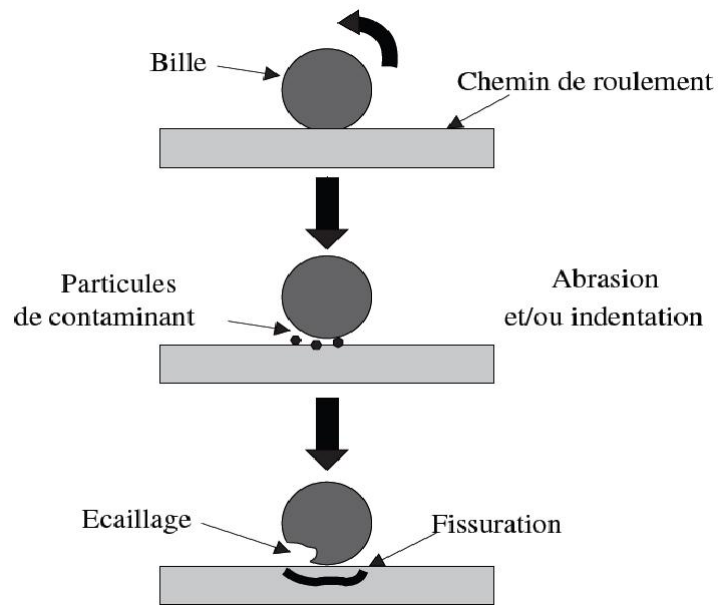


Figure IV.10 – Mécanismes généraux d'usure en présence de contaminant

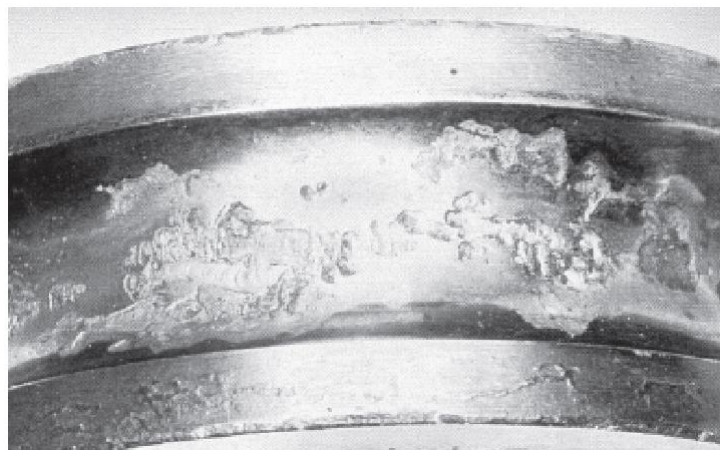


Figure IV.11 – Ecaillage d'un chemin de roulement



Figure IV.12 – Eclatement de billes

2.2.6 Température de fonctionnement

En conditions normales d'utilisation, la température de fonctionnement d'un roulement est généralement comprise entre $-20^{\circ}C$ et $+120^{\circ}C$. La géométrie générale d'un roulement est dépendante de la température de fonctionnement de celui-ci. En effet, les éléments métalliques se dilatent avec la température. Dans ce cas, le jeu naturel du roulement est compensé et le lubrifiant ne peut plus s'insérer au niveau des points de contacts bille/bague, ce qui entraîne les memes effets que ceux précédemment cités. Par ailleurs, la graisse est composée d'une huile lubrifiante et d'un savon liant [53]. 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 (IV.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 IV.1 – Causes de défauts dans les roulements.

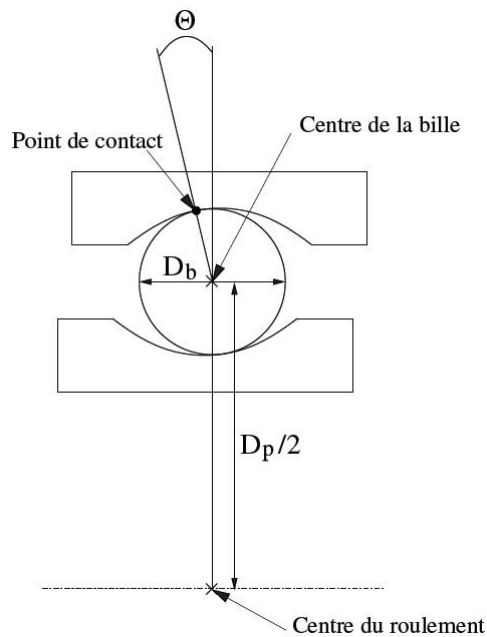


Figure IV.13 – Représentation schématique de la géométrie interne d'un roulement à une rangée de billes

3 Influence des défauts sur la structure des signaux

Lorsqu'un défaut localisé apparaît en un point d'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, 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**]) :

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

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

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

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

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

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

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

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 (IV.13)).

5 Bases des données

Dans la section 6 , nous allons discuter l'application de notre méthodologie sur une première base de données des roulements.

Nous démontrerons dans la section 7 l'efficacité de notre approche en utilisant une deuxième

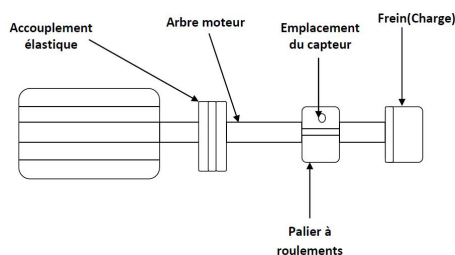


Figure IV.14 – Figure représentative du banc d'essais

base de données.

6 La première application

Une scématisation de l'installation du premier banc d'essai est illustrée en figure(IV.14).

Le banc d'essais est composé de :

- Un moteur électrique triphasé, tournant à une vitesse de 1500 RPM.
- un palier à roulements à billes.
- La transmission du couple de rotation du moteur vers le réducteur est assurée par l'intermédiaire d'un accouplement élastique.
- Les accélérations sont mesurées à l'aide d'accéléromètres piézo-électrique, dont leurs fixation se faite par un goujon vissé sur le palier à roulements.

Le banc est constitué d'un ligne d'arbre très rigides. L'entraînement est assuré par un moteur électrique triphasé. Le frein joue le role d'une charge.

Les roulements sur lesquels seront effectuées les mesures de dégradation, sont des roulements de type 6204 provenant du fabricant SKF. L'enregistrement du donnée est assurée par un analyseur de Fourier Bicanale de type B&K 2035 avec une fréquence d'échantillonnage de 16384 Hz [54]. La première base de donnée est mise sous forme de fichier ou de tableau (matrice). Ce dernier est représenté par des classes de vecteurs, ou chacune des classes représente un type de défaut. Et chaque vecteur représente un signal vibratoire, tableau IV.2.

6.1 Prétraitement des signaux et calcul des paramètres

6.1.1 Données temporelles

La représentation temporelle des signaux vibratoires est illustrée par La figure (IV.15).

Après une réduction de dimension par l'algorithme de l'Analyse en Composantes Principales(ACP), La base de donnée est représentée en trois dimensions afin de visualiser la distri-

Catégorie	Type de défaut	Symbole
1	Sain	C_1
2	Défaut bague extérieure	C_2
3	Défaut cage avec un élément dénoué	C_3
4	Défaut cage avec quatre éléments dénoués	C_4
5	Roulement très usés	C_5

Tableau IV.2 – Les types des défauts.

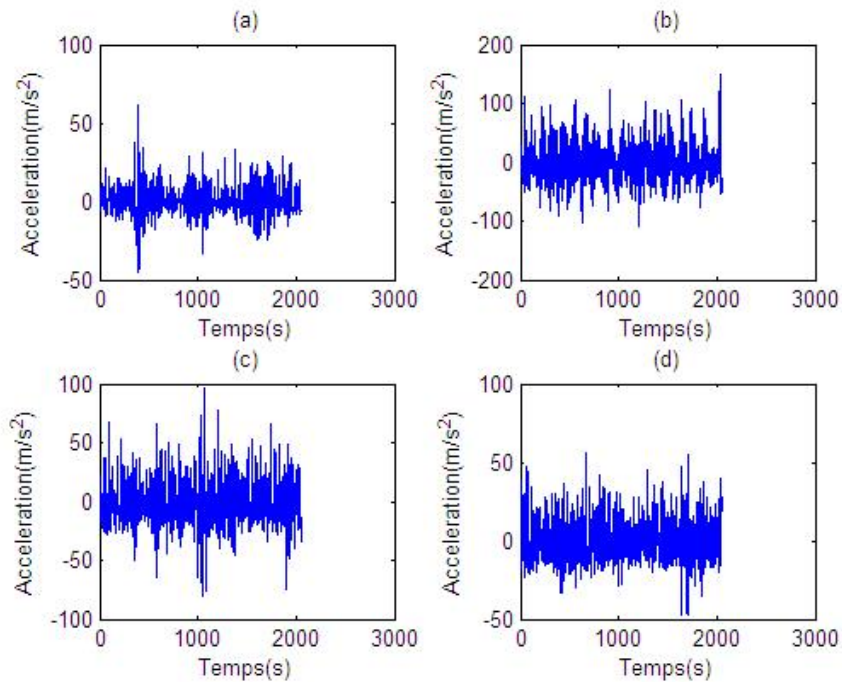


Figure IV.15 – Représentation temporelle des signaux vibratoires d'un roulement :(a) sain, (b) avec défaut bague extérieure, (c) avec défaut cage avec un élément dénoué, (d) avec défaut cage avec quatre éléments dénoués

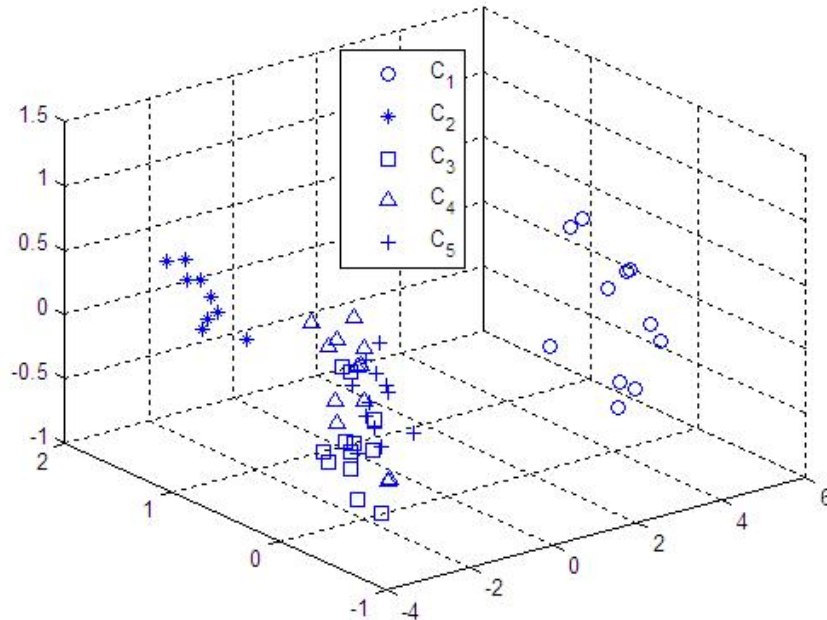


Figure IV.16 – Représentation des données en 3D

bution des cinq classes, figure(IV.16) . On remarque qu’il y a un chevauchement entre les trois classes C_3 , C_4 et C_5 .

6.1.2 Recherche de vecteur forme

La recherche de signatures consiste, en reconnaissance de formes, à construire le vecteur forme pour définir de manière pertinente les observations effectuées sur le système. En outre, le vecteur forme (qui est, à proprement parler, la signature du processus) doit permettre de bien distinguer les classes.

Deux algorithmes d’extarction de paramètres sont utilisés :

1. L’extraction basée sur l’algorithme de paquets d’ondelettes présenté en quatre étapes comme suit[7] :
 - Décomposition du signal vibratoire par une ondelette Daubechies (db8) en trois niveaux, et extractions des coefficients des huit bandes de fréquence a partir de la troisieme couche, (figure(IV.17) et figure(IV.18))
 - 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}

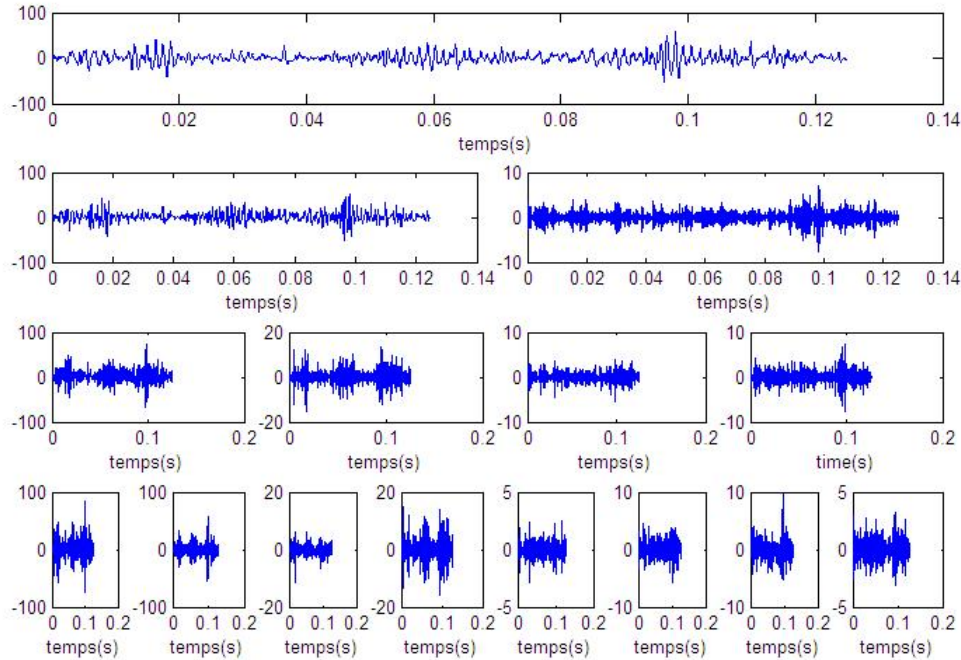


Figure IV.17 – Paquet d’ondelette d’un roulement sain

présente le signal reconstruit de d_{31} et ainsi de suite, Le signal composé 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 calculé par : $E_{3j} = \sum_{k=1}^n |d_{jk}|^2$.
 - Le vecteur d’entrée du classifieur "x" a été construit en fonction de l’énergie de chaque bande de fréquence : $x = [E_{30}, E_{31}, E_{32}, E_{33}, E_{34}, E_{35}, E_{36}, E_{37}]$
2. L’algorithme basé sur la modelisation autoregressive (AR) avec estimation des paramètres par la méthode de covariance, qui est une des méthodes consistant à rechercher des paramètres pertinents.

Le vecteur d’entrée du classifieur "x" a été formé a partir des coefficients du modèle AR comme suit : $x = [a_2, a_3, \dots, a_p]$, où p est l’ordre du modèle AR.

6.2 Méthodes de la prise de décision

Le diagnostic automatique nécessite au moins deux étapes principales, la première consiste à l’extraction de l’information pertinente du signal à traiter, chose qui a été faite précédemment via l’algorithme de paquets d’ondelettes ou par la modelisation autoregressive. La seconde étape

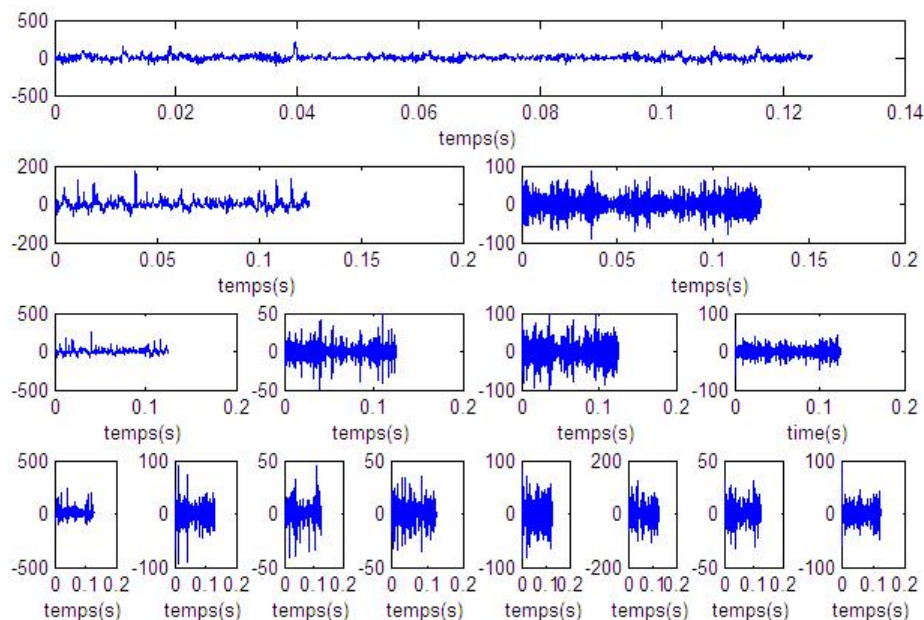


Figure IV.18 – Paquet d’ondelette d’un roulement avec un défaut bague extérieure

consiste à la prise de décision, pour cela Nous avons utilisé deux méthodes de prise de décision, à savoir les réseaux de neurones (MLP) et les séparateur a vaste marge(SVM).

6.3 Prise de décision par les réseaux de neurones (MLP)

Avant la construction du système de RNA pour diagnostic des défauts dans les roulements (identification des signaux acquis à partir d’un accéléromètre). On doit tout d’abord accéder à la phase de prétraitement des données (basée sur modelisation autoregressif suivie par une analyse de composante principale) à partir de laquelle le RNA pourra apprendre.

Le réseau utilisé est un réseau multi couches, comportant une couche d’entrée , une couche de sortie qui correspond à la décision, et une couche cachée, tableau(IV.3).

RNA	Couche d’entrée	Couche cachée	Couche de sortie
Nombre de neurones	3	10	5

Tableau IV.3 – Architecture de RNA utilisé

Le réseaux sélectionné est entraîné par l’algorithme de rétro propagation. la simulation donne

un taux de validation de 100% et un taux de test de 86.3636%.

6.4 Prise de décision par les SVM

6.4.1 Méthode classique

Un classifieur SVM issu de l'approche un contre tout est utilisé pour la classification de cinq défauts défectueux de roulement.

comme nous l'avons déjà vu, dans le chapitre 2, parmi les paramètres de noyaux influençant notablement la performance du SVM, on citera le paramètre " σ " de la fonction noyau, et la variable de pénalisation " c " du SVM qui est un hyper-paramètre additionnel qui influence la généralisation du classifieur lorsque les classes sont fortement enchevêtrées où un bruit considérable subsiste.

Un système de classification par SVM nécessite donc le réglage de paramètres tels que le paramètre de pénalisation " c " et le choix d'un type de noyau et de sa paramétrisation. Ces paramètres sont typiquement réglés au moyen d'une procédure de validation. Cette procédure nécessite l'exécution de plusieurs instances d'apprentissage et de test en explorant un ensemble de valeurs possibles des différents paramètres, pour retenir celles donnant les meilleurs résultats de classification.

La base de données est divisée en deux parties ; l'une est utilisée pour faire l'apprentissage de SVM, et l'autre est utilisée pour le test.

Pour les différents choix du paramètre, en mesurant le taux de classification sur les deux bases de données construites. Dans la suite nous présentons respectivement les résultats de classification obtenus avec la procédure AR, dans un premier temps, et dans un deuxième temps, avec les paquets d'ondelettes.

Nous commençons par étudier le comportement des SVM vis à vis du paramètre de pénalisation " c ". De nombreux tests préliminaires ont été effectués en utilisant les deux noyaux polynômial et gaussien et des valeurs de " c " que nous avons fait varier par puissances de 10, en prenant $c = \{1; 10; 100; 1000; 10000\}$. Cela nous a conduit à restreindre l'étude à des valeurs de " c " dans l'ensemble $c \in [50 : 100 : 2000]$, figure(IV.19).

Nous souhaitons maintenant, en nous fixant une valeur de " c ", sélectionner le noyau et sa paramétrisation les plus appropriés.

Les choix intéressants de " σ " se situent dans l'intervalle $[1 : 1 : 50]$; Nous présentons les cas $\sigma \in [1 : 1 : 10]$, figure(IV.20). L'influence de l'ordre de modélisation sur le taux de classification est illustrée dans la figure (IV.21).

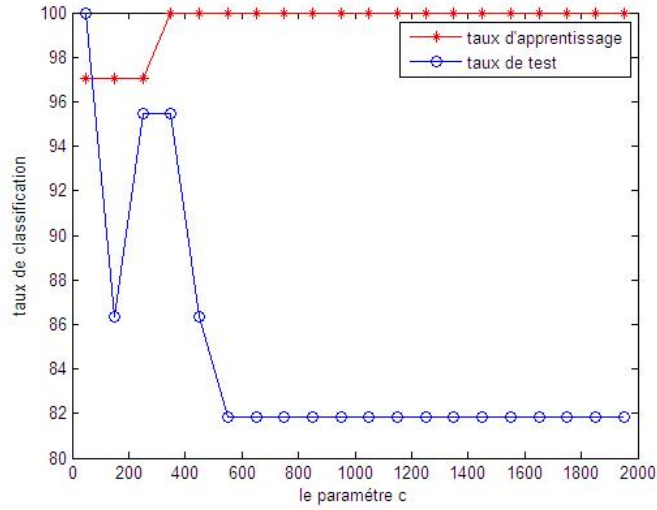


Figure IV.19 – Taux de test et d'apprentissage avec $\sigma = 1, c \in [50 : 100 : 2000]$.

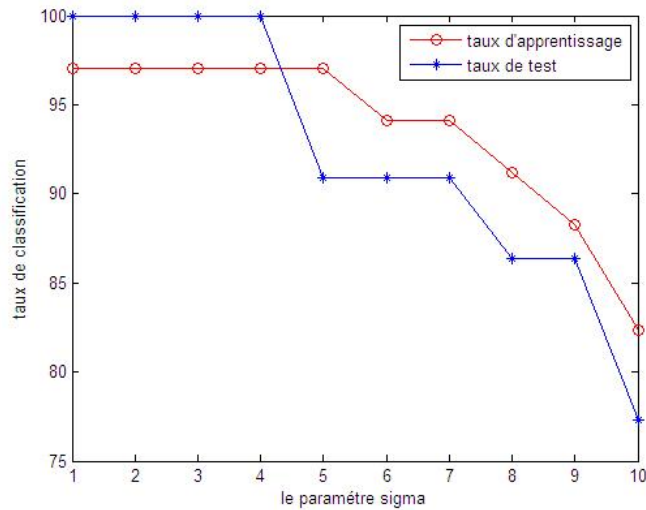


Figure IV.20 – Taux de test et d'apprentissage avec $c = 50, \sigma \in [1 : 1 : 10]$.

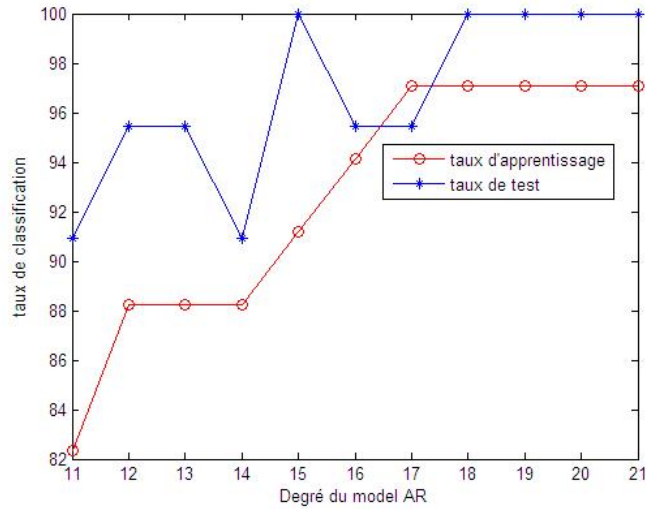


Figure IV.21 – Taux de test et d’apprentissage avec $c = 50, \sigma = 1$, degré du modèle AR $\in [11 : 1 : 21]$.

Les figures (IV.22) et (IV.23) illustrent les résultats obtenus avec le noyau polynomial, pour le deuxième expert basée sur un prétraitement par paquets d’ondelettes. Le taux de classification est donné pour le couple de paramètres $(\sigma; c)$.

Finalement; le tableau(IV.4) résume Les performances de classification par les SVM tout en variant l’ordre de modélisation et le paramètres de régularisation (c).

Nous remarquons que l’ensemble des paramètres qui donnent les meilleurs résultats pour le

Ordre du modèle AR	Paramètre de regularisation	Taux de validation	Taux de test
20	1000	97.06%	100%
18	1000	97.06%	100%
18	50	97.06%	100%
17	1000	97.06%	95.45%
16	1000	97.06%	95.45%
15	1000	91.18%	95.45%
14	1000	88.24%	90.91%
18	10000	100%	86.36%
18	100	94.12%	100%

Tableau IV.4 – Effet de l’ordre de modèle et de paramètre de noyau sur le taux de classification.

premier et le deuxième experts sont respectivement $(\sigma = 1; c = 50; \text{degré du model AR}=18)$ et $(\sigma = 2; c = 2400)$.

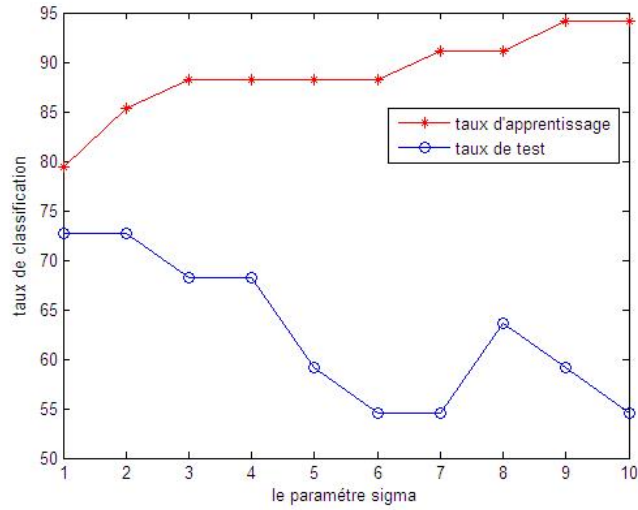


Figure IV.22 – Taux de test et d'apprentissage avec $c = 2500, \sigma \in [1 : 1 : 10]$.

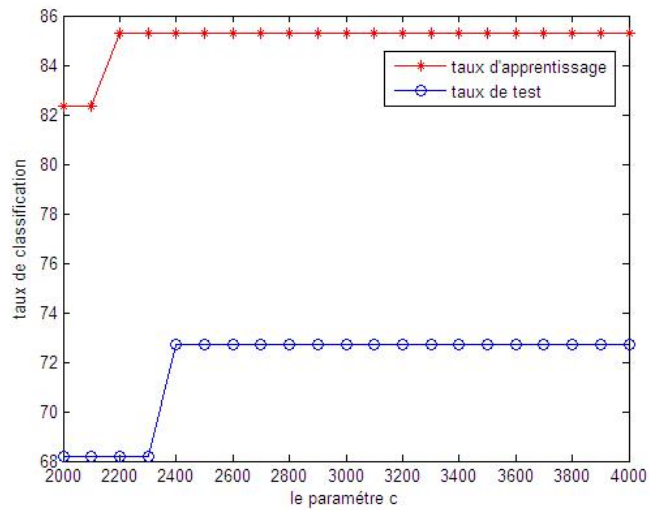


Figure IV.23 – Taux de test et d'apprentissage avec $\sigma = 2, c \in [2000 : 100 : 4000]$.

IV.6 La première application

Les meilleurs résultats de classification pour les deux experts sont représentés dans le tableau (IV.5). Une matrice de confusion de dimension 5×5 a été construite pour montrer les performances de classification. Les éléments diagonaux représentent les défauts des roulements correctement classés, alors que celles non-diagonaux représentent les défauts mal classés. On

SVM+	Fonction noyau	Résultat de validation	Resultat de test	Taux de validation	Taux de test
AR	Gaussienne	7 0 0 0 0 0 6 0 0 0 0 0 7 0 0 0 0 1 6 0 0 0 0 0 7	4 0 0 0 0 0 3 0 0 0 0 0 5 0 0 0 0 0 5 0 0 0 0 0 5	97.06%	100%
Paquets d'ondelette	Polynomial	7 0 0 0 0 0 6 0 0 0 0 0 5 1 1 0 0 2 5 0 0 0 1 0 6	4 0 0 0 0 0 3 0 0 0 0 0 1 4 0 0 0 0 1 4 0 0 0 1 4	85.29%	72.73%

Tableau IV.5 – Résultats de classification

remarque que le taux de classification obtenu avec le premier expert est meilleur que celui obtenu avec le deuxième. Il y a une faible erreur de classification due à la similarité entre la classe 3 et 4 qui représente pratiquement le même défaut, avec une petite différence dans le nombre de billes dénouées.

la taille de l'ensemble de validation peut affecter les résultats et les performances du classifieur. Il est évident qu'une validation croisée ou une procédure de " k-fold cross validation " par exemple peut améliorer la précision de la probabilité d'erreur estimée.

Ces expériences nous ont permis de déduire que la capacité de généralisation des SVM dépend essentiellement du bon choix des deux paramètres à savoir celui de la fonction noyau ainsi que de la régularisation, nous nous orientons alors pour leurs résolutions vers l'optimisation de ces paramètres par un algorithme récent : l'optimisation par essaim de particule.

6.4.2 Optimisation par essaim particulaire (OEP)

Kennedy et Eberhart, proposent en 1995 une nouvelle méthode d'optimisation nommée Optimisation par Essaim de Particule PSO, (en anglais : Particle Swarm Optimization). PSO est une méthode d'optimisation stochastique basée sur une population de particules, un Essaim

Chapitre IV. Détection des défauts des roulements par analyse vibratoire

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 [55],[56],[57] :

$$Z_{id}^k = Z_{id}^{k-1} + V_{id}^k \quad (\text{IV.5})$$

Tel que [55] :

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[55],[56] :

$$V_{id}^k = \omega.V_{id}^{k-1} + c_1.r_1.(P_{id} - Z_{id}^{k-1}) + c_2.r_2.(P_{Gd} - Z_{id}^{k-1}) \quad (\text{IV.6})$$

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,

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 [47] :

$$\omega(k) = \omega_{max} - \frac{(\omega_{max} - \omega_{min})}{k_{max}}.k \quad (\text{IV.7})$$

Tel que :

ω_{max} : Poids initial,

ω_{min} : Poids final,

k_{max} : Nombre d'itérations maximum,

k : Itération courante.

La fonction de pondération "w" 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 :

- **Etape 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*.
- **Etape 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.
- **Etape 3** : Modification du point de recherche ;
La modification du point de recherche se fait en moyennant les équations (IV.5),(IV.6).
- **Etape 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 [56],[57],[58],[59].

Maintenant, nous allons procéder à l'optimisation par essaims de particules des paramètres des SVM, à savoir : le paramètre de régularisation "c", et celle de la fonction kernel " σ ".

Formulation de la fonction objective

Notre objectif est d'arriver, via la procédure d'optimisation par l'OEP, au meilleur taux de classification basée sur la 10-fois validation croisée (10-fold cross validation) et ce, en explorant les valeurs admissibles des paramètres des SVM existant dans l'espace de recherche.

6.4.3 Choix des paramètres de l'OEP

Le choix des paramètres reste en grande partie empirique mais suit une certaine logique. Après un grand nombre d'essais, nous avons fixé les facteur $c_1 = 1.3$, $c_2 = 1.3$, $\omega_{max} = 1$ et $\omega_{min} = 0.00001$. Nous avons aussi fixé la population à 20 et le nombre d'itérations à 10.

6.4.4 Résultats de simulation

Le tableau (IV.6) présente les paramètres des SVM obtenus par l'optimisation par l'OEP ainsi que le taux de classification. Dans cette expérience, l'approche basée sur la modélisation

SVM+	Fonction noyau	c	σ	Taux de validation%	Taux de test%
Modélisation AR	Gaussien	8.1317	0.5789	98.00	96.00
Paquets d'ondelette	Polynomial	27.1447	17.2956	91.33	62.00

Tableau IV.6 – Résultats de simulation

AR donne 96% pour un couple de (σ, c) égale (0.5789, 8.1317), alors que l'approche basée sur le paquage d'ondelette donne 62.00% pour un couple de (σ, c) égale (17.2956, 27.1447).

Nous avons aussi pu établir l'efficacité de l'optimisation des paramètres de SVM par PSO où nous avons pu obtenir des meilleures performances que celles obtenues avec la sélection manuelle.

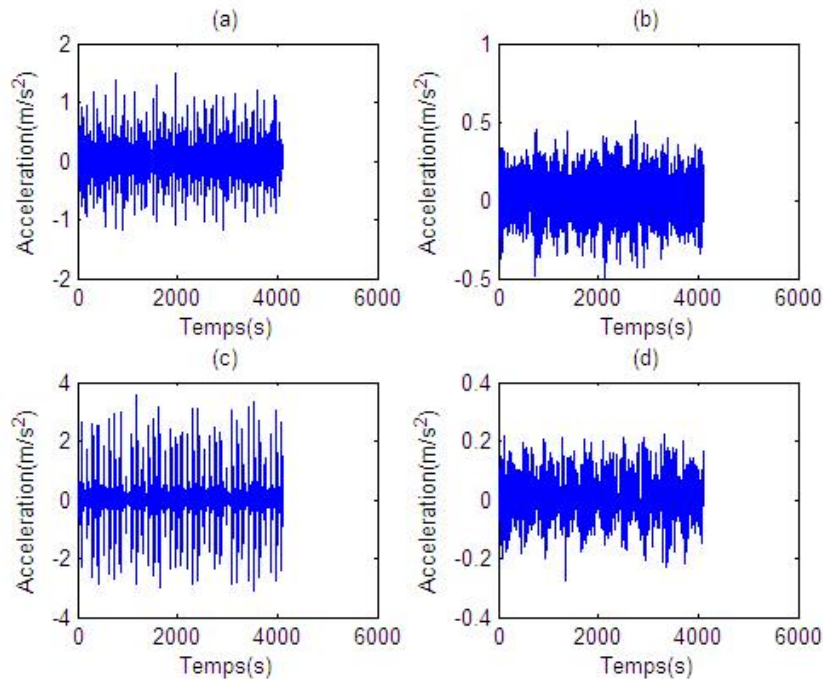


Figure IV.24 – Représentation temporelle des signaux vibratoire d’un roulement : (a) avec défaut de bague intérieure, (b) avec défaut de bille, (c) avec défaut bague extérieure, (d) sain

7 La seconde application

Le banc d’essais de l’université de Case Western Reserve a pour but de réaliser les expériences nécessaires à la détection des défauts de roulements. Comme le montre la figure (IV.25) ci-dessus, 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) [60].

Les défauts ont été provoqués pour les roulements rigides à bille de type 6205-2RS-EM-SKF, soit sur la bague intérieure, extérieure ou 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 défaillants ont été réinstallés dans le moteur de test et les données de vibration ont été enregistrées pour les charges de moteur de 0 à 3 chevaux (vitesse moteur de 1720 à 1797RPM). 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.

Des signaux de vibration ont été recueillies à l’aide d’un enregistreur DAT 16 canaux, et ont été traitées 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 de donnée est présentée dans le tableau (IV.7).

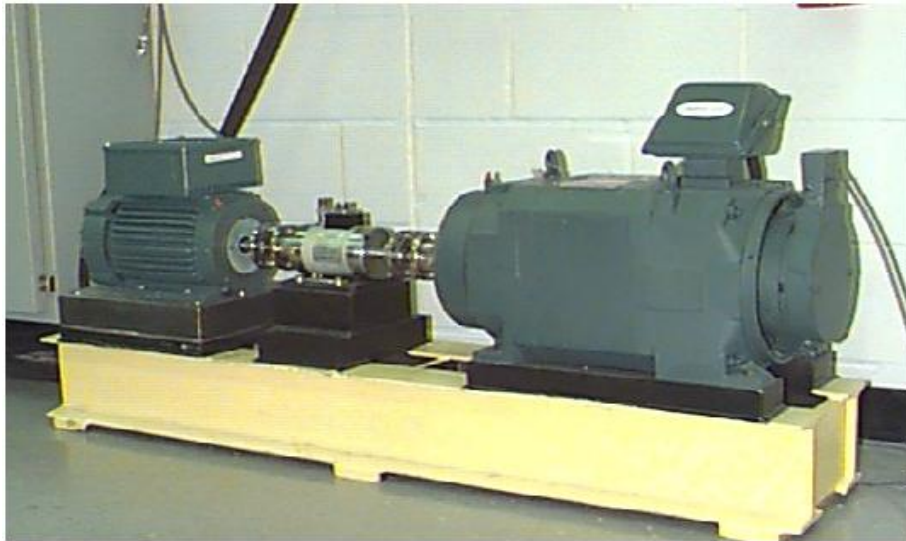


Figure IV.25 – Banc d'essai

Après une réduction de dimension par l'algorithme de l'Analyse en Composantes Principales(ACP), La représentation des signaux en 3D permet de visualiser la complexité de cette base de donnée, figure(IV.26).

7.1 Résultat de simulation

Dans un premier temps nous avons divisée la base de donnée en deux parties, dont la première est utilisée pour l'apprentissage (tous les données sauf celles marquées avec étoile) et la deuxième est consacrée pour le test (données marquées par étoile).

le résultat de simulation est illustrée dans le tableau (IV.8).

Un partitionnement différent des ensembles d'apprentissage et test donne des résultats de test et de validation différents, ce qui se traduit par une variance importante. C'est la raison pour laquelle on a opté d'appliquer, dans un second temps, la méthode "Leave-One-Out (LOO)" sur les données marquées par le signe (+). le tableau (IV.9) expose les résultats de simulation obtenus.

D'une manière générale, l'ensemble des résultats obtenus, montrent l'intérêt de l'approche AR-SVM-PSO, Ils confirment ainsi que cette dernière permette de répondre aux problèmes complexes de diagnostic des défauts des roulements.

IV.7 La seconde application

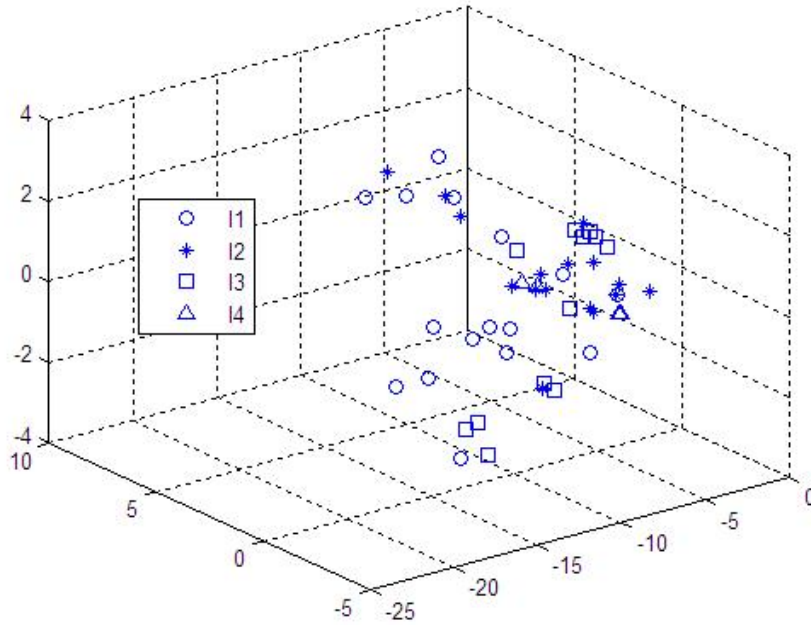


Figure IV.26 – Représentation des données en 3D

Diamètre de défaut (pouces)	Charge (HP)	Vitesse du moteur (rpm)	Défaut bague intérieure	Défaut bille	Défaut bague extérieure	Roulement sain
0.007	0	1797	l_{1+}	l_{2+}	l_{3+}	l_{4+}
	1	1772	l_{2+}	l_{2*+}	l_{3+}	l_{4+}
	2	1750	l_{3*+}	l_{2+}	l_{3+}	l_{4+}
	3	1730	l_{4+}	l_{2*+}	l_{3*+}	l_{4*+}
0.014	0	1797	l_{15}	l_{25}	l_{35}	
	1	1772	l_{16*}	l_{26}	l_{36}	
	2	1750	l_{17}	l_{27*}	l_{37}	
	3	1730	l_{18}	l_{28}	l_{38*}	
0.021	0	1797	l_{19}	l_{29*}	l_{39}	
	1	1772	l_{110*}	l_{210}	l_{310}	
	2	1750	l_{111*}	l_{211}	l_{311}	
	3	1730	l_{112}	l_{212}	l_{312*}	
0.028	0	1797	l_{113*}	l_{213*}		
	1	1772	l_{114}	l_{214}		
	2	1750	l_{115}	l_{215}		
	3	1730	l_{116}	l_{216*}		

Tableau IV.7 – Base de donnée

SVM+	Fonction noyau	c	σ	Taux de Validation%	Taux de test%
Modelisation AR	Gaussien	9.9965	1.7788	100	100

Tableau IV.8 – Résultats de simulation

SVM+	Fonction noyau	c	σ	Taux de Validation%	Taux de test%
Modelisation AR	Gaussien	8.1317	0.5789	100	100

Tableau IV.9 – Résultats de simulation

8 Conclusion

Dans ce chapitre, nous avons présenté les performances des différentes méthodes testées pour le diagnostic de l'état d'un roulement.

Tout d'abord, une présentation des roulements à billes, de leurs processus de dégradation ainsi que des effets vibratoires associés, nous a permis de souligner l'importance de la mise en place de structures de surveillance pour détecter les possibles pannes.

L'expérimentation sur des données multiclasse est effectuée. Celle-ci est scindée en deux parties. Dans la première, nous présentons les résultats d'expérience sur une première base de données, cela est décomposé en deux. En premier, les performances du système sont calculées en choisissant de façon manuelle les paramètres de SVM. Ensuite, nous montrons les résultats de l'optimisation automatique par PSO. Dans la deuxième partie des expériences, nous exécutons la procédure sur la base de données de l'université de Case Western Reserve.

Dans notre système, nous avons proposé pour la phase de prétraitement des données, deux algorithmes pour l'extraction de caractéristiques à savoir le paquage d'ondelette et la modélisation autorégressive.

Pour effectuer l'étape de classification, les MLP et l'approche multi-classes "un-Contre-tous" pour les SVM ont été mise en oeuvre, les performances de deux classifieurs ont été ainsi comparées. Le temps de calcul est nettement à l'avantage des SVM : le temps de calcul d'un RNA est de 24.1229 secondes alors que le SVM met 7.5068 secondes. Pour l'optimisation des SVMs, nous avons effectué dans une première approche, un test croisé sur les paramètres : de régularisation "c" et de noyau " σ ". Cette étape consistait à faire un balayage d'un intervalle de valeurs pour le couple de paramètres (c et σ) pour pouvoir retenir celui qui donne les meilleurs

résultats de classification. Cette méthode conduit à des calculs lourds, et dont l'efficacité n'est pas garantie ; c'est pour ces raisons que nous avons opté la deuxième approche où les paramètres des SVM sont obtenus par l'optimisation par essaim de particules.

Nous avons observé que les taux de classifications étaient systématiquement plus élevés dans le cas d'utilisation d'un classifieur SVM basé sur l'approche "un-Contre-tous" optimisé par PSO et combiné avec la modélisation AR en les comparant avec ceux obtenus en utilisant les autres approches des diagnostics présentés.

Conclusion générale

L'objectif de cette étude était d'apporter des éléments de diagnostic et de traitement du signal pour permettre de détecter et diagnostiquer les défauts mécaniques d'un entraînement et plus particulièrement les défauts des roulements. Le travail mené au cours de cette thèse a permis d'obtenir un système de diagnostic automatique de l'état d'un roulement. Les objectifs de l'automatisation du système de diagnostic sont de détecter précisément le roulement qui cause l'anomalie et d'identifier le type de défaut. Nous avons choisi d'utiliser une approche de surveillance et de diagnostic par reconnaissance de formes à partir de l'enregistrement de signaux vibratoires sur le roulement.

Le premier chapitre concernait l'étude des méthodes de diagnostic par reconnaissance des formes. Nous avons présenté, dans un premier temps, les deux étapes principales sur lesquelles repose tout système de surveillance, à savoir la perception et le diagnostic. Nous avons ensuite dressé un état de l'art des méthodes de diagnostic en nous concentrant plus particulièrement sur les méthodes de diagnostic par reconnaissance des formes. La dernière phase du processus est une étape de classification supervisée où les classes de défauts étant supposées connues.

Le mécanisme de mise en place d'un système de diagnostic à base de reconnaissance des formes nécessite l'existence d'une bibliothèque de mesures réalisées sur le processus et totalement étiquetées. Un ensemble d'apprentissage et un ensemble de test, totalement disjoints, en sont extraits.

Les trop grandes dimensions des matrices de chaque signal, essentiellement dûes à une fréquence d'acquisition trop grande, les rendaient peu intéressantes en vue de la classification. Le chapitre 3 revient sur la sélection et/ou l'extraction de variables. Nous nous sommes intéressés aux méthodes d'extraction de caractéristiques et plus particulièrement au paquage d'ondelette et au modélisation autorégressive.

La dernière partie de la thèse est consacrée à la présentation de notre système complet de diagnostic automatique et d'aide à la décision. De nombreuses expérimentations menées sur nos données ont permis de montrer que notre méthode, la modélisation autorégressive, est capable de fournir des variables pertinentes pour la classification des signaux vibratoires.

deux outils de classification ont été testés sur nos données. Un intérêt très particulier a été réservé aux machines à vecteurs de support (SVMs). Leurs performances ont été comparées à celles des méthodes connexionnistes. Cette étude démontre la supériorité des SVM sur les méthodes discriminantes classiques telles que le MLP. L'influence du paramètre de régularisation c et du noyau s'est avérée importante dans notre contexte, alors nous nous sommes intéressés à l'optimisation de ces paramètres par PSO pour la sélection du meilleur modèle.

Références bibliographiques

- [1] AFNOR. *Maintenance industrielle, méthodes et outils*. AFNOR **1** (1996). [1](#), [6](#)
- [2] R. BIGRET, J. L. FÉRON ET C. PACHAUD. *Diagnostic, maintenance, disponibilité des machines tournantes - modèles, mesurages, analyses des vibrations*. Masson. [1](#)
- [3] P. O'DONNELL. *Report of large motor reliability survey of industrial and commercial installations*. IEEE Transactions on Industry applications **1A-21** Juillet-Aout (1985). [2](#)
- [4] R. H. ENGELMANN ET W. H. MIDDENDORF. *Handbook of electric motors*. Marcel Dekker, New York (1995). [2](#)
- [5] 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). [2](#)
- [6] S. A. MCINERNEY ET Y. DAI. *Basic vibration signal processing for bearing fault detection*. IEEE Transactions on Education, pages 149–156. (2003). [2](#)
- [7] 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). [2](#), [74](#)
- [8] P. K. KANKAR, C. S. SATISH ET S. P. HARSHA. *Fault diagnosis of ball bearings using machine learning methods*. ELSEVIER (2011). [2](#)
- [9] L. SHUANG ET L. MENG. *Bearing fault diagnosis based on pca and svm*. IEEE International Conference on Mechatronics and Automation August (2007). [2](#)
- [10] A. GATTAL, S. CHENIKHER, K. M. PEKPE ET J. P. CASSAR. *Bearing faults diagnosis using artificial network with dimensionality reduction by pca*. 1st International Conference on Electrical Engineering CIGET09 (2009). [2](#)

- [11] M. BASSEVILLE ET M. O. CORDIER. *Surveillance et diagnostic de systèmes dynamiques, approches complémentaires du traitement de signal et de l'intelligence artificielle*. Technical Report IRISA 1(1004) (1996). 4
- [12] G. ZWINGELSTEIN. *Diagnostic des défaillances : Théorie et pratique pour les systèmes industriels, traité des nouvelles technologies*. Série Diagnostic et Maintenance. Hermès (1995). 5
- [13] P. M. FRANK. *Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy : A survey and some new results*. Automatica **26(3)**, pages 459–474. (1990.). 6
- [14] J. N. CHATAIN. *Diagnostic par système expert, traité des nouvelles technologies, série diagnostic et maintenance*. Hermès (1993). 6, 8
- [15] J. J. GERTLER. *Fault detection and diagnosis in engineering systems. electrical engineering and electronics*. Marcel Dekker. **104** (1998). 7
- [16] R. ISERMANN ET P. BALLE. *Trends in the application of model-based fault detection and diagnosis of technical processes*. Control Engineering Practice **5(5)**, pages 709–719. (1997). 7
- [17] J. L. ALTY ET L. I. MIKULICH. *Industrial applications of artificial intelligence*. Kluwer Academic Publishers **2** (1991). 8
- [18] R. P. W. DUIN, A. K. JAIN ET J. MAO. *Statistical pattern recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence **22(1)**, 4–37. (2000). 8, 9, 33
- [19] L. HOLMSTRÖM, P. KOISTINEN, J. LAAKSONEN ET E. OJA. *Comparison of neural and statistical classifiers theory and practice*. research reports **a13** (1996). 9
- [20] M. MILLGRAM. *Reconnaissance de formes, méthodes numériques et connexionnistes*. Armand Colin, Dunod (1993). 9
- [21] J. SOTTILE ET L.E. HOLLOWAY. *An overview of fault monitoring and diagnosis in mining equipment*. IEEE Transactions on Industry Applications **30(5)**, pages 1326–1332. (1994). 9
- [22] M. BERAU. *Contribution à la théorie des sous-ensembles flous à la règle de discrimination des k plus proches voisins en mode partiellement supervisé*. Thèse de doctorat, Université de Technologie de Compiègne (1986). 9
- [23] R. B. OUAGHRAM. *Contribution au diagnostic des machines tournantes complexes. application à un laminoir*. Thèse de doctorat, Université de Technologie de Compiègne (1999). 9
- [24] S. ZIEBA. *Une méthode de suivi d'un système évolutif. application au diagnostic de la qualité d'usinage*. Thèse de doctorat, Université de Technologie de Compiègne (1995). 9

Références bibliographiques

- [25] H. M. ZAHRA. *Analyse temps echelle et reconnaissance des formes pour le diagnostic du système de guidage d'un tramway sur pneumatiques*. Thèse de doctorat, Université Blaise Pascal - Clermont 2 (2008). 9, 33
- [26] P. E. HART, R. O. DUDA ET D. G. STORK. *Pattern classification*. John Wiley and Sons 2 (2001). 9, 12
- [27] B. DUBUISSON. *Diagnostic, intelligence artificielle et reconnaissance des formes*. Editions Hermès, paris (2001). 12
- [28] V.N. VAPNIK. *The statistical learning theory*. Springer (1998). 12, 15, 27
- [29] P. BURMAN. *A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated Learning-Testing methods*. Biometrika. pages 503–514. (1989). 14
- [30] R. C. DUBES, A. K. JAIN ET C. C. CHEN. *Bootstrap techniques for error estimation*. IEEE Trans. Pattern Anal. Mach. Intell. pages 628–633 (1987). 14
- [31] B. EFRON ET R. TIBSHIRANI. *An introduction to the bootstrap*. Chapman & Hall. pages 628–633 (1993). 14
- [32] V.N. VAPNIK. *The nature of statistical learning theory*. Springer Verlag (1995). 20
- [33] L. WANG. *Support vector machines :theory and applications*. Springer-Verlag (2005). 20, 22, 23, 25, 29, 30
- [34] V. KECMAN. *Learning and soft computing : Support vector machines, neural networks, and fuzzy logic models*. The MIT Press Cambridge, Massachusetts London, England (2001). 21, 22, 24
- [35] A. SHIGEO. *Support vector machines for pattern classification*. Springer-Verlag (2005). 22, 23, 24, 25, 30, 31
- [36] V. N. VAPNIK. *Statistical learning theory*. New York : Wiley (1998). 25
- [37] C.CORTES ET V.N.VAPNIK. *Support vector networks. machinery learning*. Machinery Learning (1995). 27
- [38] B.SCHOLKOPF, A.SMOLA, R. C.WILLIAMSON ET P. L.BARTLETT. *New support vector algorithms*. Neural Computation (2000). 27
- [39] L. HAMEL. *Knowledge discovery with support vector machines*. JohnWiley & Sons (2009). 30
- [40] 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). 33

- [41] S.J. RAUDYS ET V. PIKELIS. *On dimensionality, sample size, classification error and complexity of classification algorithms in pattern recognition*. PAMI **2(3)**, pages 243–252. (1980). 33
- [42] F. COTTET. *Traitement des signaux et acquisition de données*. Dunod, Paris (2009). 34
- [43] L. DEBNATH. *Wavelet transforms and their applications*. Birkhauser Boston (2002). 34, 35, 38, 40, 41
- [44] S. MALLAT. *A wavelet tour of signal processing*. Elsevier Inc (2009). 34, 35, 38, 39, 40
- [45] Y. MEYER. *Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs*. Séminaire Bourbaki, (1985-1986). 37
- [46] Y. MEYER. *Ondelettes et opérateurs : Ondelettes*. Editions Hermann **1** (1990). 37
- [47] S. MALLAT. *A theory for multiresolution signal decomposition : The wavelet transform*. IEEE Trans. Pattern Anal. Machine Intell **11**, pages 674–693. (1989). 37
- [48] A. N. AKANSU ET R. A. HADDAD. *Multiresolution signal decomposition : Transforms, subbands, and wavelets*. Academic Press (2001). 38, 42, 44, 45
- [49] P. STOICA ET R. MOSES. *Spectral analysis of signal*. Pearson Education (2005). 53
- [50] P. STOICA ET R. MOSES. *Introduction to spectral analysis*. Labrary of Congress Cataloging-in-publication Data (1997). 53
- [51] T. A. HARRIS. *Rolling bearing analysis*. Wiley, New-York **3** (1991). 62
- [52] R. A. GUYER. *Rolling bearings handbook and troubleshooting guide*. Chilton Book Company, Radnor, Pennsylvania (1996). 63, 64, 65, 67
- [53] M. MORET. *Roulements et butées à billes et à rouleaux*. Techniques de l'Ingénieur Mai (1990). 64, 69
- [54] M. TABASZEWSKI ET C. CEMPEL. *Ball bearing dataset, structural integrity and damage assessment. network (sidanet)*. available at : www.sidanet.org (1998. (consulté en Juillet 2006)). 72
- [55] B. K. PANIGRAHI, Y. SHI ET L. M. HIOT. *Handbook of swarm intelligence concepts, principles and applications*. Springer (2011). 82
- [56] R. YUAN ET B. GUANGCHEN. *Determination of optimal svm parameters by using ga/psa*. Journal of computer **5(8)** August (2010). 82, 84
- [57] H. CHENG-LUNG ET D. JIAN-FAN. *A distributed pso-svm hybrid system with feature selection and parameter optimization*. ELSEVIER (2008). 82, 84

Références bibliographiques

- [58] 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). [84](#)
- [59] J. LI. *A combination of pso and svm for road icing forecast*. Journal of computers **5**(9) September (2010). [84](#)
- [60] K. A. LOPARO. *Bearings vibration data set.case western reserve university*. [/http ://www.eecs.case.edu/laboratory/bearing/download.htm](http://www.eecs.case.edu/laboratory/bearing/download.htm) (consulté en septembre 2012). [85](#)
- [61] M. ZANI. *Guide d'achat, mesures mécaniques, la surveillance des machines tournantes*. pages 77–84. (2003). [108](#)
- [62] P. LYONNET. *La maintenance : Mathématique et méthodes, techniques et documentation*. (1992). [108](#)
- [63] I. MOUMENE. *Identification des défauts mixtes d'engrenage et de roulement par analyse vibratoire*. Memoire magistère, Guelma (2010). [108](#)
- [64] B. PINKOWSKI. *Principal component analysis of speech spectrogram images*. Pattern recognition **30**, pages 777–787. (1997). [111](#)
- [65] B. DUBUISSON. *Diagnostic et reconnaissance des formes. traité des nouvelles technologies*. Série diagnostic et Maintenance **30** (1990). [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 :

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

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

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 $c_1, c_2, \dots, c_n \in \mathbb{C}$ 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 : La maintenance

La maintenance des machines a une grande signification économique pour toute entreprise. La continuité dans la production, la sauvegarde du capital, le respect de l'environnement et le succès économique ne peuvent être garantis que par une maintenance qui anticipe les dégâts et les catastrophes. Pour la maintenance des machines, on distingue deux stratégies : la maintenance reactive et celle préventive.

1 Maintenance reactive

Appelée maintenance corrective selon la norme AFNORNF60-010, maintenance effectuée après altération d'un bien à accomplir la fonction requise (donc après la défaillance). Ce type de maintenance est le plus répandu, mais il consiste à intervenir une fois que le mal est fait, en effectuant des réparations : il est donc le plus coûteux. De plus, les défaillances survenant de manière non prévisible, l'organisation de cette activité est limitée.

2 Maintenance préventive

Appelé maintenance systématique selon la norme AFNORNF60-100 effectuée selon des critères prédéterminés, dans l'intention de réduire la probabilité de défaillance d'un bien ou la dégradation d'un service rendu (donc avant la défaillance). Cette maintenance est systématique, dans le sens où les visites d'entretien vont être programmées selon une exploitation statistique des données recueillies sur les artefacts. Le surcoût entraîné par des visites systématiques est pondéré par la planification qui peut en être faite, et par la diminution des arrêts de productions due à la limitation de la maintenance corrective. Dans certains cas, cette maintenance est vitale pour assurer la sécurité du système et de son environnement (transports aériens, centrales nucléaires), elle comprend trois types :

- **La maintenance préventive systématique :**

Les dates de révisions sont déterminées à partir des durées de vie statistiques des différentes pièces d'usure le constituant, indépendamment de l'état réel de la machine.

– **La maintenance préventive prévisionnelle :**

La décision d'intervention et son degré d'urgence reposent sur l'identification préalable du ou des défauts et sur l'estimation de leur gravité. Ce type de maintenance est exécuté en suivant les prévisions extrapolées de l'analyse et de l'évaluation de paramètres significatifs de la dégradation du bien.

– **La maintenance préventive conditionnelle :**

Elle consiste à surveiller et à analyser de façon continue l'état de fonctionnement et son évolution dans le temps des installations ou des machines en pratique; ce qu'on appelle une surveillance de l'état des machines et ceci pendant les phases de démarrage, de fonctionnement et d'arrêt. La connaissance et le traitement de ces données permet de déceler les anomalies de fonctionnement et de décider les interventions nécessaires. Un arrêt pour inspection ou remise en état n'est décidé que si les valeurs des indicateurs de surveillance retenus ont dépassé des valeurs limites ou seuils prédéfinis. Cette technique s'appuie sur des mesures de quelques paramètres ou descripteurs physique tels que la mesure des vibrations et du bruit, la mesure des températures, l'analyse des huiles, etc. Cette méthode présente donc l'avantage d'éviter la surprise, de permettre généralement la planification de l'intervention et de limiter au strict minimum les arrêts des machines. En outre, la gérance des pièces de rechange est facilitée et la durée des arrêts est limitée.

3 Pratique de la maintenance préventive conditionnelle

son principe consiste, à partir de différentes techniques basées sur la mesure des paramètres physiques (les vibrations, analyse d'huile, ...etc.), à suivre l'état de marche (et son évolution) du matériel en fonctionnement, de manière à diagnostiquer et programmer à l'avance les interventions de maintenance.

3.1 les techniques d'analyse

le tableau (A.1) indique les différentes techniques utilisées dans la maintenance conditionnelle qui sont les suivantes :

3.1.1 L'analyse vibratoire

est la plus connue et la plus largement employée. Il faut dire qu'elle permet de détecter pratiquement tous les défauts susceptibles d'apparaître dans les machines tournantes. Un balourd, un jeu, un défaut d'alignement, un roulement usé ou endommagé, se traduisent par une variation des efforts internes que subit la machine, et donc à une modification de son comportement vibratoire. En plaçant des accéléromètres aux endroits où se transmettent ces efforts (c'est-à-dire sur les paliers des machines), on peut alors suivre l'état de santé de l'équipement. L'analyse vibratoire est une technique utilisée pour le suivi des machines tournantes en fonctionnement, cette dernière engendre des efforts qui seront souvent la cause des défaillances ultérieures (turbulence, chocs, instabilité), ces efforts sont à leur tour causes de vibrations qui vont endommager les structures et les composants des machines. La mesure de ces vibrations va permettre d'identifier les efforts dès qu'ils apparaissent, avant qu'ils causent de dommages irréversibles. Ces vibrations recueillies lors des campagnes de mesures sont porteuses d'informations qui caractérisent l'état de fonctionnement de certains composants mécaniques, constituant la machine analysée. L'analyse vibratoire a comme avantages : Les mesures sont faciles à prendre, les défauts sont détectés à un stade précoce, et il est possible de réaliser un diagnostic approfondi pour en connaître l'origine.

3.1.2 L'analyse des lubrifiants

appliquée à toutes les machines pourvu qu'elles contiennent un fluide (huile ou graisse) que l'on peut prélever, la méthode consiste à détecter les particules d'usure drainées par le lubrifiant afin d'en déduire l'état de l'équipement. L'analyse d'un échantillon de quelques dizaines de millilitres d'huile (Prélevé à l'aide de petites seringues ou lors d'un vidange) peut alors suffire à révéler l'usure anormale de l'un des composants, le mauvais état d'un filtre, ou encore la dégradation du lubrifiant (suite à une oxydation, à une pollution, à une température d'utilisation trop élevée). À côté de la surveillance vibratoire et de l'analyse des lubrifiants, la thermographie infrarouge.

3.1.3 La thermographie infrarouge

suscite un intérêt encore récent dans le domaine de la maintenance. Jusqu'alors réservée au contrôle d'installations électriques, elle est depuis peu utilisée dans la surveillance des machines tournantes pour détecter les défauts qui se manifestent par un échauffement anormal à la surface de l'équipement. La thermographie permet de réaliser des mesures à distance, et d'obtenir instantanément une image thermique de la zone inspectée.

Enfin, il existe d'autres méthodes, moins répandues que les précédentes mais néanmoins intéressantes. Citons par exemple :

3.1.4 L'analyse acoustique

qui permet de détecter tout bruit anormal à l'aide de microphones placés (le plus souvent) à distance de l'équipement.

3.1.5 Le contrôle ultrasonore

qui permet de détecter des défauts de faible amplitude à haute fréquence tel que l'initiation de la dégradation d'un roulement, par exemple. **Remarques**

- L'analyse vibratoire, l'analyse d'huiles, la thermographie infrarouge et l'analyse acoustique sont les principales méthodes de surveillance des machines tournantes.
- Toutes ne permettent pas de détecter les mêmes types de défauts, ni de réaliser un diagnostic approfondi pour en connaître l'origine et la gravité.
- Avant de choisir une telle méthode d'analyse, il faut bien cerner la criticité de la machine, et définir le type de surveillance qu'elle nécessite.

4 Choix d'une technique d'analyse

Entre ces différentes méthodes, le choix n'est pas si difficile. Chaque méthode a son champ d'application privilégié, (tableau A.2). C'est pourquoi elles entrent rarement en concurrence. L'analyse vibratoire, par exemple, convient aux défauts liés à la cinématique et à la structure de la machine, mais dans une plage de fréquence déterminée (située généralement entre quelques Hertz et plusieurs dizaines de kHz). Elle couvre alors les défauts structurels (basse fréquence), les défauts spécifiques aux roulements (à plus haute fréquence), ainsi que la majorité des dysfonctionnements les plus courants. En dehors de cette plage de mesure, on utilise d'autres méthodes. Par exemple, au-delà de 20 kHz, il est souvent préférable de réaliser un contrôle par l'analyse ultrasonore ou par la thermographie infrarouge. Il en est de même avec l'analyse acoustique, qui se limite à la détection de bruits dans des fréquences audibles. Donc tout dépend de défauts que l'on souhaite détecter, Ainsi, dans le cas d'un défaut de roulement, qui commence à se manifester à haute fréquence puis se déplace à une fréquence de plus en plus faible (avec une énergie croissante) au fur et à mesure de sa dégradation, les ultrasons vont permettre de réaliser une détection à un stade plus précoce que l'analyse vibratoire ou l'analyse

4 Choix d'une technique d'analyse

Types de défauts	Analyse vibratoire	Analyse d'huiles	Thermographie IR	Analyse acoustique
Déséquilibres, balourd	Oui	Non (sauf s'il y a usure)	Oui (s'il y a un échauffement)	Non
Jeux, défauts de serrage	Oui	Non (sauf s'il y a usure)	Oui (s'il y a un échauffement)	Oui (s'il y a une résonance)
Défauts spécifiques aux roulements	Oui	Oui	Oui	Oui
Défauts spécifiques aux engrenages	Oui	Oui	Oui	Oui
Défauts de courroies	Oui	Non	Oui	Oui
Défauts d'alignement	Oui	Non	Oui	Non
Défauts liés à la lubrification (Dégradation de la qualité d'huile)	Non (sauf paliers fluides)	Oui	Non	Oui
Manque d'huile	Oui	Oui	Oui	Oui
Défauts de nature électrique ou électromagnétique	Oui	Non	Oui	Oui
Défauts liés aux écoulements (pour pompes et machines hydrauliques)	Oui	Non	Non	Oui

Tableau A.1 – Principales méthodes de détection.

acoustique. Et lorsque la dégradation du roulement se manifeste en zone audible, il est souvent trop tard. De même, les défauts liés au phénomène de cavitation ou aux écoulements de fluides, qui se manifestent à basse fréquence, sont détectés par un contrôle ultrasonore. Du côté de l'analyse d'huiles, c'est un peu la même démarche, elle permet d'abord de s'assurer que l'huile joue bien son rôle, autrement dit que ses caractéristiques physico-chimiques (Ex. : La viscosité) ne sont pas altérées, et ne vont pas nuire au bon fonctionnement de l'équipement. D'autre côté, les particules présentes dans l'huile vont révéler l'usure anormale d'un ou de plusieurs organes, ou la présence d'une pollution extérieure (Ex. : Des poussières dues au mauvais état d'un filtre) avant que celle-ci n'entraîne l'usure d'un élément. Dans ce cas, la détection est plus précoce que des méthodes telles que l'analyse vibratoire, qui ne détecte que la conséquence de la pollution. Il faut qu'elle soit appliquée à des machines peu polluées dans leur état normal (c'est le cas par exemple des machines hydrauliques). Sinon, comment être sûr que les particules que l'on compte sont bien dues à une pollution anormale.

Tout dépend aussi de la vitesse de dégradation du composant. En théorie, il est possible d'utiliser l'analyse d'huile pour détecter un défaut de roulements, puisqu'il se manifeste par des débris d'usure, mais en pratique, il faut savoir que la défaillance d'un roulement se produit en deux ou trois heures. Donc même si l'on effectue l'analyse au bon moment, il est souvent trop tard pour agir à temps. Il en est de même pour la thermographie infrarouge. En règle générale, si la température d'un roulement augmente, c'est qu'il est déjà très dégradé, on a donc très peu de temps devant soi. Ce qui explique d'ailleurs que certains fournisseurs aient développé leur propre outil de surveillance des roulements, par exemple, la préconisation de la méthode dite "d'onde de choc", son principe est basé sur l'utilisation d'un capteur dont la bande passante est centrée sur la fréquence des chocs émis par le roulement. On ne surveille alors que les défauts provenant de ce composant. On le voit, il n'existe pas de méthode universelle permettant de détecter de manière assez précoce tous les types de défauts que l'on est susceptible de rencontrer sur une machine tournante. La méthode idéale consiste alors souvent à réaliser une combinaison des techniques pour en exploiter la complémentarité.

Marie-Line Zani [61], a montré que l'analyse vibratoire, l'analyse d'huiles, la thermographie infrarouge et l'analyse acoustique sont les principales méthodes de surveillance des machines tournantes, en suite, il a donné quelques critères de choix entre telle méthode ou technique et leur domaine d'application. P. Lyonnet a détaillé les techniques d'analyse et leurs applications dans l'industrie, dans son livre [62] et [63]. L'objectifs d'une telle analyse est de :

- Réduire le nombre d'arrêts sur casse ;
- Fiabiliser l'outil de production ;
- Augmenter son taux de disponibilité ;

4 Choix d'une technique d'analyse

Techniques d'analyse	Principaux avantages	Principales limitations	Champ d'applications privilégié
Analyse vibratoire	<ul style="list-style-type: none"> -Détection de défauts à un stade précoce -Possibilité de réaliser un diagnostic approfondi -Autorise une surveillance continue - Permet de surveiller l'équipement à distance (télémaintenance) 	<ul style="list-style-type: none"> -Spectres parfois difficiles à interpréter -Dans le cas de la surveillance continue, installations relativement coûteuses 	Détection des défauts de tous les organes cinématiques de la machine (balourd, défauts d'alignement, jeux, etc.) et de sa structure
Analyse d'huiles	<ul style="list-style-type: none"> -Détection d'une pollution anormale du lubrifiant, avant que celle-ci n'entraîne une usure ou un échauffement -Possibilité de connaître l'origine de l'anomalie par analyse des particules 	<ul style="list-style-type: none"> -Ne permet pas de localiser précisément le défaut -Nécessite de prendre nombreuses précautions dans le prélèvement de l'échantillon 	Contrôle des propriétés physico-chimiques du lubrifiant, détection d'un manque de lubrification, analyse des éléments d'usure, analyse de contamination par le processus (étanchéité), etc.
Thermographie IR	<ul style="list-style-type: none"> - Permet de réaliser un contrôle rapide de l'installation -Interprétation souvent immédiate des résultats 	<ul style="list-style-type: none"> -Détection de défauts à un stade moins précoce que l'analyse vibratoire - Ne permet pas de réaliser un diagnostic approfondi 	Détection de tous les défauts engendrant un échauffement (manque de lubrification, en particulier)
Analyse acoustique	<ul style="list-style-type: none"> -Permet de détecter l'apparition de défauts audibles -Autorise une surveillance continue 	<ul style="list-style-type: none"> -Sensibilité au bruit ambiant -Diagnostic souvent difficile à réaliser - Problèmes de répétabilité des mesures 	Détection d'un bruit inhabituel pouvant ensuite être analysé par analyse vibratoire

Tableau A.2 – Choix d'une technique d'analyse.

- Mieux gérer le stock de pièces détachées.

Parmi les techniques citées au dessus, la plus importante, la plus fiable et performante surtout pour les phénomènes d'usure et de fatigue est l'analyse vibratoire.

On distingue communément deux principales activités pour cette technique d'analyse qui sont : la surveillance «la détection» et le diagnostic.

4.1 La surveillance

Le but est de suivre l'évolution d'une machine par comparaison des relevés successifs de ses vibrations. Une tendance à la hausse de certains indicateurs par rapport à des valeurs de référence constituant la signature alerte généralement le technicien ou l'ingénieur de maintenance, sur un dysfonctionnement probable. La signature est établie à partir d'une première campagne de mesures sur la machine neuve ou révisée. D'après, on distingue deux types de surveillance : surveillance «On line», les mesures des indicateurs soit de manière continue, et soit d'une manière périodique qui est la surveillance «Off line».

4.2 La détection

Pour détecter les défaillances d'un système, il faut être capable de classer les situations observables comme étant normales ou anormales. Cette classification n'est pas triviale, étant donné le manque d'information qui caractérise généralement les situations anormales, amené à une simplification communément adaptée, qui consiste à considérer comme anormale toute situation qui n'est pas normale.

4.3 Le diagnostic

Il permet de désigner l'élément de la machine défectueux suite à une évolution anormale des vibrations constatée lors de la surveillance (mesures). Le diagnostic n'est réalisé que lorsque la surveillance a permis de détecter une anomalie ou une évolution dangereuse du signal vibratoire. Le diagnostic dans la majorité des cas est établi en interprétant les signaux vibratoires émis par les machines en fonctionnement.

Annexe D : 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 [64] :

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

S une matrice carrée ($d \times d$), réelle et symétrique, c'est aussi la matrice d'inertie du nuage

Annexe D : Analyse en Composantes Principales

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 (1) 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 (2), sont les valeurs propres.

$$V=[v_1, v_2, \dots, v_{d'}, \dots, v_d] \quad (1)$$

$$\lambda = \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \dots & & \\ & & & \lambda_{d'} & \\ & & & & \dots \\ 0 & & & & & \lambda_d \end{bmatrix} \quad (2)$$

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 :

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

La nouvelle matrice des données s'écrit :

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

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 [65] 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.

RESUME

La surveillance vibratoire constitue une technique de contrôle non destructif très utilisée dans le secteur industriel, l'automatisation de son utilisation représente un gain important pour une meilleure maintenabilité des machines. Dans ce travail nous avons proposer une démarche pour améliorer le diagnostic automatique des défauts de roulements à billes par une sélection du vecteur d'entrée, à partir d'indicateurs extraits de signaux vibratoires. Deux techniques ont été utilisée au niveau de l'extraction du vecteur d'entrée, la modelisation autoregressive et le paquage d'ondelette. la phase de classification est assuré par des classifieurs intelligents a savoir, le perceptron multicouches et les support vecteur machine . l'algorithme des essaim de particule est utilisé pour l'optimisation des paramètre des SVM . Les performances des différentes combinaisons ont été comparées et discutées. La méthode proposée basée sur la modélisation autorégressive combinée avec SVM-PSO donne les résultats les plus satisfaisante.

Mots clés : la modelisation autoregressive, Support vector machine, le paquage d'ondelette, optimisation par essaim de particule, diagnostic des défauts, Roulement à billes.

ABSTRACT

Bearing fault diagnosis has attracted significant attention over the past few decades. It consists of two major parts : vibration signal feature extraction and condition classification for the extracted features. in This study we propose two approaches : Autoregressive Modeling and wavelet packet for feature extraction. The classification step is realised using neural network or support vector machine. To improve the classification accuracy for bearing fault prediction, particle swarm optimization (PSO) is employed to simultaneously optimize the SVM kernel function parameter and the penalty parameter. This work has extended the application of Autoregressive Modeling feature extraction algorithm as a powerful method for finding an effective subset of features in ball-bearing fault detection. Compared with other methods, AR-SVM-PSO is simpler and easier to realize. The combined AR modeling and SVM-PSO based technique is tested for bearing faults and provides satisfactory results.

Keywords : autoregressive modeling, Support vector machine, Wavelet Packet, Particle Swarm Optimization, Fault Diagnosis, roller bearing.