

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE FERHAT ABBAS – SETIF
FACULTE DES SCIENCES DE L'INGENIEUR
DEPARTEMENT D'OPTIQUE ET DE MECANIQUE DE PRECISION

THESE

Présentée pour l'obtention du diplôme de

DOCTORAT D'ETAT

Option : optique et mécanique de précision

Par

FELKAOUI Ahmed

Thème

**Diagnostic vibratoire au service de la maintenance :
de l'acquisition à la décision**

Soutenue le 12/12/2007 devant le Jury Composé de :

ZEGADI R.	Professeur UFA Sétif	Président
BEKKA R.E.	Professeur UFA Sétif	Rapporteur
SIDAHMED M.	Professeur UT Compiègne FRANCE	Co-Rapporteur
CHIKOUCHE Dj.	Professeur U.M ^{ed} Boudiaf M'Sila	Examineur
KHANFER N.	Professeur UFA Sétif	Examineur
BENMOUSSA H.	Maître de conférence U. E L. Batna	Examineur

*A mes chers parents,
A toute ma famille,
A ma femme et mes enfants
A mes amis.*

Remerciements

Cette étude a été effectuée au Laboratoire de Mécanique de Précision Appliquée (LMPA) Département d'Optique et de Mécanique de précision, Faculté des Sciences de l'Ingénieur (FSI) Université Ferhat Abbas- Sétif.

J'exprime ma profonde gratitude et ma reconnaissance envers Messieurs Raïs El-hadi BEKKA, Professeur à l'Université de Sétif et Ménad SIDAHMED, Professeur à l'Université de Compiègne France, pour l'aide précieuse qu'ils m'ont apporté dans l'élaboration de ces travaux. Je les remercie pour leurs conseils, leur soutien, leur disponibilité et les discussions fructueuses.

Je voudrais adresser mes sincères remerciements à monsieur Rabah ZEGADI, Professeur à l'université de Sétif, directeur du laboratoire LMPA, d'avoir accepté de présider l'honorable jury de cette thèse.

Je remercie vivement Messieurs Djamel CHIKOUCHE, Professeur à l'Université de M'sila, Nabil KHANFER, Professeur à l'Université de Sétif et Hocine BENMOUSSA, maître de Conférences à l'Université de Batna, de m'avoir fait l'honneur d'accepter d'être les Examineurs de cette thèse.

Je tiens à remercier sincèrement le Maître de Conférences Jérôme ANTONI du Laboratoire Roberval UT Compiègne qui m'a accueilli dans le laboratoire pour de nombreux stages. Qu'il trouve ici ma profonde sympathie pour son amabilité.

Je remercie également tous les membres du personnel du département de Génie mécanique de Compiègne pour leur accueil chaleureux et leur aide toujours très spontanée.

Mes remerciements vont également à tous les membres du LMPA et à mes collègues enseignants de l'Institut d'Optique et de Mécanique de précision.

Je ne saurai enfin oublier les jeunes Magisters, docteurs et futurs docteurs du laboratoire qui, durant ces années, ont entretenu une très bonne ambiance et ont fait preuve de solidarité. Je leurs souhaite à tous du succès dans leurs recherches ainsi que dans leurs vies professionnelles.

Résumé

La reconnaissance automatique des défauts mécaniques et particulièrement la détection des niveaux de dégradation suscitent de nos jours un grand intérêt. En effet, on aurait à gagner en prédisant la durée de vie des éléments des machines pour une maintenance conditionnelle.

Le travail que nous présentons dans cette thèse s'inscrit dans le cadre de l'automatisation du diagnostic moyennant les méthodes de reconnaissance de formes et de l'intelligence artificielle.

Dans un premier temps, nous exposons la méthodologie adoptée pour la conception du système de reconnaissance automatique des défauts mécaniques pour ensuite, détailler les différentes étapes en allant de l'acquisition à la décision. En seconde partie, nous nous intéressons à la sélection d'indicateurs les plus pertinents et non redondants qui constituent une étape importante de l'élaboration du système. De plus, nous montrons l'avantage qu'il y a à utiliser le vecteur forme (VF) réduit par le biais des signaux réels et de la visualisation par analyse en composantes principales (ACP). Nous proposons alors, une méthode basée sur une hybridation des algorithmes génétiques (AGs) et de l'analyse discriminante linéaire (ADL). Enfin, trois algorithmes de classification ont été mis au point à savoir :

- ✓ les k plus proches voisins (k-ppv).
- ✓ le perceptron multicouches (PMC)
- ✓ les séparateurs a vaste marge (SVM).

Ces choix sont dictés par le fait que le problème de diagnostic constitue un parfait exemple de problème de classification.

Des tests de performance pour les cas sans et avec sélection d'indicateurs sont effectués. L'évaluation des résultats a été réalisée en se basant sur le taux d'observations bien classées et des signaux réels de roulements et d'engrenages d'un hélicoptère. Les meilleurs résultats ont été obtenus par l'utilisation des SVM. De bonnes performances ont aussi été obtenues par le couplage des AGs-ADL.

Mots clés : machines tournantes, diagnostic vibratoire, reconnaissance de formes, intelligence artificielle, classification, traitement du signal, traitement de données

Sommaire

Chapitre 1 Introduction

1.1	Position du problème et Motivation	1
1.2	Conception d'un système automatique de reconnaissance de défauts	4
1.2.1	Signaux vibratoires	4
1.2.2	Phase d'analyse	4
1.2.3	Phase de décision (classification)	7
1.2.4	Phase d'exploitation	10
1.3	Plan de la thèse	10
1.4	Tavaux réalisés	11

Chapitre 2 Préliminaires

2.1	Introduction	15
2.2	Défauts dans les machines tournantes	15
2.3	Acquisition vibratoire	16
2.3.1	Grandeur à mesurer	17
2.3.2	Capteur et dispositions	19
2.3.3	Multiplexage	20
2.3.4	Conditionnement du signal	21
2.3.5	Numérisation du Signal	24
2.3.6	L'Ordinateur	25
2.3.7	Dispositions pratiques et prétraitement	27
2.4	Méthodes d'analyse des signaux	40
2.4.1	Analyse statistiques et temporels	41
2.4.2	Analyse fréquentielle	43
2.4.3	Analyses temps-fréquence et temps-échelle	45
2.4.4	Analyse Spectrale d'Ordre Supérieur à 2	47
2.4.5	Analyse temps multi-fréquences	48
2.4.6	Analyse cyclostationnaire	48
2.4.7	Analyse Spectrale Cyclique d'Ordre Supérieur	49
2.4.8	Signaux multidimensionnels	49
2.5	Génération d'indicateurs	51
2.5.1	Classification des indicateurs	52
2.5.2	Indicateurs temporels et statistiques	52
2.5.3	Indicateurs fréquentiels	55
2.5.4	Indicateurs du modèle AR ou AR évolutif	55
2.5.5	Indicateurs temps-fréquence	56
2.5.6	Indicateurs de la fonction de modulation	58
2.5.7	Indicateurs des statistiques d'ordre supérieur	58
2.5.8	Indicateurs des spectres d'ordre supérieur	59
2.5.9	Indicateurs des spectres cycliques d'ordre supérieur	60
2.6	Réduction du nombre d'indicateurs	60
2.6.1	Les algorithmes de réduction	60
2.6.2	Les approches de sélection	61
2.6.3	Les fonctions d'évaluation	63
2.7	Règles de décision	64
2.7.1	Règle des k-ppv	66
2.7.2	Règle du PMC	67
2.7.3	Règle des séparateurs à vaste marge (SVM)	68
2.8	Conclusion	68

Chapitre 3

Sources de Données Vibratoires

3.1	Introduction	69
3.2	Statistiques sur les défauts de machines	69
3.3	Les signaux d'engrenage	71
3.3.1	Origine des défauts	71
3.3.2	Types de défauts	72
3.3.3	Traitement du signal d'engrenage	76
3.4	Les signaux de roulement	78
3.4.1	Origine des défauts	81
3.4.2	Types de défauts	82
3.4.3	Traitement du signal de roulement	85
3.5	Banc d'engrenage (réducteur à engrenages)	88
3.6	Banc de roulement (moteur asynchrone)	89
3.7	Banc de la boîte à vitesse de l'hélicoptère	91
3.8	Conclusion	97

Chapitre 4

Applications des méthodes de classification par RdF

4.1	Introduction	97
4.2	Principe du diagnostic par RdF	97
4.3	Importance de la sélection d'indicateurs (signaux de roulement)	99
4.3.1	Calcul d'indicateur	99
4.3.2	Visualisation par analyse en composante principale (ACP)	102
4.3.3	Sélection d'indicateurs par le critère de fisher	102
4.3.4	Visualisation du vecteur forme réduit	105
4.4	Importance de la sélection d'indicateurs (signaux d'hélicoptère)	106
4.4.1	Calcul d'indicateurs	106
4.4.2	Visualisation par ACP du vecteur forme initial	108
4.4.3	Sélection d'indicateurs par le critère de Fisher	109
4.4.4	Visualisation par ACP du vecteur forme réduit	110
4.5	Sélection d'indicateurs par couplage (AG_ADL)	111
4.5.1	Principe des algorithmes génétiques	112
4.5.2	Principe de l'analyse discriminante linéaire (critère de la trace)	113
4.5.3	Couplage des algorithmes	116
4.5.4	Conduite des différentes étapes du système de sélection	116
4.5.5	Analyse des Résultats et interprétation	123
4.5.6	Conclusion	130
4.6	Performance de l'algorithme des k-ppv	130
4.6.1	Principe de la classification par les k-ppv	130
4.6.2	Implémentation de l'algorithme	132
4.6.3	Performance de la classification sans sélection	133
4.6.4	Performance de la classification avec sélection	135
4.6.5	Conclusion	138
4.7	Performance de l'algorithme du PMC	140
4.7.1	Principe de la classification par les PMC	140
4.7.2	Implémentation de l'algorithme	141
4.7.3	Performance de la classification sans et avec sélection	150
4.7.4	Conclusion	153
4.8	Performance de l'algorithme des SVM	154
4.8.1	Principe de la classification par les séparateurs à vaste marge	154
4.8.2	Implémentation de l'algorithme	155
4.8.3	Performance de la classification sans et avec sélection	156
4.8.4	Conclusion	168
4.9	Conclusion	168

Chapitre 5 Conclusion et perspectives

5.1 Principaux résultats	169
5.1.1 Importance de la sélection d'indicateurs	169
5.1.2 Couplage AGs-ADL	170
5.1.3 Performances du classifieur k-ppv	170
5.1.4 Performances du classifieur PMC	170
5.1.5 Performances du classifieur SVM	171
5.2 Perspectives	171
5.2.1 Notion de rejet	171
5.2.2 Logique floue	171
5.2.3 Réseau de neurones dynamique	172
5.2.4 autres algorithmes d'extraction de connaissance	172
5.2.5 Autres méthodes hybrides pour la reconnaissance automatique	172
Références	173
Annexes	

Chapitre 1

Introduction

1.1 Position du problème et Motivation

Le concept de "*maintenance*" est apparu pour la première fois aux Etats-Unis d'Amérique dans l'armée américaine en 1954. Depuis, diverses techniques ou méthodologies se sont mises en place afin de mieux répondre à cette attente. Nous pouvons distinguer : maintenance curative, maintenance systématique, et maintenance conditionnelle. Dans l'annexe 1.1, est présenté le schéma synoptique des différents cas de la maintenance.

La "*maintenance conditionnelle*" permet de mettre en place la logistique de façon optimale (stock de pièces de rechange, nombre de réparateurs, organisation du service entretien). Ainsi, elle présente un investissement étalé dans le temps puisqu'elle permet avec des méthodes modernes de diagnostic technique de surveiller l'état des machines de manière prédictive et continue; et ainsi elle détermine les dates optimales de révision et d'entretien.

Le "*diagnostic*" des défauts est aujourd'hui primordial pour la fiabilité, la disponibilité et l'efficacité des systèmes de production ainsi que pour la sécurité ou la protection de l'environnement et la réduction des coûts de maintenance.

Un système de diagnostic classique de défauts se fait en trois étapes :

- (i) La détection : décider que le système est en défaut ou non à partir d'une mesure,
- (ii) la localisation : déterminer quelle partie du système est affectée en observant les symptômes de la défaillance, et
- (iii) l'identification : estimer l'ampleur et le type du défaut qui revient à évaluer les risques de pannes en discernant de ce qui est « acceptable » de ce qui ne l'est pas et alors prendre une décision.

Pour établir le diagnostic d'une machine en fonctionnement, il faut s'appuyer sur le fait qu'une machine émet vers l'extérieur de nombreuses signatures, qu'on peut mesurer, et qui sont symptomatiques de son fonctionnement, tels que : Débit, pression, puissance ou courant consommés, température des fluides qui traversent la machine, vitesse de rotation en fonction de la charge, niveau vibratoire ou bruit, analyse chimiques des effluents, etc.

Les "*vibrations*" sont aujourd'hui le moyen privilégié. Elles présentent des avantages considérables dus à la richesse des informations qu'elles véhiculent et qu'on peut mettre sous forme "*d'indicateurs*" ou descripteurs. Le changement du niveau vibratoire d'une machine intervient dans la plupart des défauts observés, contrairement à la pression, au débit, au courant et à la température, etc. (Arques,1996 p.87).

Les "*méthodes de traitement*" des vibrations sont nombreuses et variées car elles correspondent à la diversité des problèmes rencontrés. Sur un problème un peut complexe, il n'est d'ailleurs pas rare de devoir utiliser à la fois plusieurs méthodes et de plus en plus avancées. Nombreux sont les ouvrages et articles de synthèse présentant cette approche incluant des tableaux détaillés de l'analyse des défauts de machines, voir (Braun, 1986 ; Mitchell, 1993 ; Eisenmann et al., 1997 ; Boulenger et al., 2003).

"*L'approche Classique*" de la surveillance et du diagnostic des machines tournantes (MT) est donnée par le Figure 1.1. Cette approche traditionnelle consiste à prendre un certain nombre d'indicateurs aussi grands que possible et d'observer si l'évolution ou la transformation de chacun est significative de l'apparition ou de l'aggravation du défaut. Bien sur, cette approche est acceptable pour des études académiques peu nombreuses. Mais pour des applications industrielles, elle peut devenir extrêmement **impraticable** pour un nombre important d'exigences à savoir :

- la présence d'un expert pour ajuster une série de paramètres qui dépendent virtuellement du signal,
- le tri de l'information de diagrammes ou de figures qui conduit souvent à des décisions ambiguës sur les signaux vibratoires concrets,
- l'utilisation de plusieurs types d'indicateurs dans des situations plus complexes ce qui conduit à effectuer des décisions dans un espace multidimensionnel,
- la prise de décision humaine est souvent difficile devant la complexité des indicateurs calculés, de leur apparent désordre et du nombre important de défauts susceptibles d'intervenir.

La progression logique des techniques de surveillance et de diagnostic est "*l'automatisation de ce processus*" par les méthodes de Reconnaissance de Formes (Zwinglstein, 1995). Elle peut nous fournir des solutions alternatives ou complémentaires. Plusieurs approches sont alors disponibles, entre autres : l'approche paramétrique telle que la règle de Bayes ou non paramétrique telle que la règle des k plus proches voisins (Dubuisson, 1990 ; Duda et al., 2001), l'approche neuronale (RNA) (Bishop, 1995), l'approche séparateurs à vaste marge SVM (Vapnik, 2000), l'approche floue (LF) (Zadah, 1987) et une autre approche, qui a été un peu délaissée, basée sur les systèmes experts (Liu et al., 1996). Mais, face à l'accroissement rapide de la taille des bases de données, en particulier celle des signaux vibratoires, il a été nécessaire, de penser à utiliser des "*algorithmes de sélection*" d'un sous-ensemble d'indicateurs

pertinents et non redondants facilitant ainsi l'exploitation du système de reconnaissance. La recherche sur la sélection d'indicateurs a été largement effectuée, voir les références de Kudo et Abe (Kudo et al., 2000 ; Abe et al., 2006), mais peu de travaux traitent de ceci dans le domaine de la détection et du diagnostic des machines tournantes (Jack, 2000), (Jack et al., 2000a,b), (Jack et al., 2002), (Samanta, 2004a,b,c).

La motivation principale de cette thèse est d'essayer d'explorer les algorithmes avancés de traitement de signaux, des méthodes de reconnaissance de formes et de déterminer l'applicabilité de certaines techniques de diagnostic développées au cours des dernières années dans le domaine des machines tournantes. Afin de développer **une stratégie automatisée et intelligente de diagnostic et de surveillance** rassemblant des connaissances et des compétences pluridisciplinaires (souvent non disponible dans nos entreprises). Cette stratégie constituera une aide précieuse pour optimiser les interventions des techniciens de la maintenance. Dans un but :

- d'améliorer les conditions de travail de l'utilisateur par l'allègement des désagréments auditifs,
- d'augmenter la durée de vie de la machine et sa fiabilité,
- de réduire les coûts de l'entreprise par une anticipation des pannes car ces derniers coûtent cher et occasionnent, à savoir,
 - des coûts d'intervention et de réparation,
 - des coûts de non-qualité du produit, et
 - des coûts indirects (frais fixes et variables, dépenses supplémentaires pour pallier les pertes de production, etc.).

Cette étude n'est bien sûr pas exhaustive, notre but étant surtout de présenter et d'appliquer les différents outils nécessaires à un diagnostic automatique des éléments mécaniques de machines industrielles.

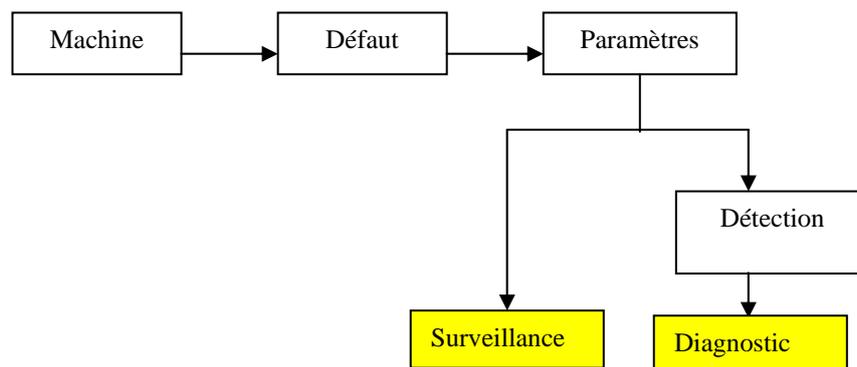


Figure 1.1 Schéma Classique de l'approche de surveillance et diagnostic

1.2 Conception d'un système automatique de reconnaissance de défauts

L'efficacité de l'approche automatique du diagnostic par RdF repose sur plusieurs étapes à savoir (Duda et al., 2001) (Dubuisson, 1990): une bonne acquisition du signal vibratoire, un meilleur prétraitement des données, l'extraction optimale de caractéristiques et enfin la reconnaissance de ces caractéristiques par les méthodes de reconnaissance de formes. Il n'est évidemment pas possible de définir une stratégie universelle, ce n'est d'ailleurs pas notre objectif. Nous résumons dans la figure 1.2 le chemin suivi dans notre cas pour l'**élaboration du système de diagnostic par reconnaissance de forme**, et nous donnons également quelques explications. Les détails des différentes étapes seront abordés au chapitre 2.

1.2.1 Signaux vibratoires

Plusieurs signaux vibratoires sont à notre disposition :

- des signaux simulés, que nous avons généré par différents modèles,
- des signaux de **courroie** d'un dispositif réalisé au département OMP,
- des signaux de **roulement** d'un dispositif réalisé, également, au département,
- des signaux d'**engrenage** du Centre CETIM, France,
- des signaux combinés d'**engrenages** et de **roulements** de INSA, Lyon,
- des signaux d'une **boîte de vitesses** de l'hélicoptère CH-46 de la marine américaine de Westland, provenant de Univ. d'état de Pennsylvanie, USA,
- des signaux de **roulements** de l'université « the Case Western Reserve university, Bearing data center », Cleveland, Ohio. USA

1.2.2 Phase d'analyse

La phase d'analyse est une étape très importante qu'il convient de mener avec méthodologie et rigueur. Son objectif est de déterminer d'une part l'espace de représentation au sein duquel sont décrites les observations sur le système à surveiller, et d'autre part de l'espace de décision correspondant aux différents modes de fonctionnement caractérisés par des classes. Elle comprend 4 sous-phases.

1.2.2.1 Pré-traitement et Traitement du signal

Cette sous-phase d'analyse est lourde en terme de calcul et nécessite toute la connaissance du système étudié pour rechercher les indicateurs adéquats en utilisant les méthodes de traitement adaptées.

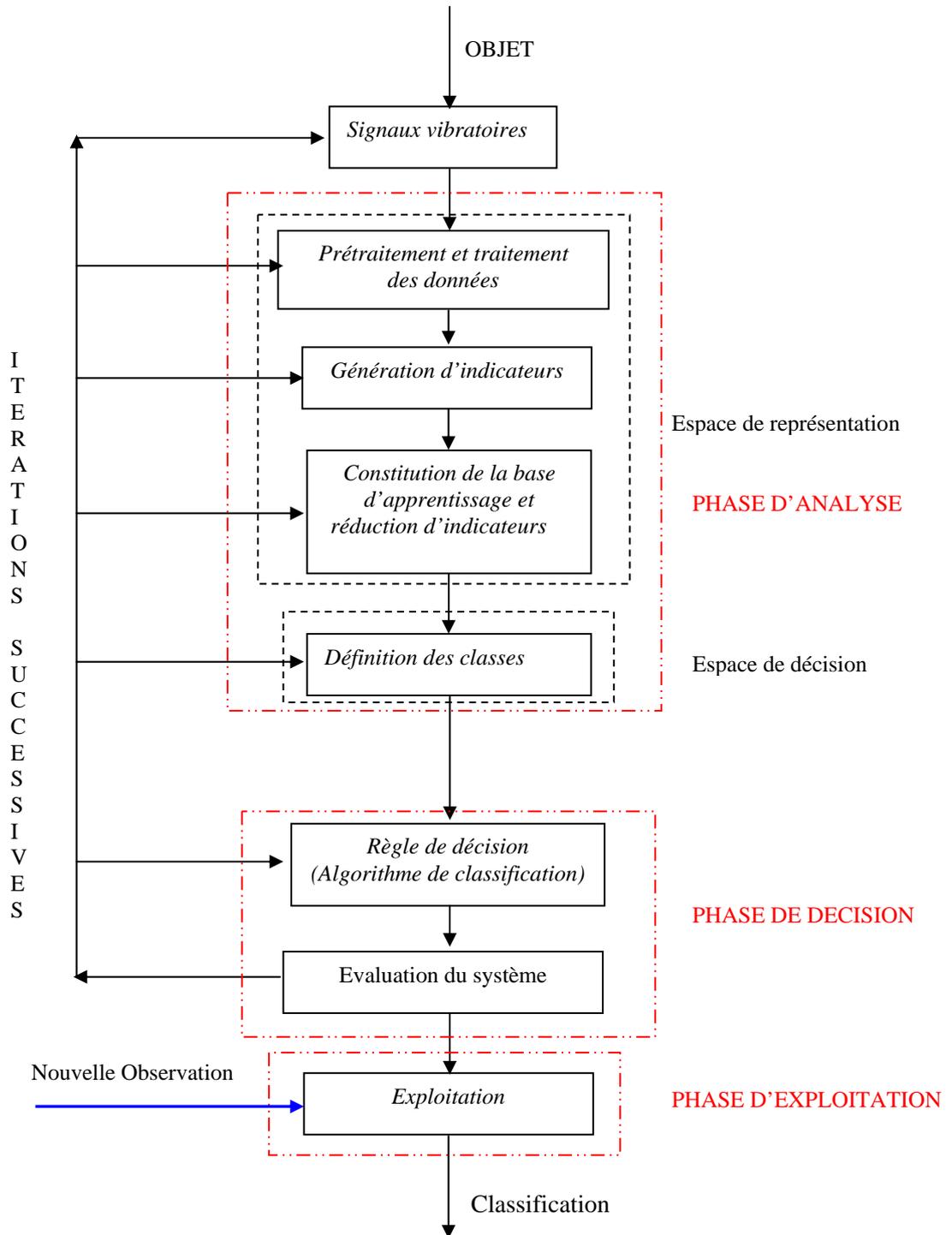


Figure 1.2 Elaboration d'un système de diagnostic automatique par RdF

Un intérêt particulier a été accordé à cette phase d'analyse pour réaliser tous les programmes sous la forme d'un menu interactif de traitement du signal : des indicateurs statistiques conventionnels, à la Transformée de Fourier Rapide (TFR), à la densité spectrale de puissance (DSP), à la Transformée en Ondelettes (TO), aux Statistiques d'ordre supérieurs (SOS) ou aux méthodes spécifiques comme la corrélation spectrale, le kurtosis spectral, etc. Nous ne présenterons pas ce menu dans cette thèse.

Comme les données issues de ce traitement du signal ne sont pas toutes utilisées directement. Le choix de l'espace de représentation est essentiel, car le traitement du signal peut engendrer des représentations non utilisables directement par les méthodes de décision comme le cepstre, les DSP et les représentations temps-échelle, les SOS, etc. Le traitement du signal doit être suivi par une paramétrisation qui consiste à trouver une représentation **parcimonieuse** mais informative du processus, c'est l'extraction d'indicateurs.

1.2.2.2 Extraction ou Génération d'indicateurs

L'indicateur est un résultat de mesure ou de calcul représentant un ou plusieurs aspects de l'état ou de la performance d'un équipement et dont l'évolution ou la transformation est significative de **l'apparition ou de l'aggravation** d'un défaut. L'utilisation d'indicateurs est la base de la surveillance et du diagnostic vibratoire. Nous avons choisi de calculer nos indicateurs sur des bandes étroites d'énergie car les indicateurs large bande présentent plusieurs inconvénients majeurs qui de fait limitent considérablement leur fiabilité d'utilisation et leur capacité de détection de défauts à un stade précoce, ils sont également entachés de l'effet de masque (Boulenger et al., 2003). Nous avons choisi d'utiliser des indicateurs énergétiques et de comportement : énergie dans les bandes spectrales, paramètres AR, valeur spectrale, valeurs maximales ou pics dans les spectres, valeur scalaire temporelles, etc.

1.2.2.3 Constitution de la base d'apprentissage (l'espace de décision)

Il s'agit de l'ensemble de toutes les classes possibles. Si on connaît la classe d'origine de chaque observation de l'ensemble d'apprentissage, l'espace de décision est parfaitement connu et l'**apprentissage** des classes peut être fait en mode dit **supervisé** par des **méthodes dites d'affectation**, voir figure 2.15. Dans le cas contraire, l'apprentissage des classes est fait en mode **non supervisé** par des méthodes de classification automatique appelées **méthodes de clustering**. Ces techniques passent par deux étapes : une étape de partitionnement automatique de l'ensemble des formes en classes et une étape de validation de ces classes soit par un expert, soit par des critères de validité de la partition. Il existe de nombreuses méthodes de classification non supervisée nommées méthodes de coalescence, voir par exemple figure 2.15 : les méthodes hiérarchiques telles que les méthodes des nuées dynamiques dont l'algorithme des

Centres Mobiles ou la carte de cohérence. Le principe général de ces méthodes est de détecter dans l'ensemble d'apprentissage une structure de classes disjointes au sens d'un critère de similarité, de telle façon que les observations d'une même classe soient semblables et que celles de classes différentes soient dissemblables.

Nous développerons plus loin les différentes méthodes d'affectation utilisées dans notre travail en soulignant certains de leurs avantages et inconvénients. Nous n'aborderons pas ici les méthodes clustering, ce n'est pas notre objectif, le lecteur intéressé pourra consulter les références suivantes (Duda et al, 2001), (Théodoridis et al., 2003).et (Fukunaga K., 1990).

1.2.2.4 Réduction de l'espace de représentation

Les données issues de la sous phase du traitement du signal ne sont pas toutes utilisées. Le nombre souvent élevé d'indicateurs :

- peut contenir des informations redondantes, et correspondre à du bruit
- peut ne pas être forcément pertinents vis à vis des modes de fonctionnement étudiés,
- peut être pénalisant en terme de temps de calcul, etc.

La réduction de la dimension de l'espace de représentation, par l'une des approches de la section 2.6.2, est donc essentielle dans la reconnaissance de formes. La dimension doit être faible devant la taille de la base d'apprentissage afin que les surfaces de séparation entre classes soient plus précises, voir figure 2.16. Faute de quoi les performances du classifieur sur les données non apprises (**test**) se détériorent (problème de généralisation). Pour toutes ces raisons et pour ne garder que les paramètres les plus pertinents, nous avons proposé un couplage de l'analyse discriminante linéaire (ADL) et des algorithmes génétiques (AGs).

L'analyse en composante principale (ACP) a été utilisée pour la projection du vecteur forme sélectionné sur un repère de dimension 3 dans le but de visualiser la séparabilité des classes et de juger, ainsi, de la performance des différentes méthodes de sélection.

1.2.3 Phase de décision (classification)

Les problèmes de classification ont été traités à travers plusieurs ouvrages dont : Duda et al., 2001) (Dubuisson, 1990) (Théodoridis et al., 2003).et (Fukunaga K., 1990).

La classification est l'élaboration d'une *règle de décision* qui transforme les indicateurs caractérisant les formes en appartenance à une classe (passage de l'espace de codage vers l'espace de décision). Mais, avant qu'un modèle de décision ne soit intégré dans un système de reconnaissance de défaut, il faut avoir procédé auparavant aux deux étapes : l'étape d'*apprentissage* et l'étape de *test*.

Ces étapes d'apprentissage et de test sont réalisées préalablement à l'intégration du module de décision dans le système de reconnaissance. Dans tous les cas, on peut permettre au système de reconnaissance **d'itérer** les étapes d'apprentissage et de test tant que l'on n'a pas atteint les performances désirées. Le calcul de cette *performance* est le résultat du classifieur utilisé.

1.2.3.1 Règle de décision

Pour construire un classifieur inductif, il existe deux approches : l'approche statistique et l'approche analytique. Dans l'approche analytique, une 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 d'une mauvaise classification. Le choix de la fonction mathématique dépend de la complexité de la frontière de décision. En pratique, on commence par des fonctions linéaires puis si la probabilité de mauvaise classification n'est pas satisfaite, on passe à des fonctions mathématiques plus complexes comme par exemple des fonctions quadratiques.

Dans notre cas, nous ne considérons que les méthodes d'affectation (classification supervisée, voir figure 2.15) suivantes : les réseaux de neurones artificiels (RNA), les Séparateurs à Vaste Marge (SVM), l'approche statistique des k-plus proches voisins.

1.2.3.2 Evaluation des performances

Quelle que soit l'approche utilisée pour la mise en place du classifieur, il est primordial de justifier sa robustesse.

Pour obtenir des valeurs significatives, on pratique généralement la moyenne des performances sur un nombre statistiquement pertinent d'apprentissage, avec à chaque fois, une Base d'apprentissage (BA) et une Base de test (BT) différente. L'ensemble d'apprentissage est utilisé pour déterminer les paramètres du modèle de classification, par exemple les poids dans le cas d'un réseau de neurones, le nombre k pour la méthode des k plus proches voisins et le coefficient C dans le cas de la méthode des séparateurs à vaste marge. L'ensemble de test sert à tester les performances de la méthode en calculant le taux de classification correcte de l'ensemble des cas. Ce taux est déterminé en divisant le nombre de cas bien classés sur le nombre des cas testés.

Le taux observations bien classés (T_B):
$$T_B = \frac{N_{Bc}}{N} \times 100$$

Avec N_{Bc} : Nombre de bien classées

N : Nombre total d'échantillons (d'observations)

Cette quantité à beaucoup d'importance. Si le taux d'erreur est proche de 0 %, alors on peut estimer que non seulement, "en amont", le vecteur forme caractérise convenablement

l'ensemble d'apprentissage, mais aussi que la règle de décision est bien adaptée. Parallèlement, il est souhaitable d'avoir une valeur très faible de la probabilité de mauvaise classification lors de la phase d'apprentissage afin d'avoir un minimum d'erreur de diagnostic concernant les nouvelles observations. Le système de décision n'a pas à "apprendre par coeur" la base de données mais plutôt à savoir bien classer de nouvelles observations non apprises.

La méthode de division de la base de données en 2 sous-ensembles requiert un nombre d'observations important, car en réservant des échantillons de la base de données à la phase de test, la représentation des classes que reflète la base d'apprentissage risque de se dégrader ainsi que les performances en classification. De plus, on ne dispose pas toujours d'une base de données de taille suffisante et on a toujours recours à des méthodes de **rééchantillonnage** (resampling techniques) telles que :

La validation croisée avec ses variantes : leave one out cross-validation, K fold cross-validation, et stratified cross-validation (Dubuisson, 2001).

- le *bootstrap* (Duda et al., 2001)
- le *jackknife* (Duda et al., 2001)

Nous donnons ci-dessous le principe de la méthode de *leave one out cross-validation* (Weiss et al., 1991) :

Soit N le nombre d'observations (X_1, X_2, \dots, X_N)

Notons X_{-i} le sous-ensemble $(X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_N)$, obtenu à partir de l'ensemble initial X en supprimant la $i^{\text{ème}}$ observation, cela revient à considérer que l'on fait un sondage dans l'échantillon base en tirant $N-1$ observations sans remise. Nous avons remarqué que plusieurs autres variantes de cette méthode sont parfois utilisées. Breiman dan (Breiman, 1996) a montré que le *K fold cross-validation* donne de bons résultats.

Pour plus de détails concernant les techniques de rééchantillonnages le lecteur peut se référer à (Weiss et al., 1991) et (Kohavi, 1995).

Cette étape de test permet d'évaluer la performance du classifieur pour un apprentissage donné. C'est une étape importante car elle peut mettre en cause le choix des indicateurs ou le choix de la méthode d'apprentissage (règle de décision). En effet, il est difficile de trouver *a priori* les indicateurs pertinents et la méthode d'apprentissage la plus adaptée au problème posé d'où l'utilité de procéder par **itérations successives**. Ces itérations consistent à extraire des indicateurs jugés utiles au problème de reconnaissance à résoudre et de tester la performance du système avec cet ensemble d'indicateurs.

1.2.4 Phase d'exploitation

Dans cette phase, le système de diagnostic par RdF peut être mis en service. Il 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 de décision. La détermination de cette classe permet de connaître le mode de fonctionnement actuel du système.

1.3 Plan de la thèse

Il s'articule autour de cinq chapitres.

Après avoir posé le problème et établi un bref aperçu dans le premier chapitre introduction.

Le chapitre 2 présente des préliminaires utiles pour la suite du travail. Il est divisé en 6 sections. Dans la section 2.2, nous proposons un recensement non exhaustif des défauts dans les machines tournantes. La section 2.3 détaille la procédure *d'acquisition du signal vibratoire*. La section 2.4 donne un aperçu général sur les méthodes avancées du traitement du signal utilisées en diagnostic vibratoire. La section 2.5 est consacrée à la présentation des procédures de génération d'indicateurs à partir des représentations temporels, fréquentielles et temps fréquences, etc. La section 2.6 traite les méthodes de sélection et d'extraction d'indicateurs. Elle consiste à choisir les indicateurs les plus pertinents et non redondants pour la classification. La dernière section 2.7 expose un panorama général des règles de décision ou méthodes de classification.

Le chapitre 3 fournit une synthèse des sources de données (bancs) sélectionnées pour être utilisées dans nos expériences. Il est divisé en 6 sections. La section 3.2 expose des statistiques sur les défauts de machines tournantes. Les sections de 3.3 à 3.5 sont respectivement consacrées à la présentation des bancs d'essais d'engrenage, de roulement et de la boîte à vitesse d'un hélicoptère. Les sections 3.6 et 3.7 présentent, d'une manière générale, respectivement les signaux de roulement et d'engrenage, leurs traitements et les différentes allures temporelles et spectrales.

Le quatrième chapitre présente les applications expérimentales du système de diagnostic. Dans la première section 4.2, nous proposons de donner le principe du diagnostic par reconnaissance de forme. Les sections 4.3 et 4.4 exposent l'importance de la sélection d'indicateurs. Dans la quatrième section 4.5, nous proposons une nouvelle méthode de sélection d'indicateurs basée sur une hybridation des Algorithmes génétiques (AGs) et de l'Analyse Discriminante Linéaire. La cinquième section 4.6 est consacrée à l'application de l'algorithme des k plus proches voisins (k -ppv). La sixième section 4.7 est dédiée à l'application des réseaux de neurones et principalement l'algorithme du

perceptron multicouches (PMC). La dernière section 4.8 de ce chapitre traite l'application de la récente méthode des séparateurs à vaste marge (SVM).

- La règle des k – plus proches voisins qui est **une approche statistique** non paramétrique dans la mesure où elle réalise une estimation des lois de probabilité régissant les classes.
- Les algorithmes des PMC et du SVM, qui déterminent directement des frontières entre les classes, visent à calculer les coefficients d'une fonction mathématique de manière à séparer au mieux les modes de fonctionnement, i.e. à minimiser le taux d'une mauvaise classification. Cette **approche** est donc **analytique**

Pour chaque phase du système de diagnostic, nous avons voulu revenir sur les différents travaux en présentant respectivement les acquis et les perspectives et notre contribution afin de situer notre travail. Chaque chapitre contient une introduction et une conclusion.

Finalement, nous concluons en synthétisant les différentes contributions et en discutant des perspectives envisagées pour poursuivre cette recherche.

1.4 Travaux réalisés

Les travaux suivants ont été faits dans le cadre de cette thèse

➤ Publications

A. **Felkaoui**, R. E. Bekka, J. Antoni, M. Sidahmed, S.Fedala (2007)
Bearing Fault Diagnosis of an Induction Motor: Feature Selection Importance in Pattern Recognition Analysis, International Review of Electrical Engineering (I.R.E.E.), Vol. 2, n.4, pp.587-592, July-August 2007.

A. **Felkaoui**, R. E. Bekka, J. Antoni, M. Sidahmed
Etude des performances de la sélection d'indicateurs par couplage algorithmes génétiques – analyse discriminante linéaire : Application au diagnostic des défauts d'engrenage. Accepté dans le Journal Européen des systèmes automatisés (J.E.S.A.).

➤ Communications internationales

Mahgoun H, **Felkaoui A.** & Bekka R-E- (1999). *Application de la transformée en ondelettes à la détection des chocs et impulsions*. SSA2'99 colloque international sur les Signaux, Systèmes et Automatique, vol.1 168-173, BLIDA 10-12 mai 99.

Chiter A., **Felkaoui A.** & Bekka R-E- (1999). *Détection et Diagnostic Automatique des défauts de roulement par le test de wald*. Conférence Maghrébine en Génie Electrique, CMGE'99, pp. 286-293 Constantine 4-6 Déc. 99.

Haloui N., Chikouche Dj. & **Felkaoui A.** (2001) *Application de la méthode spectrale AR et la méthode de Prony à la détection des défauts d'engrenages : comparaison de performances*. Colloque International sur la maintenance Assistée par ordinateur : Outils,

méthodes et technologies " MAO2001", ENIM, RABAT (MAROC), 7, 8 et 9 Novembre 2001, pp. A18/1-A18/7.

Chiter A., **A Felkaoui** & al. (2004). *Efficacité des indicateurs de surveillance dans les problèmes de la détection et du diagnostic des défauts dans les machines tournantes : Mesures en mode spécifique* Proceeding 1èr Cong. Inter. Sur le Génie Electrique, CIGE'04, 10-12 Octobre 2004, SETIF,

Fedala S., **Felkaoui A.**, & Zegadi R.(2006). *Diagnostic vibratoire automatisé : extraction et sélection des paramètres du vecteur forme*. 3ISHC'2006, 3rd Int. Symp. On Hydrocarbons & chemistry, 27-29 Mars 2006, Ghardaïa. Algérie.

Felkaoui A., Bekka R.E., Sidahmed M.(2006) *Etude des performances de la sélection d'indicateurs par couplage Algorithmes génétiques (AG)-Analyse discriminante linéaire (ADL) : Application au diagnostic des défauts d'engrenage*, ICMM06, Sétif 04-06 Novembre 2006

Ziani R., **Felkaoui A.**, Zegadi R., Fedala S. (2006) *Diagnostic des défauts d'un réducteur à engrenage par les séparateurs à vaste marge (SVM)*, ICMM06, Sétif 04-06 Novembre 2006

MahgounH. , **Felkaoui A.**, Bekka. R.E. (2006). *La prévention des défauts d'engrenages par les paquets d'ondelettes*. ICMM06, Sétif 04-06 Novembre 2006

Chikouche D., Haloui N., Bekka, R.E. **Felkaoui A.** (2006) *Diagnostic précoce d'un réducteur à engrenage par le Cepstre Synchrone*, ICMM06, Sétif 04-06 Novembre 2006

Fedala, S. **Felkaoui A.**, Zegadi R. et Ziani R. (2006) *Optimisation des paramètres du vecteur forme : Application au Diagnostic vibratoire automatisé des défauts d'une boite de vitesse d'un hélicoptère*, ICMM06, Sétif 04-06 Novembre 2006

Fennineche H, Felkaoui A, Zegadi.R (2006) *Application des réseaux de neurones artificiels pour l'automatisation du diagnostic vibratoire des engrenages*, ICMM06, Sétif 04-06 Novembre 2006

Felkaoui A., S. Fedala et R. Zegadi, (2006) *Diagnostic vibratoire automatisé des défauts de roulement d'un moteur asynchrone : Sélection des paramètres du vecteur forme*. CEE'06, 4th International Conference on Electrical Engineering, Batna 07 - 08 November 2006

Ziani R., **Felkaoui A.**, Zegadi R. (2007) *Application des séparateurs à vaste marge (SVM) au diagnostic vibratoire des engrenages*. Accepté au CIP2007 (Conf . Inter. Sur la Productique). Sétif 03-04 Nov. 2007.

Djouada M., Ziani R., **Felkaoui A.**, Zegadi (2007) *Diagnostic des défauts par un Couplage réseaux de neurones artificiels - algorithmes génétiques*. Accepté au CIP2007 (Conf . Inter. Sur la Productique). Sétif 03-04 Nov. 2007.

S. Fedala, **Felkaoui A.**, et R. Zegadi (2007) *diagnostic vibratoire automatisé par reconnaissance de forme des défauts d'une boite de vitesse d'un hélicoptère*. MOAD'07 Béjaia 2007

➤ **Communications nationales**

Felkaoui A. (1997), *Analyse des signaux vibratoires : Une méthode moderne d'analyse spectrale*, à 6^{ème} Séminaire National de Mécanique de M'sila. M'SILA, 16-18 Fév. 1997.

Mahgoun H, **Felkaoui A.** & Bekka R-E- (1999) *Analyse des défauts d'engrenages par la transformée en ondelettes*. 2^{ème} Séminaire National en génie Electrique, SNGE'99, pp. 145-148, Biskra, 22-24 Nov. 99.

Bekka R.E. **Felkaoui A.** Mahgoun H. (2002) *La détection précoce des défauts d'engrenage par les ondelettes orthogonales*. 2^{ème} conférence sur le Génie Electrique, Organisée par Ecole Militaire Polytechnique, CGE'02, Bordj El Bahri, Alger 17-18 Décembre 2002

Chikouche D., Bekka R.E. Haloui N., & **Felkaoui A.** (2004), *Application des techniques spectrales autorégressives à l'analyse des signaux d'engrenages*. 3^{ème} conf. Sur le génie Electrique CGE'03, Bordj El Bahri, 15-16 février 2004.

Mahgoun H, **Felkaoui A.** & Bekka R-E- (2007) *Apport du scalogramme de phase au diagnostic des engrenages*, séminaire sur les techniques et le management de la maintenance, SIMM 07 07-08 mais 2007 EMP Alger

➤ **Co –encadrement de thèses de Magister**

1. MAHGOUNE Hafida (D.O.M.P. Nov. 2000)
Analyse et caractérisation des signaux vibratoires cyclostationnaires : application des ondelettes

2. CHITER Amar (D.O.M.P. Juin 2001)
Thème : Détection et Diagnostic des défauts de roulements : contribution à la maintenance conditionnelle.

3. M. FEDALA Samechedine (nov.2005).
Le diagnostic vibratoire automatisé : comparaison des méthodes d'extraction et de sélection du vecteur de forme. (Encadré par le chef du Labo. MPA Pr. Zegadi et M. FELKAOU I A.).

4. R. ZIANI (D.O.M.P. Sept. 2007)
Etude des performances de la classification par les séparateurs à Vaste Marge (SVM) : Application au diagnostic vibratoire.

5. M. FENINECHE Hocine (en cours),
Application des Réseaux de Neurones Artificiels à la Détection et au Diagnostic des défauts dans les machines tournantes. (Encadré par le chef du Labo. MPA Pr. Zegadi et M.FELKAOU I A.).

6. M. HARROUCHE Fateh (en cours)
Application de la Logique Floue au Diagnostic des défauts des machines Tournantes. (Encadré par le chef du Labo. MPA Pr. Zegadi et M.FELKAOU I A.).

7. M. DJOUADA. (en cours).

Etude des performances de la classification par couplage Réseaux de Neurones Artificiels (RNA) – Algorithmes génétiques (AG) : Application au diagnostic vibratoire.

➤ **Projet de recherche**

1. Mise en place d'un système expert pour le suivi vibratoire au service de la Maintenance (D. O.M.P., U. de Sétif , code J 1901/03/01/96).
2. Surveillance et diagnostic des défauts de machines tournantes et de structures (D. O.M.P., U. de Sétif Janv.2002 code : J 1901/03/02/02).
3. Application des techniques de séparation aveugle pour l'automatisation du diagnostic des défauts mécaniques dans les machines tournantes. (D. O.M.P., U. de Sétif Janv.05 code : J 1901/03/53/06).

Chapitre 2

Préliminaires

2.1 Introduction

Les vibrations peuvent être d'origine mécanique, aérodynamique ou magnétique. Le choix des techniques de séparation adaptées à chaque type de machine permet généralement de reconnaître l'origine de la vibration. L'étude des vibrations nécessite le choix d'un capteur approprié, d'une chaîne de mesure de qualité et d'un logiciel d'analyse et de traitement pour tirer des indicateurs performants et fiables à la décision.

Dans ce chapitre, nous donnons en premier lieu un aperçu sur les différents défauts dans les machines tournantes et sur l'acquisition du signal vibratoire. Ceux-ci font introduire le choix de la grandeur à mesurer, du capteur et des différentes dispositions à prendre pour éviter les premières erreurs d'acquisition et de mise en forme du signal. En deuxième lieu, nous proposons de donner une synthèse bibliographique des méthodes d'analyse du signal vibratoire.

Comme l'information donnée par les méthodes du traitement du signal ne peut être utilisée directement, la section 2.5 expose les différentes techniques de génération d'indicateurs et la section 2.6 les approches utilisées pour la réduction de la dimension (l'espace de représentation). Dans La section 2.7, nous revenons sur la constitution des classes (l'espace de décision) et l'élaboration des règles de décision. Nous y présentons les deux formes d'apprentissage supervisé et non supervisé et nous discutons de différentes techniques de classification.

C'est dans le prochain chapitre 3 que nous insisterons plus particulièrement sur les méthodes de diagnostic d'engrenages et de roulement avec leurs limitations et leur potentiel.

2.2 Défauts dans les machines tournantes

En général, les vibrations d'origine mécanique sont causées, par exemple, à (Braun, 1986), (Bourgain et al., 1988) :

- un déséquilibre de pièces rotatives (de forces, de couples, en porte à faux, excentrique, etc.),
- une flexion de l'arbre ou cassure de celui-ci,
- un mauvais alignement d'accouplement (angulaire, parallèle, etc.),
- un engrenage usé, endommagé, excentrique, etc.,
- un jeu entre dents et déversement des roues d'engrenage (angulaire),

- une dent fissurée, cassée ou erreurs de fabrication,
- une mauvaise courroie ou chaîne de transmission (courroie usée, mal ajustée ou résonance, tension incorrecte, déversement angulaire ou axiale des poulies, excentricité des poulies, crevasses, points durs, points mous, bosses ou morceaux arrachés sur la courroies),
- un mauvais serrage,
- une semelle douce, élastique et résonance de celle-ci,
- un frottement du rotor et perte d'ajustement,
- un palier lisse défectueux (usure, jeu, instabilité tourbillonnaire des huiles),
- un roulement défectueux (bagues interne et externe, bille, cage),
- une résonance d'élément. etc.

Nous soulignons que les niveaux de vibration de certains éléments (engrenages, poulies, arbres, ...) peuvent être supérieurs aux niveaux de vibration d'autres éléments (roulements, désalignement d'arbres de poulies, etc.), ce qui masquent ces derniers et ne se manifestent pas ainsi dans les courbes de suivi (ou de tendance) de l'accélération, de la vitesse et du déplacement (Boulenger et al., 2003).

Les vibrations d'origine hydrodynamiques et aérodynamique sont dues, par exemple, à :

- des forces hydrodynamiques et aérodynamiques (passage de lame d'ailettes),
- un écoulement turbulent
- une cavitation d'une pompe, etc.

Les vibrations d'origine électrique et magnétique sont dues, par exemple, à (Bachir 2002), (Casimir, 2003) :

- pour un à moteur à induction (une excentricité du stator ou du rotor, un mal ajustement du bobinage statorique, une rupture de la barre rotorique, etc.),
- pour un moteur à courant continu (une armature de ventilation cassée, un système de réglage ou d'allumage défectueux, une carte de comparateur défectueuse, un passage d'un courant électrique à travers les enroulements, etc.)
- une irrégularité du couple moteur,
- des forces électromagnétiques, etc.

Cette liste n'est bien sûr pas exhaustive, nous y avons recensé le maximum de défauts d'autres peuvent y être rajouter.

2.3 Acquisition vibratoire

Pendant longtemps, l'acquisition représentée par la chaîne de mesure de la Figure 2.1, le traitement du signal représenté par les différents algorithmes de calcul fut l'apanage des électroniciens et des automaticiens. Aujourd'hui la donnée a radicalement changée avec la prolifération des moyens de traitement de l'information à bas prix qui gagne toutes les

disciplines, si bien qu'il est devenu inconcevable pour un *ingénieur*, lié de près ou de loin à des essais ou à des mesures, de ne pas savoir interpréter les outils de base de l'acquisition et du traitement du signal. Plutôt que d'être le spectateur de cette évolution, il devrait en être l'acteur.

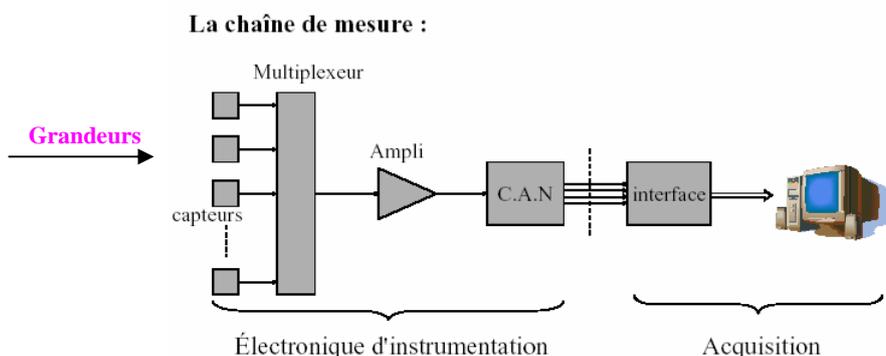


Figure 2.1 Exemple d'une chaîne de mesure

L'acquisition des données comme une étape du processus de surveillance et de diagnostic implique la sélection de la grandeur à mesurer, le nombre et le type de capteurs à utiliser, les dispositions à prendre, et le choix du système Conditionnement /Numérisation du signal.

2.3.1 Grandeur à mesurer

Pour s'affranchir des défauts des instruments, on aura intérêt à utiliser la grandeur qui représente le mieux ce que l'on voit. La plage de fréquences dans laquelle le défaut recherché est susceptible de s'exprimer, voir Figure 2.2, détermine la grandeur à mesurer.

Nous avons généralement trois modes de mesure des vibrations (Thomas, 2003) :

- La mesure en *mode déplacement* n'est généralement utilisée que pour mettre en évidence des phénomènes de basses fréquences, <10Hz, et dont les composantes prépondérantes se situent entre 1 et 2 fois la fréquence de rotation tels que : déséquilibre, déformation, désalignement, desserrage, etc. ou voire en dessous de la fréquence de rotation tels que : instabilité, frottement, défaut de courroie, etc.
On l'utilise également pour la mesure des déplacements absolus des arbres ou des déplacements relatifs arbres-coussinets. La mesure de l'amplitude crête ou crête à crête pour le déplacement permet de mieux estimer l'ampleur du mouvement vibratoire. Par contre la mesure de l'amplitude efficace pour la vitesse ou l'accélération permet de mieux apprécier l'énergie dépensée,
- La mesure en mode vitesse permet d'observer correctement des phénomènes dont la fréquence n'est pas trop élevée c'est-à-dire ne dépassant pas le 1000Hz et couvrant une très large plage de fréquences tels que : passage des pâles, défauts d'engrènement, défauts de roulement dans le cas d'écaillage localisé, etc. et bien

évidemment les défauts traditionnels des lignes d'arbres tels que : le déséquilibre, le désalignement, la déformation, l'instabilité, le frottement. Cette mesure en mode vitesse est utilisée par les normes pour l'évaluation de la sévérité ou de l'intensité vibratoire en utilisant la valeur efficace du signal qui est proportionnelle à l'énergie dissipée (voir la norme AFNOR E90-300).

- La mesure en mode accélération permet de mettre en évidence des phénomènes dont les fréquences sont élevées, >1000Hz, tels que : le défauts d'engrènement sur un multiplicateur, le passage d'encoches sur un moteur, ou des défauts qui génèrent des signaux impulsionnels de courte durée riche en composantes hautes fréquences tels que : l'écaillage de roulement, les jeux, les cavitations, etc.

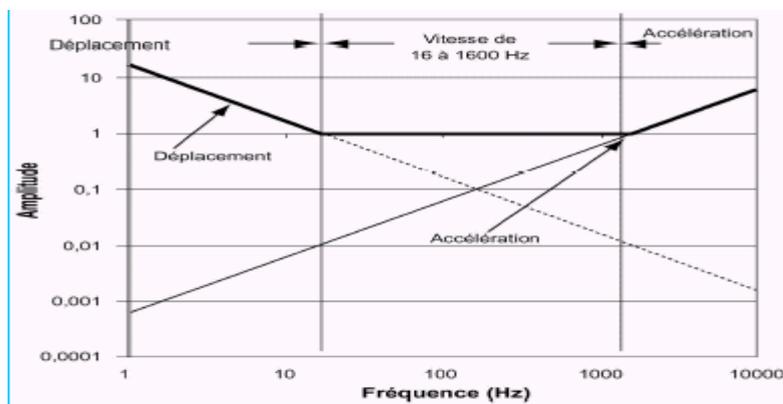


Figure 2.2 Choix du descripteur ou de la grandeur à mesurer (extraite de Thomas, 2003)

Notre choix s'est porté sur la **grandeur accélération** car nous voulons tester les roulements et les engrenages caractérisés par des efforts à moyennes et hautes fréquences. Les plages de fréquences requises, pour détecter les défauts, sont résumées sur le tableau 2.1.

Tableau 2.1 Plages de fréquences des défauts

Plage de fréquences	faible	moyenne	haute
Défaut	Déséquilibre, arbres non alignés, flexion d'arbre, tourbillon d'huile, excentricité	Usure, défaut d'engrenage.	Défaut de roulement

Dans les paragraphes suivants sont décrites brièvement les composantes principales des chaînes d'acquisition utilisées en mécanique

2.3.2 Capteur et dispositions

Le choix de l'accéléromètre a été guidé par l'étendue de sa plage de fréquence et sa robustesse, voir annexe 2.1. Un accéléromètre peut mesurer aussi bien l'amplitude de l'accélération que, par intégrations successives, celle de la vitesse et du déplacement. Notons qu'inversement, les proximètres et les vélocimétries ne permettent pas d'obtenir une accélération car l'opération de dérivation est génératrice d'un bruit de fond non négligeable. Bien utilisé, l'accéléromètre permet d'améliorer les mesures issues d'autres techniques, voire impossibles à obtenir différemment. Cependant, il est important de considérer différents critères avant d'entreprendre le réglage d'un accéléromètre.

La connaissance des caractéristiques du capteur permet de faire un choix judicieux en fonction de l'utilisation que l'on compte en faire. Ainsi, il est important de noter qu'il ne peut être question de traiter tous les problèmes afférents à l'utilisation des capteurs. Il est donc nécessaire de se limiter à un certain nombre de dispositions telles que (Morel, 1992) (Bourgain et al., 1988) (Norme AFNOR, E90-152) :

Orientation : on constate assez fréquemment que les niveaux de vibrations sont plus élevés en horizontal qu'en vertical sur les machines car les raideurs sont plus faibles.

Support : il faut faire des mesures sur des points significatifs et des éléments rigides, massifs ou plaques de base, plutôt que sur des tôles des capotages pouvant présenter des résonances locales. Nos mesures ont été effectuées sur les paliers.

Fixation du capteur : On utilise la fixation par goujon pour négliger les déformations et les modes propres des liaisons. Les surfaces doivent être propres et polies avec des axes de fixations normaux aux surfaces et film d'huile léger ou une graisse entre les surfaces. Pour un bon contact il faut utiliser une clé dynamométrique avec un serrage de 0.6 m kg. Ce type de fixation par goujon donne une réponse linéaire sur une assez large gamme de fréquences. L'expérience montre que la flexibilité ajoutée par une mauvaise fixation filtre les hautes fréquences. Nous avons une plage de fréquence jusqu'à :

- 100 Hz dans le cas d'un adhésif double face ;
- 1 kHz dans le cas d'un collage à l'araldite ;
- 10 kHz dans le cas où l'accéléromètre est vissé avec un goujon rapporté ;
- 100 kHz si l'accéléromètre est vissé avec un goujon solidaire.

Fixation des câbles de liaison : Tous les câbles de liaison génèrent des contraintes mécaniques qui entraînent des parasites. Afin de limiter ces contraintes, il est souhaitable pour les mesures de précision ou pour les installations permanentes de fixer les câbles.

Etendue de mesure : L'intervalle de variation de la grandeur d'entrée où la mesure fournie par le capteur est correcte de 0 à 10g.

Bande passante, linéarisation et précision : la gamme de fréquence du capteur pour laquelle la réponse présente une erreur inférieure à un seuil déterminé par rapport à la

valeur réelle de 1 Hz à 12.5 kHz (dans notre cas de l'acquisition des bancs réalisés au département)

Sensibilité : c'est le rapport entre les variations des grandeurs de sortie et d'entrée dans leurs unités respectives. Pour les capteurs KD35 du laboratoire, elle est autour de 5mV/ms^{-2} pour la tension et 6pC/ms^{-2} pour la charge.

Sensibilité transverse : on pourra observer une sensibilité de 5 à 10 % dans les autres directions de mesure.

Tenue en température : au-delà d'une certaine température qui est fonction des matériaux constituant le capteur, la température peut faire chuter rapidement la précision de la mesure et même endommager définitivement le capteur

Impédance de sortie : l'accéléromètre a une très haute impédance de un Méga Ohms.

Vérification de la mise à la masse : La mise à la masse contribue à rendre le système très silencieux ou très bruyant. Si vous ne prenez pas cette précaution, il se formera une boucle de masse qui produira des sources externes de bruit. Dans le cas d'un accéléromètre isolé, reliez l'entrée de l'amplificateur à la terre ; sinon, cela peut provoquer un signal fortement bruité ou une rupture de signal du fait que l'entrée de l'amplificateur flottante peut se trouver en dehors des limites de réglage. En bref, le système doit toujours être mis à la masse, mais en un seul et unique point.

2.3.3 Multiplexage

Un multiplexeur est un circuit logique à "*n*" **entrées dites de données** (capteurs) et **une sortie** qui transmet les informations parvenant à l'entrée sélectionnée.

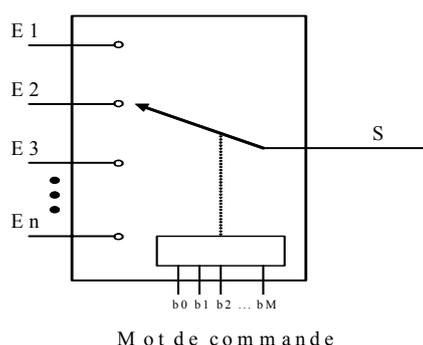


Figure 2.3 *Multiplexeur*

Pour sélectionner une entrée (un capteur), le multiplexeur doit recevoir un ordre qui provient de *M* entrées complémentaires appelées entrée d'adresse : ces entrées aiguillent les données d'entrée choisies vers la sortie.

Position du multiplexeur

Nous pouvons distinguer deux sortes de multiplexage (Schenck, 1994, pp 122-123) :

Le multiplexage analogique

Dans ce cas, typiquement exploité avec n capteurs de même catégorie, le *multiplexeur* est placé en sortie des capteurs et l'on n'aura besoin que d'un seul Convertisseur – CAN- mais un ou plusieurs échantillonneur bloqueur -E/B-. Cependant il faut noter que la transmission d'un signal de faible amplitude n'est possible que sur de courtes distances, on supposera donc que l'ensemble des capteurs sont proches les uns des autres à la fois en terme de position géographique que d'amplitude de sortie et que le conditionneur l'est également. Dans le cas contraire on utilisera avantageusement un système décentralisé, plus onéreux mais plus fiable.

Le multiplexage digital

On profitera le plus souvent des progrès de l'intégration sur silicium de l'ensemble de la chaîne de conditionnement (y compris éventuellement le processeur digital) pour réaliser des systèmes performants. Cette structure permet évidemment l'emploi de capteurs de nature fort différente dans la mesure où l'on dispose de modules de conditionnement adaptés et assurant en sortie (c'est à dire sur chaque entrée du multiplexeur) des signaux compatibles. Nous avons dans ce cas plusieurs E/B et plusieurs CAN. *Le multiplexeur est placé en sortie des CAN.*

Pour plus de détails, voir les références suivantes (techniques de l'ingénieur, etc.).

2.3.4 Conditionnement du signal

Les **conditionneurs**, premier élément de la chaîne d'acquisition après le capteur, ils ont un rôle très important sur la qualité de la mesure. Le module de conditionnement tient un rôle essentiel dans la chaîne de mesure. Les conditionneurs sont des dispositifs électroniques qui permettent la **conversion** de la grandeur électrique, charge ou tension, fonction de la grandeur physique à mesurer en une grandeur électrique normalisée (tension 0-10 V ou courant 4-20 mA). Si nécessaire, le conditionneur du signal réalise la **linéarisation** de la grandeur électrique normalisée en fonction de la grandeur physique à mesurer. Les capteurs modernes intègrent de plus en plus fréquemment la fonction "conditionneur" du signal. Il suffit alors de les alimenter sous une tension continue comprise dans une fourchette donnée par le constructeur.

Cette section a pour but au travers d'un descriptif sommaire des principales caractéristiques de ces modules de préciser leur limite et leur principe de fonctionnement.

2.3.4.1 Conversion charge/tension et Amplification

L'effet piézoélectrique se caractérise par l'apparition d'une charge "Q" aux bornes du matériau sollicité qui se comporte comme une source de courant "I" en parallèle avec un condensateur C_1 . Le capteur est généralement relié à la station d'acquisition par un câble coaxial dont la capacité C_2 dépend de la longueur. La **tension mesurée en statique** aux bornes du capteur s'écrit alors :

$$V = \frac{Q}{C_1 + C_2} \quad (2.1)$$

Cette tension, fortement diminuée par la longueur du câble utilisé (C_2 est en général très supérieur à C_1), est aussi sensible aux variations de C_2 dues à des facteurs parasites comme :

- la température ;
- les contraintes mécaniques : elles créent des charges parasites par triboélectricité (dans le câble) ;
- les couplages électromagnétiques, bien qu'il soit possible de s'en prémunir en utilisant des câbles blindés.

L'ensemble de ces inconvénients explique que les amplificateurs de tension ont été abandonnés pour le conditionnement du signal et remplacés par les **amplificateurs de charge** qui permettent l'élimination de l'influence du câblage.

2.3.4.2 Adaptation d'impédance (isolement)

L'amplificateur suiveur ou d'isolation permet de passer d'une impédance élevée ($1 M\Omega$) à l'entrée à une impédance faible à la sortie sans affecter la valeur du signal.

Le signal de sortie V_s «suit» le signal d'entrée V_e .

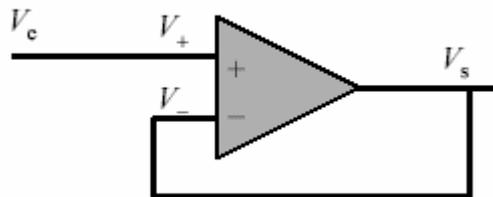


Figure 2.4 Amplificateur suiveur ou d'isolation

2.3.4.3 Linéarisation et compensation des grandeurs d'influence

Lorsque la courbe d'étalonnage de l'ensemble capteur (détecteur-conditionneur) n'est pas linéaire (ceci empêche de considérer la sensibilité comme constante), il peut être intéressant de chercher à la modifier par un traitement spécifique. Une méthode générale

utilise des multiplieurs analogiques et des additionneurs analogiques pour réaliser un signal V_1 , fonction polynomiale du signal V qui compense les non-linéarités.

Donc, En général, on introduit dans la chaîne une fonction de transfert qui compense la non linéarité de la réponse du capteur. La compensation s'effectue en insérant un composant non linéaire dans la boucle de la réaction d'un amplificateur.

2.3.4.4 Filtrage analogique anti-repliement

On démontre quand on échantillonne un signal, sa transformée de Fourier est conservée mais elle devient périodique de période égale à la fréquence d'échantillonnage. Cela veut dire physiquement que l'échantillonnage implique des composantes fréquentielles supplémentaires : c'est l'effet de **repliement ou aliasing** voir Figure 2.5. C'est à dire, si une courbe d'une fréquence donnée passe par un certain nombre de points, les courbes à toutes les harmoniques de cette fréquence passe également par ces points. Pour annuler cet effet indésirable, on utilise un filtre analogique anti-repliement jouant le rôle d'un filtre passe-bas dont la fréquence de coupure correspond à la fréquence de Nyquist (f_N) = $\frac{f_e}{2}$ et $f_{\max} < f_N$. Le filtre doit présenter une pente très importante de 110 dB/ Octave. Un facteur $f_e \geq 3f_{\max}$ permet de couvrir une gamme de fréquence très importante.

Pour une pente de coupure de 72dB/Octave (ce qui sous entend un filtre du deuxième ordre), la gamme d'exploitation sera de 25% de la fréquence d'échantillonnage.

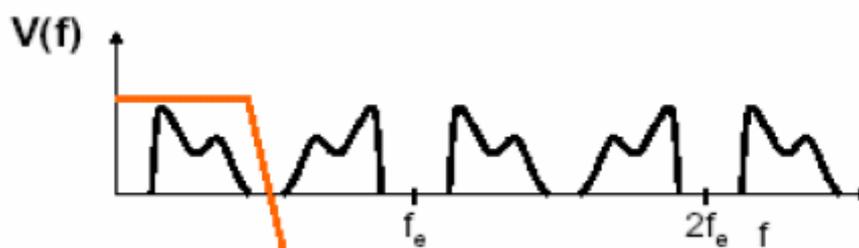


Figure 2.5 Densité spectrale d'un signal avec repliement

Le type de filtre anti-repliement adapté dépend de l'application. Nous distinguons : Le filtre de Butterworth, Le filtre de Bessel et le filtre elliptique de Chebyshev, etc.

Les accéléromètres se caractérisent par leur diversité et leur sensibilité propres. En respectant les précautions d'usage, nous pouvons leur accorder toute notre confiance pour relever des données au sein de nombreux systèmes.

Ce filtre analogique passe-bas est indispensable et doit être réalisé avant l'analyse. La fréquence de coupure est prise généralement $f_c < \frac{f_e}{2}$.

2.3.5 Numérisation du Signal

Généralement intégrée dans une centrale de mesure ou dans un micro-ordinateur, elle transforme les grandeurs analogiques de chaque voie de mesure en grandeurs numériques, voir figure 2.6.

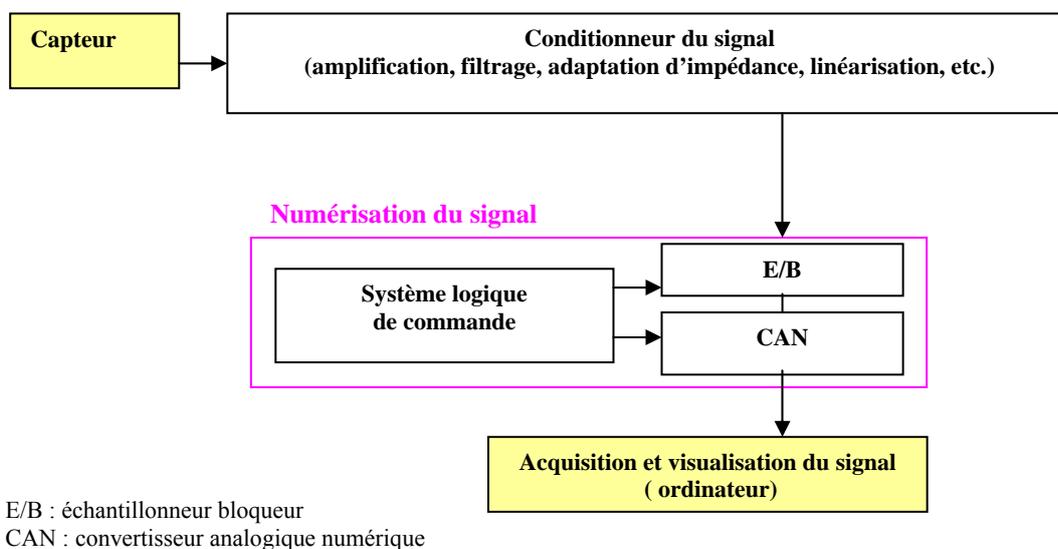


Figure 2.6 Position de la numérisation du signal

2.3.5.1 Echantillonneur-Bloqueur (en anglais : Sample and Hold)

Les notions d'échantillonnage et de codage sont souvent indûment confondues.

L'échantillonnage est l'opération qui consiste à prendre une toute petite partie du signal périodiquement. L'échantillonnage peut très bien se faire de manière analogique, la tension du signal étant par exemple stockée dans une capacité. Les filtres à capacités commutées fonctionnent de cette manière : les échantillons successifs sont des tensions analogiques transférées de capacités en capacités.

2.3.5.2 Convertisseur analogique numérique

Pour numériser le signal, on passe par un CAN, ce convertisseur analogique/numérique met un certain temps pour faire sa conversion (codage); pendant ce temps la valeur analogique doit rester stable d'où l'échantillonneur/bloqueur.

La conversion est l'opération qui consiste à convertir un échantillon analogique en grandeur numérique. Si l'échantillonnage est linéaire, on va le quantifier en donnant soit

le nombre de bits n_1 , soit le nombre total de niveaux N_1 , on a $N_1 = 2^{n_1}$. Par exemple, un convertisseur linéaire 8 bits fournit 256 niveaux de reconnaissance.

Caractéristique principale des CAN

- **Résolution** (exprimée en nombre de bits ou en % de la pleine échelle) : définit la valeur de la variation de la tension d'entrée donnant lieu à une variation d'une unité de la donnée numérique présente en sortie.

Ex: Un convertisseur à une résolution de $n= 8, 12, 16, \dots$ bits, c'est-à-dire un nombre de mesure de 2^n et une précision de $\frac{1}{2^n}$. La dynamique

$$D = 20 \log_{10}(2^n) \approx 6n .$$

On estime qu'un CAN ayant une dynamique de 72 dB (12 Bits) est largement suffisant

- **Temps de conversion:** durée écoulée entre l'instant d'apparition de l'impulsion de début de conversion et l'instant où la donnée est disponible sur le bus de sortie
- **Polarité:** Un convertisseur peut être **bipolaire** ou **unipolaire** selon qu'il peut accepter ou non des signaux d'entrées symétriques par rapport à zéro. Dans le cas d'un convertisseur bipolaire, le code de sortie correspond soit à un code binaire décalé, soit à un code complément à deux, ou un code appelé amplitude signe (le bit de signe est inversé par rapport au complément à 2).

Le Principe de fonctionnement, les applications, les avantages et les inconvénients de l'échantillonneur bloqueur et du CAN sont donnés dans les références.

2.3.6 L'Ordinateur

L'architecture d'un ordinateur constitue l'ensemble des sous-systèmes réalisant différentes fonctionnalités, la manière de les relier et de les faire communiquer.

Au premier sens du terme, une architecture informatique est définie par des paramètres tels que les instructions du microprocesseur, les jeux de registres, la méthodologie de gestion de la mémoire, et d'autres fonctions.

Un simple ordinateur est composé de :

- mémoires (RAM/ROM) ;
- CPU : Central Processeur Unit (UC : Unité Centrale, MPU : MicroProcesseur Unit)
- quelques circuits d'entrée/sortie.

Les trois parties sont connectées par les trois bus (**adresse, données, contrôle**).

2.3.6.1 Chargement et Visualisation des données

Le CPU est le maître unique dans les ordinateurs simples, et c'est lui qui contrôle tout le système. La séquence des instructions (ou programme) est stockée en mémoire à des

emplacements successifs et en code binaire. Le rôle du CPU se décompose en quatre étapes :

1. aller **chercher** l'instruction suivante en mémoire. C'est le cycle de chargement des lignes du programme d'acquisition;
2. la **décoder** pour déterminer l'action qu'il doit effectuer ;
3. l'**exécuter** ! Cycle d'exécution ;
4. revenir à l'étape 1.

Cycle de chargement

Les instructions sont lues une par une, puis décodées et exécutées. Pour aller chercher une instruction en mémoire, il faut exécuter les opérations suivantes:

1. charger le contenu du pointeur de programme dans le registre d'adresse mémoire qui est relié au bus ;
2. demander à la mémoire de lire la donnée, et de la placer sur le bus (*signal read memory*) ;
3. ranger la valeur lue sur le bus de données, dans le buffer de l'UC ;
4. transférer le buffer dans le registre d'instruction ;
5. incrémenter le pointeur de programme d'une unité pour aller chercher l'instruction suivante.

Cycle d'exécution :

L'ensemble des opérations précédentes se répète à chaque cycle de chargement. Mais, selon l'architecture et la teneur de chaque instruction, la façon dont le cycle d'exécution se déroule est très variable.

Prenons comme exemple l'addition de l'accumulateur et d'un autre registre avec le résultat dans l'accumulateur.

Puisque les registres sont dans l'UC, la séquence d'exécution est :

1. transfert du contenu de l'accumulateur dans l'UAL ;
2. transfert du contenu du registre adressé vers l'autre entrée de l'UAL ;
3. déclenchement de l'addition (en envoyant les signaux nécessaires dans l'UAL) ;
4. retour du résultat dans l'accumulateur.

Remarque : si une instruction fait appel à la mémoire le cycle est plus long.

2.3.6.2 Puissance d'un microprocesseur

La notion de puissance est la capacité de traiter un grand nombre d'opérations par seconde sur de grands nombres et en grande quantité.

Intrinsèquement la puissance se joue donc sur les trois critères suivants:

- ✓ La longueur des mots : données et instructions (on parle de largeur du bus des données).

- ✓ Le nombre d'octets que le microprocesseur peut adresser (on parle de largeur du bus des adresses).
- ✓ La vitesse d'exécution des instructions liée à la fréquence de fonctionnement de l'horloge de synchronisation exprimée en *MHz*.

2.3.6.3 Architecture interne d'un microprocesseur

On distingue 3 éléments logiques principaux :

- Une Unité Arithmétique et Logique (U.A.L.)
- Un Accumulateur et
- Des registres que l'on nomme couramment :
 - Le Compteur d'Instructions (C.I.)
 - Le Registre d'état
 - Le Registre d'Instructions (R.I.)
 - Le Registre d'Adresses (R.A.)
 - Le Registre temporaire des données

2.3.7 Dispositions pratiques et prétraitement

Les forces générés par un défaut mécanique, électromécanique ou hydraulique sur une machine surveillée ne "s'expriment" sur un appareil de mesure ou d'analyse qu'après avoir traversées la structure de la machine et la chaîne de mesure (capteur, conditionneur, ...). Le signal induit est donc modifié (amorti, amplifié, ...) par un coefficient qu'on appelle "fonction de transfert". Nous avons, par exemple, pour chaque fréquence constituant le signal :

$$\boxed{\text{Amplitude affichée sur un appareil}} = \boxed{\text{Amplitude de la force générée}} \times \boxed{\text{Valeur de la fonction de transfert}}$$

La fonction de transfert dépend :

- ↪ des caractéristiques propres à la structure (masse, rigidité, amortissement),
- ↪ des caractéristiques de la fixation de la machine sur le sol qui oppose des réactions aux vibrations et amplifie ainsi leur intensité,
- ↪ des caractéristiques du capteur utilisé,
- ↪ de la position et de la fixation de ou des capteurs sur la machine,
- ↪ des performances des amplificateurs, filtres et enregistreurs ou analyseurs.

La surveillance ou le diagnostic ne peut donc être fiable que si la fonction de transfert reste constante dans le temps et les conditions de fonctionnement (charge et vitesse) seront sensiblement identiques à chaque prise de mesure. Il faut aussi que les fréquences des défauts redoutés soient situées à l'intérieur de la plage d'utilisation de chacun des éléments de la chaîne de mesure.

Il conviendra aussi de se prémunir de toutes les erreurs décrites ci-dessous.

2.3.7.1 Erreur liées au choix du pas d'échantillonnage

Lorsqu'on veut traiter un signal par ordinateur on doit d'abord l'échantillonner correctement. Ceci revient à prélever des échantillons de ce signal à des instants bien précis. C'est ce temps de prélèvement qui est appelé période d'échantillonnage $\Delta t = \frac{1}{f_e}$.

Echantillonner correctement un signal signifie choisir la bonne période d'échantillonnage Δt . Le signal est alors approximé par $x(k\Delta t)$, $k = 1, 2, \dots$. Cette approximation induit une perte d'information sur le signal échantillonné par rapport au signal continu. Cette perte d'information dépend de la valeur attribuée à Δt .

Notons que les points d'échantillonnage ne sont pas représentatifs exclusivement d'un seul signal. Un grand nombre de signaux différents peuvent passer par ces points. Le choix de la période d'échantillonnage est imposé par le **théorème de Shannon** qui exprime pour des spectres d'ordre 2 que : Pour pouvoir restituer le signal analogique à partir des échantillons il faut que la fréquence d'échantillonnage soit supérieure ou égale à deux fois la fréquence maximale du spectre. Les conditions d'échantillonnage sont plus strictes lorsque l'on s'intéresse à des multispectres d'ordre supérieur à 2.

A l'ordre 2 :

Si le signal à temps continu a une largeur de bande limitée par $B = f_{\max}$ c'est-à-dire $X(f) = 0$ si $|f| > f_{\max}$ et si la période d'échantillonnage vérifie la condition de Shannon $\Delta t \leq \frac{1}{2B} = \frac{1}{2f_{\max}}$ alors $f_e \geq 2.f_{\max}$ à l'ordre 2. Dans ces conditions on peut reconstituer la corrélation à temps continu à partir de la corrélation à temps discret. Par transformation en fréquences réduites de la corrélation à temps continu on obtient exactement la *densité spectrale* du signal continu. On ne perd pas d'information sur le signal étudié s'il vérifie les conditions du théorème d'échantillonnage.

L'expérience en mécanique montre que la fréquence d'échantillonnage doit être égale approximativement de 3 à 10 fois la fréquence du signal (Morel, 1992). Il est recommandé de choisir une fréquence d'échantillonnage de l'ordre de 10 fois la première fréquence d'intérêt.

A l'ordre supérieur :

On montre dans (Lacoume & al. 97 pp.53-54) que la fréquence d'échantillonnage doit être pour le bispectre $f_e \geq 3.f_{\max}$. Ce résultat peut être, bien sur, généralisé aux multispectres. Pour le calcul d'un multispectre d'ordre « p » la fréquence d'échantillonnage doit être supérieure ou égale à p fois la fréquence maximale d'analyse de ce signal $f_e \geq p.f_{\max}$.

Nous pouvons donc énoncer le théorème d'échantillonnage pour les multispectres, sous la forme suivante : Pour que le multispectre d'ordre « p » du signal échantillonné soit égal au multispectre du signal à temps continu $x(t)$ dans l'hypercube $|f_i| \leq \frac{f_e}{2}$, ($1 \leq i \leq p-1$), il faut : Que le signal $x(t)$ ait une longueur de bande limitée par f_{\max} , et que la période d'échantillonnage Δt vérifie la condition $\Delta t \leq \frac{1}{p.f_{\max}}$ (c'est-à-dire $f_e \geq p.f_{\max}$).

Echantillonnage et phénomène de recouvrement

La période d'échantillonnage Δt doit être suffisamment petite par rapport à la période du signal recherché, pour ne pas perdre l'information. Lorsque cette période d'échantillonnage est trop grande par rapport à la période du signal recherché, on voit apparaître un signal apparent, semblable au phénomène stroboscopique, dont la fréquence apparente est égale à : $f_{apparent} = f_e - f_{réel}$

Or, en général, on ne connaît pas toutes les fréquences recherchées et il est donc inévitable que ce phénomène apparaisse aux hautes fréquences si on ne filtre pas le signal. On appelle ce phénomène recouvrement.

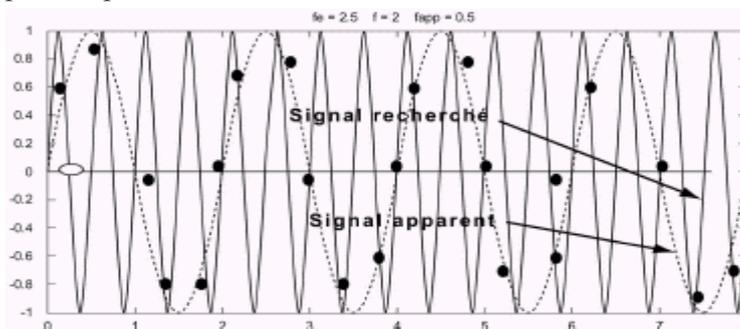


Figure 2.7 Signal recherché et signal apparent (extraite de Thomas, 2003)

Echantillonnage et temps maximal d'observation

La précision en fréquence est égale à l'inverse du temps maximal d'observation

$$\Delta f = \frac{1}{T_M} \quad (T_M : \text{temps maximal d'observation}, M : \text{nombre d'échantillons})$$

Le principe d'*incertitude de Heisenberg* prescrit qu'il est impossible d'avoir à la fois un temps d'observation court et une bonne résolution en fréquence.

$$\Delta f (\text{Hz}) = \frac{fe}{M} = \frac{1}{M \cdot \Delta t} = \frac{1}{T_M} \quad (2.2)$$

Il est impossible d'avoir à la fois une bonne précision temporelle et fréquentielle. Si on désire réaliser une analyse avec une bonne précision fréquentielle, le temps d'observation maximal T_M doit être grand et on devra choisir une largeur de bande en fréquence faible pour un nombre d'échantillons M donné.

Si on désire analyser un signal de haute fréquence (f max grand), le temps d'observation T_M sera faible pour un nombre d'échantillons M donné et par conséquent la précision de la mesure en fréquence sera faible. Pour augmenter la précision de la mesure, il faut donc augmenter le nombre d'échantillons.

Echantillonnage et densité spectrale de puissance

La transformée de Fourier fournit une suite de valeurs discrètes qui correspondent sur l'axe des fréquences, à une fonction échantillonnée tous les $\frac{1}{T_M}$, si T_M est la durée du signal.



Figure 2.8 Transformée de Fourier

Le problème que l'on doit résoudre maintenant se limite aux deux questions suivantes :

1. la suite de valeurs $X(n)$ sortant de la transformée de Fourier rapide (FFT), est-elle une représentation rigoureuse de la transformée de $F(fi)$ du signal analogique $f(t)$? autrement dit, risque-t-on de perdre de l'information sur le signal si on l'échantillonne ?
2. sous quelles conditions, toute l'information utile contenue dans le signal est-elle préservée au cours de ce traitement ?

Sous quelles conditions peut-on retrouver la fréquence du signal échantillonné ?

1. Le signal $x(t)$ doit être limité en fréquence, ce qui veut dire que la TF $X(f)$ doit être bornée. Sinon, le chevauchement multiple exclura toute possibilité d'extraire correctement les composantes fréquentielles.

2. Cette condition n'est pas suffisante, il faut encore que la fréquence d'échantillonnage f_e soit suffisamment grande pour éviter le chevauchement de deux périodes voisines dans le domaine fréquentiel. En bref, $F(f)$ et $F(f + f_e)$ ne doivent avoir aucun point commun (Figure 2.2). Cette condition s'écrit directement $f_{\max} < \frac{f_e}{2}$ soit $f_e > 2.f_{\max}$, il s'agit de la condition de Shannon. Dans ces conditions on peut reconstituer la corrélation à temps continu à partir de la corrélation à temps discret. Par transformation en fréquences « réduites » de la corrélation à temps discret on obtient exactement la densité spectrale du signal à temps continu.

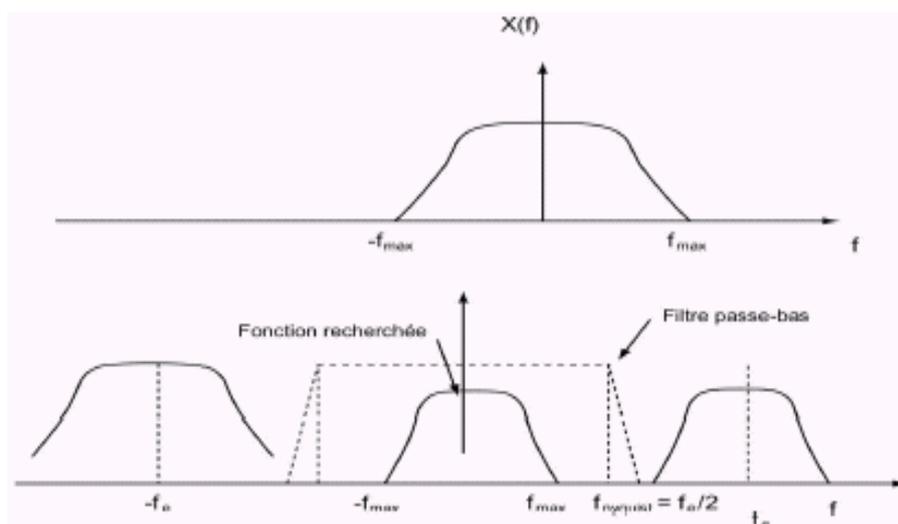


Figure 2.9 Densité Spectrale de Puissance représentant le repliement du spectre (extraite de Thomas, 2003)

2.3.7.2 Erreurs liées à la mise en conditionnement

Trois précautions doivent être prises dans le conditionnement du signal (Bourgain & al 88), (Morel, 1992) :

- ✓ réglage des niveaux d'entrée (étendu de mesure) : lorsque les niveaux d'entrée sont trop faibles, le signal à analyser est noyé dans le bruit de fond de l'appareillage électrique. Pour éviter les imprécisions dues aux faibles niveaux et les saturations liées aux niveaux trop élevés, on procède par une augmentation du gain jusqu'à ce que le voyant de saturation, de l'oscilloscope placé en parallèle, s'allume puis en revient très légèrement en arrière.
- ✓ élimination des fréquences indésirables : placé un filtre analogique à double effet.
 1. Une fonction passe bas, la fréquence de coupure ' f_c ' est fonction de la gamme à analyser dans tous les cas elle doit être inférieure ou égale à la

fonction de résonance du capteur f_r sur 2 pour garder la linéarité et avoir, alors, un gain nul.

2. Une fonction anti-repliement ou « anti-aliasing »

- ✓ adaptation d'impédance (Morel, 1992) : l'accéléromètre à l'inconvénient d'être à très haute impédance de plusieurs Méga Ohms, et de nécessiter des adaptateurs d'impédance.

Pour les acquisition du laboratoire, nous avons utilisé l'appareillage SM231 (voir annexe 2.2) et un oscilloscope placé en sortie.

2.3.7.3 Erreurs dans le spectre dues à la troncature

On n'observe le signal que durant un temps limité qui est la durée de mesure. Cette opération, imposée par les conditions pratiques, nous conduit naturellement à négliger toute information qui se situe en dehors de l'intervalle de mesure T_M . Pour la transformée de Fourier, une fonction tronquée est une fonction qui se répète périodiquement en dehors de l'intervalle de troncature (par référence à la décomposition en série de Fourier).

Mathématiquement, le signal tronqué comme le signal transitoire dont on ne considère qu'un intervalle de temps donné $\left[-\frac{1}{2}T, \frac{1}{2}T\right]$ est obtenu en multipliant ce signal par la fonction porte $\prod_T(t)$. La fonction tronquée $x_T(t)$ n'est plus limitée en fréquence et il est impossible en toute rigueur de l'échantillonner sans erreur, c'est à dire dans les conditions de Shannon.

Quelle erreur commet-on en faisant quand même l'échantillonnage d'un signal tronqué ou transitoire ?

La troncature du signal amène deux sortes d'altérations dans le domaine fréquentiel (Bourgain et al., 1988) :

a) une altération locale : autour de la raie principale du signal. La raie s'élargit pour former ce qu'on appelle le lobe principal, sa largeur à la base est égale à $\frac{2}{T}$, si T est la durée de prise en compte du signal. Plus la durée est faible plus le lobe principal sera large et plus la représentation fréquentielle sera altérée. Dans le cas d'un signal comportant plusieurs fréquences ou un spectre continu de fréquences, ce phénomène se généralisera sur toute l'étendue fréquentielle avec pour conséquence un lissage de la fonction dans le domaine fréquentiel.

b) une altération globale : produite par ce qu'on appelle les rebondissements multiples ou lobes secondaires du signal. Il en résulte des composantes fréquentielles là où théoriquement il ne doit pas en exister. Si le signal comporte plusieurs fréquences ou un spectre continu de fréquences, on assistera, du fait de ces lobes secondaires multiples à

une erreur d'évaluation sur toute l'étendue du spectre. Cet élargissement de lobes a des conséquences pratiques importantes.

Supposant qu'on a un défaut d'engrenage sous forme d'impulsion à $f_0 = 1kHz$. Une largeur d'impulsion de $T_{imp} = 100 ms$ élargit le signal en fréquence dans le lobe principal $f_0 = 1kHz$ de $\Delta f_{imp} = \frac{1}{T_{imp}} = 10Hz$. Ce qui représente 10Hz de part et d'autre de la raie centrale. Cela veut dire que l'on risque de perturber d'autres défauts existant dans les bandes voisines situées entre $f_0 \pm \Delta f_{imp}$. On pratique, on considère la largeur du lobe, non à partir de la base, mais à mi hauteur, ce qui revient à diviser par deux la largeur d'interférence.

Le fait de tronquer un signal introduit, donc, des erreurs qui génèrent de fausses fréquences. La correction réalisée consiste à multiplier le signal par une fonction de pondération qui dépend du type de signal et du régime des essais.

Fenêtrage ou pondération pour la représentation spectrale

On choisira, pour diminuer les effets néfastes, les fenêtres de pondération suivantes (Morel, 1992), (Bourgain et al., 1988):

Tableau 2.2 *Les principales fenêtres en analyse vibratoire*

Signal	Fenêtre
- aléatoires (ou inconnus),	- Hanning
- force d'impacts	- Rectangulaire
- Transitoire et choc	- Exponentielle
- Périodique ou ayant des amplitudes différentes	- Flat Top et Kaiser Bessel

Le fenêtrage permet, donc, de limiter les erreurs du traitement du signal de durée finie, traitement qui devrait s'appliquer à un signal de durée infinie. Elle permet également dans le domaine fréquentiel d'éliminer les composantes indésirables du spectre, les bruits, ... Pour un signal périodique, on peut montrer que si l'on veut obtenir un résultat satisfaisant, il est impératif que le temps d'intégration (fenêtre, $T = N dt$), soit égal à la période de la fondamentale (par analogie aux séries de Fourier) ou un multiple entier de cette période.

Dans le cas des fonctions presque périodique ou aléatoire, il faut que la durée d'intégration soit aussi longue que possible et théoriquement infinie.

La longueur de la fenêtre de pondération ou le temps de N points est inversement proportionnel à la longueur de la plage de fréquences dans laquelle on veut calculer le spectre.

$$T = N dt = \frac{N}{f_e} = \frac{N}{2.56 f_{\max}} \quad (2.3)$$

Plus la de fréquences f_{\max} est petite, plus le temps nécessaire à l'intégration est long. Il faut donc un temps de prise de mesure, très long pour analyser les vibrations induites par les machines tournantes à basse vitesse (Boulenger & al., 1998, p.37)

La fréquence maximale " f_{\max} " est choisie en fonction non seulement de la fréquence de rotation de la machine, mais aussi et évidemment des fréquences de défauts que l'on peut redouter, et donc du type et de la complexité de la machine à surveiller. Ces fréquences des défauts redoutés doivent se situées à l'intérieur de la plage d'utilisation de chacun des éléments de la chaîne mesure.

Fenêtrage pour le domaine temps multifréquences :

Si on ne considère pas le fenêtrage, on s'aperçoit que tous les événements contenus dans un signal, qu'ils soient éloignés ou non de la date de référence t , participent avec le même « poids » à la représentation en t . Or, il est légitime d'accorder plus d'importance aux instants proches de l'instant d'analyse.

Une façon de réaliser cela est de pondérer la multicorrélation locale avant de d'effectuer la T.F. Puisque la fonction de multicorrélation locale dépend à la fois de t et de τ , la pondération la plus générale doit dépendre également de t et de τ .

Un autre point de vue pour comprendre cette pondération est de considérer la fonction de multicorrélation comme un estimateur de la multicorrélation des moments d'un signal aléatoire. Pour passer aux multispectres, il faut alors « apodiser » en temps et/ou lisser en fréquence. Une fenêtre de pondération sera alors $F(t, \tau)$ où la dépendance en t correspond au *lissage* et la dépendance en τ concerne l'apodisation.

Pour la T.F. de la fonction de multicorrélation locale ainsi pondérée conduit à la définition de la représentation temps-multifréquences (Lacoume & al., 1997 p.159).

$$R_{x(p)}^q(t, f) = \iint F(t - \theta, \tau) C_{x(p)}^{l(q)}(\theta, \tau) e^{-2\pi j f \tau} d\tau d\theta \quad (2.4)$$

L'ensemble de ces représentations (paramétrées par F) définit une classe générale des représentations temps-multifréquences d'ordre $p + q$

2.3.7.4 Erreurs liées aux bruits de fond

Quel que soit le système étudié, et quelle que soit la qualité de l'instrument, il existe toujours des bruits de fond d'origine électrique ou autre. Il arrive fréquemment que le niveau des bruits soit inacceptable. Les dispositions à prendre pour une réduction des bruits de fond sont (Bourgain et al., 1988) (Arquès, 1996) :

- l'adjonction d'un filtre passe bas à l'entrée de la carte. Ce filtre est tout simplement le **filtre anti-repliement**,
- adopter un rapport ‘deux’ entre la **fréquence de coupure** et la fréquence maximale du signal à analyser,
- l'adoption de la **bande de fréquence** la plus faible possible compatible avec la nature des phénomènes étudiés,
- l'augmentation du **nombre de points** pour le calcul de spectres. L'énergie des bruits de fond est alors répartie sur un plus grand nombre de points et le niveau dans chaque canal s'en trouve réduit,
- Si la période d'intégration n'est pas infinie, il faut faire un moyennage fréquentiel c'est à dire stocker un certain nombre de spectres successifs et en faire la moyenne. On peut faire également un moyennage synchrone, c'est-à-dire, la moyenne des signaux temporels successifs avant qu'ils ne soient traités numériquement.

Les deux méthodes principales qui permettent d'augmenter le rapport signal sur bruit sont les moyennages fréquentiel et synchrone (temporel).

2.3.7.5 Erreurs dues aux grandeurs physiques parasites

Le capteur peut se trouver soumis non seulement à la grandeur physique que l'on veut mesurer, mais à d'autres grandeurs physiques que l'on ne veut pas mesurer. Leurs variations sont susceptibles d'avoir une influence non négligeable sur la justesse des mesures. Parmi ces grandeurs, on peut citer :

- Le champ acoustique,
- Les vibrations transversales,
- Les champs électromagnétiques,
- La circulation de courant d'arbre (boucle de masse),
- Les effets triboélectriques, ...

Il conviendra d'éviter les effets de champs électromagnétiques et la présence éventuelle de circulation de courants d'arbre par élimination des boucles de masse et l'isolation électrique du capteur. Pour le filtrage de bruit on utilise le filtrage adaptatif (Arquès, 1996). La technique utilise un signal auxiliaire de référence détecté par un ou plusieurs capteurs localisés en des points où le champ du bruit est important, alors que le signal à détecter est faible ou indétectable.

Le signal bruité capté à proximité de l'élément susceptible de présenter un défaut, et le signal de référence, composé essentiellement des bruits originaux de la machine, sont

introduits en entrée du système d'annulation adaptative du bruit. La sortie du système est composée du signal en provenance de l'élément débruité et du bruit caractéristique de la machine.

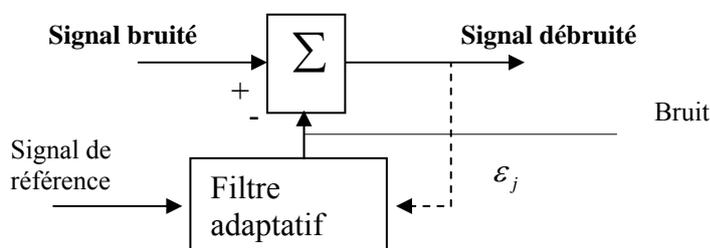


Figure 2.10 Cas général d'un système d'annulation adaptatif du bruit
(extraite de (Arquès, 1996))

Le filtre adaptatif est ajusté pour que la moyenne de l'erreur ε_j entre le signal bruité et le signal de référence soit minimale.

Exemple : dans le cas d'un réducteur de vitesse pour lequel la détection des défauts de roulement est souhaitée, le signal bruité du roulement est capté à proximité de la bague extérieure du roulement et comprend le signal crée par les chocs du au défaut de roulement associé aux bruits en provenance de l'engrenage, additionné au bruit aléatoire. Le signal de référence capté sur le bâti du réducteur à une grande distance du roulement sera constitué du bruit des engrenages additionné du bruit aléatoire. En sortie du filtre adaptatif, le signal comprend le bruit des engrenages additionné du bruit aléatoire, et en sortie de l'annulation adaptative du bruit, le signal comprend uniquement le signal en provenance du roulement.

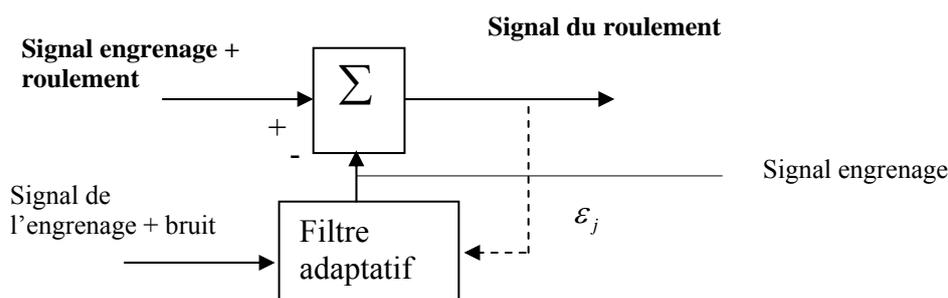


Figure 2.11 Application à la séparation du signal vibratoire en provenance du roulement et de l'engrenage (Arquès, 1996)

2.3.7.6 Erreur dans le spectre liée à la périodisation

Une autre erreur liée au choix de la longueur d'intégration et le calcul de la FFT et que la périodisation du signal a pour effet d'échantillonner le spectre de ce signal à une fréquence Δf , appelée "résolution" ou " finesse d'analyse" et telle que :

$$\Delta f = \frac{1}{T} = \frac{1}{N dt} = \frac{f_e}{N} \quad (2.5)$$

Δf : est appelée également fréquence de discrétisation du spectre,

T : Longueur de traitement ou de calcul.

Le spectre du signal n'est alors connu de manière parfaite que pour une suite de points espacés de " Δf ". Il en résulte des erreurs parfois importantes sur l'amplitude et la fréquence des composantes élémentaires du signal si la fréquence de ces dernières ne correspond pas à un multiple entier de la résolution Δf . Notons que, théoriquement, deux raies espacées de deux fois la résolution devraient parfaitement être identifiables.

Le pouvoir de résolution sera très réduit également, si les deux composantes à dissocier sont très différentes en amplitude. Pour une fenêtre de Hamming qui est la plus fréquemment utilisée pour nos applications, $f_2 - f_1 = \frac{\Delta f}{8}$ pour que les deux raies de fréquences f_1 et f_2 et d'amplitudes très différentes soient dissociées (Boulenger & al., 1998 p. 38). La durée d'intégration, est donc inversement proportionnelle au pouvoir de séparation, il est donc logique de choisir $T = \frac{1}{f_2 - f_1} = 8 \frac{1}{\Delta f}$. La durée d'intégration est 8 fois plus grande que dans les conditions normales.

2.3.7.7 Erreur due à la fréquence de discrétisation du spectre ou du multispectre

Dans le cas du calcul d'un spectre l'une des erreurs effectuée provient de la non coïncidence entre les fréquences détectées et celles calculées dans la transformation. L'amplitude de cette erreur peut être diminué par une acquisition **synchrone du signal** à la vitesse de rotation du moteur. Cette acquisition synchrone peut être effectuée par (Arquès, 1996 p.197) :

- un codeur comportant 2^n impulsion par tour,
- En mesurant la vitesse moyenne d'un tour et en générant une fréquence d'échantillonnage appropriée pour digitaliser le signal.

La première solution nécessite une sortie de l'arbre du moteur libre. La deuxième nécessite la détection d'un top tour de synchronisation et une compensation implicite des irrégularités de la vitesse de rotation du moteur.

2.3.7.8 Erreur due à la composante continue du signal

La présence d'une valeur moyenne introduit dans la corrélation une branche infinie (ici constante) qui par transformation de Fourier, « sature » le spectre (Lacoume et al.

1997) et introduit dans le spectre une composante singulière à l'origine. Par suite de l'élargissement spectral, issu de la durée finie des signaux traités, cette composante singulière masque les composantes basses fréquences. Dans cette situation, il est nécessaire de centrer le signal traité, selon une "bonne procédure", avant de procéder à l'analyse spectrale (Lacoume surveillance4 p.66).

Quand on estime les multispectres la même difficulté se présente c'est à dire une branche infinie dans les multicorrélations. Pour le bispectre, il suffit de centrer le signal, par contre, pour le trispectre la situation est plus complexe car des produits de termes d'ordre 2 apparaissent dans la tricorrélation et perturbent fortement l'estimée du trispectre (Lacoume et al., 1997).

Pour l'estimation des spectres et des bispectres, les valeurs du signal seront centrées pour ne pas avoir une composante singulière à l'origine.

2.3.7.9 Erreur due à l'effet de masque

L'effet de masque entache fortement la sensibilité des indicateurs "large bande" lorsque l'énergie vibratoire introduite par l'apparition d'un défaut n'est pas prépondérante devant celle introduite par l'ensemble des forces dynamiques générées par le fonctionnement de la machines et des défauts déjà existants. L'augmentation de la valeur d'un indicateur risque d'être considérée comme non significative, alors que l'état de cet élément vient considérablement de se dégrader (Voir un défaut de roulement devant un défaut d'engrenage).

Il est préférable que les indicateurs soient calculés dans des **bandes fréquentielles**, ni trop larges ni trop étroites, définies en fonction des caractéristiques cinématiques de la machine (Boulenger et al., 2003), (Pachaud et Canetto, surveillance04).

2.3.7.10 Erreur dans le spectre liée à la vitesse de rotation

Dans une analyse vibratoire, on cherche principalement à mettre en évidence la périodicité des phénomènes pour ensuite la corrélérer avec la cinématique de la machine surveillée. Il est alors impératif que la vitesse de la machine reste relativement constante pendant la prise de mesure. Dans le cas contraire, l'analyse FFT n'est pas suffisante et il faut avoir recours à une analyse temps/fréquence, analyse en ordre et le rééchantillonnage angulaire (Antoni, 2000) et (Bonnardot, 2003).

2.3.7.11 Résonance et choix de la vitesse de rotation :

Pour un diagnostic fiable, les vitesses de rotation du moteur doivent être choisies de telle sorte qu'il n'existe aucune résonance entre les fréquences propres des différents éléments de la machine (engrenages, courroies, roulements, etc.) et les multiples entiers de celle-ci. En cas de résonance, l'amplitude vibratoire va

augmenter très fortement dans une bande de fréquence très étroite qu'il est nécessaire d'éviter pour faire une évaluation exacte de la dégradation.

2.3.7.12 Représentations graphiques des spectres.

Souvent, pour plus de lisibilité, on adopte une échelle en décibel. Il s'agit bien évidemment d'une représentation graphique,

$$L(dB) = 10 \log\left(\frac{A}{A_0}\right)^2 = 20 \log\left(\frac{A}{A_0}\right) \quad (2.6)$$

A_0 : est une valeur de référence dans la même unité, qui peut être fixée conventionnellement par des normes ou prise égale à une unité physique.

\log : est le logarithme à base de 10.

$$\text{Si } \frac{A}{A_0} = 2 \Rightarrow L = 6 \text{ dB}.$$

Trois types de représentation peuvent être adoptées :

- ↪ Représentation linéaire des amplitudes et des fréquences
- ↪ Représentation en décibel des amplitudes et linéaire des fréquences
- ↪ Représentation linéaire des amplitudes et en décibel des fréquences

L'adoption d'une échelle des fréquences logarithmique privilégie la lisibilité des composantes "basses fréquences" correspondant notamment aux premiers ordres de la fréquence de rotation.

Une dynamique élevée associée à une représentation des amplitudes en décibels est fondamentale pour le diagnostic puisque ce mode de représentation permet la visualisation de composantes de fréquences dans le rapport d'amplitude peut varier de 1 à 10000, sans que la composante de plus faible niveau ne se trouve noyée dans le bruit de fond.

2.3.7.13 Enregistrement de données

Pour les applications à la surveillance et au diagnostic de *machines tournantes*, nous avons dans notre cas un système avec :

- une plage de fréquences (bande passante) qui s'étend de "2" à plus de 15 kHz c'est une caractéristique du capteur,
- une dynamique de 70 dB (résolution 12 bits), une caractéristique du convertisseur analogique numérique de notre carte d'acquisition, cette valeur permet de limiter le bruit de fond plein échelle au maximum de 1/3160.

Ces deux qualités sont largement suffisantes pour le diagnostic comme il est souligné dans (Boulenger & al., 1998) et (Schenck, 994).

- les mesures sont enregistrées pour un traitement différé, ce qui a pour avantages
 1. la prise d'un grand nombre de mesures en simultané,
 2. le traitement est fait en un lieu plus calme, à un moment choisi,
 3. la prise de mesures par un personnel non diagnosticien,
 4. le stockage des mesures et leur conservation dans le temps,
 5. la comparaison des mesures dans le temps ou les unes avec les autres,
 6. la recherche en parallèle de tous les renseignements cinématiques,
 7. la formulation du diagnostic en un endroit plus propice à la réflexion.

Remarque : Toutes ces dispositions ont été prises en considérations au niveau de notre laboratoire lors de l'acquisition des signaux des défauts de roulement et de courroie.

2.4 Méthodes d'analyse des signaux

Les Capteurs de Mesure délivrent des signaux qui sont généralement des tensions en fonction du temps. Ces dernières peuvent être envoyées sur l'écran d'un Oscilloscope et interprétées directement. Cette technique était couramment utilisée dans les années 1940-1960. Dans la majorité des cas le graphe est difficilement exploitable.

Pour tirer des informations sur l'état de santé du système et établir des indicateurs puis formuler des critères de décision, on a fait appel aux méthodologies de traitement du signal utilisant soit les **propriétés énergétiques** soit les **statistiques** d'ordre 1, 2, ou d'ordre supérieur.

Bien sur, Il est de pure utopie de vouloir croire en l'existence d'une méthode capable de caractériser tous les types de signaux. En plus, la qualité même des indicateurs dépend très étroitement, d'une part de la qualité des signaux bruts et d'autre part des procédures utilisées pour le traitement de l'information. Il est clair que les signaux complexes demandent bien plus qu'un simple spectre FFT. Certains signaux rencontrés sur des processus industriels ne peuvent être analysés avec les outils classiques de traitement du signal, soit parce que celles-ci s'avèrent être inadaptées, soit parce que les **hypothèses de gaussienneté, de stationnarité, d'ergodicité, de linéarité**, etc., qui conditionnent la validité d'emploi ou d'utilisation de tel ou tel outil, ne sont pas valables et sont inadéquates. Nous allons nous intéresser également à ces situations, nous parlerons alors de non-gaussienneté, non-linéarité et non-stationnarité ou plus génériquement des non-propriétés. En plus, sur un problème un peu complexe, il n'est d'ailleurs pas rare de devoir utiliser plusieurs méthodes.

Compte tenu du nombre important d'outils de traitement de données. Le choix de l'algorithme de traitement représente certainement la difficulté majeure car il induit plusieurs questions sur :

- le système, est-il linéaire, bilinéaire ou non linéaire? Stable ou non? causal ou non? dynamiques ou à mémoire? instantanés? continu et discret? Stationnaire, non-stationnaire ou pseudo-périodique?
- le signal, est-il à temps continu ou discret? centré ou non centré? déterministe ou aléatoire? aléatoire stationnaire, cyclostationnaire ou non stationnaire? composite? à valeurs complexes ou réelles? à énergie finie ou à énergie infinie? à bande limitée? à mémoire limitée? circulaire? ergodique ou non ergodique? gaussien ou non gaussien? linéaire, bilinéaire ou non linéaire? monodimensionnel ou multidimensionnel de sources indépendantes ou non? à une variable ou multivariées?
- le bruit, est-il aléatoire ou déterministe? gaussien ou non gaussien? blanc, rose ou coloré (rose, ...)? additif ou non? indépendant du signal ou non?

Le choix de l'algorithme de traitement le mieux approprié est aussi fonction :

- de l'organe à surveiller (roulement, engrenage, alignement d'arbre, palier lisse, etc.),
- de la nature des vibrations (sinusoïdale, impulsionnelle, ...),
- de la plage de fréquences ou le défaut est le plus facilement identifiable.
- de certain nombre de traitements précoces (vue dans la section 2.4) à effectuer tel que, l'élimination des parties qui obscurcissent l'interprétation des signaux : bruit aléatoire, modulation d'amplitude ou de phase, phénomène de résonance, etc.

Dans cette section, nous allons faire une synthèse bibliographique des méthodes et principaux outils utilisés en traitement du signal. L'objet, bien sur, n'est pas de donner des études complètes celles-ci peuvent être trouvées dans les livres, thèses et articles que nous citerons au fur et à mesure de cette section.

2.4.1 Analyse statistiques et temporels

L'analyse statistique du signal est un outil intéressant, les vibrations de type sinusoïdal ou impulsionnel génèrent des paramètres différents. Mais l'effet de masque est l'écueil le plus sérieux de suivi en mode global "large bande" et en limite considérablement le domaine d'utilisation et la fiabilité (Boulenger et al., 2003).

L'étude visuelle du signal en fonction du temps n'est pas toujours simple, dans quelques circonstances, associée à d'autres méthodes, elle peut apporter une aide à la compréhension du message vibratoire que délivre la machine et générer un traitement spécifique. Cette analyse traditionnelle calcule les indicateurs des signaux temporels

comme les statistiques descriptives telles que : moyenne, crête, crête-crête, écart type, facteur de crête, statistiques d'ordre supérieur : valeur efficace, skewness, kurtosis, etc.

Une approche populaire d'analyse dans le domaine temporel est le moyennage synchrone en anglais : Time Synchronous Average (TSA). L'idée est de réaliser la moyenne de plusieurs réalisations temporels à l'aide d'une référence de phase liée à la partie tournante (encodeur optique) afin d'essayer de réduire ou d'éliminer le bruit et l'effet d'autres sources et d'augmenter les composantes de signal d'intérêt. TSA est donné par

$$\tilde{s}(t) = \frac{1}{N} \sum_{n=0}^{N-1} s(t + nT), \quad 0 \leq t \leq T \quad (2.7)$$

Où $s(t)$ dénote le signal, T est la période du moyennage et N est le nombre d'échantillons de tranches du signal pour faire la moyenne. Les détails sur TSA ne seront pas discutés ici. Ils peuvent être trouvés dans les références de (Dalpiaz et al., 2000), (Miller, 1999a,b). Un bref examen de TSA a été donné par Dalpiaz et al., dans (Dalpiaz et al., 2000) et quelques inconvénients de TSA ont été précisés par Miller (Miller, 1999a,b).

Des approches plus avancées de l'analyse temporelle appliquent des modèles de série chronologique aux données du signal. L'idée principale est de modéliser le processus aléatoire de vibration par un modèle linéaire paramétrique dit ARMA (Auto-Regressif à Moyenne Ajustée), tiré d'un modèle plus générale dans l'identification des séries temporelles appelé modèle de Box et Jenkins ou un modèle non-linéaire de Prony (Key et al., 1981) et (Felkaoui, 1995), où l'on suppose que la série temporelle $x(i) = \{x_1, x_2, x_3, \dots, x_N\}$ est générée par l'excitation d'un bruit blanc $e(i)$. Le modèle sera :

$$x(i) = - \sum_{m=1}^p a_m x(i-m) + \sum_{m=1}^q b_m e(i-m) \quad (2.8)$$

$e(i)$: est un bruit blanc centré, de variance σ_e^2 . Ce modèle est souvent noté ARMA (p,q) où p et q sont respectivement les ordres des parties AR et MA et sont, respectivement, les nombres des pôles et des Zéros dans une fonction de transfert.

Dans la plupart des cas, on préfère mettre les b_m à zéro. On aboutit donc à la modélisation Auto-Regressive notée AR(p). Le modèle sera

$$x(i) = - \sum_{m=1}^p a_m x(i-m) + e(i), \quad (2.9)$$

où $x(i)$: une séquence d'observation,

$e(i)$: les échantillons d'un bruit blanc de moyenne nulle et de variance σ_e^2 .

dans lequel, il faut déterminer p_0 (le nombre d'harmonique sera alors $p_0 / 2$), et les \hat{a}_m ($m=1, \dots, p_0$); C'est-à-dire que, si l'ordre est connu, la solution de ce problème passe par la méthode du maximum de vraisemblance ou la méthode des moindres carrés et leurs variantes pour estimer les \hat{a}_m .

Poyhonen et al. dans (Poyhonen et al., 2004) ont appliqué le modèle d'AR aux signaux de vibration d'un moteur à induction et emploient les coefficients modèles AR comme indicateurs. Baillie et Mathew (Baillie et al., 1996) ont comparé les performances de trois techniques de modélisation autoregressive dans le cas du diagnostic de roulements, Garga et al., (Garga et al., 1997) ont proposé une approche qui emploie la modélisation AR suivi d'une réduction de la dimension. Récemment, (Zhan et al., 2003) ont utilisé la représentation d'état d'un modèle AR pour analyser les signaux vibratoires. (Zhan et al., 2006) ont proposé une nouvelle technique pour la détection de l'état d'une boîte de vitesse, qui ajuste un modèle autorégressif AR non stationnaire aux signaux résidus du mouvement d'un engrenage appliquant un filtrage de Kalman du bruit-adaptatif, dans la cas sans défaut de l'engrenage considéré. L'ordre optimal du modèle qui donne le meilleur compromis modèle/données du résidu de la roue sans défaut et sous différentes charges est déterminé avec l'aide d'une méthode de sélection spécifique proposée dans cette étude.

Il y a beaucoup d'autres techniques d'analyse dans le domaine temporel pour le diagnostic de défaut de machines. (Wang et al., 2001a) ont discuté trois méthodes de diagnostic non linéaires, connues sous le nom 'pseudo-phase portrait', 'singular spectrum analysis' et 'correlation dimension', basées sur l'analyse des séries temporelles du signal et la théorie d'analyse des séries temporelle. D'autres travaux sur l'application de ces méthodes sont : (Wang et al., 2003b) pour 'pseudo-phase portrait' et (Koizumi et al., 2000) (Wang et al., 2001b) pour la 'correlation dimension'. Zhuge et Lu (Zhuge et al., 1991) ont proposé un algorithme de moindres carrés modifié pour modéliser les signaux non-stationnaires 'impulse-like'. (Baydar et al., 2001a,b) ont exploré l'Analyse en Composantes Principales (ACP) pour étudier les signaux temporels dans le diagnostic de défaut d'engrenage.

2.4.2 Analyse fréquentielle

L'analyse spectrale permet de décomposer un signal complexe en composantes sinusoïdales caractérisées par leur amplitude et leur fréquence. L'interprétation d'un spectre sera autant plus précise et pertinente que le diagnosticien sera en possession d'un maximum de renseignements concernant les conditions de fonctionnement (vitesse et couple) et la cinématique de la machine. La qualité d'un spectre est principalement liée :

- à la résolution ou l'aptitude à dissocier deux composantes de fréquences proches,
- à la possibilité d'observer deux composantes d'amplitudes très différentes.

La qualité du spectre dépend du respect d'un nombre maximal de points cités dans la section 2.3.7 de ce chapitre.

L'analyse conventionnelle la plus répandue est l'analyse du spectre au moyen de transformée de Fourier rapide (FFT). L'idée principale de l'analyse du spectre est de regarder le spectre entier ou regarder étroitement à certaines composantes fréquentielles

utiles et extraire ainsi des indicateurs à partir du signal (voir, par exemple, (Schoen et al., 1995), (Almeida et al., 2002), et (Liu et al., 2004)). Généralement, l'outil le plus utilisé dans l'analyse du spectre est le périodogramme. IL est définie comme $E[X(f).X^*(f)]$, où $X(f)$ est la transformée de Fourier du signal $x(t)$, E dénote l'espérance et le $*$ dénote le conjugué complexe.

Sur le papier la transformée de Fourier représente l'instrument idéal pour rechercher les fréquences caractéristiques de défauts. Dans la pratique il en va, en fait, autrement et l'utilisation simple du spectre de Fourier, pour certains cas de diagnostic, ne suffit plus. C'est le cas en particulier des défauts qui engendrent des :

- faibles variations d'amplitude et de fréquence,
- faibles modulations,
- chocs et événements impulsionnels.

L'emploi de méthodes spécifiques mais complémentaires et auxiliaires à l'analyse de Fourier seront très utiles. Nous trouvons :

L'analyse d'enveloppe, la démodulation d'amplitude, de fréquence ou de phase (Ho et al., 2000), (Randall et al., 2001) et (Stack et al., 2004a,b), l'analyse des bandes latérales (Blankenship et al., 1995), ont été également employée pour la détection et le diagnostic des défauts de machine (Ho et al., 2000) et (Minnicino et al., 2004). Les détails des techniques mentionnées ci-dessus (section 2.3.7) et du spectre basé sur le FFT peuvent être trouvées dans les livres (Goldman, 1999) et (Harris et al., 2002) . Le rééchantillonnage angulaire et le Kurtosis spectral ont été suggérés par (Bonnardot, 2003) et (Antoni et al., 2000).

Le Cepstre a les possibilités pour détecter des harmoniques et des modèles de bande latérale dans le spectre de puissance. Il y a plusieurs versions du cepstre (El Badaoui, 1999) et (Harris et al., 2002). Parmi elles, le cepstre de puissance, qui est défini comme la transformée de Fourier inverse du logarithmique du spectre de puissance, c'est généralement le plus employé. Une analyse d'un cepstre modifié est proposée dans (Van der Merwe et al., 2002).

Une nouvelle technique appelée le 'holospectrum' a été présentée par Qu et al. (Qu et al., 1989) pour intégrer toute l'information de phase, d'amplitude et de fréquence d'un signal d'un signal temporel. L'application du 'holospectrum' aux signaux de vibration a été également étudiée dans (Yu et al., 1997) et (Chen et al., 1998).

D'une manière générale, Il y a deux classes d'approches pour l'évaluation du spectre de puissance. La première classe inclut les approches non paramétriques qui estiment la séquence d'autocorrélation du signal et puis appliquent la transformée de Fourier à l'autocorrélation estimée. Pour des détails, voir (Hayes, 1996). La deuxième classe inclut les approches paramétriques qui établissent un modèle paramétrique pour le signal et puis estiment le spectre de puissance basé sur le modèle adapté. Parmi elles, le spectre AR

(Mechefske et al., 1992), (Felkaoui et al., 1994), (Felkaoui, 1995), (Dron et al., 1998) (Stack et al., 2004) et le spectre ARMA (Salami et al., 2001).

2.4.3 Analyses temps-fréquence et temps-échelle

Les analyses des défauts dits simples et stationnaires (les vibrations dues aux balourds, les défauts de courroies, de désalignement, etc.) sont pour l'essentiel basées sur l'analyse de Fourier qui donne de bons résultats. Cependant, l'analyse de Fourier devient souvent inefficace en présence de défauts caractérisés par des variations complexes du spectre. En particulier s'il s'agit de signaux non stationnaires composés de chocs multiples et des signaux transitoires dont la durée dépend de la nature du choc (engrenage, roulement, moteur diesel et thermique, cavitation, desserrage de paliers, etc.) et dont l'instant d'apparition se déduit de la position de l'élément défectueux. Il est souhaitable de connaître par exemple le contenu fréquentiel et les évolutions des caractéristiques fréquentielles et temporelles d'un choc composé de signaux transitoires. La représentation comme fonction du temps du signal vibratoire exhibe mal le spectre des fréquences en jeu, alors qu'au contraire son analyse de Fourier masque l'instant d'émission et la durée de chacun des chocs. Également, le spectre d'un signal modulé en fréquence, par exemple chirp, montre une partie large bande donc il n'est pas révélateur de la structure précise du signal. Une philosophie s'est alors développée à partir de 1948 par J.Ville proposant de regarder l'évolution du contenu fréquentiel (ou énergétique) d'un signal en fonction du temps. Le but est d'obtenir une répartition de l'énergie du signal dans un plan temps-fréquence. La transformée de Fourier à court terme (TFCT) ou le spectrogramme (la puissance de la TFCT) (Wang et al., 1993a,b), (Andrade et al., 1999) et la distribution de Wigner-Ville (Meng et al., 1991), (Pan et al., 1998), (Koo et al., 2000), (Baydar et al., 2001a) sont les distributions de temps-fréquence les plus populaires. Cohen (Cohen, 1989) a passé en revue une classe des distributions de temps-fréquence qui incluent le spectrogramme, la distribution de Wigner-Ville, le Choi-Williams et d'autres. L'idée de STFT est de diviser le signal entier en segments avec la fenêtre à court terme et puis d'appliquer la transformée de Fourier à chaque segment. On fait alors l'hypothèse d'une quasi-stationnarité à l'intérieur de la fenêtre. Le spectrogramme a une certaine limitation dans la résolution de temps-fréquence due à la segmentation de signal. Il peut seulement être appliqué aux signaux lentement variables. La transformée Bilinéaire comme la distribution de Wigner-Ville ne sont pas basés sur la segmentation du signal et ne surmontent pas ainsi la limitation de résolution de temps-fréquence du spectrogramme. Cependant, il y a un inconvénient principal de la transformée bilinéaire c'est la présence des interférences. Ce qui réduit la lisibilité dans le cas de signaux à grand nombre de composantes (Bonato et al., 1997). Différentes versions de cette représentation peuvent être données pour remédier, principalement, à cet inconvénient. Elles découlaient toutes de la même formule (Chiollaz et al., 1987, 1992), (Flandrin, 1993). On cite, par exemple, la pseudo Wigner Ville et La Pseudo Wigner Ville lissée, la représentation de Rihaczek,

de Cohen, de Choï-Williams, etc. Gu et al., (Gu et al., 2002) ont appliqué la décomposition en valeurs singulières pour extraire des indicateurs de la distribution temps fréquence. Loughlin et al. (Loughlin et al., 2000) ont employé une série de moments en tant qu'indicateurs pour le diagnostic de défaut.

Une autre transformée pour l'analyse de signaux non-stationnaires et transitoires est la transformation en ondelettes. À la différence des distributions ci-dessus qui sont des représentations de temps-fréquence d'un signal, la transformation en ondelettes est une représentation temps-échelle d'un signal. Cette théorie a été développée depuis 1984, (soit près de 200 ans après les théories de Fourier) notamment par J. Morlet, M. Barrat, S.G.Mallat, Y. Meyer et I. Daubechies, voir (Daubechies, 1992). La transformée en ondelettes est un outil alternatif en traitement de signaux non-stationnaires. Plutôt qu'un **ensemble de sinusoides** de durée théoriquement infinie dans le cas TFCT, l'analyse en ondelettes utilise une collection de **fonctions** $\psi(t)$ localisées en temps, appelées ondelettes. Une infinité d'ondelettes est à notre disposition. Mais rien ne nous empêche de créer notre propre ondelette. Le choix de l'ondelette peut être délicat mais très important car il conditionne la qualité des résultats obtenus. Une transformée en ondelettes continue est définie dans le domaine temporel et dans le domaine fréquentiel. Comme pour la transformée de Fourier, la transformée en ondelettes a également sa forme discrète, et sa forme rapide Fast Wavelet Transform.

Un avantage principal de transformation par ondelettes est sa capacité de produire une très grande résolution fréquentielle à de basses fréquences et une très grande résolution temporelle à de hautes fréquences pour des signaux longue durée avec de basses fréquences et de courte durée avec hautes fréquences. Un autre avantage de la transformée en ondelettes est sa capacité de réduire le bruit des signaux.

La transformée en ondelettes a été appliquée avec succès à l'analyse de données temporelles pour le diagnostic des défauts d'engrenage (Staszewski et al., 1994), (Wang et al., 1996) de roulements (Rubini et al., 2001), (Luo et al., 2003) et d'autres systèmes mécaniques (Aretakis et al., 1997), (Chandroth et al., 1999). Dalpiaz et Rivola (Dalpiaz et al., 1997) (Mahgoun et al., 1999) ont évalué et ont comparé l'efficacité et la fiabilité de la transformée en ondelettes à d'autres techniques d'analyse de signal de vibration. Baydar et Ball (Baydar et al., 2003) ont appliqué la transformée en ondelettes aux signaux de vibration et d'acoustique. Addison et al., (Addison et al., 2002) ont étudié l'utilisation des ondelettes complexes, le chapeau mexicain et l'ondelette de Morlet, entant qu'outils de détection. L'analyse en ondelettes employant l'ondelette de Haar a été considérée dans (Xu et al., 1991) (Tonshoff et al., 2003). Miller et Reichard (Miller et al., 1999a,b) utilisent une base d'ondelettes comme peigne de filtres pour décomposer les signaux vibratoires. Un outil graphique appelé les cartes polaires d'ondelette pour montrer l'amplitude et la phase de l'ondelette a été proposé dans (Boulahbal et al., 1999) et a été appliqué au diagnostic de défaut d'engrenage dans (Meltzer et al., 2004). La transformée

en ondelettes combinée avec la transformée de Fourier pour améliorer les possibilités d'extraction d'indicateurs a été proposé dans (Wang et al., 2003). Une transformée plus avancée, connu comme la transformée en paquets d'ondelettes, a été étudié dans (Yen et al., 2000), (Zhang et al., 2005), (Toliat et al., 2003). Une nouvelle technique connue sous le nom 'basis pursuit' basée les paquets d'ondelettes a été appliquée pour analyser les signaux de vibration pour le diagnostic de défaut de roulement (Yang et al., 2005). On montre que la 'basis pursuit' a quelques avantages par rapport à d'autres approches généralement utilisées en analyse d'ondelettes. Lin et al., (Lin et al., 2004) utilisés la technique de débruitage de seuil d'ondelettes pour extraire l'information sur les défauts à partir des signaux bruités. Une revue récente avec des discussions plus étendues et plus de références sur les applications de la transformée en ondelettes pour le traitement des signaux dans la surveillance d'état de machine et le diagnostic de défaut a été donnée dans (Pen et al., 2004).

2.4.4 Analyse Spectrale d'Ordre Supérieur à 2

L'observation du fonctionnement de nombreux systèmes industriels tels que les engrenages et les roulements a fait sortir l'existence de séquence non linéaires pouvant être liées à des phénomènes de glissement ou à l'apparition de certains types de défauts notamment les défauts de profil. Le spectre d'ordre supérieur c.-à-d. bispectrum ou trispectrum (outil d'analyse des signaux stationnaires) peut fournir plus d'information de diagnostic pour les signaux non gaussiens et non linéarités. Il révèle d'éventuels couplages (interactions) existants entre les fréquences composants un signal. Par exemple, le bispectre est l'outil idéal pour la recherche de couplages quadratiques, il fait apparaître une fréquence qui est la somme des deux fréquences concernées est donc aisément identifiable. Il y a donc une fois encore un gain en terme d'information.

Dans la littérature, le spectre d'ordre supérieur s'appelle également les statistiques d'ordre supérieur (McCormick, 1998a,b, 1999), (Wang et al., 2002). Ce nom vient du fait que le bispectrum et le trispectrum sont réellement la *transformée de Fourier des statistiques de troisième et quatrième ordre* du signal temporel respectivement (bicorrrelation et tricorrrelation). L'analyse bispectrale a été largement appliquée dans le diagnostic de machines pour différents systèmes mécaniques tels que les engrenage (Xiong et al., 2002), (Bouillaut, 2000), (Raad, 2003) et les roulements (Yang et al., 2002), les machines rotatives (Parker et al., 2000) et les machines à induction (Chow et al., 1995), (Arthur et al., 1997). Li et al. (Li et al., 2003) ont étudié l'application de la partie diagonale du bispectre pour le diagnostic de défaut d'engrenage. Yang et al. (Yang et al., 2002) ont employé à la fois la partie diagonale du bispectre et la diagonale de la bicoherence (bispectre normalisé à la puissance du signal.). Ils sont des indicateurs des relations statistiques entre les fréquences). L'application du bispectre et du trispectre au diagnostic de défaut de roulement a été discutée dans (McCormick et al., 1999).

2.4.5 Analyse temps multi-fréquences

L'analyse des signaux stationnaires examine le contenu fréquentiel des signaux, ou la répartition de l'énergie sur les fréquences éternelles contenues dans le signal. D'autre part, une analyse multispectrale (corrélation spectrale, bispectre, trispectre, ...) révèle d'éventuels couplages (des non linéarités) existants entre les fréquences composants un signal. Dans un cas non-stationnaire, des fréquences peuvent s'évanouir dans le temps ou apparaître. Effectuer une analyse supposant le signal stationnaire peut alors conduire à des interprétations erronées, et il faut incorporer à l'analyse la **dimension temporelle** pour espérer interpréter correctement le contenu du signal. Il est donc très intuitif pour effectuer des analyse d'ordre supérieur des signaux non-stationnaires d'essayer de rendre les objets déjà définis dépendant du temps.

Les articles traitants le sujet sont nombreux nous pouvons trouver entre autres (Swami et al., 1997) et (Amblard, 1994) mais les applications en mécanique manquent, voir la thèse de (McCormick, 1999).

2.4.6 Analyse cyclostationnaire

L'analyse des signaux cyclostationnaires a pour origine le domaine des télécommunications où l'on peut transmettre des signaux dont les propriétés statistiques sont périodiques. Cette analyse a été largement exploitée pour la détection de défauts de roulement (Antoni, 2000), (Antoni, 2007). Cette hypothèse correspond à la nature physique des signaux issus des machines tournantes (engrenages roulement, etc.).

Généralement, on a coutume d'utiliser l'analyse synchrone. En fait, on exploite la cyclostationnarité d'ordre 1, c'est-à-dire, de la moyenne. Cette approche est réellement prometteuse pour l'analyse et le diagnostic du bruit et des vibrations, par exemple pour l'étude des corrélations entre harmoniques, et bien sûr pour l'étude des phénomènes cycliques dans les machines alternatives.

Dans le cadre de l'analyse des signaux d'*engrenages* (défauts d'écaillage ou des fissures) de nombreuses études montrent que les vibrations sont cyclostationnaires; notamment (Capdessus, 1992), (Bouillaut, 2000) et (Raad, 2003). Plusieurs études sont menées actuellement dans ce domaine, y compris le lien avec le bispectre et la bicohérence qui exploitent des liens entre les composantes. Dans (Capdessus, 1995), (Li et al., 2003), il a été également montré que les liens qui apparaissent dans la **corrélation spectrale**, outil cyclostationnaire, permettaient de détecter de manière plus précoce un défaut d'engrenage (les défauts apparaissant à certaines fréquences cycliques et rendent le signal de plus en plus cyclostationnaire).

Les principaux outils rencontrés dans les études sont l'autocorrélation cyclique, les indices et les degrés de cyclostationnarité mais surtout la corrélation spectrale.

2.4.7 Analyse Spectrale Cyclique d'Ordre Supérieur

Le spectre cyclique d'ordre supérieur c'est-à-dire les multispectres cycliques sont encore très peu explorés. Elles résultent de la combinaison des classes cyclostationnaires et d'ordre supérieur. Cet outil d'analyse généralise les multispectres utilisés dans le cas stationnaire classique au cas cyclostationnaire. La force de cet outil est qu'il est compact et permet de joindre les propriétés cyclostationnaires ainsi que celles non-linéaires des processus. Des estimateurs du multispectre cyclique sont proposés par Gardner et Giannakis et bien d'autres, voir pour plus de détails la thèse de Raad (Raad, 2003).

En 2001, Bouillaut a introduit l'idée d'utiliser le bispectre cyclique pour des applications mécaniques (Bouillaut, 2001). Raad dans (Raad, 2003) a marié les propriétés du bispectre et de la corrélation spectrale pour aboutir à une formulation générale, la première dans son genre, où toutes les informations de base sur le bispectre cyclique sont explicitées.

Une nouvelle terminologie est proposée par analogie au cas du bispectre et aux outils temporels telle que : la bicohérence cyclique, le skewness, le kurtosis cyclique, le cepstre cyclique. Suite aux nouveaux aspects du bispectre cyclique, une partie cruciale de la thèse de Raad est consacrée à une application mécanique complexe qui concerne une boîte de vitesse des signaux d'hélicoptères provenant de la Navy américaine. Les défauts de roulement ont été étudiés par de Li et Qu dans leur article (Li et al., 2003).

2.4.8 Signaux multidimensionnels

Dans de nombreuses situations en surveillance vibratoire, les signaux de contrôle reçus sur un réseau de capteurs sont constitués par la somme de contributions diverses dues à différents éléments du système étudié et à différents phénomènes.

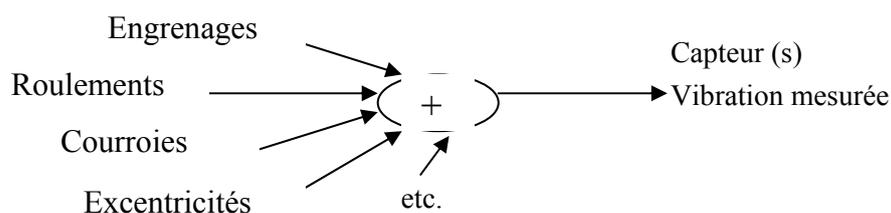
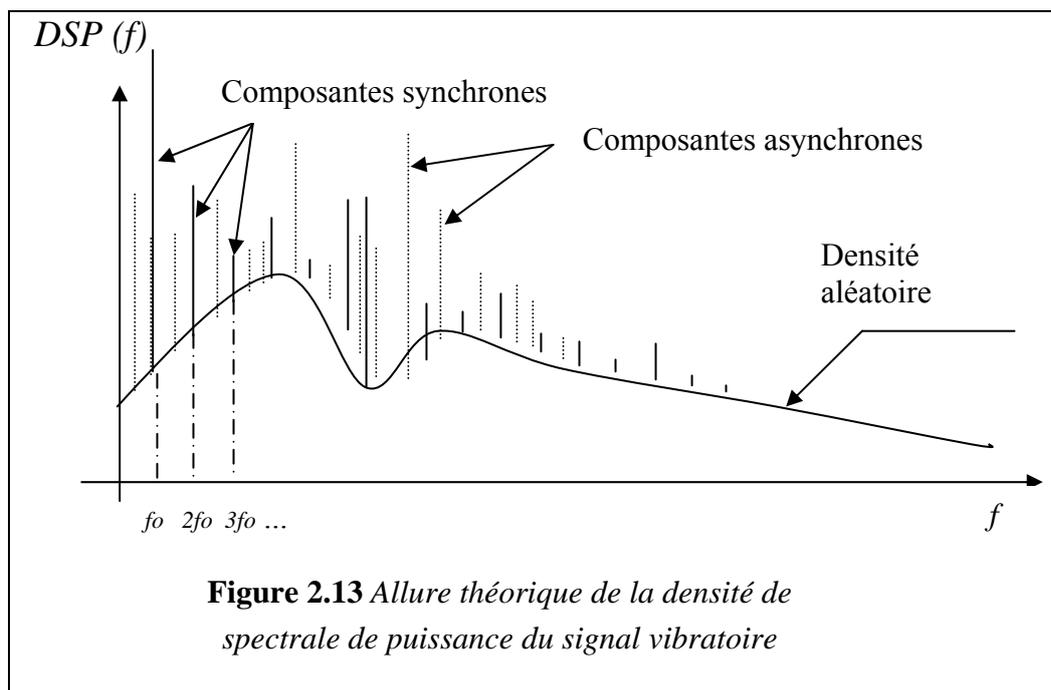


Figure 2.12 *Signal capté sur une machine*

L'allure générale de la densité de puissance d'un tel signal est de la forme figure 2.13.



On trouve dans cette DSP, une partie continue correspondant à la composante large bande, et une partie discrète formée d'un ensemble de distributions de Dirac, correspondant à la composante déterministe du signal de vibrations traité.

Dans le contexte des signaux vibratoire, il est souvent très difficile, sinon impossible de faire un diagnostic fiable et de suivre l'évolution d'un des défauts. Dans cette situation, en littérature, on a appliqué et développé, des techniques de séparation dites "aveugles" puisqu'elles ne supposent aucune connaissance préalable sur le système.

Les hypothèses généralement retenues pour la séparation de sources sont :

- La linéarité : on suppose que la transmission entre les émetteurs et les récepteurs est linéaire.
- L'indépendance statistique : on admet que les excitations émises par les différentes sources sont assez différentes pour pouvoir être considérées comme des signaux aléatoires statistiquement indépendants.

Différentes méthodes et algorithmes de calcul ont été développées. Nous citons entre autres les algorithmes les plus répandus :

- FICA : Fast ICA (Independent Component Analysis),
- JADE : Joint Approximate Diagonalization of Eigen Matrix
- TICA : Thin QR/SVD algorithm for ICA (extension of Fixed Point ICA)

Pour plus de détails le lecteur est invité à consulter le numéro spéciale dédié à la Séparation aveugle de source (BSS) de la revue « Mechanical Systems and Signal

Processing volume 19 de l'année (2005) » 16 articles sont présentés, les thèses de (Bonnardot, 2005), (Ypma, 2001), etc.

Remarque : Nous soulignons que cette étude bibliographique des méthodes de traitement du signal a été soldée par la mise en place, au niveau de notre laboratoire d'un logiciel interactif de calcul de la plupart de ces outils travaillant sous environnement MATLAB. Et que le nombre très élevé de paramètres définis jusqu'ici donne un grand intérêt à certains travaux pour l'analyse de données et la prise de décision automatique. Il est donc préférable de concevoir des indicateurs pouvant remplacer l'inspection visuelle des différentes figures par une simple procédure. Les techniques de génération d'indicateurs fait l'objet de la section suivante

2.5 Génération d'indicateurs

L'étape précédente du traitement de signal sert principalement à extraire et à mettre sous forme explicite l'information contenue dans le signal. Cette dernière donne dans la plupart des cas des courbes de deux ou trois dimensions qui peuvent servir comme test visuel pour la présence ou non de défauts. Dans la présente étape, il s'agit de transformer l'état qualitatif des différentes représentations en indicateurs simples permettant une représentation parcimonieuse mais informative de l'état de santé de la machine. On doit générer des indicateurs simples à mettre en œuvre et qui assurent la distinction entre les différents niveaux de dégradation (classes). Donc, établir sans aucune équivoque, l'évolution des indicateurs avec l'importance de la dégradation. Dans cette section, nous nous proposons donc de donner les principaux outils qui sont utilisés pour extraire des indicateurs. Ces outils sont censés fournir des paramètres suffisamment pertinents vis à vis d'une ou plusieurs dégradations. Cette pertinence, dans le cas de la surveillance et du diagnostic, doit s'effectuer en reliant de façon monotone l'importance de la dégradation avec l'évolution de l'indicateur.

L'indicateur doit posséder un fort potentiel descriptif tout en restant *insensible à certaines opérations sur les signaux définies comme des invariants du problème et sensibles à la puissance du signal*. La qualité d'un indicateur est grande lorsque ce descripteur est indépendant des grandeurs d'état qui caractérisent la machine (température, fluctuation de vitesse, ou de couple, etc.) et de l'environnement. L'efficacité du diagnostic dépendra de l'efficacité des différents **critères ou symptômes** évoqués et de la façon dont on sera capable d'en tenir compte. Il est nécessaire de choisir les traitements les plus adaptés et de les réaliser d'une façon pratique. Les sections 2.5.2 à 2.5.8 présentent plusieurs types d'indicateurs en fonction des hypothèses adoptées. Elles présentent un calcul pratique de ceux-ci.

2.5.1 Classification des indicateurs

La grande diversité des indicateurs, leur mode d'élaboration, les grandeurs physiques et phénomènes qu'ils peuvent représenter et leur niveau de pertinence nécessite une classification. On distingue deux grandes familles selon leur nature et leur degré d'élaboration, voir (Pachaux & Canetto surveillance 4 pp.539). :

- les indicateurs énergétiques globaux qui représentent l'énergie du signal vibratoire mesurée dans la bande fréquentielle plus ou moins étendue sans relation identifiée avec les forces dynamiques dont la machine est le siège, et
- les indicateurs topologiques qui sont en relation directe avec ces machines et, de ce fait, en bien meilleure adéquation avec les défauts susceptibles de l'affecter.

Ces indicateurs doivent permettre une surveillance d'anomalies à un **stade précoce**, déterminant la nature et les états de dégradation.

2.5.2 Indicateurs temporels et statistiques

Les indicateurs temporels et statistiques sont de bons candidats en vue de la caractérisation du signal vibratoire par un nombre réduit d'indicateurs. Les paramètres sont alors l'écart type, la variance et les différents cumulants d'ordre 3, 4, 5, etc. incluant aussi les outils conventionnels tels que le skewness et le Kurtosis. On peut calculer autant d'indicateurs quand le veut. Il s'agit de transformer l'état qualitatif de la machine en indicateurs statistiques simple, d'ordre supérieur 2 ou cycliques d'ordre supérieur 2. Pour le calcul de ces indicateurs l'utilisation des cumulants d'ordre n est plus appropriée que celle des moments d'ordre n (Voir les propriétés des cumulants (Lacoume et al., 1997)).

Dans le cas des variables aléatoires centrées (de moyennes nulles) et si on considère les termes des écarts i, j, k nuls (valeurs à l'origine) les expressions des multivariations se simplifient en :

$$R(0) = C_2(0) = E[x^2] = \frac{1}{N} \sum_{n=1}^N x(n)x(n) \quad (2.10)$$

C'est le cumuland d'ordre 2 à l'origine ou le moment d'ordre 2 à l'origine ou la variance notée par σ_x^2 .

$$R(0,0) = C_3(0,0) = E[x^3] = \frac{1}{N} \sum_{n=1}^N x(n)x(n)x(n) \quad (2.11)$$

C'est le cumuland d'ordre 3 à l'origine ou le moment d'ordre 3 à l'origine

$$R(0,0,0) = C_4(0,0,0) = E[x^4] - 3E[x^2]^2 = \frac{1}{N} \sum_{n=1}^N x(n)x(n)x(n)x(n) - 3C_2(0)C_2(0) \quad (2.12)$$

C'est le moment d'ordre 4 à l'origine moins 3 fois le carré du cumulants d'ordre 2 à l'origine.

Il est à noter que plusieurs **indicateurs** biaisés de retard nul issus de la normalisation de ces facteurs de cumulants sont couramment rencontrés dans l'étude des chocs et vibrations. Nous trouvons :

$$V = C(0) = E[x^2] : \text{La variance ou la valeur efficace}$$

$$\sigma = \sqrt{V} = \sqrt{C_2(0)} = \sqrt{E[x^2]} : \text{L'écart type}$$

$$Fc = 20 \log \frac{x_{\max}}{\sqrt{C_2(0)}} = 20 \log \frac{x_{\max}}{\sigma} : \text{Le facteur de Crête} \quad (2.13)$$

$$Sk = \frac{C_3(0,0)}{\sqrt{C_2^3(0)}} = \frac{E[x^3]}{\sqrt{E^3(x^2)}} = \frac{E[x^3]}{\sqrt{V^3}} : \text{Le skewness}$$

$$K = \frac{C_4(0,0,0)}{C_2^2(0)} = \frac{E[x^4]}{E^2[x^2]} - 3 = \frac{E[x^4]}{V^2} - 3 : \text{Le Kurtosis}$$

Les grandeurs de dispersion Sk et K sont appelées les coefficients de Fisher ou respectivement les paramètres de dissymétrie et d'aplatissement par rapport à la normale. Dans la référence (Lacoume & al., 1997) on donne les valeurs numériques pour les différentes variables aléatoires.

- V.A.Uniforme : $Sk = 0$ $K = -6/5$
- V.A.Gaussienne : $Sk = 0$ $K = 0$
- V.A.Exponentielle : $Sk = 2$ $K = 6$
- V.A. de Laplace bilatérale : $Sk = 2$ $K = 3$
- V.A. Sinusoïdale : $Sk = 0$ $K = -3/2$
- V.A. Chi-deux dans le cas central : $K = 48\nu$, ν : le degré de liberté de la variable.

Dans le cas pratique pour des signaux gaussiens le skewness et le kurtosis sont théoriquement nuls : leur valeur proche de zéro est issue des erreurs d'estimation. Dans le cas des signaux non gaussiens, ils sont significativement différents.

Remarque sur le Kurtosis

Dans le cas d'une forte détérioration du roulement, l'allure de la distribution de l'amplitude redevient gaussienne avec $K \approx 0$, mais une augmentation importante de l'amplitude efficace est constatée. Par conséquent, comme pour le facteur crête, il y a lieu de tenir compte simultanément de l'évolution de deux critères : Le Kurtosis et l'amplitude efficace.

Remarque sur les Cumulants

Les cumulants d'ordre supérieur à 2 ont été introduits par les statisticiens dans le but d'extraire de nouvelles informations à partir de données. Les principales motivations sont qu'elles offrent des possibilités supplémentaires non disponibles à l'ordre 2. En permettant notamment (Lacoume et al., 2007):

- la mesure de l'écart à la gaussienneté,
- la suppression de bruit gaussien de densité spectrale inconnue,
- la caractérisation et la détection de phénomènes non linéaires.

1. Le cas d'un processus Gaussien :

Les cumulants d'un processus aléatoire Gaussien sont nuls pour des ordres supérieurs à 2, exemple,

$$\begin{aligned} C_3(i, j) &= 0, \\ C_4(i, j, k) &= 0, \end{aligned} \tag{2.14}$$

etc.

Cette propriété des statistiques d'ordre supérieur mesure l'écart à la gaussienneté. Elle est utilisée par les chercheurs pour tester la non-gaussienneté d'un processus et pour la détection et la classification des signaux non-gaussiens. Elle peut être utilisée pour séparer les signaux non-gaussiens indépendants et le bruit additif gaussien.

Nous soulignons, également, que les statistiques d'ordre 2 ne permettent pas de retrouver l'information de phase sur un système linéaire (phase minimale, phase polynomiale, phase quadratique, ...). Qui peut être porteuse d'information, sur l'état du système, non disponible sur l'amplitude ou la fréquence.

2. Le cas du Bruit Additif Gaussien :

Si $z(n) = S(n) + g(n)$, où le signal $S(n)$ est statistiquement indépendant du 'bruit' $g(n)$ et si $g(n)$ est gaussien, alors

$$C_2(z(i)) = C_2(S(i)) + C_2(g(i)) \tag{2.15}$$

mais pour les cumulants d'ordre supérieur :

$$\begin{aligned} C_3(z(i, j)) &= C_3(S(i, j)), \\ C_4(z(i, j, k)) &= C_4(S(i, j, k)), \end{aligned} \tag{2.16}$$

etc.

Noter ainsi, qu'au lieu de considérer l'hypothèse que le bruit additif est négligeable, en utilisant les SOS, les effets du bruit gaussien sont éliminés dans les domaines du cumulants d'ordre supérieur sur les différentes méthodes d'estimation. Noter également que $C_3(z(i, j)) = C_3(S(i, j))$ est valable si le bruit $g(n)$ est symétriquement distribué (exemple, gaussien, laplace, uniforme, etc.).

Il apparaît, donc, intéressant d'utiliser les SOS si le signal est entaché d'un bruit gaussien inconnu car il perturbe l'estimation par statistiques d'ordre 2, mais n'influence pas les estimations d'ordre supérieur. Cette propriété est utilisée pour supprimer le bruit additif dont la fonction d'autocovariance peut ne pas être connue.

3. Le cas d'un processus non-linéaire :

Le cumulants d'ordre 2 ne peut donner une évidence de la non-linéarité. Par exemple, la présence de couplages quadratiques (QPC) ou cubiques (CPC) de phases ne peuvent être détectés en utilisant l'autocorrélation. Mais, elles peuvent être détectés et quantifiés en utilisant les cumulants d'ordre 3 et d'ordre 4.

On peut étudier des cas particuliers des cumulants (les diagonales $i=j$ ou $i=j=k$).

2.5.3 Indicateurs fréquentiels

Pour des machines tournantes, on sait qu'il est intéressant de transformer les données temporelles sous la forme d'une représentation fréquentielle (densité spectrale de puissance (DSP)) qui donne directement la part d'énergie présente pour la composante spectrale recherchée. Les indicateurs sont alors le calcul de l'énergie " e_i " par bande spectrale " i ", dont les bornes fréquentielles sont choisies au vue des évolutions de cette énergie en fonction des modes de fonctionnement de la machine (bon état, peu dégradé, fortement dégradé, etc.).

$$e_i = \sum_{n=K_i}^{K_i+l} DSP[n] \left\{ \begin{array}{l} i = 1, \dots, p \\ n = [K_1, \dots, K_1 + l], [K_2, \dots, K_2 + l], \dots \end{array} \right\} \quad (2.17)$$

On obtient ainsi " p " valeurs qui représentent les parts d'énergies totales présentes dans les " p " zones de la DSP.

2.5.4 Indicateurs du modèle AR ou AR évolutif

Les paramètres d'un modèle paramétrique d'un processus objet peuvent jouer le rôle d'identificateur de celui-ci. En particulier si le processus évolue dans le temps, soit par vieillissement soit par usure progressive. Les paramètres de son modèle comportemental évoluent également et peuvent inversement être un indicateur de "l'état de santé" du

processus à surveiller. Nous présentons ci-dessous, la méthodologie utilisée pour estimer les paramètres d'un modèle AR et AR évolutif.

Nous

2.5.4.1 Modélisation AR

1. recherche de l'ordre
2. estimation des paramètres a_i ($i = 1, 2, \dots, p$) qui représenteront un vecteur de forme pour la classification.

On peut utiliser également les coefficients de réflexion du filtre de la structure en treillis K_i ($i = 1, 2, \dots, p$) ou les coefficients cepstraux C_i ($i = 1, 2, \dots, p$) calculés à partir des paramètres a_i .

2.5.4.2 Modélisation AR évolutif

Dans le cas d'un de système à évolution lente, on peut utiliser les paramètres a_{ij} pour créer un vecteur de forme pour la classification.

2.5.5 Indicateurs temps-fréquence

Les méthodes temps fréquences fournissent des représentations temps-fréquence (RTF) qui permettent de caractériser plus efficacement les signaux vibratoire non stationnaires. Afin d'utiliser ces outils, il est nécessaire d'appliquer d'autres techniques sur ces représentations *cartographiques* afin de condensée l'information et extraire des indicateurs pertinents. On peut citer sans exhaustivité, les techniques de traitement d'images, la décomposition en valeurs singulières et le calcul des énergies dans certaines zones qui permettent de générer des indicateurs en nombre restreint. L'outil du traitement d'image ne sera pas présenté dans ce mémoire.

2.5.5.1 La décomposition en valeurs singulières

La décomposition en valeurs singulières (SVD) (Felkaoui, 1995), (Benoughram, 1999) appliquée à la matrice d'une Représentation temps-fréquence ou temps échelle permet de donner un nombre significatif d'indicateurs. L'Algorithme du SVD est l'un des plus stable du point de vue calcul en présence des erreurs de données bruitées dans la théorie de l'algèbre des matrices (Golub et al., 1996).

Dans la théorie du SVD, toute matrice $m \times n$ réelle, évaluée Q de rang p , où $p \leq \min(m, n)$, peut être décomposée comme :

$$Q = URV \quad (2.18)$$

Où U est la matrice $m \times n$, V est une matrice diagonale, orthogonale $n \times n$, et

R est une matrice diagonale, les éléments diagonaux α_i de R sont les valeurs singulières de Q arrangées dans un ordre croissant.

Toutes les valeurs singulières ne sont pas négatives et exactement p éléments entre eux sont strictement positifs $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_p \geq \alpha_{p+1} = \dots = \alpha_n = 0$.

En théorie, le rang de Q peut être déterminé en comptant le nombre de ces valeurs singulières. En pratique, on utilise toujours la matrice observée réelle Q_1 constituée de la matrice Q perturbée par les erreurs de troncature dues à la précision finie des opérations numériques et de la matrice E des bruits aléatoires.

La plus simple procédure jusqu'ici utilise un modèle de perturbation additif donné par $Q_1 = Q + E$, tandis que la matrice Q ($m \times n$) est supposée avoir le rang $p \leq \min(m, n)$, la matrice de perturbation E à de valeurs petites, mais elle est de rang plein. Alors, la décomposition en valeurs singulières de Q_1 donnera une matrice $R = \text{diag}(\beta_1, \dots, \beta_n)$, où $\beta_1 \geq \beta_2 \geq \dots \geq \beta_p \geq \beta_{p+1} \geq \dots \geq \beta_n \geq 0$, et $\beta_{p+1}, \dots, \beta_n$ seront d'habitude petites, mais pas nécessairement nulles, ainsi β_p peut être également petite, le problème est de déterminer quelle valeur singulière est significative. En général, nous voulons avoir une méthode précise de détermination du rang effectif en utilisant les valeurs singulières observées. Plusieurs critères ou règles ont été proposés pour trouver le rang effectif exact (p_0 valeur limite de l'ordre) significatif, le plus prometteur, est le critère suivant (Golub et al., 1996) :

$$\beta_p \ll \beta_{p+1}, \quad (2.19)$$

Où le rang est estimé égal à k si :

$$\frac{\beta_k}{\beta_{k+1}} = \max_{p=1, \dots, p_{\max}} \frac{\beta_p}{\beta_{p+1}}, \quad (2.20)$$

Pour revenir à ce qui nous intéresse ici, on peut considérer les " p " valeurs singulières comme une condensation de l'information et une extraction de paramètres pertinents, utilisables par des outils de reconnaissance de formes. On obtient ainsi, par ce biais, le moyen de caractériser une représentation tridimensionnelle (RTF) avec seulement quelques paramètres.

2.5.5.2 Calcul énergétique sur les représentations temps fréquence

En se basant sur les propriétés des énergies des RTF (Bardou, 1994 p.41) et (Benoughram, 1999 p. 24), On peut générer des paramètres énergétiques e_i en découpant les axes temporel et fréquentiel en plusieurs morceaux, l'énergie sera dans une portion donnée par :

$$e_i = \sum_{m=l_i}^{l_i+l} \sum_{n=K_i}^{K_i+l} RTF[m, n] \left\{ \begin{array}{l} i = 1, \dots, p \\ m = [l_1, \dots, l_1 + l], [l_2, \dots, l_2 + l], \dots \\ n = [K_1, \dots, K_1 + l], [K_2, \dots, K_2 + l], \dots \end{array} \right\} \quad (2.21)$$

On obtient ainsi "p" valeurs qui représentent les parts, de l'énergie totale, présentes dans les "p" zones de la RTF. Dans chacune des zones un calcul énergétique est effectué en sommant une à une toutes les valeurs comprises "RTF [m,n]" dans la zones considéré.

2.5.6 Indicateurs de la fonction de modulation d'Amplitude, de Phase et de Fréquence (APF)

Très adaptés à l'identification de défauts induisant essentiellement des fluctuations de couple ou de vitesse, voir (Pachaud & Canneto, p.633). Pour cela :

- il faut faire un spectre de la fonction de modulation de l'amplitude, de la phase ou de la fréquence de la composante d'ordre 1, 2, 3, ... de la vitesse de rotation, du courant d'alimentation, du couple, de la pression, du débit, ...
- utiliser un critère énergétique ou scalaire comme paramètre, par exemple, le Taux de modulation APF qui est un indicateur scalaire. La démodulation (APF) s'avère particulièrement bien adaptée au diagnostic des machines tournantes et à la détection **précoce** d'un certain nombre de défauts.

2.5.7 Indicateurs des statistiques d'ordre supérieur

Quand on utilise les multicorrélations on obtient des fonctions dépendant de plusieurs variables : deux pour la bicorrélation $R_{x(3)}(i, j)$, trois pour la tricorrélation $R_{x(4)}(i, j, k)$, etc. En général l'ensemble de ces vecteurs, matrices, etc. donnent trop d'informations. Si pour la bicorrélation l'interprétation est facile pour la tricorrélation il apparaît une difficulté d'ordre pratique qui vient de la représentation d'une fonction à trois variables. On se contente souvent de présenter des coupes à deux dimensions ou des lignes à une dimension de la tricorrélation.

On se limitera donc à certaines parties de l'espace des retards ne dépendants que d'un seul paramètre. On appelle ces parties de la multicorrélation des lignes de multicorrélation. On démontre également l'intérêt particulier des diagonales (Lacoume & al., 97 p.80).

Pour la multicorrélation d'ordre « p » d'un signal $x(n)$, la diagonale est :

$$R_{x(p)}(k) = R_{x(p)}(k, k, \dots, k) \quad (2.22)$$

On utilise également des lignes de multicorrélation obtenues en fixant toutes les valeurs des retards sauf une, la ligne est :

$$R_{x(p)}(k) = R_{x(p)}(k, l_2, \dots, l_{p-1}), \text{ les } l_2, l_3, \dots, l_{p-1} \text{ sont successivement fixés.}$$

On peut ainsi définir « p » corrélations de lignes. Mais étant données les symétries des multicorrélations de signaux à valeurs réelles, ces « p » corrélations de lignes sont égales.

2.5.8 Indicateurs des spectres d'ordre supérieur

Les spectres d'ordre supérieur (SOS), permettent de caractériser des non linéarités. Pour un signal aléatoire, réel, centré et stationnaire à l'ordre 3, le bispectre est défini par la double transformation de Fourier (transformée de Fourier à deux dimensions) du cumulatif d'ordre 3,

$$B(f_1, f_2) = \sum_m \sum_n C_3(m, n) e^{-j2\pi(mf_1 + nf_2)} \quad (2.23)$$

Le bispectre peut être calculé aussi par la formule fréquentielle suivante

$$Bx(f_1, f_2) = X(f_1)X(f_2)conj(X(f_1 + f_2)) \quad (2.24)$$

Où $X(f_i)$: représente la transformée de Fourier de $x(n)$

Noter que le bispectre est une fonction de deux variables fréquentielles. D'une façon similaire, le trispectre est défini comme la transformée de Fourier à 3 dimensions du cumulatif d'ordre 4, est c'est une fonction de 3 variables fréquentielles.

Dans la littérature, On définit les multicohérences comme des multispectres normalisés. La bicohérence et la tricohérence sont de bons indicateurs de la gaussiannité et de la linéarité du signal.

Si $x(n)$ est le signal, $DSP_1(x(n))$ le spectre, $DSP_2(f_1, f_2)$ le bispectre. Alors, le carré de la bicohérence de $x(n)$, $c_2(f_1, f_2)$, est donné par :

$$c_2^2(f_1, f_2) = \frac{DSP_2(f_1, f_2)}{DSP_1(f_1)DSP_1(f_2)DSP_1(f_1 + f_2)} \quad (2.25)$$

et le carré de sa tricohérence est :

$$c_2^2(f_1, f_2, f_3) = \frac{DSP_2(f_1, f_2, f_3)}{DSP_1(f_1)DSP_1(f_2)DSP_1(f_3)DSP_1(f_1 + f_2)} \quad (2.26)$$

Ces indicateurs normalisés sont ***insensibles à la puissance du signal*** : il sont un indicateur des relations statistiques existant entre les différentes fréquences.

Pour un signal gaussien, les muticohérences sont nulles.

Pour un signal linéaire, non gaussien, les multicohérences sont constantes.

Pour les autres signaux les multicohérences dépendent de la fréquence

2.5.9 Indicateurs des spectres cycliques d'ordre supérieur

Les spectres cycliques d'ordre supérieurs (SCOS) permettent de mettre en relief l'aspect cyclostationnaire et l'aspect bilinéaire (Bouillaut, 2000), (Prieur et al., 1995) et (Hurd et al., 1991) et (Breneur, 2002 chapitre 1 parag. 5.4).

Il n'est pas imaginable d'utiliser les multispectres cycliques d'ordre supérieur dans le cas d'une surveillance automatique en raison du coût élevé du calcul qu'ils impliquent et le nombre gigantesque de données. Afin de pallier cette limitation, Raad (Raad, 2003) a proposé des indicateurs brefs et globaux qui mesurent la cyclostationnarité de l'ordre 1 à 4. L'idée est basée sur les travaux avant-gardistes de W.A. Gardner et G.B. Giannakis). Les indicateurs proposés sont exprimés en termes de cumulants et sont normalisés. Ils généralisent les degrés de cyclostationnarité présentés par Zivanovic et Gardner en 1991. Plusieurs applications sont détaillées dans ce mémoire où la répartition de la cyclostationnarité aux différents ordres a été orientée pour surveiller des systèmes mécaniques tels que des engrenages et des roulements.

Remarque : Il est important de signaler qu'il n'existe pas une méthode systématique pour choisir la meilleure méthode de traitement du signal, dès lors que l'on utilise une base réelle de données vibratoires. Par contre, en s'appuyant sur certains travaux conduits dans le domaine de la réduction du nombre d'indicateurs, on peut soit faire une sélection ou une extraction de paramètres pertinents et en nombre restreints. Ces dernières techniques sont exposées dans la section suivante et interviennent directement dans la méthodologie utilisée, pour résoudre le problème de l'automatisation du diagnostic.

2.6 Réduction du nombre d'indicateurs

Dans l'étape précédente (génération d'indicateurs), les données collectées ne sont pas toutes aussi informatives : certaines peuvent correspondre à du bruit, présentent des redondances, sont peu significatives, sont corrélées ou non pertinentes pour la tâche à réaliser de diagnostic. Il est donc nécessaire d'effectuer une réduction parmi le très grand nombre de données, et de ne conserver que les plus pertinents pour la décision relativement au problème de séparabilité entre classe.

2.6.1 Les algorithmes de réduction

Deux grandes classes d'algorithmes ou de possibilités sont disponibles pour la diminution du nombre d'indicateurs (Ciupke, 2005), (Casimir, 2003), (Hallinan, 2001) et (Malhi et al., 2004) :

- la *sélection* qui consiste à trouver une représentation parcimonieuse mais informative d'un sous-ensemble d'indicateurs non redondants et pertinents, et
- l'*extraction* qui consiste à projeter via une transformation « A » l'ensemble des M indicateurs originaux dans un autre espace d'indicateurs, de dimension M' inférieure. Nous trouvons entre autre, l'ACP, l'entropie, les distances

euclidiennes ou de Mahalanobis, etc. Le sens des indicateurs peut être perdu puisque les indicateurs statistiques et énergétiques sont combinés par l'algorithme.

Nous nous intéressons dans ce travail aux algorithmes de sélection d'indicateurs. Ces algorithmes font l'objet d'une littérature abondante depuis une dizaine d'années (Kudo et al., 2000), (Duda et al., 2001), (Théodoridis et al., 2003), (Abe et al., 2006), (Dash et Liu, 1997) et (Yang et al., 1997). Le problème est que si l'ensemble des caractéristiques contient L indicateurs défini initialement alors le nombre total de possibilités ou de candidats est de 2^L . Ce nombre est très élevé et le but est de trouver une solution parmi ce 2^L possibilités.

2.6.2 Les approches de sélection

Dans l'article d'Abe et Kudo (Abe et al., 2006), on divise les méthodes de sélection en 2 approches :

Les algorithmes dits 'filters' (*classifier-independent feature selection*) qui exploitent les propriétés intrinsèques des indicateurs utilisés, sans référence à une quelconque application. Dans cette approche peu d'algorithmes sont proposés. Dans ces algorithmes, on cherche un sous ensemble qui est universellement efficace pour n'importe quel classifieur. En pratique, puisque on ne sait pas la règle de décision ou le classifieur à utiliser, il est désirable d'utiliser cette approche. Ceux appelés 'wrappers' (*classifier-specific feature selection*), au contraire, définissent la pertinence des indicateurs par l'intermédiaire d'une prédiction de la performance du système final (ce sont des algorithmes supervisés, un sous-ensemble d'indicateurs est généré et évalué par l'algorithme d'apprentissage dans l'approche).

(Dash et Liu, 1997), (Yang et Honavar, 1997) et (Jack, 2000) divisent les méthodes de sélection en trois approches : *complète*, *heuristique* et *aléatoire*. La figure 2.14 illustre ces approches avec l'ensemble des méthodes qui s'y rattachent.

Complète : Ce procédé fait une recherche complète du sous-ensemble optimal d'indicateurs selon la *fonction d'évaluation* utilisée (section 2.6.3). Différentes *fonctions d'évaluation* sont employées pour réduire l'espace de recherche sans compromettre les chances de trouver le sous-ensemble optimal. Par conséquent, bien que l'espace de recherche soit énorme, peu de sous-ensembles sont évalués. L'optimalité du sous-ensemble d'indicateurs, selon la *fonction à optimiser*, est garantie parce que le procédé peut faire des retours arrière (backtrack). Parmi les techniques utilisées: Branch and Bound (B&B), Best First Search (BFF), Beam Search (BS), Approximate Branch and Bound (AB&B), Minimum Description Length Method (MDLM), Focus, Méthode de Shlimmer et Min-Features (MIFES_1).

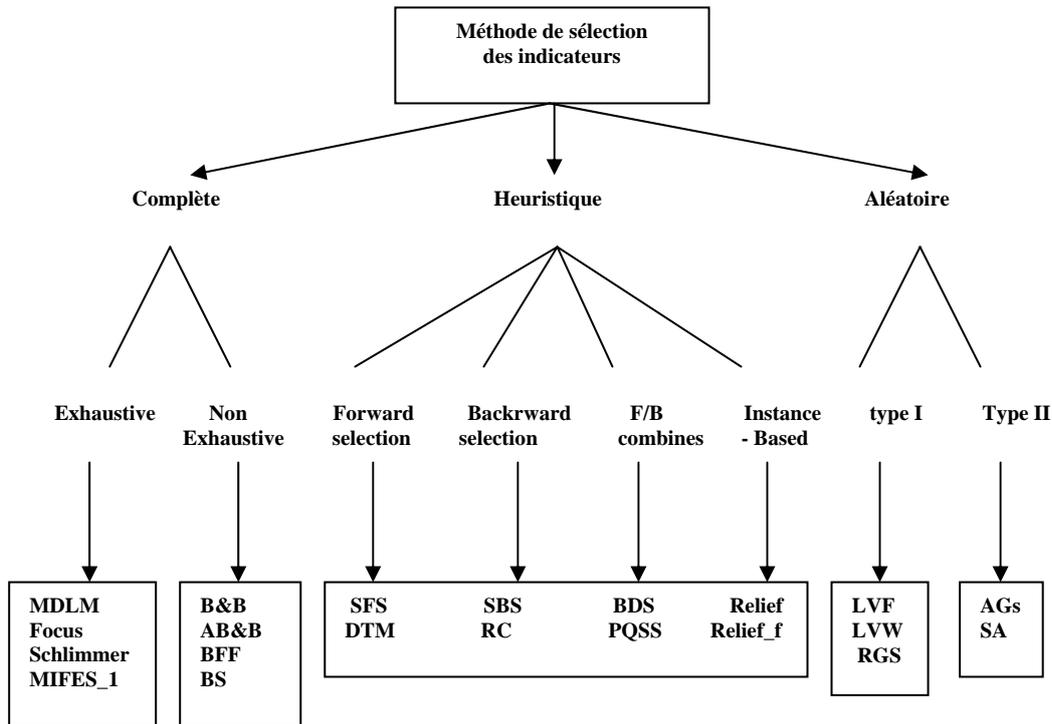


Figure 2.14 Méthodes de sélection d'indicateurs (extraite de Dash et Liu, 1997)

Heuristique : À chaque itération de la procédure de recherche, toutes les caractéristiques qui restent à être sélectionnées sont considérées pour la sélection d'indicateurs. Ces procédés sont très simples et rapides car l'espace de recherche est quadratique en terme du nombre d'indicateurs. Parmi les techniques utilisées : Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), bi-directional search (BDS), RELIEF, relevance in context (RC), (p,q) sequential search (PQSS).

Aléatoire : Ce procédé de recherche est plutôt nouveau dans son utilisation dans des méthodes de sélection d'indicateurs en comparaison aux deux autres catégories. Bien que l'espace de recherche soit 2^L , ces méthodes recherchent des sous-ensembles en effectuant un maximum d'itérations. Chaque procédure de recherche aléatoire exige des valeurs de quelques paramètres pour l'appliquer au problème à résoudre. La tâche d'attribution de valeurs appropriées à ces paramètres «fonction d'évaluation , section 2.6.3 » est importante afin d'obtenir des résultats satisfaisants. Parmi les techniques utilisées, on rencontre les algorithmes génétiques (AGs), le recuit simulé (Simulated Annealing (SA)), random generation plus sequential selection (RGS).

Kudo et Sklansky dans (Kudo et al., 2000) ont étudié plusieurs algorithmes d'optimisation et ont trouvé que les AGs conviennent parfaitement lorsque le nombre d'indicateurs est très élevé. Ces auteurs ont rajouté que les AGs ont une grande probabilité de trouver les meilleures solutions par rapport aux autres techniques

d'optimisation.

L'application des AGs dans la sélection d'indicateurs a été utilisée dans différents domaines de recherche tels que l'ingénierie des réacteurs multiphasés par Laurentiu (Laurentiu, 2005), la vérification de signatures par Ramesh et Murty (Ramesh et al., 1999), le diagnostic médical par Hallinan (Hallinan, 2001), la reconnaissance d'image par Bhanu et al. (Bhanu et al., 1999), la reconnaissance de la parole par Hong et Kwong (Hong et al., 2005), la reconnaissance de l'empreinte digitale par Codrea et al. (Codrea et al., 2003) et le diagnostic des défauts mécaniques par (Chen et al., 2003 ; Samanta, 2004 ; Guo et al., 2005 ; Zhang et al., 2005). D'autres algorithmes récents d'optimisation peuvent être envisagés, tels que les colonies de fourmis et la recherche tabou (voir Haupt et al., 2004) et (Jourdan, 2003).

2.6.3 Les fonctions d'évaluation

La **fonction d'évaluation** peut être soit mono critère ou multicritère. Une fonction d'adaptation mono critère signifie que la fonction **dépend d'une seule fonction objective** par contre la fonction d'adaptation multicritère dépend d'une combinaison de plusieurs fonctions objectives d'où la notion objective unique et objectives multiples.

Les fonctions d'évaluations mesurent l'ensemble des solutions candidates générées par les procédures de recherche. Ces valeurs sont comparées aux valeurs de l'itération précédente. Les meilleures valeurs seront gardées.

Il existe plusieurs types d'évaluation :

- la mesure de la distance qui permet de mesurer les capacités de discrimination, Les distances les plus utilisées sont : la distance euclidienne classique, la distance de Manhattan ou city – block, la distance du "Max", enfin, la distance de Mahalanobis,
- la mesure d'information qui permet d'estimer le gain d'une caractéristique (Fisher),
- la mesure de dépendance qui permet de quantifier la corrélation des caractéristiques (la trace des matrices intraclasse et interclasse),
- la mesure de consistance de l'ensemble des caractéristiques, et
- la mesure du taux d'erreur lors de la classification.

Remarque : L'application des algorithmes de sélection avant l'exploitation par les règles de décision permet d'obtenir des sous-ensembles d'indicateurs plus discriminants et offre plusieurs avantages lors de l'utilisation des méthodes de Reconnaissance de forme, les plus importants sont (Kudo et al., 2000), (Duda et al., 2001) et (Abe et al., 2006):

- La réduction du temps de calcul et la complexité des algorithmes de classification mis en jeu lors de l'exploitation.
- Le dépassement du phénomène de malédiction de la dimension qui se manifeste lorsque nous traçons les performances du classifieur en fonction de la dimension de

l'espace d'entrée et du nombre d'observations de la base d'apprentissage : la dimension de l'espace de représentation doit être faible devant la taille de la base de données afin que les surfaces de séparation entre classes que l'on calcule soient précises, voir section 4.6.3.2.

- la visualisation et la compréhension des données seront plus faciles.

Dans le chapitre 4 section 4.5, nous proposons d'étudier un couplage des **algorithmes génétiques** comme approche de recherche **et du critère de la trace** comme fonction d'évaluation. Dans ce même chapitre, nous exposons le principe de ce couplage ou de cette hybridation.

2.7 Règles de décision

Les méthodes de décision définissent une règle de décision pour la classification de 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 le taux de bonne classification (voir section 1.2.3.2).

On distingue deux approches de classification représentées par la figure 2.15 :

- *Les méthodes de classification non supervisées* (appelées méthodes de *clustering*) : méthodes basées sur la notion d'apprentissage non supervisé, laquelle consiste à regrouper des objets appartenant à un ensemble T en classes restreintes de telle sorte que les objets d'une même classe soient le moins dispersés possible. Dans ce cas, on fournit au système de reconnaissance un grand nombre de formes **non étiquetées**. L'étape de la classification va se charger d'identifier automatiquement les formes appartenant à une même classe. Ce cas ne sera pas traité ici, le lecteur désireux de faire connaissance avec ces méthodes pourra se reporter à (Duibuisson, 1990), (Duda et al., 2001).
- *Les méthodes de classification supervisées* (appelées méthodes d'affectation) : méthodes basées sur la notion d'apprentissage supervisé. Ces méthodes utilisent un ensemble d'exemples où les classes d'appartenance sont connues au préalable (à priori). La phase d'apprentissage utilise soit l'apprentissage déductif ou inductif. Les algorithmes d'apprentissage **inductif** dégagent un ensemble de règles de classification à partir d'un ensemble d'exemples déjà classés. Le but de ces algorithmes est de produire des règles de classification afin de prédire la classe d'affectation d'un nouveau cas. Dans les algorithmes d'apprentissage **déductif**, les règles d'affectation sont déterminées à priori par l'interaction avec l'expert. A partir de ces règles on détermine les classes d'affectation de nouvelles observations.

Pour plus d'informations, les ouvrages de (Duda et al., 2001), (Fukunaga, 1990), traitent ces approches en détails.

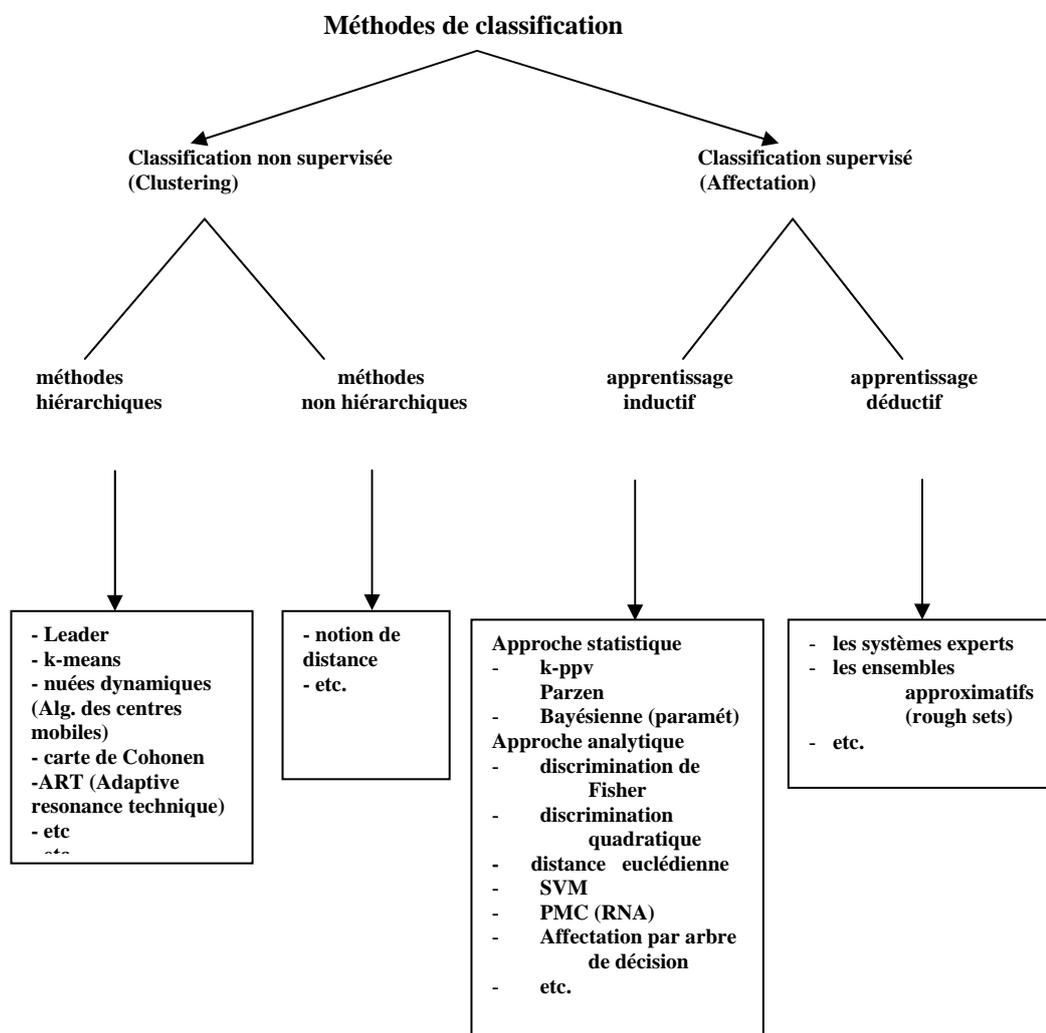


Figure 2.15 Méthodes de classification des modes de fonctionnement

Dans cette thèse, on s'intéresse particulièrement aux méthodes basées sur la notion d'apprentissage supervisé inductif et plus précisément à l'approche statistique non paramétrique des k – plus proches voisins (k-ppv), et aux deux méthodes de calcul direct des frontières : le perceptron multicouche (PMC) et les Séparateurs à Vaste Marge (SVM) que nous introduisons ci-dessous en soulignant certains de leurs avantages et inconvénients. Des détails complémentaires seront exposés au chapitre 4.

2.7.1 Règle des k -ppv

Au contraire des méthodes statistiques paramétriques (exemple la règle de Bayes) qui font usage des lois de probabilités a priori des classes sur une base d'apprentissage. Les méthodes statistiques non paramétriques des k -ppv et de Parzen sont utilisées si les lois de probabilités régissant les observations et les classes ne sont pas connues ou incomplètes. Le lecteur désireux faire de plus ample connaissance avec le classifieur des K -ppv pourra se reporter à (Dubuisson, 1990), (Fukunaga, 1990).

Le principe général de la méthode des k -ppv consiste à rechercher parmi l'ensemble l'apprentissage T , contenant l'ensemble des individus et leurs classes d'affectation, un nombre k d'individus parmi les plus proches possibles de l'individu à classer. Puis, l'individu est affecté à la classe majoritaire parmi ces k individus trouvés. Le nombre k est fixé a priori par l'utilisateur.

Si $k = 1$, alors l'individu est affecté à la classe du plus proche voisin de l'ensemble T . En général, la manière la plus simple d'élaborer cette règle est de mesurer la distance entre le nouveau vecteur \mathbf{x} et chacun des vecteurs de l'ensemble d'apprentissage. Le vecteur \mathbf{x} sera alors affecté à la classe la mieux représentée parmi ses k observations \mathbf{x}' les plus proches (figure 2.16). Dans cet exemple, l'élément \mathbf{x} sera affecté donc à la classe majoritaire représenté ici par des cercles pleins.

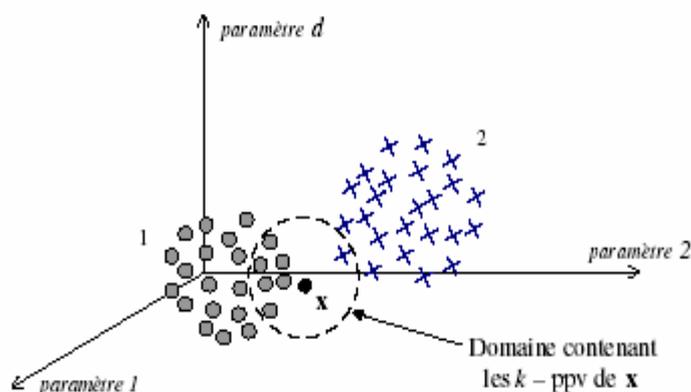


Figure 2.16 Illustration de l'effet de l'affectation par la règle des k -ppv

Il existe plusieurs types d'indices. Parmi les plus connus, dans le cas des données numériques : la métrique de Minkowski. Où nous pouvons distinguer : la *distance euclidienne*, La *distance de Manhattan* ou city – block, la distance du "Max", etc. Enfin, il existe une mesure particulière de distance appelée *distance de Mahalanobis*, voir [annexe 2.3](#).

2.7.2 Règle du PMC

Les réseaux de neurones sont des outils de l'intelligence artificielle, capables d'effectuer des opérations de classification. Leur fonctionnement est basé sur les principes de fonctionnement des neurones biologiques. Le principe général consiste à définir des unités simples appelées neurones, chacune étant capable de réaliser quelques calculs élémentaires sur des données numériques. On relie ensuite un nombre important de ces unités formant ainsi un outil de calcul puissant (Bishop, 2005) et (Jack, 2000). La figure 2.15 montre la Structure d'un réseau de neurones artificiels, le perceptron multicouches (PMC).

La phase d'apprentissage supervisé du réseau de neurones (adaptation ou ajustement des poids w_{ij}) dépend de son analyse des modes de fonctionnement du système. Chaque mode est caractérisé par un ensemble de données recueillies sur le système. A chaque mode, on associe une expertise faite par l'expert. Cette association (ensemble de données - modes de fonctionnement) sera apprise par le réseau de neurones. Après cette phase d'apprentissage, le réseau de neurones associera les classes représentant les modes de fonctionnement aux formes d'entrée caractérisées par les données du système (Bishop, 1995).

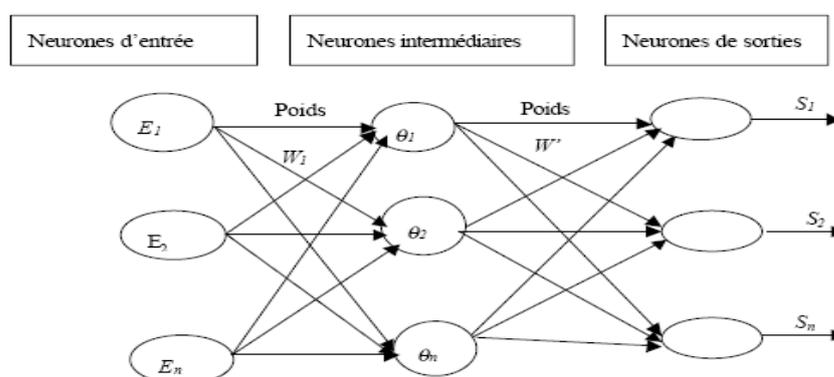


Figure 2.17 Structure d'un réseau de neurones artificiels

Leur principal avantage par rapport aux autres outils est leur capacité d'**apprentissage** et de **généralisation** de leurs connaissances à des entrées inconnues. Les réseaux de neurones peuvent être également implémentés en circuits électroniques, offrant ainsi la possibilité d'un traitement temps réel. Le processus d'apprentissage est donc une phase très importante pour la réussite d'une telle opération.

Une des qualités de ce type d'outil est son adéquation pour la mise au point de systèmes de surveillance modernes, capables de s'adapter à un système complexe avec reconfigurations multiples. Les réseaux PMC permettent d'approcher des frontières non-linéaires.

2.7.3 Règle des séparateurs à vaste marge (SVM)

Les SVM sont des algorithmes de classification récemment développés par Vapnik qui procèdent par un calcul direct des frontières « les hyperplans » entre classes. (Vapnik, 1998), (Vapnik, 2000). L'idée essentielle des SVM est d'utiliser des fonctions noyaux (kernels) et de projeter les données de l'espace d'entrée, non linéairement séparables, dans un espace de plus grande dimension appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables. La figure 2.18 illustre l'effet du changement de l'espace par une fonction noyau. Elle montre également que les données non linéairement séparables dans l'espace d'entrée "X" sont à présent séparables dans l'espace augmenté " $\Phi(x)$ ". Notons " Φ " une transformation non linéaire de l'espace d'entrée "X" en un espace de redescription " $\Phi(x)$ ".

Les algorithmes SVM tentent de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes, voir figure 4.31. Le comportement des SVM est conditionné par le type de noyau utilisé dans la projection et les valeurs accordées à ses paramètres. Pour plus de détail (voir chapitre 4)

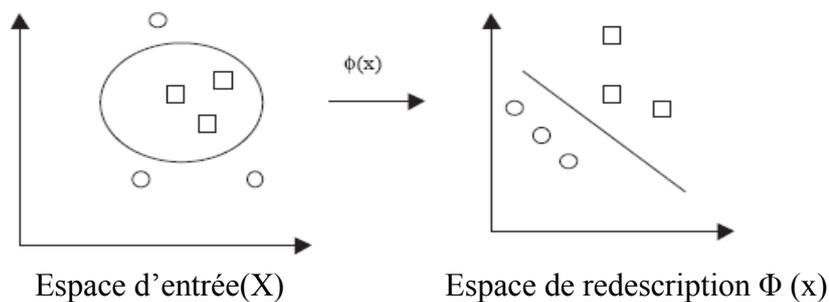


Figure 2.18 Illustration de l'effet de changement d'espace par une fonction noyau (extraite de (Jack et al., 2002)).

2.8 Conclusion

Ce chapitre nous a permis d'avoir un aperçu général sur le principe du diagnostic par reconnaissance de formes en mentionnant les principales phases pour la conception de ce système. La première phase est liée à la construction des bases de données d'apprentissage et de test, elle va de l'acquisition avec ses dispositions à prendre à la formation de l'espace de représentation. La deuxième phase est la plus importante, là on doit choisir une méthode de décision qui offre les meilleures performances. La troisième phase est celle de l'exploitation. Dans notre étude, nous avons choisi d'étudier les performances des k-ppv, du PMC et des SVM. Nous avons donné une première description des algorithmes de sélection (voir figure 2.14) et des méthodes de classification (voir figure 2.15), plus de détails sur ces méthodes seront présentés dans le chapitre 4. Une synthèse des sources de données (les bancs) que nous avons utilisé dans notre étude est exposée dans le chapitre suivant.

Chapitre 3

Sources de Données Vibratoires

3.1 Introduction

Nous avons été principalement motivé à choisir et concentrer nos investigations sur les engrenages et les roulements par :

- le nombre gigantesque d'ouvrages et d'articles traitant cette application, et
- les études statistiques réalisés sur ces défauts, que nous présenterons ci-dessous dans la section 3.2.

Dans ce chapitre, nous présentons au début quatre exemples de statistiques réalisées par des sociétés étrangères, ensuite, nous donnons des indications sur les engrenages et les roulements, sur l'origine de leurs défauts et sur les interprétations des résultats des différents traitements. Enfin, nous nous limitons à exposer seulement les bancs d'essais utilisés dans les travaux présentés dans cette thèse.

3.2 Statistiques sur les défauts de machines

Les études statistiques sont nombreuses et variées, nous présentons ici quelques exemples pour justifier notre choix des éléments étudiés :

- Le tableau 3.1 présente des statistiques sur les causes de défaillance et la localisation des défauts dans les systèmes à engrenages (réducteurs et multiplicateurs). Cette enquête, réalisée par la société des assurances Allianz, permet de conclure que les organes les plus sensibles sont les engrenages et les roulements. Les défaillances résultant d'autres éléments viennent ensuite ([Alattas, 1994](#)), ([Morel, 1992](#)), ([Sidahmed et al., 1990](#)) et ([Breneur, 2002](#)).

Tableau 3.1 Localisation des défauts dans les transmissions de puissance à engrenage (extrait de ([Breneur, 2002](#)))

Localisation des défauts	%
Dentures	60
Paliers	19
Arbres	10
Carters	7
Autres	4

- Afin également d'identifier le composant le plus faible dans les moteurs électriques qui est soumis à la défaillance, 3 études statistiques sont présentées dans la référence de (singh et al., 2003). Ces études ont été entreprises par General Electric Co (EPRI), USA, IEEE-IGA USA et Brüel et Kaer Danemark. On a constaté que la principale cause de défaut auquel la machine à induction est soumise est le défaut de roulement. Ce dernier constitue entre 40-50% des défauts de moteurs à induction, voir les figures 3.1 à 3.3 extraites de l'article de singh et al. 2003. La même remarque a été également faite par (Thorsen, 1995) bien avant.

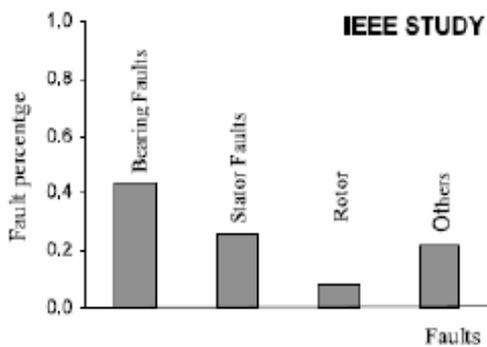


Figure 3.1 Résultats d'étude de IEEE

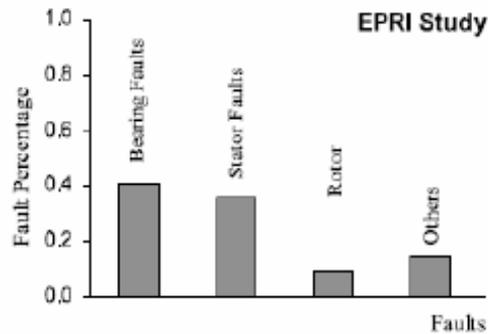


Figure 3.2 Résultats d'étude de EPRI

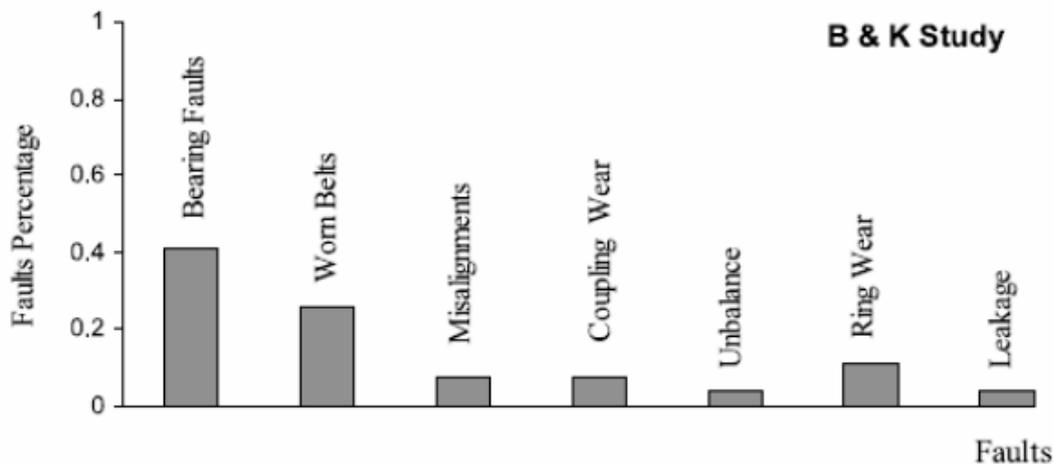


Figure 3.3 Résultats d'étude conduite par B&K sur les défauts dans un moteur asynchrone

Ces statistiques concernant les causes de défaillances et la localisation des défauts dans les transmissions de puissance (boîte de vitesses) et les moteurs asynchrones permettent de conclure que les organes les plus sensibles sont les engrenages et les roulements. Les causes d'avaries sont multiples, et donnent lieu à des anomalies plus ou moins graves. Dans la prochaine section, nous discutons les défauts d'engrenages et de

roulements. Ces défauts apparaissant par fatigue, phénomène inévitable, auquel la maintenance préventive remédie par la détection précoce avant destruction au cours du fonctionnement du système.

3.3 Les signaux d'engrenage

Les engrenages sont des éléments de construction directement générateurs de chocs en cas d'usure ou de mauvais fonctionnement. Les signaux mesurés sur les machines qui en sont équipées sont très riches en harmoniques. L'énergie contenue dans ces signaux est en général très importante. Les chocs peuvent en effet exciter les fréquences propres de la structure et faire monter le bruit de fond des spectres (Schenck, 1994). Il est donc parfois utile de recourir à des méthodes plus avancées du traitement du signal, voir section 3.3.2. Les vibrations que produit un engrenage, figure 3.4, sont par nature composées d'un signal de base produit par le signal d'engrènement et de modulations qui sont liées aux phénomènes plus lents leur superposition se traduit par des battements dont la période est la différence des deux fréquences des roues. Les défauts ont des manifestations larges bandes dans les spectres associés aux signaux (nombre important de raies de modulations en amplitude et en phase, phénomènes non linéaires, etc.).

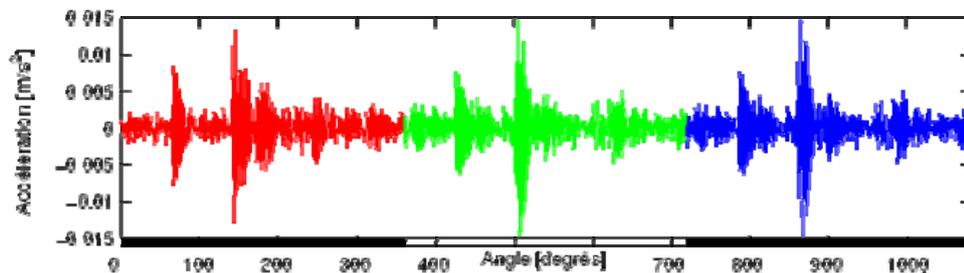


Figure 3.4 *Signal vibratoire issu d'un réducteur à engrenages (extraite de Bonnardot, 2003)*

3.3.1 Origine des défauts

Les bruits et les vibrations sont d'origines diverses, voir (Bourgain & al., 1998) et (Breneur, 2002). Ils peuvent être dus aux modes de flexion ou de torsion des dentures, aux défauts d'usinage, aux chocs de contact entre dents. Ils peuvent être, aussi, des défauts de montage ou dus à des paramètres opérationnels, dans ce qui suit nous examinons chacun des aspects :

- a. **modes de flexion et de torsion** : Nous trouvons des bruits dus aux phénomènes d'amplification résultant des caractéristiques modales du système quand la fréquence d'excitation est proche de celle d'un des modes de flexion ou de torsion des dentures. Les pignons et les dentures sont des milieux déformables qui possèdent donc une infinité de modes propres. Le bruit est un bruit à bande étroite.

- b. défauts d'usinage** : un engrenage fonctionnent en ayant toujours une paire de dents en contact. Il est évident que tout défaut de denture, d'espacement ou de division des dents produira du bruit.
- c. ondes de contraintes** : les chocs qui se répètent à chaque contact entre pignons menant et mené, et qui peuvent à la longue entraîner des ruptures par fatigue. Les contacts des dents provoquent des contraintes de compression aux points des lignes de contact et des contraintes de flexion dans les autres parties, en particulier à la base des dents. Quand la dent est désengagée, ces contraintes cessent soudainement en causant une onde de contrainte qui se propage à travers le pignon. A ces cycles de contraintes alternées, correspondent des *usures par fatigues* qui s'avèrent être la principale cause de la baisse de longévité des pignons.
- d. poches d'air** : les poches d'air sont des sources mineurs de vibration. Elles se produisent quand une dent engrène entre deux dents du pignon complémentaire. Quand la dent se désengage, il y a un effet de succion. Ce bruit de sirène se produit généralement sur les pignons à grande vitesse. Le bruit produit est un bruit aléatoire à bande étroite à la même fréquence que la fréquence de passage des dents.
- e. poche d'huile** : L'effet est similaire à celui des poches d'air, mais plus sévère car l'huile est presque incompressible.
- f. friction** : la friction est probablement la source principale de bruit dans les pignons. Quand deux pignons sont en rotation, il y a une action poussée qui cesse brutalement quand l'action inverse commence. Il y a inversion soudaine des forces et il en résulte une vibration à la fréquence de passage des dents. Ceci se produit inévitablement à chaque engrenement quelle que soit la qualité de la fabrication.
- g. impact** : ces impacts sont causés par une dent qui est heurtée par une autre dent à la fréquence d'engrenement. Ces bruits se produisent généralement dans les pignons bon marché pour lesquels l'épaisseur du film d'huile est insuffisante, en liaison avec la rugosité des surfaces.

3.3.2 Types de défauts

Quel que soit le type d'engrenage rencontré, droits, coniques, ou Hélicoïdal, voir figures 3.5a et 3.5b, on peut distinguer pour les dentures d'engrenages deux types de défauts, les défauts uniformes et les défauts localisés dus à des aspect opérationnels. Un exemple des deux cas est donnée sur les figures 3.6a et 3.6b (El badaoui, 1999) et (Morel, 1992) :



Figure 3.5a Engrenage conique (document Wikipedia)



Figure 3.5b Engrenage parallèle (document Wikipedia)

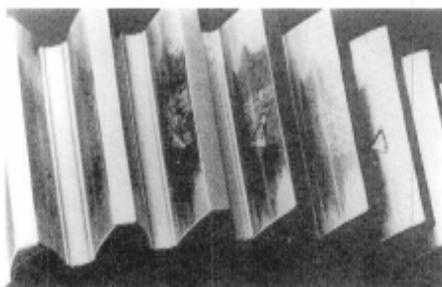


Figure 3.6 a Début d'écailage local sur les dents (document CETIM)

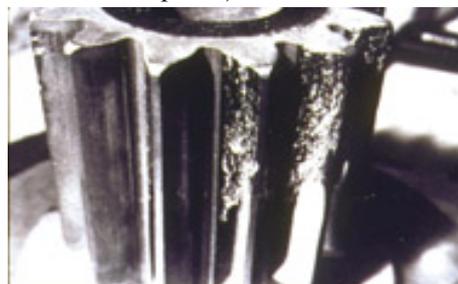


Figure 3.6 b Défaut local d'écailage (document CETIM)

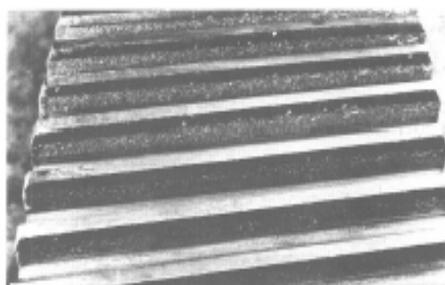


Figure 3.7a Piqûres (pitting), défaut réparti sur les dents (document CETIM)



Figure 3.7b Défaut de fissure au pied d'une dent d'engrenage

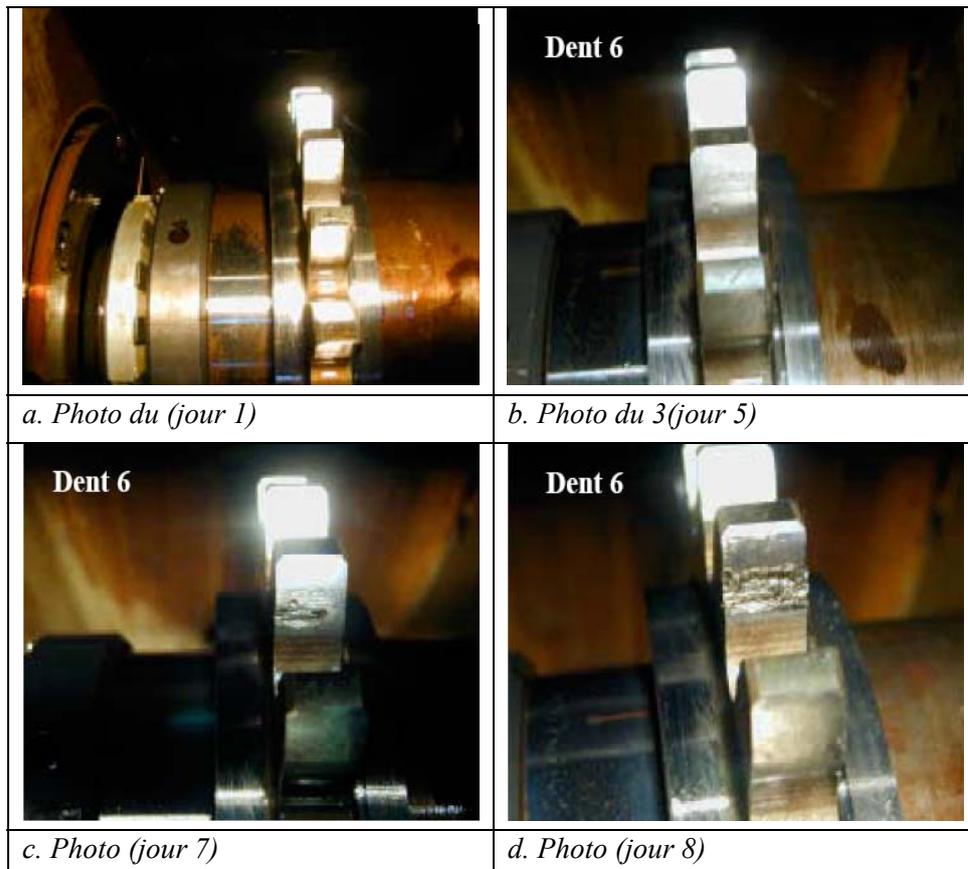


Figure 3.8 Evolution d'un défaut d'écaillage dans le temps (extraite de (Raad, 2003))

3.3.2.1 Défauts uniformes

Les défauts uniformes concernent en général toutes les dents. On peut en citer quatre exemples:

- a. **l'usure** (wear) peut modifier le profil des dents et conduit à un mauvais fonctionnement de l'engrenage (perte de puissance,...) voire à sa mise hors service. Les défauts d'usure, considérés comme des défauts répartis, engendrent une augmentation du niveau RMS et une modulation d'amplitude dans le signal temporel de vibration qui se traduit par une augmentation de la raie d'engrènement et des harmoniques ainsi que des **raies latérales dans le spectre**. Le nombre et l'amplitude des composantes harmoniques par rapport à la fréquence d'engrènement traduiront l'état d'avancement des dentures.
- b. **le pitting** (piqûres réparties de manière homogène sur toutes les dents) est considéré également comme un défaut uniforme, il peut être détecté sur le signal temporel de la même façon que le défaut d'usure, mais cette détection est meilleure par **analyse d'huile**. Il peut évoluer suivant les matériaux vers la destruction de l'engrenage. Il s'agit de trous plus ou moins profonds qui affectent toutes les dents, en particulier dans la zone du cercle primitif. Cette avarie se produit surtout sur des engrenages en

acier de construction relativement dur, et donc plus fragile, sensible aux nombre de cycles de contrainte et résultent de la fissuration de la surface par fatigue avec perte de particules créant des trous de 0,5 à 1mm de diamètre. Les piqûres sont moins à craindre lorsque la viscosité du lubrifiant est importante, car dans ce cas, le film d'huile séparant les surfaces en contact est plus épais. Cette avarie peut apparaître à la suite de légers désalignements d'axes par exemple, à cause de surpressions locales.

- c. **L'usinage** d'un profil parfait pour les dents d'un engrenage est très difficile à réaliser. Il existe donc, pour chacune d'entre elles, des écarts non aléatoires par rapport à ce profil idéal. Ces modifications se traduisent par des **phénomènes de glissement** (qui sont **non linéaires**) lors de l'engrènement des différentes dents. Des méthodes appropriées ont donc été appliquées avec succès telles que : les bispectres et les bicohérences,
- d. **Le défaut de surface**, survenant dans un système d'engrenage, il est considéré comme des phénomènes de nature aléatoire. En effet, les conséquences de tels défauts sont généralement différentes d'une dent à l'autre. De plus l'altération de la surface peut également varier sur une même dent. Or, il a été montré que les vibrations engendrées par ces défauts étaient **non stationnaires**

3.3.2.2 Défauts localisés

Les défauts localisés ne concernent que quelques dents et se repartissent en trois types : la fissuration, l'écaillage et la fracture:

- a. **la fissuration** d'une dent, généralement par fatigue qui conduit rapidement à la rupture de la dent, est caractérisée par des impulsions et une **modulation de phase** dans le signal temporel. Elle progresse à chaque mise en charge, et elle est située en pied de dent. Elle apparaît sur des aciers fins et durcis, qui sont sujets aux concentrations de contraintes. Leur apparition est due à un dépassement de la limite élastique en contrainte au pied de la dent du côté de la dent en traction.
- b. **l'écaillage** (spalling) est caractérisé par des impulsions et des chocs (faible à l'état précoce) dans le signal temporel. ce type de défauts existe dans les engrenages cimentés trempés (dans la couche superficielle est durcie) et qui sont plus répandus à l'heure actuelle, car ils permettent de passer des couples importants avec des dimensions (i.e. des masses) réduites, ce qui minimise le coût de fabrication. Ce type d'engrenage ne présente pratiquement pas de phénomène de d'usure : l'écaillage est le principal défaut qui évolue également très vite vers la rupture. L'écaillage se manifeste aussi sous forme des trous, mais ceux-ci sont beaucoup moins nombreux, plus profonds et plus étendus que les piqûres.
- c. **la fracture** est un état de dégradation de l'engrenage très avancé et génère des chocs importants sur le signal temporel.

Cette liste n'est pas exhaustive. Les cas les plus fréquents sont en **général une combinaison**. Par ailleurs dans les systèmes à engrenages, on observe des oscillations de vitesse de rotation qui peuvent être liées à des torsions des arbres, à la flexibilité des dents ou à l'imperfection de l'engrenage. De plus, ces fluctuations de vitesse engendrent des modulations de fréquence. Cette modulation se traduit par des raies latérales de distributions complexes, autour des fréquences d'engrènement. En outre, même sur des engrenages «sains » on retrouve ces peignes de raies. Tous ces phénomènes contribuent à la complexité des signaux d'engrenage.

3.3.3 Traitement du signal

Les méthodes de traitement des vibrations d'engrenage sont nombreuses et variées car elles correspondent à la diversité des problèmes rencontrés (section 3.3.2). Il est important de signaler qu'on devrait, sur un problème un peu complexe, utiliser plusieurs méthodes et de plus en plus avancées. L'utilisation d'un seul type d'indicateurs, par exemple, temporels ou fréquentiels, signifie qu'on ne surveille que les défauts dont l'apparition ou l'aggravation a une incidence directe sur la valeur des indicateurs retenus. Autrement dit, l'apparition d'un défaut préalablement non envisagé et qui n'influent sur aucun des indicateurs de surveillance retenus, ne pourra pas être détecté.

Afin de brosser un bref panorama sur la diversité des cas quand peut avoir dans la surveillance des engrenages, nous allons exposer plusieurs situations, cela permettra ainsi de montrer la nécessité d'utiliser un grand nombre de méthodes du traitement du signal qui seront, bien sûr, complémentaires.

- Il est connu que l'apparition de **fissurations** en pied de denture provoquent des impulsions et une **modulation de phase** (Bardou, 1994), (Mc Fadden, 1986) et (Randall, 1997b). La démodulation de phase est bien adaptée à ce type de défauts qui entraîne une **variation de phase** dans le signal d'engrènement.
- alors que la présence de **chocs faibles** à l'état précoce dus à un **écaillage** produit une **modulation en amplitude**. Cette modulation se traduit par la présence de bandes latérales (raies) de part et d'autres des harmoniques de la fréquence d'engrènement. L'analyse cepstrale et ses différentes variantes permettent de seulement de localiser ce type de défaut (Elbadaoui, 1999a, b, 01), (Antoni, 2000a).
- **La fracture** est un état de dégradation de l'engrenage très avancé et génère des chocs importants sur le signal temporel qui produit une **modulation d'amplitude** se traduisant par des raies latérales autour des harmoniques de la fréquence d'engrènement. La démodulation d'amplitude permet de localiser le défaut et de l'identifier.
- Un **défaut d'usinage** du profil, **l'usure** et les **piqûres** réparties (pitting) se traduisent toujours par des phénomènes de glissement, qui sont **non linéaire et une modulation d'amplitude** dans le signal temporel, lors de l'engrènement des

différents dents (Benoughram, 1999), (Bouillaut, 2001). L'étude de ces phénomènes non-linéaires est généralement basée sur l'utilisation des multispectres (bispectre, trispectre, ...). L'utilisation également de la corrélation spectrale permet de détecter précisément les liens non-linéaires (couplages quadratiques de phases ou de fréquences : c'est-à-dire une corrélation entre les harmoniques ou les composantes du signal).

- Il a été montré également que les vibrations engendrées par les défauts de surface étaient **non-stationnaire** (Capdessus, 1992), (Wang et MacFadden, 1993). Les méthodes temps fréquences ont permis de mettre en évidence la modulation de fréquence linéaire due en partie au défaut d'**écaillage**. L'analyse temps fréquence par la transformée en ondlettes semble donner de bons résultats (OEH, 1995). Cette dernière permet une bonne localisation en temps et en fréquence des différents phénomènes et avec une excellente résolution (Mahgoun et al., 1999).
- De nombreuses études ont été réalisées et il est ressorti que les méthodes **cyclo-stationnaires, telles que** : la corrélation spectrale, la cyclostationnarité à l'ordre 3 et 4 fournissent de très bon résultats (MacCormick, 1998a), (Bouillaut, 1998, 1999), (Raad, 2003). Elles permettent de détecter de manière plus précoce les défauts dites cyclostationnaires (Gardner, 1991), (Capdessus, 1992), (Bouillaut, 2000), (Bonnardot, 2003), (Boustany, 2005),
- Un jeu de montage entre pignons (i.e., désalignement) se traduit par une amplitude élevée à la **deuxième harmonique** de la fréquence d'engrènement dans le spectre associé (Benoughram, 1999), (Bouillaut 2001),

Il existe encore de nombreux travaux sur l'étude des engrenages d'un point de vue connaissance (modélisation, physique) et méthodes d'analyses spécifiques.

Par ailleurs dans les systèmes à engrenages, on observe des oscillations de vitesse de rotation qui peuvent être liées à des **torsions des arbres**, à la **flexibilité des dents** ou à l'imperfection de l'engrenage. De plus, ces fluctuations de vitesse engendrent des **modulations de fréquence**. Cette modulation se traduit par des raies latérales de distributions complexes, autour des fréquences d'engrènement.

Dans le tableau 3.2, nous donnons une synthèse des différents défauts généralement rencontrés dans les engrenages et les modes de détection correspondants.

Tableau 3.2 Synthèse des défauts d'engrenages (extrait de (Benoughram, 1999))

Type de défaut	SIGNAL TEMPOREL DE VIBRATION	SPECTRE ASSOCIE
Défauts uniformes		
Usure	<ul style="list-style-type: none"> •Augmentation du niveau RMS •Modulation d'amplitude 	<ul style="list-style-type: none"> •Augmentation du nombre et l'amplitude des raies d'engrènement et des harmoniques •Raies latérales
Pitting (piqûres)	Idem mais la détection semble meilleur par analyse d'huile	
Défauts localisés		
Fissure	<ul style="list-style-type: none"> •impulsion •modulation de phase 	<ul style="list-style-type: none"> • larges bandes •raies latérales
Ecaillage	<ul style="list-style-type: none"> •Impulsion /chocs à l'état Précoce 	<ul style="list-style-type: none"> •Larges bandes
Fracture	<ul style="list-style-type: none"> •chocs importants 	<ul style="list-style-type: none"> • Larges bandes
Jeux de montages entre deux Pignons (i.e. désalignement)		<ul style="list-style-type: none"> • Amplitude élevée à la deuxième harmonique de la fréquence d'engrènement

Remarque : En conclusion de cette revue bibliographique des origines et des types de défauts, il ressort que lorsque l'on veut détecter de façon précoce les phénomènes à caractère impulsion, exemple le choc des engrenages, qui est caractérisé par un signal vibratoire d'allure particulière couvrant une large bande de fréquence. Le spectre obtenu peut être complexe et il est difficile de suivre les évolutions de raies, il est alors délicat de caractériser la raie en cause par des méthodes classiques. Le Cepstre ne permet pas d'évaluer la sévérité du défaut repéré. Le Kurtosis et le Facteur de crête ne servent qu'à détecter les défauts et ne sont expérimentalement pas fiables pour des plages étendues de la vitesse de rotation. Nous avons souligné également dans le chapitre 2, section 2.4 l'inefficacité de certaines méthodes à détecter certains comportements, Voir l'utilisation du spectre dans la section 2.4.3 et l'utilisation de la représentation temps fréquence dans la section 2.4.4, etc. Par conséquent, l'utilisation de plusieurs méthodes est une nécessité.

3.4 Les signaux de roulement

La surveillance des roulements est l'une des applications les plus mûres. Les roulements, voir figure 3.11a, sont des éléments de machines très sollicités. Les vibrations que produit les roulements sont par nature composées d'un signal de base produit par le signal choc du défaut et des modulations qui sont liées aux phénomènes plus lents de vitesse de rotation. La figure 3.9a montre un défaut local sur la bague

externe. Lorsque la bille passe par un défaut, elle génère un choc sur la bague. Ce choc provoque une vibration qui se propage sur la bague, puis sur la cage externe du roulement. On parle d'impulsion.

En mettant en évidence la répétition de cette *impulsion* on définit les fréquences caractéristiques ou des défauts, voir **annexe 3.1**. L'*amplitude de l'impulsion* quantifie l'intensité du choc, elle est fonction de la vitesse de rotation des jeux mécaniques, de la sévérité du défaut et de la charge (figure 3.9). La *fréquence* typique du défaut est fonction de la géométrie du roulement et de la vitesse de rotation. L'ordre de grandeur du temps de montée et d'impulsion est de quelques dizaines de microsecondes. Le signal de roulement (une série de chocs) sera modulé par le passage de l'élément roulant dans la zone de charge (figure 3.10). Une analyse détaillée de ce phénomène peut être trouvée dans (Ypma, 2001), (Randall, 2004), (Antoni, 2007), (Antoni et al., 2003), (Kiral et al.2006).

L'idéal serait que l'on puisse mesurer directement la fréquence des chocs. Mais comme nous venons de le souligner, en pratique, entre la source du choc et le capteur, on va trouver d'autres éléments de transmissions, comme la bague extérieure, l'élément roulant, la fixation du capteur, tous caractérisés des fonctions de transfert, qui vont influencer la forme du signal. De plus, la machine génère d'autres signaux vibratoires et des bruits de fond, qui vont venir se superposer à ceux émis par les roulements.

L'effet de la zone de charge sur les signaux de roulement est montré sur la figure 3.9b et 3.10. Cette figure montre des signaux typiques et des signaux d'enveloppe des défauts d'un roulement à bille situé sur les différents éléments du roulement. Ces chocs répétés (impulsions) seront modulés par les autres fréquences.

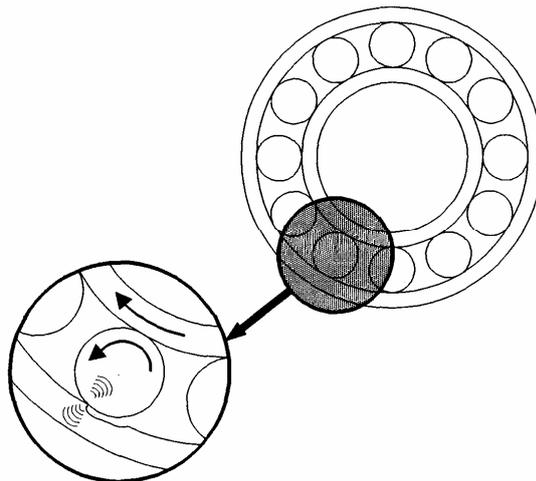


Figure 3.9a Roulement avec bague extérieure endommagé (Extraite de Schenck, 1994)

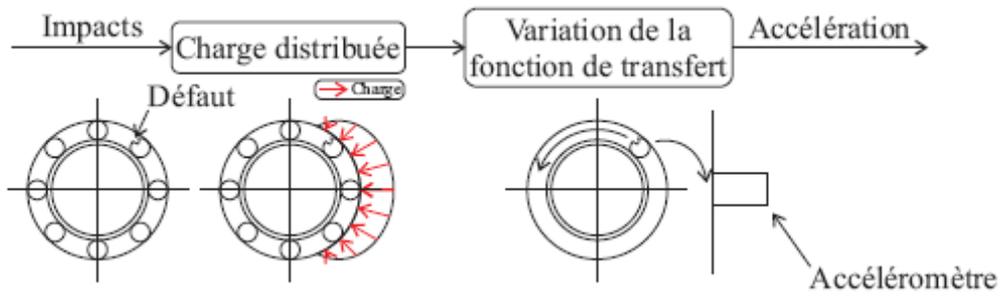


Figure 3.9b: *Modèle du signal vibratoire (extraite de bonnardot, 2005) (Antoni et al. 2003). Roulement avec élément roulant endommagé*

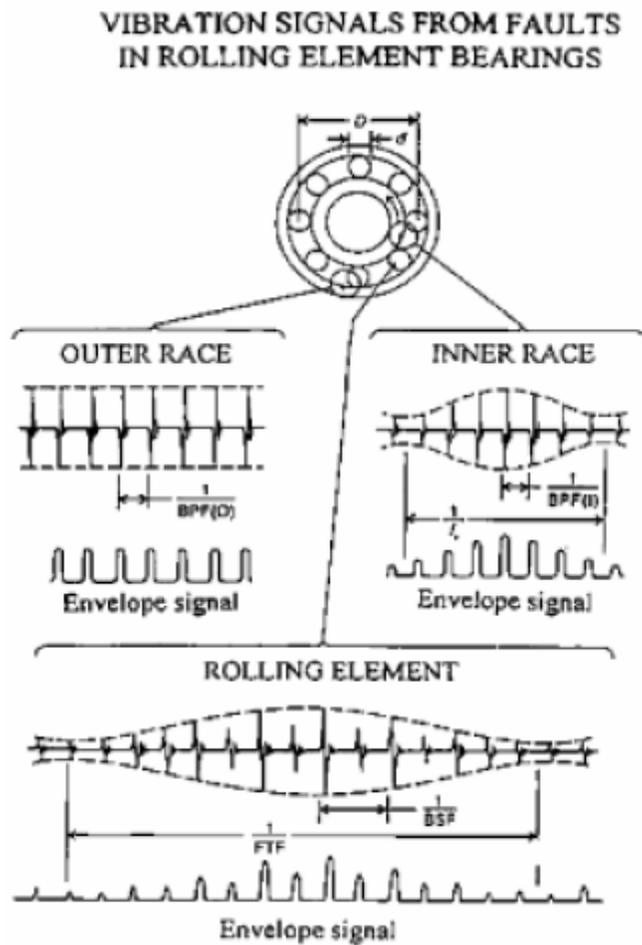


Figure 3.9c *Signaux typiques et signaux d'enveloppe d'un défaut local d'un roulement à bille (Extraite de (Randall, 2004))*

BPFO (Ball Pass Frequency, Outer race) : fréquence du défaut sur la bague externe,
 f_r : fréquence de rotation,

BPFI (Ball Pass Frequency, Inner race) : fréquence du défaut sur la bague interne,

FTF (Cage frequency) : fréquence du défaut sur la cage

BSF : (Ball Spin Frequency, fréquence du défaut sur l'élément roulant.

3.4.1 Origine des défauts

Les causes d'incidents sur les roulements sont très nombreuses. Elles doivent être recherchées dans le procédé de fabrication, le transport, le stockage, le montage et le fonctionnement. La durée de vie du roulement dépend fortement des conditions de charge (axiale et radiale), de l'évolution des vitesses, de la lubrification et du montage.

Nous donnons ci-dessus un récapitulatif des causes de dégâts les plus fréquents (Schenck, 1994). Ensuite, nous détaillerons comment les différentes détériorations se produisent (Bourgain & al., 1988).

3.4.1.1 Causes de dégâts prématurés

Les causes les plus fréquentes sont dues principalement :

A la **fabrication** : matériaux non homogènes, Tolérance sur les cotes,...

Au **stockage/transport** : emballage insuffisant, vibrations, ...

Au **montage** : chocs, mauvaise précontrainte, erreurs de cote, erreur de lignage, ...

Au **fonctionnement** : Surcharge, manque ou excès de graisse, corps étrangers (poussière), substance étrangère (gaz agressif, humidité), charge thermique, ...

3.4.1.2 Exemples détaillés des causes

Nous donnons ci-dessous quelques *détails* des causes de défauts :

- a. **la propreté** : toute intrusion de particules étrangère entraîne des détériorations qui évoluent inéluctablement vers l'écaillage.
- b. **la chauffe** : il est impératif d'éviter les chauffes, à plus de 120 degrés, il y a alors destruction de la graisse avec tous les ennuis qui en résultent (écaillage)
- c. **les chocs** : les roulements sont très sensibles aux chocs. Une bague mal engagée sur l'arbre ne doit pas recevoir de coups pour sortir. Un rapide passage au séchoir permettra de la sortir aisément (fissure)
- d. **la géométrie** : il faut éviter les ovalisations, les conicités, les déversements, les jeux trop faibles ou trop forts,

- e. **la vaseline de protection** : l'épaisse couche de vaseline qui couvre certains roulements doit être éliminée avant le garnissage en graisse. Sinon en fonctionnement, il y aura fusion de la vaseline, entraînement de la graisse et grippage.
- f. **la charge** : une charge trop élevée ou mal appliquée conduit à des destructions rapides. Les chocs se traduisent par des écaillages, des empreintes et la destruction finale.
- g. **la vitesse** : chaque roulement admet une vitesse maximale au delà de laquelle il y a échauffement exagéré, usure et grippage.
- h. **la température** : pour les roulements courants, elle ne doit pas dépasser les 120°C.
- i. **les vibrations en rotation** : les roulements supporte aisément un certain niveau vibratoire, il n'en va pas de même à l'arrêt où elles donnent naissance à la "corrosion de contact".
- j. **la pollution** : nécessite parfois l'adjonction de dispositif d'étanchéité.
- k. **la détérioration par étincelage** : résulte d'une mauvaise mise à la masse et provoque de la corrosion de contact.

3.4.2 Types de défauts de roulement

Les phénomènes de fatigue des surfaces qui roulent l'une sur l'autre sont complexes. Ils résultent de multiples paramètres liés au mouvement lui-même d'une part, aux caractéristiques des matériaux en présence et du lubrifiant d'autre part. On peut résumer en disant que, lors de la rotation, la surface et les sous-couches superficielles sont soumises à un ensemble de contraintes variables et cycliques de compression, de cisaillement et de traction. Ces derniers engendrent des défauts telles que (figure 3.11) : l'écaillage, la corrosion, les fractures, fissures profondes, l'usure d'adhésive, les piqûres (prick) etc.

Les défauts peuvent être classés en deux catégories: défauts indépendants de la fatigue, et défauts dépendants de la fatigue. Dans les tableaux 3.3 -3.4, nous avons récapitulé les différents défauts généralement rencontrés dans les roulements et les causes s'y rapportant. Nous donnons ci-dessous quelques détails sur certains défauts les plus fréquents :

1. **les écaillages** : Il existe deux zones où les contraintes sont maximales et où le système cristallin pourra céder :

- très près de la surface, à quelques dizaines de microns au maximum.
- un peu plus profondément, à quelques dixièmes de millimètres, au point de contraintes maximales de cisaillement de Hertz.

Dans la première zone de contraintes maximales, les *pertes initiales de matériaux* sont minimales et conduisent à l'apparition de *piqûres (prick)*.

Dans la seconde zone de contraintes, les *pertes* sont plus importantes, sous formes de particules fines, de contour en général tourmenté, ressemblant à des *écailles (scales)*.

Les causes de l'écailage sont multiples :

- défaut de lubrification,
- épaisseur du film d'huile insuffisante, graisse insuffisante,
- défaut métallurgique,
- formation d'oxyde qui facilite l'amorce d'une fissure qui se développe ensuite, etc.

2. les fissures profondes : ces fissures profondes sont la conséquence d'un *écailage* profond. Elles peuvent résulter :

- de chocs au montage,
- de chocs brutaux en service,
- de traitements thermiques défectueux,
- de jeu trop serré,
- de jeu trop grand sur la bague tournante qui provoquant des glissements,
- des phénomènes de corrosion.

3. l'usure d'adhésive : elle se produit dans le cas d'un graissage insuffisant.

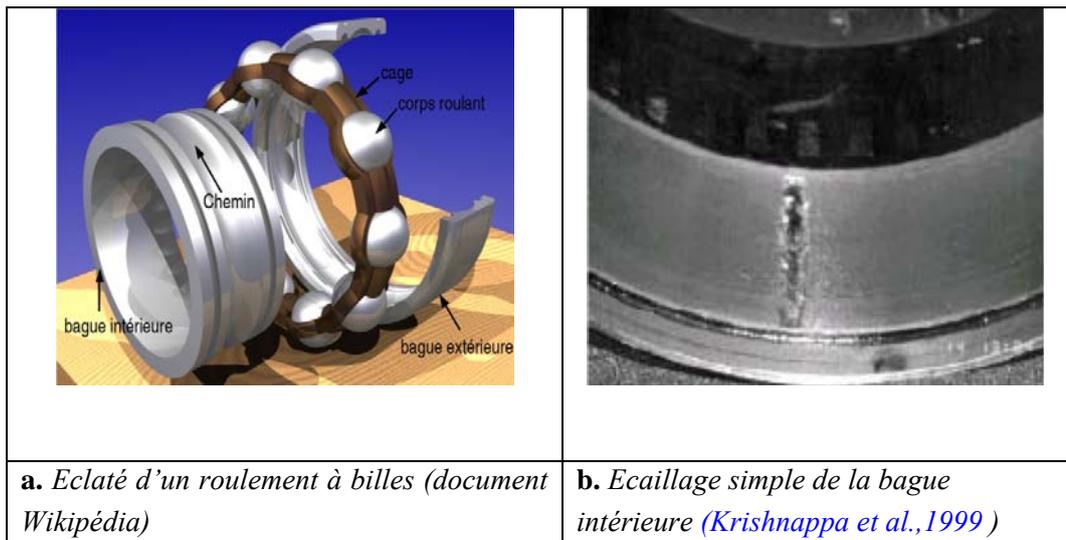




Figure 3.10 Quelques Types de défauts de roulement

Tableau 3.3 Synthèse des défauts indépendant de la fatigue touchant les roulements et leur causes

Type de défaut	Origine du défaut
Usure	Choix du lubrifiant Contamination de l'huile
Grippage et échauffement	Faible épaisseur de film d'huile Charge et vitesse
Corrosion de contact	vibrations

Tableau 3.4 Synthèse des défauts dépendant de la fatigue touchant les roulements et leur causes

Type de défaut	Origine du défaut
Ecaillage par inclusion	Vitesse et charge Zone de charge Faible épaisseur de film d'huile
Ecaillage par concentration de contraintes	Défaut d'alignement Zone de charge Faible épaisseur de film d'huile
Ecaillage par origine ponctuelle en surface	Aspérités de surface Faible épaisseur de film d'huile
Ecaillage superficiel	Faible épaisseur de film d'huile
Fatigue en sous couche	Charge
Fissuration transversale	Fatigue aggravée par les conditions de fonctionnement

Remarque : Les avaries (dites de fatigue) apparaissent même dans les conditions idéales de fonctionnement. La charge dans le roulement est transmise d'une bague à l'autre par les éléments roulants. En rotation, il se crée donc un champ de contraintes cycliques qui est responsable de la fatigue du roulement. Ce type d'avaries se manifeste par l'apparition de fissures qui croît lentement jusqu'à l'écaillage, C'est un phénomène aléatoire. Il existe 2 types de fissures : les fissures liées à la fatigue du roulement en charge (couche profonde et/ou couche superficielle) et les fissures microscopiques qui existent avant la mise sous charge, sous l'effet de la finition des surfaces.

La résistance à la fatigue d'un roulement est d'autant plus grande que l'acier est résistant, fin et homogène, c'est à dire que les défauts sont petits et dispersés.

3.4.3 Traitement du signal des roulements

Le spectre émis par un roulement est soit un "spectre de raies" soit un "spectre de bande". Au début de la détérioration, par exemple, il n'existe qu'un seul écaillage. Puis les

écaillages se multiplient; d'où la présence des harmoniques de la raie fondamentale (fréquence du défaut) jusqu'au moment où les écaillages sont si nombreux que la notion de fréquence discrète disparaît pour laisser la place à la notion de bande de fréquences.

Les méthodes de traitement des vibrations de roulements sont nombreuses et variées car elles correspondent à la diversité des problèmes rencontrés.

1. **Chocs répétitifs** : Dans le cas des défauts localisés (indentation, écaillage, rayure, etc.), les fréquences caractéristiques de répétition (f_{be} , f_{bi} , f_{er} , et f_c) qui apparaissent peuvent aisément se calculer puisqu'elles correspondent à des impacts (chocs multiples ou répétitifs) à chaque fois que les défauts isolés passent dans les *surfaces de contact* des éléments. Ces chocs seront représentés par un spectre de raie très étendu, à chaque composante de ce raie est associé plusieurs paires de bandes latérales espacées de la fréquence de rotation " f_r ". Cette présence de chocs répétés peut être vue comme une dégradation de la gaussienneté du signal d'où l'application de certains indicateurs comme le **skewness** et le **Kurtosis** ou une dégradation de l'hypothèse de stationnarité, dans ce cas, il faut appliquer des méthodes **cyclo-stationnaires ou non-stationnaires**

Exemple de spectre d'un écaillage sur la bague externe, il sera un :

- Spectre très étendu
 - Un raie de raies espacées d'un pas égal à la fréquence du défauts " f_{be} ".
 - A chaque composante de ce raie est associé plusieurs paires de bandes latérales espacées de la fréquence de rotation " f_r ".
2. **Chocs aléatoires** : Dans le cas des défauts uniformément répartis (insuffisance de lubrification, usure de pistes, défaut de forme, corrosion, ...), les vibrations ne sont plus à ces fréquences calculables, mais se répartissent dans une large bande fréquentielle (spectre large), c'est la conséquence d'un choc aléatoire. L'analyse en démodulation d'amplitude centrée autour d'une haute fréquence de résonance, permet de constater que ce défaut de choc n'a pas de périodicité et qu'il s'agit donc de choc aléatoire.
 3. **Signal périodique** : Dans le cas d'un déversement de bagues, les vibrations sont de type périodique. Le calcul des fréquences théoriques (f_{be} , f_{bi} , f_{er} , et f_c) s'effectue comme dans le premier cas. Ces fréquences difficilement détectables dans une zone du spectre particulièrement riche où l'on retrouve les principales raies liées à la cinématique de la machine.
 4. **Signal Cyclostationnaire** : Les roulements sont des organes sensibles qui peuvent être suivis efficacement par l'analyse vibratoire. Les signaux émis ont la particularité d'être essentiellement cyclostationnaires à l'ordre deux (fonction d'autocorrélation

périodique) plutôt que purement périodiques, car il existe des **glissements aléatoires** inévitables au niveau des éléments roulants (Antoni, 2001). La corrélation spectrale peut être d'une grande utilité

Remarque :

En plus, pour un roulement, la réponse de la structure à un choc est une oscillation amortie (Morel, 92 p.27). La fréquence de l'oscillation est celle des modes excités. Parmi ces modes, il semble que ceux correspondant aux déformations en anneaux des cages internes et externes soient prédominants. La fréquence de ces modes est généralement comprise entre 1 et 30 KHz.

Dans la bande inférieure à 10 KHz, ces modes sont observés conjointement avec les autres modes de la structure d'accueil. Au contraire, les modes dont la fréquence est supérieure à 10 KHz sont rarement perturbés, c'est la raison pour laquelle les constructeurs essaient maintenant de travailler dans les bandes 10 à 30 KHz. D'autre part la forme de l'oscillation amortie dépend de l'amortissement induit principalement par les forces de cisaillement au droit du film d'huile. Lorsqu'il n'y a qu'une seule impulsion par tour, ces forces sont suffisamment importantes pour que la vibration s'évanouisse complètement entre deux chocs successifs. Lorsque le nombre d'écaillages augmente, la vibration n'a plus le temps de s'évanouir. On trouve ainsi une raison supplémentaire pour passer du spectre de raies au spectre de bande.

Ces fréquences de passage sont théoriques, sensiblement respectées tant qu'il y a roulement parfait des billes sans glissement. Dès qu'il y a amorces de grippage, le frottement a tendance à ralentir la partie dégradée. Le défaut se manifeste donc à une fréquence f_d inférieure à la fréquence calculée par les formules. Cette différence bien difficile à appréhender est un élément décisif pour établir l'urgence de l'intervention, le grippage étant le stade ultime d'un défaut sur le roulement et pouvant amener par blocage à la dégradation brutale du rotor.

L'amplitude du signal vibratoire est influencée par la répartition des efforts dans le plan perpendiculaire aux roulements. En effet, les contraintes axiales et radiales appliquées aux roulements du réducteur ne sont pas isotropes. Cette répartition des forces est à l'origine des zones plus ou moins chargées comme le montre la figure 3.9b. Ainsi, dans le cas d'un écaillage sur la bague interne, l'amplitude des chocs dus au passage d'une bille sur le défaut dépend de la position du défaut par rapport à la zone de charge. Également, en pratique, l'analyse des défauts de roulement s'avère souvent complexe à cause de la variation de la vitesse et des efforts radiaux et axiaux engendrés par les engrenages dans une boîte à vitesses (variation de l'angle de contact et de la répartition des charges).

3.5 Banc d'engrenage

Les premiers signaux que nous proposons d'étudier ont été fournis par le CETIM « Centre technique des industries mécaniques, France » à plusieurs laboratoires. De nombreuses études ont été faites : temps-fréquences, spectrales, cepstrales, cyclostationnaires, non linéaires et des statistiques cycliques d'ordre supérieur (Capdessus, 1992), (Bouillaut, 2000), (Felkaoui, 2007), (Elbadaoui, 1999). Ces signaux sont issus d'un banc d'essai de grandeur industrielle (figure 3.12), fonctionnant 24h/24h. Le dimensionnement des roues de l'engrenage cémenté trempé, ainsi que les conditions de fonctionnement (vitesse, couple) sont fixés de façon à obtenir un **écaillage** sur toute la largeur d'une dent (qui fixe la fin de l'essai, la rupture étant imminente) en environ 15 jours. Tous les jours après l'acquisition des signaux vibratoires, le banc est arrêté pour observer l'état des dents des roues. Un rapport d'expertise est ainsi rédigé voir tableau 3.5.

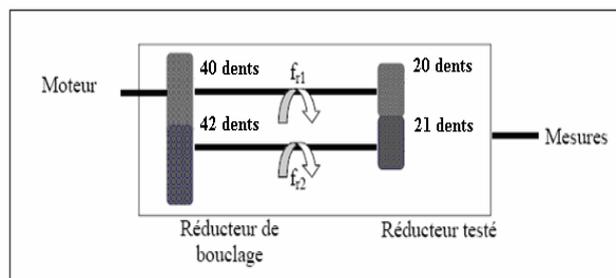


Figure 3.12 Schéma du dispositif d'engrenage

Tableau 3.5 Rapport d'expertise

Jour	Observation
2	1 ^{er} jour d'acquisition, pas d'anomalie
3, 4, 5, 6	pas d'anomalie
7	Ecaillage de la surface de la dent 1/2
8	Dent 1/2 Pas d'évolution
9	Dent 1/2 Pas d'évolution, début d'écaillage sur la surface de la dent 15/16
10, 11, 12	Pas d'évolution
13	Ecaillage sur toute la surface de la dent 15/16

Le dispositif est composé d'un moteur, d'un réducteur de bouclage de rapport 40/42 et du réducteur testé, de rapport 20/21. La vitesse du moteur est de **1000 tr/min**, soit donc une fréquence de rotation $16,67 \text{ Hz}$ et une fréquence d'engrènement de $333,48 \text{ Hz}$. Or, les deux roues composant ce réducteur ont un nombre de dents quasiment identique. Leurs fréquences de rotation sont donc très proches. Nous avons en effet : $f_1 = 16,67 \text{ Hz}$ et $f_2 = 15,88 \text{ Hz}$. La fréquence d'échantillonnage est de $f_e = 20 \text{ kHz}$, le

nombre de points par tour respectivement de la roue 1 et 2 sont : $Np_1 = 1200pts$ et $Np_2 = 1260pts$.

Un enregistrement a été fait tous les jours de la seconde journée à la treizième. Nous disposons donc d'un ensemble de douze signaux de 60160 points, échantillonnés à 20kHz, représentant chacun 3 secondes soit à peu près 47 périodes de rotation. Au cours de cette campagne de test, l'engrenage passe de l'état de bon fonctionnement à celui d'engrenage détérioré sur une durée de 13 jours, le défaut qui apparaît étant un écaillage. La figure 3.13 nous montre les signaux temporels représentant 3 périodes de rotation. L'analyse des signaux temporels nous permet de visualiser l'apparition d'un défaut très marqué le douzième jour. Malheureusement le signal d'un codeur optique n'a pas été fourni pour pouvoir effectuer des traitements de cyclostationnarité

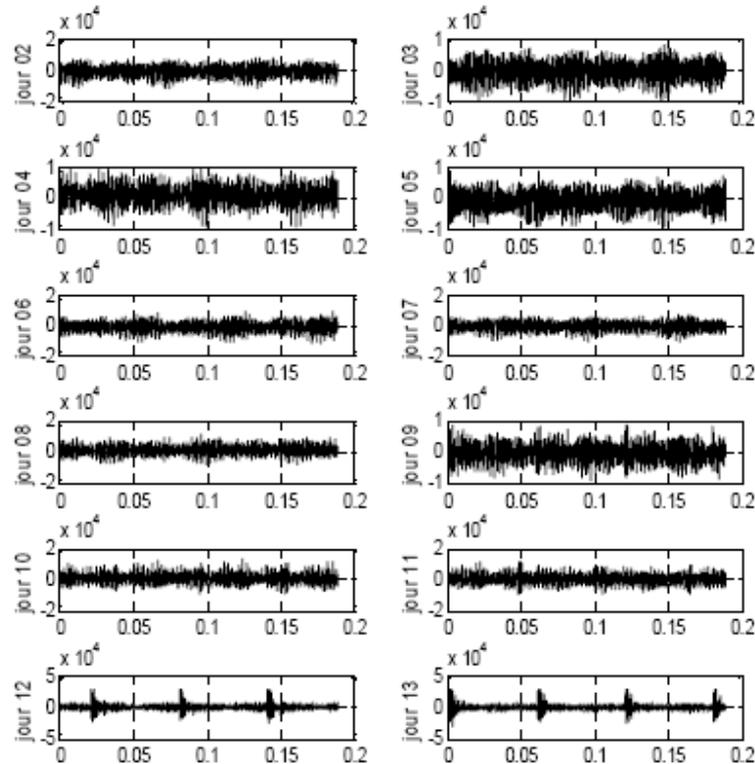


Figure 3.13 Signaux temporels du train d'engrenage représentant trois

3.6 Banc de roulements

Cette base de données provient du banc d'essai de roulements de l'université " *The Case Western Reserve University - Bearing Data Center. Cleveland, Ohio. USA.* " et peut être téléchargée librement sur l'adresse web suivante (www.eecs.cwru.edu/laboratory/bearing/welcome_overview.htm). Les roulements testés soutiennent l'arbre du moteur des deux cotés (coté entraînement et coté ventilateur). La

Figure 3.13, présente le banc d'essai qui se compose principalement d'un moteur "2 hp ⁽²⁾" (à gauche), d'un accouplement « transducteur/encodeur » (au centre), d'un dynamomètre (à droite) et des circuits de commande. Une première analyse a été faite par (Yang et al., 2005) dans laquelle des outils de transformée en ondelettes ont été appliqués.

De simples défauts sous forme de points « de différent diamètres et profondeurs » ont été créés sur les différents organes composants les roulements à billes en utilisant l'usinage électro-décharge. Les diamètres du défauts sont respectivement, 0.007, 0.014, 0.021 et 0.028 inches (1 inches=25.8mm). Ces diamètres constituent avec le cas sain, les 5 classes de reconnaissance de forme.

Les signaux ont été enregistrés à l'aide des accéléromètres, fixés à la cage du moteur par aimant, pour différents couples résistants (donc différentes vitesses de rotations). Les couples sont respectivement 0, 1, 2, 3 hp donc des vitesses respectives de $f_r = 1797, 1772, 1750, 1730$ rpm (1 rpm=1 tr/min= 1/60Hz). Ces accéléromètres ont été placés à des positions respectivement de 12 heures "verticale" et de 3 heures "horizontales", des deux cotés "accouplement et ventilateur" de la cage du moteur. La fréquence d'échantillonnage est $f_e = 12$ kHz et le nombre de points total est de $N_T = 243938$ pts

Les caractéristiques des roulements sont les multiples de la vitesse de rotation f_r :

Fréquence de la bague intérieure : $f_{bi} = 5,415 (*f_r)$

Fréquence de la bague extérieure : $f_{be} = 3,584 (*f_r)$

Fréquence de la cage : $f_c = 0,3982 (*f_r)$

Fréquence de la bille : $f_b = 4,7135 (*f_r)$

Tous les fichiers de données sont au format (*.mat) Matlab. Chaque fichier contient trois données: signaux de vibration des deux cotés et la vitesse de rotation du moteur. Pour notre étude nous avons considéré, seulement, les signaux du côté accouplement.

La figure 3.14 montre un exemple de signaux temporels pour différents diamètres de défaut, réalisés sur la bague intérieure et pour un couple résistant maximal (3 hp). Les signaux sont ceux recueillis du côté de l'accouplement. Malheureusement, nous ne disposons pas du signal d'un codeur optique pour effectuer des analyses plus poussées tels que l'analyse en ordre ou le ré-échantillonnage angulaire.

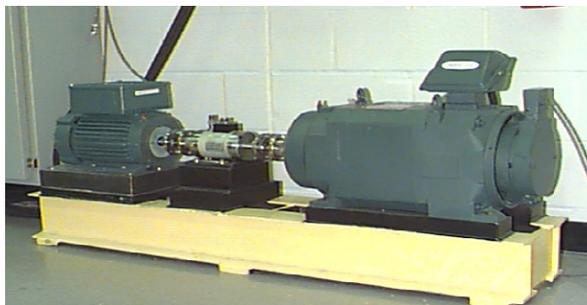


Figure 3.14 Banc d'essai de roulements

⁽²⁾ hp : horsepower, 1hp = 745, 69987 watt.

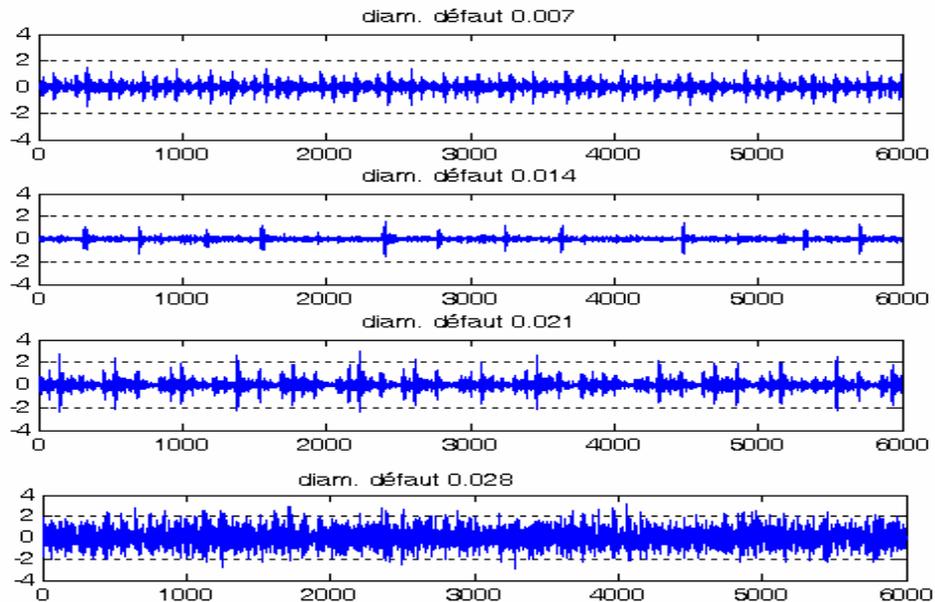


Figure 3.15 Signaux temporels, couple maximal pour différents diamètres du défaut sur la bague intérieure.

3.7 Banc de la boîte à vitesse de l'hélicoptère CH-46E

L'ensemble de signaux étudiés est publiquement disponible sur un CD-ROM. Il provient de l'instrumentation d'une boîte à vitesse CH-46E d'un hélicoptère de la marine Américaine des *USA*, par l'université d'état de Pennsylvanie. Qui souhaite comparer différentes méthodes de diagnostic. Chaque signal a été acquis à une fréquence d'échantillonnage $f_e = 103116$ Hz pendant une durée de 3.9922 s (soit 412464 échantillons), la charge n'a pas été communiqué.

De nombreuses études ont été réalisées sur ces signaux, nous citons entre autres, (Braun,1999), (Soon, 2000), (Williams,2000), (Loughlin,2000) pour l'analyses spectrale; les méthodes temps fréquences et temps échelle. Les approches cyclostationnaires et bilinéaire ont été utilisées dans les travaux de Bouillaut (Bouillaut, 2000), (Bouillaut, 2001). Raad (Raad, 2003) quand elle a appliqué principalement des méthodes cyclostationnaires à l'ordre 3. Nous trouvons également d'autres travaux sur la classification et le diagnostic (Zacksenhouse, 2000), l'analyse vibro-acoustique (Gelman et al., 2000) et (Wong, 2005), dans le but d'établir un diagnostic précoce.

Nous disposons de signaux caractérisant le système **pour huit conditions de fonctionnement**, sept défauts qui se produisent dans les différentes parties du système et un ensemble de données provenant du système sain. Nous disposons également du signal d'un codeur optique. La figure 3.16a présente une photos d'un hélicoptère de la marine

américaine. La figure 3 16b présente un schéma simplifié du module principale de la boîte à vitesses, énumérant les éléments ainsi que leur nombre de dents et leurs fréquences d'engrènement. Le tableau 3.1 récapitule les sept défauts disponibles ainsi que leur nature corrosion, piquûre, écaillage, fissure,...), les vitesses caractéristiques et les pièces concernées.

Le numéro 9 étant réservé au cas normal sans défaut. Seulement un seul défaut est considéré à la fois. Pour chaque défaut, huit accéléromètres enregistrent, en différents endroits de la boîte de vitesse, les vibrations engendrées par le système. Les différentes positions des capteurs sont les suivantes :

- (1) Starboard engine input.
- (2) Port engine input.
- (3) Aft side of mix box.
- (4) Starboard quill shaft.
- (5) Starboard planetary.
- (6) Port planetary.
- (7) Port quill shaft.
- (8) Accessory drive.

Les mesures sont effectuées à **vitesse constante** mais pour **neuf couples différents**, exercés par les pales du rotor sur l'arbre 20 (figure3.2). Le tableau 3.2 ci dessous présente les données disponibles en fonction du couple et du défaut considérés. Le défaut numéro 9 correspond en fait au système sain.

Le système complexe tel que les boîtes de vitesse d'hélicoptères, étant très riches. Il représente pour nous une base de données très importante pour tester les différents travaux sur le diagnostic.



Figure 3.16a Photos d'un hélicoptère Westland de la marine américaine

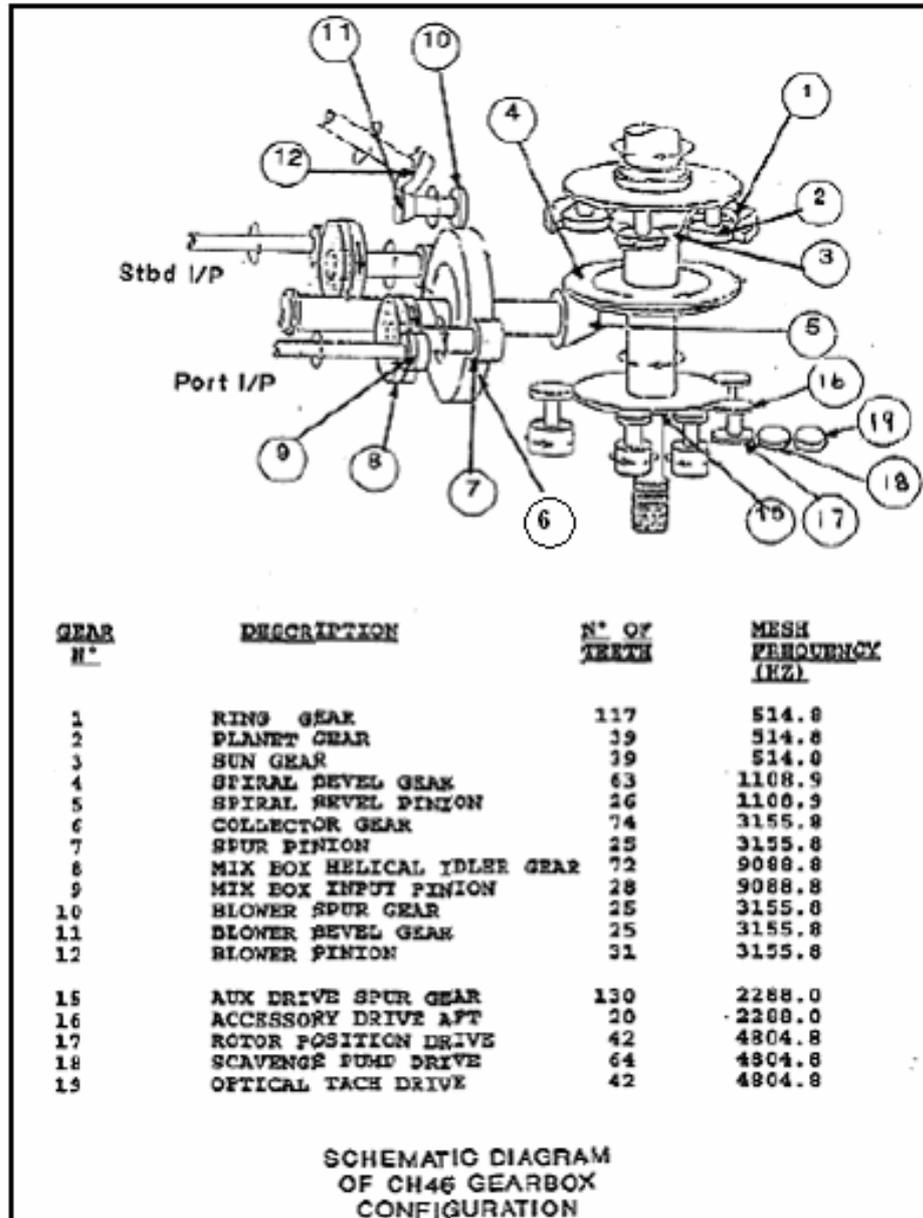


Figure 3.16b Schéma simplifié de la boîte de vitesse de l'hélicoptère CH46, US NAVY Westland Pennsylvania

Tableau 3.6 Numérotation des éléments et caractéristiques.

Numéro de défaut	Pièces concernées par le défaut	Vitesse de l'élément	Description du défaut
2	2	$f_r=13.20 \text{ Hz}$ $f_{be}=?$	Corrosion de palier de l'engrenage planétaire épicycloïdal
3	5	$f_r=42.65 \text{ Hz}$ $f_{bi}=?$	Corrosion du palier du pignon conique d'entrée
4	5	$f_r=42.65 \text{ Hz}$ $f_e=1108.9 \text{ Hz}$	Ecaillage (spalling) du pignon conique spiral d'entrée Z=26 dents
5	9	$f_r=324.60 \text{ Hz}$ $f_e=9088.8 \text{ Hz}$	Ecaillage (chipping) du pignon d'entrée Hélicoïdal Z=28 dents
6	8	$f_r=126.23 \text{ Hz}$ $f_e=9088.8 \text{ Hz}$	Fissure de l'engrenage 'Helical Idler ' Z=72 dents
7	6	$f_r=42.65 \text{ Hz}$ $f_e=3155.8 \text{ Hz}$	Fissure de l'engrenage collecteur Z=74 dents
8	5 et 6	$f_r=42.65 \text{ Hz}$	Fissure de l'arbre 'Quill Shaft'

Tableau 3.8 Ensemble des données disponibles en fonction du couple et du défaut

Couple / Défaut #	2	3	4	5	6	7	8	9
100%	X	X	X	X	X	X	X	X
80%		X	X	X	X	X	X	X
75%		X	X	X	X	X	X	X
70%		X	X	X	X	X	X	X
60%		X	X			X	X	X
50%		X	X			X	X	X
45%		X	X				X	X
40%		X	X			X	X	X
27%		X	X			X	X	X

Il est à noter que, pour chaque défaut et pour chaque couple, nous ne disposons généralement que de trois signaux : un premier enregistrement provenant du système sain un second caractérisant le défaut partiel et un dernier dû au défaut établi.

Nous soulignons que le CD-ROM présente des photos des différents éléments avant les tests ainsi que des vues des différents niveaux de défaut.

La figure ci-dessous présente un exemple simplifié des éléments 4,5, 6 et 7 atteints par les défauts. Les accéléromètres les plus proches sont 4,8

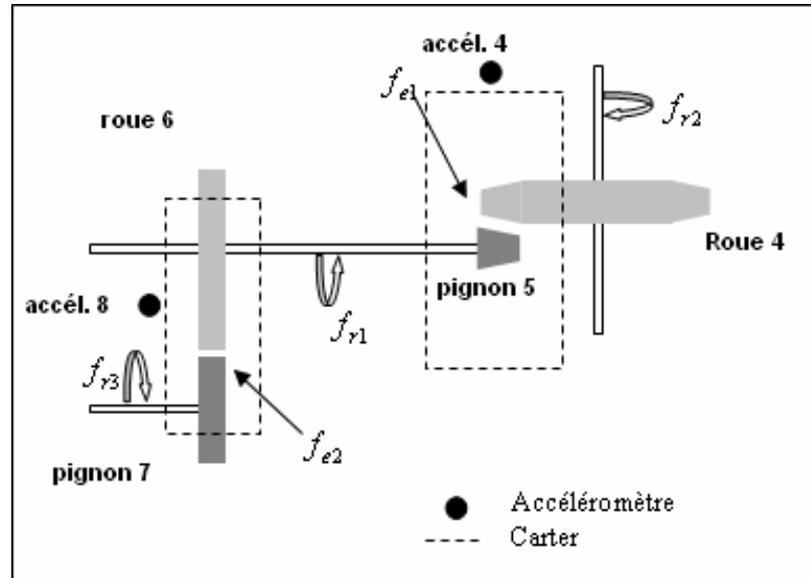


Figure 3.17 Schéma simplifié du système

Dans la suite de cette thèse, nous nous limiterons à l'étude d'un seul défauts particuliers : l'écaillage de l'engrenage conique. La figure 3.18 présente des photos de l'engrenage avant les tests ainsi que des vues des différents niveaux de défaut.

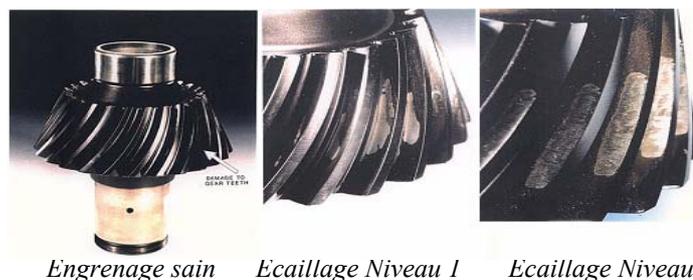


Figure 3.18. Les différents pignons coniques spirals utilisés

Ci-dessous nous présentons les différentes données nécessaires à notre traitement :

Fréquence d'échantillonnage : $f_e = 103116 \text{ Hz}$

Nombre de points total est : $N_T = 412464 \text{ points}$

Les fréquences d'engrènement : $f_{eng1} = 1108.9 \text{ Hz}$ $f_{eng2} = 3155.8 \text{ Hz}$

Nombre de dents du pignon conique 5 : $nd_1 = 26 \text{ dents}$

Sa fréquence de rotation $f_{r1} = f_{eng1} / nd_1 = 42.65 \text{ Hz}$

Le nombre de points par période N_p est : $N_p = f_e / f_{r1} = 2418 \text{ pts}$

Soit un temps de : $T = N_p * \Delta T = N_p * 1 / f_e = 23.44 \text{ ms}$

Codage des données

Les fichiers de données des signaux de Westland data sont au format "*.bin" binaire, afin de pouvoir les analyser il faut avant tout les convertir au format ".mat" matlab, à l'aide de l'utilitaire "PCCONV" présent sur le CD-ROM.

Le nombre de pièces testées (Tableau 3.6), leurs niveaux de dégradation, les différents couples résistants utilisés (Tableau 3.7.), et les huit capteurs implémentés sur la boîte vitesse CH-46, ont poussé les ingénieurs de Westland à donner un codage aux signaux acquis pour pouvoir les manipuler de manière efficace. De ce fait, on a décidé de garder le même codage.

Exemple: W4451602: W: Westland
 4: Défaut # 4
 45: Couple 45 %.
 16: Run 16. Le numéro de l'essai,
 02: la position du capteur utilisé.

L'analyse des signaux temporels pour un couple de 45%, nous permet de distinguer seulement la différence entre les amplitudes des trois signaux représentés sur la figure 3.19, on remarque que la sévérité du défaut a une influence directe sur l'amplitude du signal.

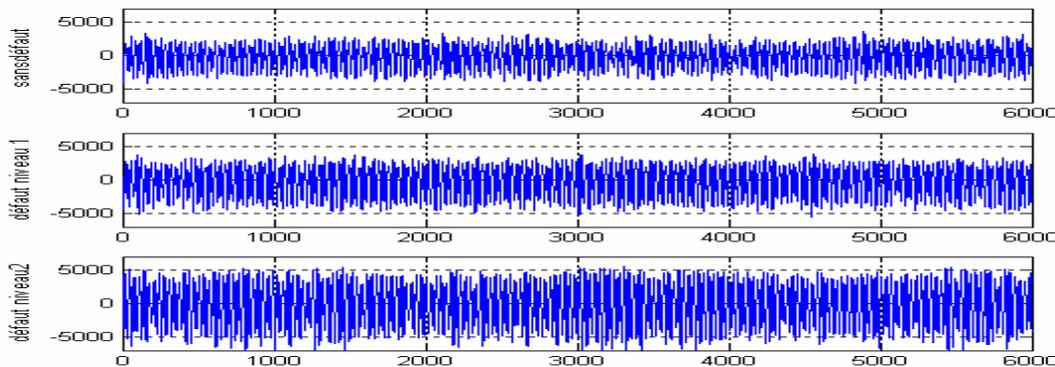


Figure 3.19 Signaux temporels pour un couple 45%,
 pour différents états du pignon conique

3.8 Conclusion

Dans ce chapitre, nous avons présenté au début des statistiques des défauts des machines tournantes. En suite nous nous sommes intéressés aux origines et natures des défauts d'engrenages et de roulements pour justifier la nécessité d'utiliser un grand nombre d'indicateurs à la fois. Enfin, nous avons exposé les trois bancs d'essais utilisés dans les travaux de cette thèse. Dans le prochain chapitre, nous appliquerons les méthodes de sélection et de classification.

Chapitre 4

Applications des méthodes de classification par RdF

4.1 Introduction

Le diagnostic d'un système est l'identification du mode de fonctionnement, à chaque instant, par ses manifestations extérieures. Son principe général consiste à confronter les données relevées au cours du fonctionnement réel du système avec la connaissance que l'on a de son fonctionnement normal ou défaillant. Si le mode de fonctionnement identifié est un mode défaillant, le système de diagnostic peut localiser sa cause.

Dans le présent chapitre, nous introduisons tout d'abord le principe de diagnostic par reconnaissance des formes, nous discutons ensuite et en détails, l'importance de la réduction du nombre d'indicateurs et puis, nous proposons une méthode de sélection d'indicateurs. Enfin, trois algorithmes de classification sont appliqués et discutés : les k-ppv, le PMC, et les SVM.

4.2 Principe du diagnostic par RdF

Effectuer un diagnostic par reconnaissance de formes c'est identifier le mode de fonctionnement et localiser éventuellement sa cause. L'application de la RDF au diagnostic est réalisée en associant un mode de fonctionnement à une classe ([Dubuisson, 2001](#)). La RDF est un des nombreux aspects de l'intelligence artificielle ou plus exactement de l'intelligence calculatoire. A partir d'un ensemble de données ou d'informations apprises, elle offre la possibilité d'interpréter toute nouvelle observation (ou forme). Les observations déjà connues sont regroupées en classes, constituant des prototypes auxquels la nouvelle observation est comparée pour être identifiée. Les algorithmes utilisés permettent donc de classer des observations dont les propriétés ont varié par rapport à une observation type. Il s'agit donc d'un outil qui a, entre autres, la capacité d'apprentissage ([Duda et al., 2001](#)), ([Casimir,2003](#)).

En reconnaissance de formes, chaque observation est représentée par un vecteur X_i de M paramètres réels appelé vecteur forme tel que $X_i = [x_1, x_2, \dots, x_M]^T$. Ce vecteur sera représenté par un point dans l'espace de dimension M , \mathfrak{R}^M , voir figure 4.1, autrement appelé espace de représentation.

Supposons maintenant que pour toute nouvelle forme, nous ayons à décider parmi N_c formes types considérées comme étant des prototypes. A cause du bruit de mesure, de la précision des capteurs, etc., une nouvelle observation sera rarement identique à l'un des prototypes (points confondus dans \mathfrak{R}^M). Les classes $(w_1, w_2, \dots, w_{N_c})$ correspondent à des zones dans l'espace, regroupant les formes semblables. Ces **zones** de l'espace de décision sont notées $(\Omega_1, \Omega_2, \dots, \Omega_{N_c})$ où $N_c \leq M$.

L'objectif en reconnaissance des formes est alors de décider à laquelle des **classes** $(w_1, w_2, \dots, w_{N_c})$ doit être associée une nouvelle forme " X ". Cela revient à rechercher dans quelle zone de l'espace se situe cette nouvelle observation (figure 4.2).

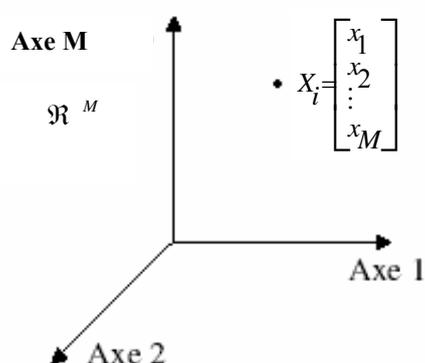


Figure 4.1 Caractérisation d'une observation par un vecteur forme : espace de représentation \mathfrak{R}^M

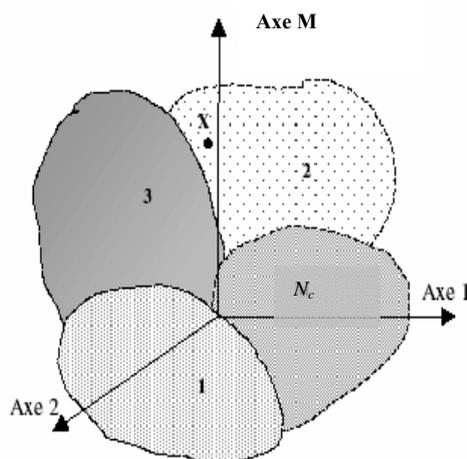


Figure 4.2 Association d'une nouvelle observation X à l'une des classes N_c : espace de décision

Si on se place dans le contexte du diagnostic, les paramètres du vecteur forme traduisent l'état de la machine étudiée. Ils sont issus d'analyses effectuées sur les signaux fournis par les capteurs implantés (vibrations, vitesse, courants, ou encore tensions par exemple). L'objectif en diagnostic est alors de définir à quel mode de fonctionnement (ou classe) correspond une observation recueillie " X " sur la machine. Cela est difficilement réalisable si la recherche de signature est basée sur un seul type d'analyse ou d'indicateur et que les données brutes sont bruitées, de natures complexes, contiennent des informations parasites, et ont subi des dégradations entre la source et le capteur. Dans la prochaine section, on discutera de l'importance à choisir un vecteur forme contenant des indicateurs pertinents et non redondants.

4.3 Importance de la sélection d'indicateurs (Banc des roulements)

Les travaux présentés dans cette section portent sur la sélection d'un nombre réduit de paramètres pertinents pour représenter les signatures vibratoires des défauts de roulements du moteur asynchrone (voir section 3.6). Nous avons pu montrer, que les performances des méthodes de reconnaissance de formes (RdF) sont étroitement liées à la pertinence des indicateurs de défauts. La procédure est validée sur des signaux réels vibratoires de roulements sains et défectueux et en utilisant des indicateurs temporels et fréquentiels.

4.3.1 Calcul d'indicateurs

Indicateurs temporels

Les indicateurs temporels sont calculés dans 5 bandes fréquentielles issues de la décomposition de la bande d'analyse "6 kHz" en 4 sous bandes adjacentes « de largeur 1500 Hz chacune » plus la bande totale. Dans chaque bande, nous avons choisi de calculer 4 indicateurs qui sont : la valeur efficace, le facteur de crête, la valeur crête à crête et le kurtosis.

Indicateurs fréquentiels

Le spectre [0-6 kHz] présenté sur la figure 4.3 montre que l'énergie de ce signal est concentrée autour des quatre résonances de fréquence respectivement 600 Hz, 1393 Hz, 2508 Hz et 3604 Hz.

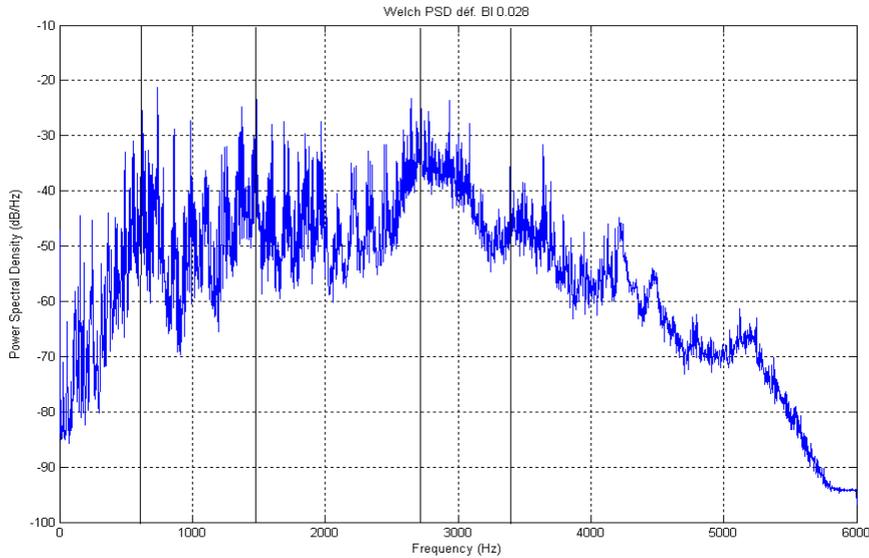


Figure 4.3 DSP des signaux de défaut de la bague intérieure, puissance moteur maximale, Ø du défaut 0,028 inches (1 inches= 25.8mm)

La figure 4.4 représente les énergies calculées dans les différentes bandes fréquentielles du spectre d'enveloppe, on peut remarquer que l'influence du diamètre du défaut est très bien détectée pour les différentes puissances utilisées.

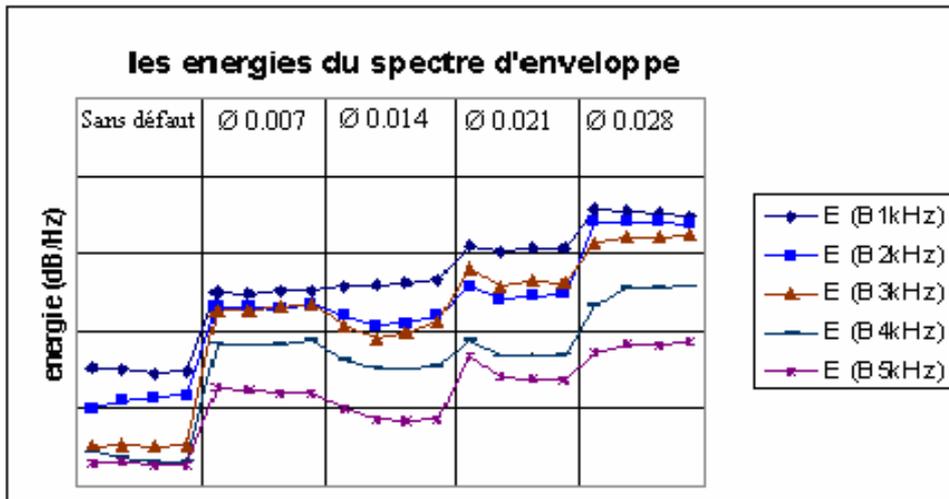


Figure 4.4 Les énergies du spectre d'enveloppe dans les différentes bandes fréquentielles de largeur 1KHz.

Définition des modes de fonctionnement et du Vecteur forme initial

Pour étudier le défaut de la bague intérieure, on a cinq classes de fonctionnement en fonction de la taille du défaut, ils sont représentés sur le Tableau 4.1.

Tableau 4.1 Modes de fonctionnement

Etiquette	Mode de fonctionnement	Puissance [HP]
1	Sans défaut	0, 1, 2, 3
2	Défaut BI Ø=0.007 [inches]	0, 1, 2, 3
3	Défaut BI Ø=0.014 [inches]	0, 1, 2, 3
4	Défaut BI Ø=0.021 [inches]	0, 1, 2, 3
5	Défaut BI Ø=0.028 [inches]	0, 1, 2, 3

La signature du système est définie par un vecteur de vingt six paramètres. Le Tableau 4.2 représente un panorama complet de l'ensemble de ces paramètres.

Tableau 4.2 Paramètres du vecteur forme initial

α_i	Domaine	Dénomination	
1	Temporel	Bande 0 -1500 Hz	
2			Valeur efficace 1
3			Facteur de crête 1
4			Valeur crête à crête 1
5		Bande 1500-3000 Hz	Kurtosis 1
6			Valeur efficace 2
7			Facteur de crête 2
8			Valeur crête à crête 2
9		Bande 3000-4500 Hz	Kurtosis 2
10			Valeur efficace 3
11			Facteur de crête 3
12			Valeur crête à crête 3
13		Bande 4500-6000 Hz	Kurtosis 3
14			Valeur efficace 4
15			Facteur de crête 4
16			Valeur crête à crête 4
17		Bande totale 6 kHz	Kurtosis 4
18			Valeur efficace 5
19			Facteur de crête 5
20			Valeur crête à crête 5
21	Fréquentiel	Energie dans la bande 0 - 1000 Hz	
22			Energie dans la bande 1000-2000 Hz
23			Energie dans la bande 2000-3000 Hz
24			Energie dans la bande 3000-4000 Hz
25			Energie dans la bande 4000-5000 Hz
26			Energie dans la bande totale 5 kHz

Quelques-uns de ces paramètres risquent d'être non pertinents ou redondants selon l'ensemble d'apprentissage à définir. Le dimensionnement de l'espace est donc revu à la baisse grâce à la sélection de paramètres " par le critère de Fisher" (Boumaza, 1999)

4.3.2 Visualisation par analyse en composante principale (ACP)

La visualisation de la base de données par ACP, figure 4.5, dont les éléments sont définis par le vecteur forme, permet d'appréhender les différentes formes (ou classes) présentes (Hardle, 2003). On en déduit de précieuses indications sur la disposition des classes et plus particulièrement, cela permet d'obtenir une première appréciation de la séparabilité de celles-ci.

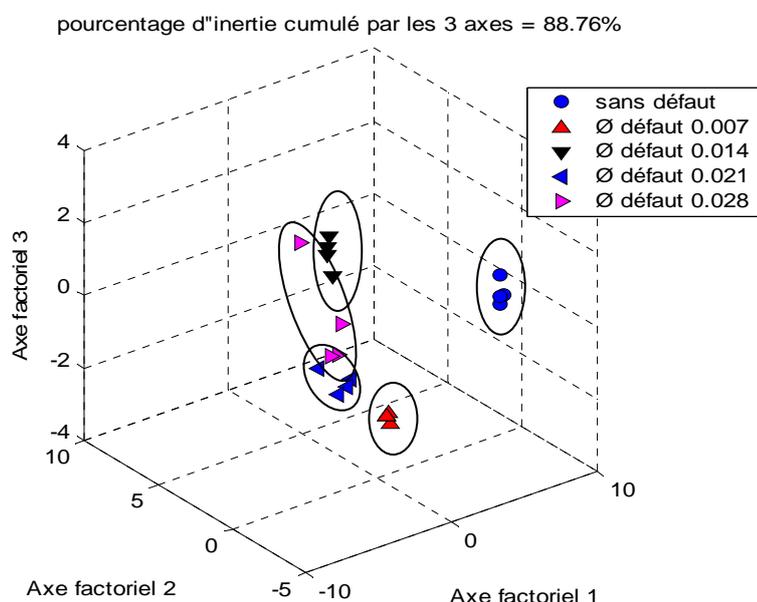


Figure 4.5 Position des classes dans le sous-espace de dimension 3

Nous remarquons que les classes ne sont pas parfaitement compactes (les observations d'une même classe) et ne sont pas disjointes (les classes entre-elles). Nous espérons améliorer la séparabilité par une méthode de sélection que nous considérons ci-dessous.

4.3.3 Sélection d'indicateurs par le critère de fisher

L'analyse de Fisher est une technique d'analyse factorielle discriminante, qui permet la projection de la base de données sur une surface informative de dimension réduite. Contrairement à l'ACP, il ne s'agit plus d'extraire les paramètres les plus représentatifs de la base de données, mais les paramètres les plus discriminants, vis-à-vis des classes représentées dans la base de données : le problème est donc la détermination d'un sous ensemble d'axes discriminants (Kudo et al., 2000), (Dubuisson, 1990).

L'objectif est d'augmenter la séparabilité des classes : la détermination des axes discriminants est effectuée de manière à ce que la variation interclasses soit maximisée (dispersion des classes entre elles) au regard de la variation intraclasse (dispersion à l'intérieur d'une classe) (casimir, 2003) et (Casimir et al., 2006).

D'une manière générale, pour M classes et pour un paramètre α , le critère de Fisher $J(\alpha)$ s'écrit :

$$J(\alpha) = \frac{\sum_{i=1}^M \sum_{j=1}^{M-1} (m_i(\alpha) - m_j(\alpha))^2}{N_i \sigma_i^2(\alpha) + N_j \sigma_j^2(\alpha)} \tag{4.1}$$

Avec : m_i : centre de gravité de la classe ω_i .

σ_i^2 : variance de la classe ω_i

N_i : nombre de vecteurs dans la classe ω_i

La figure 4.6 montre la variation du critère de Fisher pour les 26 paramètres du vecteur forme.

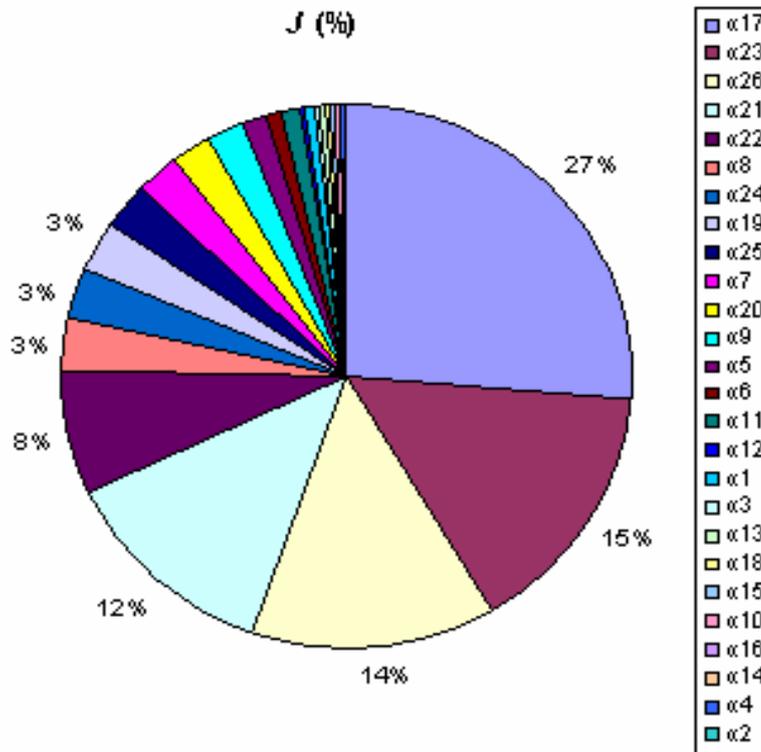


Figure 4.6 Critère de Fisher pour les différents indicateurs

Bien que le critère établi permet de classer les indicateurs calculés par ordre décroissant de puissance de discrimination, le nombre qui doit en être conservé ne peut être déterminé qu'à la suite d'essais pratiques par une règle de décision. Toutefois, Pour

remédier à ceci, on va utiliser le test d'arrêt qui consiste à déterminer la dimension correspondant à un pourcentage de perte d'informations admissible fixé a priori par l'utilisateur (l'expert) (Oukhellou, 1997). L'expression (4.2) permet de calculer le pourcentage expliqué par chaque composante du VF.

$$J_{N_j}(\%) = \frac{100 \times J(\alpha_j)}{\sum_{i=1}^d J(\alpha_i)} \quad (4.2)$$

Avec : $j=1, \dots, d$ $i=1, \dots, d$

Les résultats obtenus sont représentés sur la figure 4.7 :

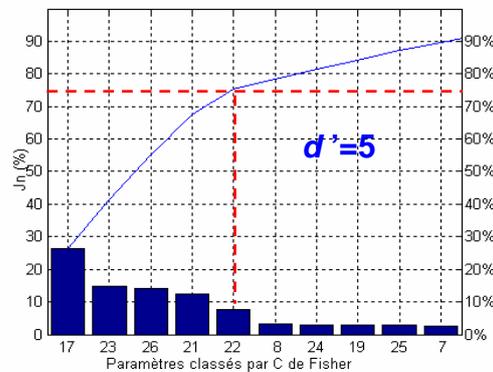


Figure 4.7 Pourcentage expliqué et cumulé des différents indicateurs classés par le critère de Fisher

Si on considère qu'un seuil supérieur à 75% est largement acceptable, le vecteur forme sera donc de dimension $d' = 5$:

$$\mathbf{x} = [\alpha_{17}, \alpha_{23}, \alpha_{26}, \alpha_{21}, \alpha_{22}]$$

Les paramètres correspondent respectivement à (Tableau 4.2):

- α_{17} : La valeur efficace dans la bande totale [0 – 6 kHz]
- α_{23} : Energie du spectre d'enveloppe dans la bande [2-3 kHz]
- α_{26} : Energie du spectre d'enveloppe dans la bande d'analyse totale [0-5 kHz]
- α_{21} : Energie du spectre d'enveloppe dans la bande [0-1 kHz]
- α_{22} : Energie du spectre d'enveloppe dans la bande [1-2 kHz]

On va réaliser une ACP du VF réduit ($d'=5$) pour pouvoir juger la pertinence des indicateurs retenus, les résultats obtenus seront comparés à ceux obtenus avec le VF total (initial).

4.3.4 Visualisation du vecteur forme réduit

La figure 4.8 montre la position et la forme des classes dans un espace de trois dimensions, à partir du vecteur forme de dimension $d' = 5$ choisi.

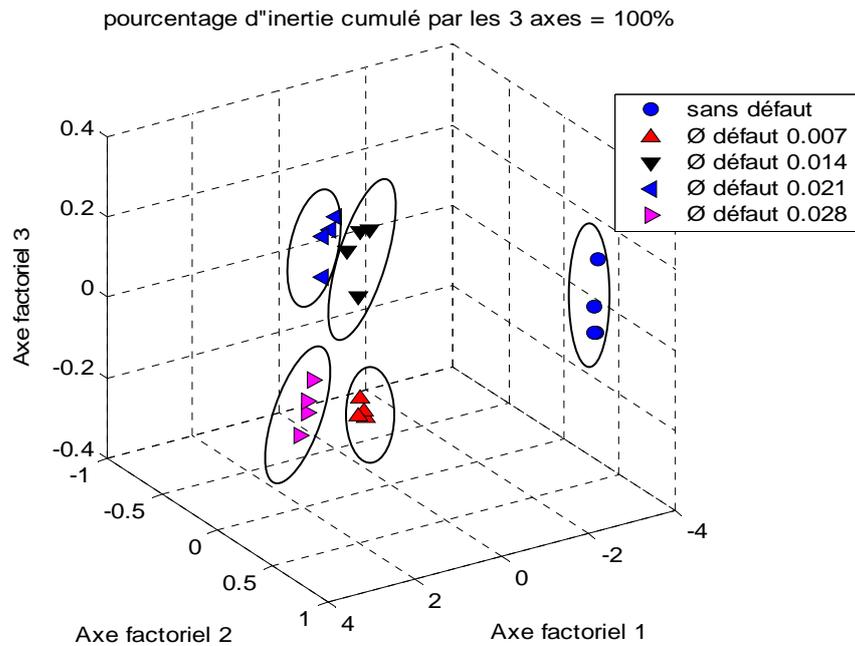


Figure 4.8 Position des classes dans le sous – espace de dimension 3

Le pourcentage d'inertie cumulé par les trois valeurs propres 100% est largement supérieur à celui obtenu par le VF initial (88,76%). De plus on peut constater visuellement sur la figure 4.8 que la séparabilité des classes obtenue avec le VF réduit est meilleure que celle obtenue avec le VF initial de la figure.4.5.

Egalement, Nous avons pu montrer dans cette section 4.3 à travers la visualisation par ACP, que la séparabilité entre modes de fonctionnement est étroitement liées à la pertinence des indicateurs de défauts. La procédure est validée sur des signaux réels vibratoires de roulements. Dans la prochaine section cette même procédure sera validée sur des signaux de l'hélicoptère de Westland.

4.4 Importance de la sélection d'indicateurs (signaux d'hélicoptère)

De nombreuses études ont été réalisées sur ces signaux de la section 3.7, pour les méthodes temps-fréquences, temps-échelles et l'analyse spectrale (Lau et al., 2000) (Williams et al., 2000), (Loughlin et al., 2000), pour l'approche cyclostationnaire (Bouillaut, 2000), (Bouillaut et al., 2001), pour la classification par RNA (Zacksenhouse et al., 2000), pour l'analyse vibro-acoustique (Gelman, et al., 2000). Nous allons montrer dans cette section et à travers la visualisation par ACP, l'importance d'une sélection d'indicateurs.

4.4.1 Calcul d'indicateurs

Indicateurs temporels

Généralement, dans le domaine temporel pour un engrenage, seules les résonances centrées autour des fréquences d'engrènement des roues véhiculent les informations relatives à la présence du défaut caractérisé par des impulsions (Boulenger et al., 2003) (Alattas, 1994). La figure 4.9 représente l'évolution des indicateurs temporels calculés dans 5 bandes fréquentielles. Les 4 premiers sont issues de la décomposition de la bande d'analyse "6 kHz" en 4 sous bandes adjacentes de 1.5 kHz. La 5^{ème} est l'énergie dans la bande totale. Dans chaque bande on a choisis de calculer 3 indicateurs qui sont : la valeur efficace, le facteur de crête et le kurtosis.

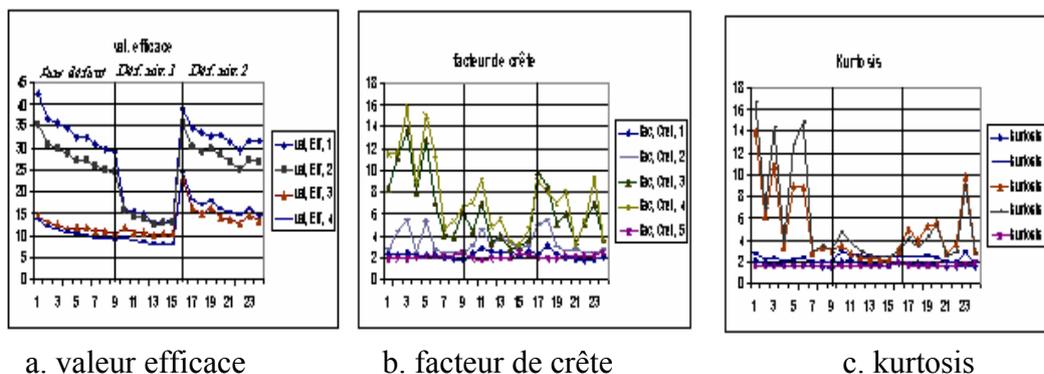


Figure 4.9 Evolution des indicateurs temporels

Indicateurs fréquentiels

Nous avons choisis de filtrer les signaux par un filtre passe bas à réponse impulsionnelle finie d'ordre 48, généré par la procédure Matlab* FIR1. Les signaux analytiques sont alors calculés pour chaque enregistrement. Nous disposons de signaux de

* MATLAB : Système interactif de calcul numérique. <http://www.mathworks.com>

412464 points, échantillonnées à 103116 Hz. Par filtrage, nous avons limité notre étude à une fréquence maximum de 6500Hz. Le signal est donc suréchantillonné. Par conséquent, nous proposons de décimer les signaux analytiques par un facteur 8. Nous obtenons ainsi des signaux de 51558 points échantillonnés à une fréquence $f_s=103116/8=12889$ HZ. Nous pouvons remarquer que la demi fréquence d'échantillonnage reste supérieure à la fréquence maximale du signal analysé. Le théorème de Shannon ($f_e \geq 2.f_{max}$) est donc bien vérifié.

La figure 4.10 montre les deux fréquences d'engrènement f_{eng1} et f_{eng2} . On constate bien une augmentation d'énergie pour le défaut niveau 2 comparer au deux autres cas " sans défaut et défaut niveau 1". De ce fait, les énergies dans les bandes et autour des fréquences d'engrènements seront d'une grande utilité.

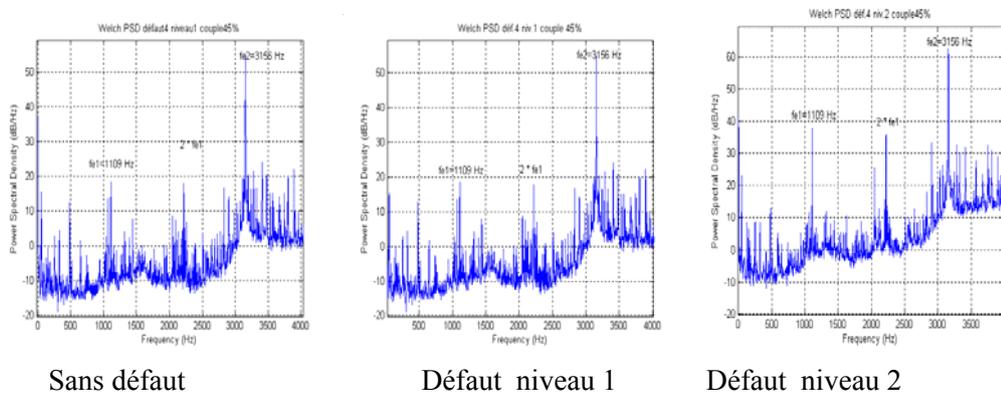


Figure 4.10 DSP après décimation et filtrage (couple 45%)

Définition des modes de fonctionnement et du Vecteur forme initial

On a trois classes de fonctionnement, ils sont représentés sur le tableau 1.

Tableau 4.3. Modes de fonctionnement.

Etiquette	Mode de fonctionnement	Taux du couple résistant (%)
1	Sans défaut	27, 40, 45, 50, 60, 70, 75, 80, 100
2	Défaut #4, pignon 5 – Niveau 1	27, 40, 45, ---, ---, 70, ---, 80, 100
3	Défaut #4, pignon 5 – Niveau 2	27, 40, 45, 50, 60, 70, 75, 80, 100

La signature de la machine est définie par un vecteur de vingt un paramètres. Nous présentons un panorama complet de l'ensemble de ces paramètres sur le tableau 4.4.

Tableau 4.4 Paramètres du vecteur forme "VF"

α_i	Domaine	Dénomination	Symbole	
1	Temporel	Bande 0 -1500 Hz	Valeur efficace 1	Vef 1
2			Facteur de crête 1	Fc 1
3			Kurtosis 1	K1
4		Bande 1500 - 3000 Hz	Valeur efficace 2	Vef 2
5			Facteur de crête 2	Fc 2
6			Kurtosis 2	K2
7		Bande 3000 - 4500 Hz	Valeur efficace 3	Vef 3
8			Facteur de crête 3	Fc 3
9			Kurtosis 3	K3
10		Bande 4500 - 6000 HZ	Valeur efficace 4	Vef 4
11			Facteur de crête 4	Fc 4
12			Kurtosis 4	K 4
13		Bande totale 6 kHz	Valeur efficace 5	Vef 5
14			Facteur de crête 5	Fc 5
15			Kurtosis 5	K5
16	Fréquentiel	Energie dans la bande 683-1533 Hz		B 1
17		Energie dans la bande 1065-1151 Hz		B 2
18		Energie dans la bande 1791-2641 Hz		B 3
19		Energie dans la bande 2173-2259 Hz		B 4
20		Energie dans la bande 683-2641 Hz		B 5
21		Energie dans la bande totale 6 kHz		B tot

4.4.2 Visualisation par analyse en composantes principales du vecteur forme initial

Le calcul de l'ACP fournit, d'une part, les corrélations entre les différents paramètres, et d'autre part, une représentation des individus qui composent la base de données.

La visualisation de la base de données par ACP (figure 4.11) du vecteur forme initial, permet d'appréhender les différentes formes (ou classes) présentes. On en déduit de précieuses indications sur la disposition des classes et plus particulièrement, cela permet d'obtenir une première appréciation de la séparabilité de celles-ci

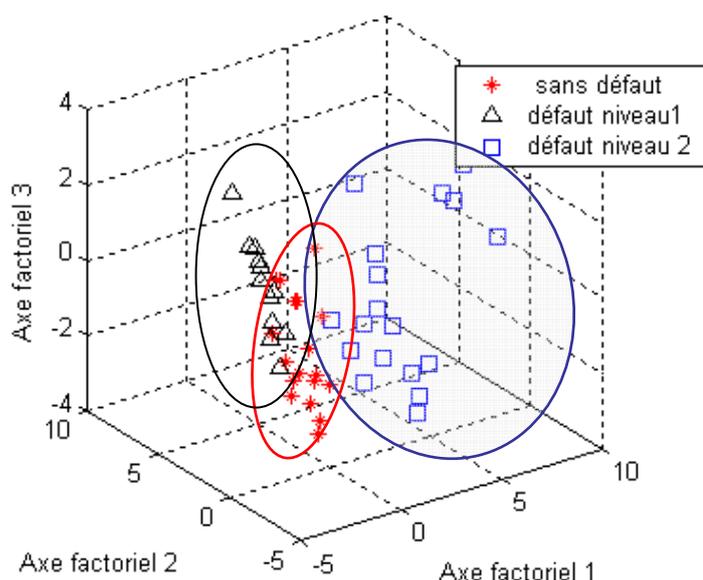


Figure 4.11 Position des classes dans le sous – espace de dimension 3 : VF initial

Nous remarquons que les classes ne sont pas parfaitement compactes (les observations d'une même classe) et ne sont pas disjointes (les classes entre-elles). Nous espérons améliorer la séparabilité par la méthode de sélection que nous considérons ci-dessous.

4.4.3 Sélection d'indicateurs par le critère de Fisher

Le même critère de Fisher défini par l'expression (4.1) est utilisé pour classer les indicateurs. On notera au passage que le terme "classe" doit être considéré au sens large ; il ne désigne pas particulièrement une classe de défaut mais il peut concerner un regroupement de plusieurs classes (par exemple : plusieurs couples résistants ou plusieurs niveaux). La figure 4.12 montre la variation du critère de Fisher pour les 21 paramètres du vecteur forme initial, voir tableau 4.4. L'expression (4.2) permet de calculer le pourcentage expliqué par chaque composante du VF.

Si on considère seulement un vecteur forme de dimension $d' = 3$, c'est-à-dire qu'un seuil de 57 % est largement acceptable, donc un vecteur forme réduit $X = [\alpha_1, \alpha_4, \alpha_{21}]$

Les indicateurs correspondent respectivement à (Tableau 4.4):

- α_1 : La valeur efficace dans la bande 0-1500Hz
- α_4 : La valeur efficace dans la bande 1500-3000Hz
- α_{21} : Energie du spectre d'enveloppe dans la bande totale 0-6000 Hz.

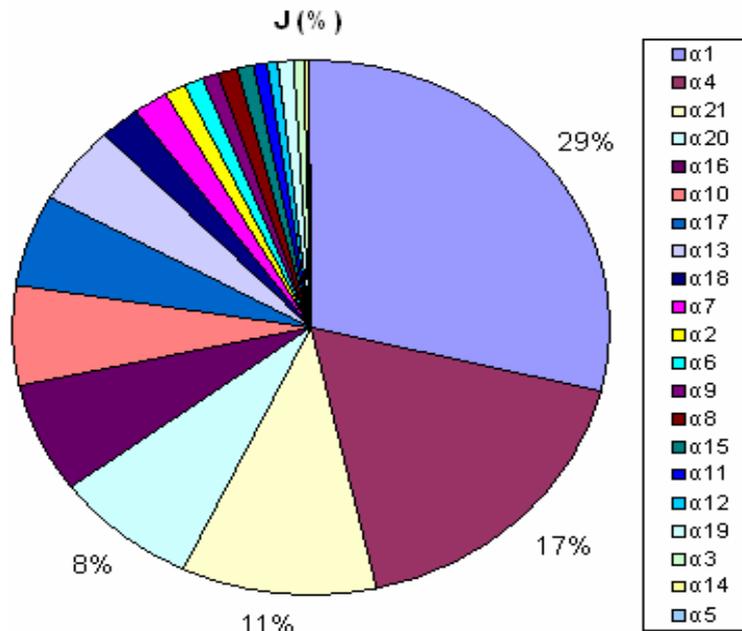


Figure 4.12 Critère de Fisher pour les différents indicateurs

On va réaliser une ACP du VF réduit ($d=3$) pour pouvoir juger la pertinence des indicateurs retenus, les résultats obtenus seront comparés à ceux obtenus avec le VF total (initial).

4.4.4 Visualisation par analyse en composantes principales du vecteur forme réduit

La projection des données dans le plan de représentation à partir du vecteur forme de dimension 3, a permis de visualiser des classes bien compactes et disjointes (Figure 4.13). L'élimination des éléments redondants et inutiles est d'une grande importance dans la séparation des modes de fonctionnement. De plus on peut constater visuellement sur la figure 4.13 que la séparabilité des classes obtenue avec le VF réduit est meilleure que celle obtenue avec le VF initial de la figure.4.11.

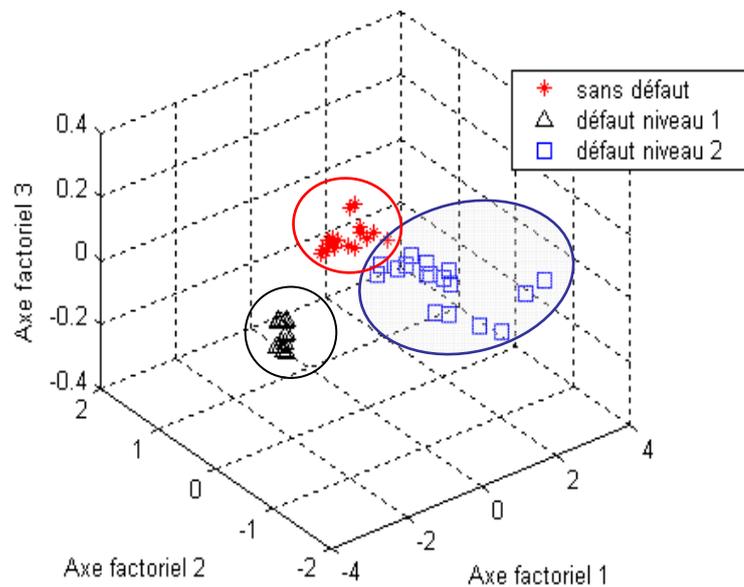


Figure 4.13 Position des classes dans le sous – espace de dimension 3 : VF réduit

Remarque : Pour définir un vecteur forme en se basant sur les analyses de sensibilité des paramètres. L'approche par expertise humaine peut donner de très bons résultats. Mais, la sélection de caractères rend la recherche de l'espace de représentation plus rapide et plus globale. Le calcul de Fisher pour la discrimination entre toutes classes, est une tâche très fastidieuse. En outre, elle devient rapidement complexe lorsque le nombre de classes augmente, et rien n'assure qu'il sera possible de n'en retenir qu'un nombre limité par la suite. Il est donc préférable d'utiliser comme critère un pouvoir discriminant « global », calculé directement sur l'ensemble des classes. Nous proposons dans ce qui suit une hybridation de deux algorithmes AGs-ADL.

4.5 Sélection d'indicateurs par couplage (AG_ADL)

Nous proposons dans cette section une méthode supervisée de sélection d'indicateurs de la famille 'filters', voir section 2.6.2. La procédure automatique de sélection est basée sur une hybridation de deux algorithmes : AGs et ADL. Cette hybridation de deux techniques permet d'exploiter les avantages de chacun et d'améliorer le système de reconnaissance. Les AGs ont une grande capacité d'optimisation et l'ADL est une méthode efficace de discrimination. Les indicateurs utilisés dans ce travail sont les paramètres autorégressifs (AR). Différents modes de réglages des AGs tels que la taille de la population, le nombre de générations, le critère d'arrêt et les différents cas d'opérateurs de sélection, de croisement et de mutation ont été effectués afin de faire des

comparaisons. Nous démontrons l'intérêt d'une telle procédure de sélection à partir des signaux d'engrenage du réducteur de vitesses du CETIM présenté dans la section 3.5.

4.5.1 Principe des algorithmes génétiques

Parmi tous les types d'algorithmes existants, certains ont la particularité de s'inspirer de l'évolution des espèces dans leur cadre naturel. Ce sont les algorithmes génétiques. Les espèces s'adaptent à leur cadre de vie qui peut évoluer, les individus de chaque espèce se reproduisent, créant ainsi de nouveaux individus, certains subissent des modifications de leur ADN, certains disparaissent, etc.

Un algorithme génétique va reproduire ce modèle d'évolution dans le but de trouver des solutions pour un problème donné. Il sera fait usage dans cette section de termes empruntés au monde des biologistes et des généticiens et ceci afin de mieux représenter chacun des concepts abordés :

- Dans notre cas, une population sera un ensemble d'individus.
- Un individu sera une réponse à un problème donné, qu'elle soit ou non une solution valide du problème.
- Un gène sera une partie d'une réponse au problème, donc d'un individu.
- Une génération est une itération de notre algorithme.

Un algorithme génétique va faire évoluer une population dans le but d'en améliorer les individus. Et c'est donc, à chaque génération, un ensemble d'individus qui sera mis en avant et non un individu particulier. Nous obtiendrons donc un ensemble de solutions pour un problème.

Le déroulement d'un algorithme génétique peut être découpé en cinq parties :

1. La création de la population initiale
2. L'évaluation des individus
3. La création de nouveaux individus (croisement et mutation)
4. L'insertion des nouveaux individus dans la population
5. Réitération du processus à partir de 2.

Nous disposons donc en première période d'une population $P(0)$ d'individus ayant des caractéristiques diverses (aléatoires au départ) et dont les chaînes de caractères représentent des solutions potentielles à notre problème. Les caractéristiques des individus sont ainsi codées par des chaînes de caractères de taille fixe et les individus n'ont aucune connaissance d'un modèle éventuel. A chaque itération, nous sélectionnons les meilleurs individus selon une méthode d'évaluation et certains individus mutent ou se reproduisent

(croisement). Nous itérons le processus jusqu'à la condition de terminaison (stabilité des caractéristiques de la population sur deux périodes par exemple).

Evaluer les individus revient à attribuer une 'note' ou un indice de qualité à chacun de nos individus. La méthode d'évaluation des individus que nous proposons est basée sur le critère de la trace tiré de l'ADL que nous présentons ci-dessous

4.5.2 Principe de l'analyse discriminante linéaire (critère de la trace)

Pour calculer le coût d'un point de l'espace de recherche, on utilise une fonction d'évaluation appelée aussi fonction d'adaptation ou fitness. L'évaluation d'un individu ne dépend pas de celle des autres individus, le résultat fourni par la fonction d'évaluation va permettre de sélectionner ou de refuser un individu pour ne garder que les individus ayant le meilleur coût en fonction de la population courante. Cette méthode permet de s'assurer que les individus performants seront conservés, alors que les individus peu adaptés seront progressivement éliminés de la population.

La fonction d'évaluation peut être soit mono critère ou multicritère. Une fonction d'évaluation mono critère signifie que la fonction dépend d'une seule fonction objective par contre la fonction d'évaluation multicritère dépend d'une combinaison de plusieurs fonctions objectives d'où la notion **objective unique** et **objectives multiples**.

Les problèmes d'optimisation doivent souvent satisfaire des objectifs multiples, dont certains sont concurrents. Une méthode classique consiste à définir des fonctions objectives, traduisant chaque objectif à atteindre, et **de les combiner au sein de la fonction d'évaluation** On établit ainsi un compromis. Le plus simple est de se ramener à une somme pondérée des fonctions objectives. A la place d'une somme, on peut également utiliser un produit des fonctions objectives ou utiliser des expressions plus complexes.

Dans ce qui suit, nous présentons un formalisme pour définir une fonction d'évaluation.

4.5.2.1 Formalisme

L'ADL est une technique d'analyse factorielle discriminante, qui permet la projection de la base de données sur une surface informative de dimension réduite. L'objectif est de rechercher, parmi les M indicateurs de départ (vecteur forme initial), un sous-ensemble de M' d'indicateurs donnant le plus d'information. Toutefois, la sélection nécessite la connaissance de la structure en classes des données pour être utilisée de la manière la plus efficace possible. En d'autres termes, l'utilisateur doit connaître le nombre de classes existantes ainsi que les observations correspondantes.

Le sous-ensemble de M' indicateurs sélectionné devra maximiser ou minimiser un critère rendant compte à la fois :

- d'une faible dispersion des points dans chaque classe, ce qui suggère une variance intra-classes minimale (dispersion des individus de la population à l'intérieur d'une même classe),
- d'un éloignement maximal entre les différentes classes, donc d'une variance inter – classes maximale (dispersion des centres de gravités c'est à dire la séparation des classes entre elles.

Les matrices de variance intra-classes et inter-classes traduisent ces deux conditions. En effet l'expression de la matrice de variance intra-classes globale, appelée en anglais *within-class scatter*, fait intervenir les échantillons et les centres de gravités de chaque classe. Elle sera définie par la somme pondérée des matrices de variance-covariance propre à chaque classe :

$$S_W = N_1 S_{w_1} + N_2 S_{w_2} + \dots + N_{N_c} S_{N_c} = \sum_{i=1}^{N_c} N_i S_{w_i} \quad (4.3)$$

S_{w_i} est la matrice de dispersion intra-classe de la classe w_i . Elle est calculée à l'aide de l'équation suivante :

$$S_{w_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij} - m_i)(x_{ij} - m_i)^t \quad (4.4)$$

avec

- N_C : Nombre de classes
- N : Nombre de vecteurs de l'ensemble d'apprentissage
- N_i : Nombre de vecteurs dans la classe w_i
- x_{ij} : $j^{\text{ième}}$ vecteur de la classe w_i
- m_i : Centre de gravité de la classe w_i :

chaque composante de ce vecteur $m_i = [m_1^i \ m_2^i \ \dots \ m_k^i \ \dots \ m_M^i]^t$ est obtenue à partir de celles des échantillons associés à la classe w_i .

Alors que la matrice de dispersion inter-classes, appelée en anglais *between-class scatter*, s'exprime en fonction des centres de gravités des classes et du centre de gravité de l'ensemble d'apprentissage, elle mesure la dispersion des classes entre elles, On définit cette matrice par :

$$S_B = \sum_{i=1}^{N_c} N_i (m_i - m)(m_i - m)^t \quad (4.5)$$

Avec :

- m : Centre de gravité de l'ensemble d'apprentissage :

c'est le vecteur moyen total $m = [m_1 \ m_2 \ \dots \ m_j \ \dots \ m_M]$ obtenu à partir de l'ensemble

des échantillons : $m_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$

m_i : Centre de gravité de la classe w_i :

Par ailleurs, la matrice variance-covariance qui traduit la dispersion totale de l'ensemble vaut :

$$S = S_W + S_B \quad (4.6)$$

L'obtention du sous-ensemble discriminant consiste alors à trouver les M' paramètres (combien et lesquels) qui permettent de maximiser la mesure de dispersion inter-classes S_B tout en minimisant la mesure de dispersion intra-classes S_W . Ce problème peut se traduire par plusieurs **fonctions objectives** représentant la trace de certaines matrices (Oukhellou, 1997) que nous présentons ci-dessous.

4.5.2.2 Définition du critère de la trace

La relation $S = S_B + S_w$ exprime le fait que pour un nuage de poids donné, la matrice S étant fixe, si la dispersion à l'intérieur des classes est faible, alors l'éloignement des classes est grand. Pour estimer le pouvoir discriminant de M' indicateurs sélectionnés parmi les M qui existent, on utilise les matrices S_w , S_B et S , mais, restreintes aux M' paramètres considérés.

Soit $V^{M'}$ un vecteur attribut arbitraire, de dimension M' , extrait de V^M (vecteur initial) et soit $S^{M'}$, $S_W^{M'}$ et $S_B^{M'}$ les matrices de variance totale, intra-classes et inter-classes associées à $V^{M'}$. Ces trois matrices sont de dimension $M' \times M'$. Pour la suite des formules nous omettons d'écrire l'exposant M' pour une cause de simplicité.

Pour obtenir des classes bien compactes et donc séparées entre elles. Le pouvoir discriminant de $V^{M'}$ peut être estimé de plusieurs façons voir pour plus de détail (Oukhellou, 1999), (Casimir, 2003).

Le premier critère est à minimiser. Il est défini comme suit :

$$J_1 = \text{trace}(S_w) \quad (4.7)$$

L'opérateur « trace » n'est malheureusement pas invariant sous l'effet de transformation linéaire non singulière de données de type changement d'échelle sur les paramètres. Un deuxième type de critère aussi à minimiser est alors défini, équivalent à J_1 mais après normalisation de la matrice intra-classes :

$$J_2 = \text{trace}(S^{-1}S_w) \quad (4.8)$$

Une autre approche inspirée des résultats de l'analyse factorielle discriminante, s'intéresse au produit $S_w^{-1}S_B$. En cherchant à maximiser la trace de ce produit, on obtient des classes bien condensées et séparées les unes des autres :

$$J_3 = \text{trace}(S_w^{-1}S_B) \quad (4.9)$$

Un autre critère est présenter dans la littérature, il consiste à maximiser :

$$J_4 = \text{trace}(S^{-1}S_B) \quad (4.10)$$

Il existe enfin un dernier critère à minimiser qui permet d'examiner la dispersion des classes entre elles utilisant le déterminant de S_w et non sa trace :

$$J_5 = \frac{\det(S_w)}{\det(S)} \quad (4.11)$$

Remarque : Il n'existe pas une méthode permettant d'orienter le choix vers un critère plutôt qu'un autre. Si les solutions obtenues par les cinq critères sont différentes, elles devraient toutes être réexaminées a posteriori au vu des résultats de classification.

Si dans les matrice de données, des colonnes sont dépendantes (proportionnelles ou l'une d'elles est une combinaison linéaire d'autres colonnes) ; les matrices S , S_w , S_B seront singulières.

4.5.3 Couplage des algorithmes

Notre stratégie est la suivante, nous utilisons les AGs comme outils d'optimisation, le critère de la trace comme fonction d'évaluation. Nous présenterons ci-dessous respectivement *l'organigramme de sélection par AGs sur la figure 4.13* et *l'organigramme de calcul de la fonction d'évaluation « le critère de la trace » sur la figure 4.14*. La conduite détaillée des différentes étapes de l'algorithme de sélection est décrite dans la section suivante.

4.5.4 Conduite des différentes étapes du système de sélection

Les AGs sont des procédures assez robustes pour résoudre un problème d'optimisation de sélection d'indicateurs. Néanmoins, ils présentent certaines difficultés qui reposent essentiellement sur le choix des bons paramètres tels que : la taille de la population, le nombre de générations, les probabilités de croisement et de mutation et les méthodes des opérateurs de reproduction. Ces paramètres dépendent aussi du problème à résoudre et d'une codification appropriée des solutions. C'est un problème de réglage qui doit être optimisé pour chaque type de fonction d'évaluation. Cela constitue une part importante du travail de l'expérimentateur.

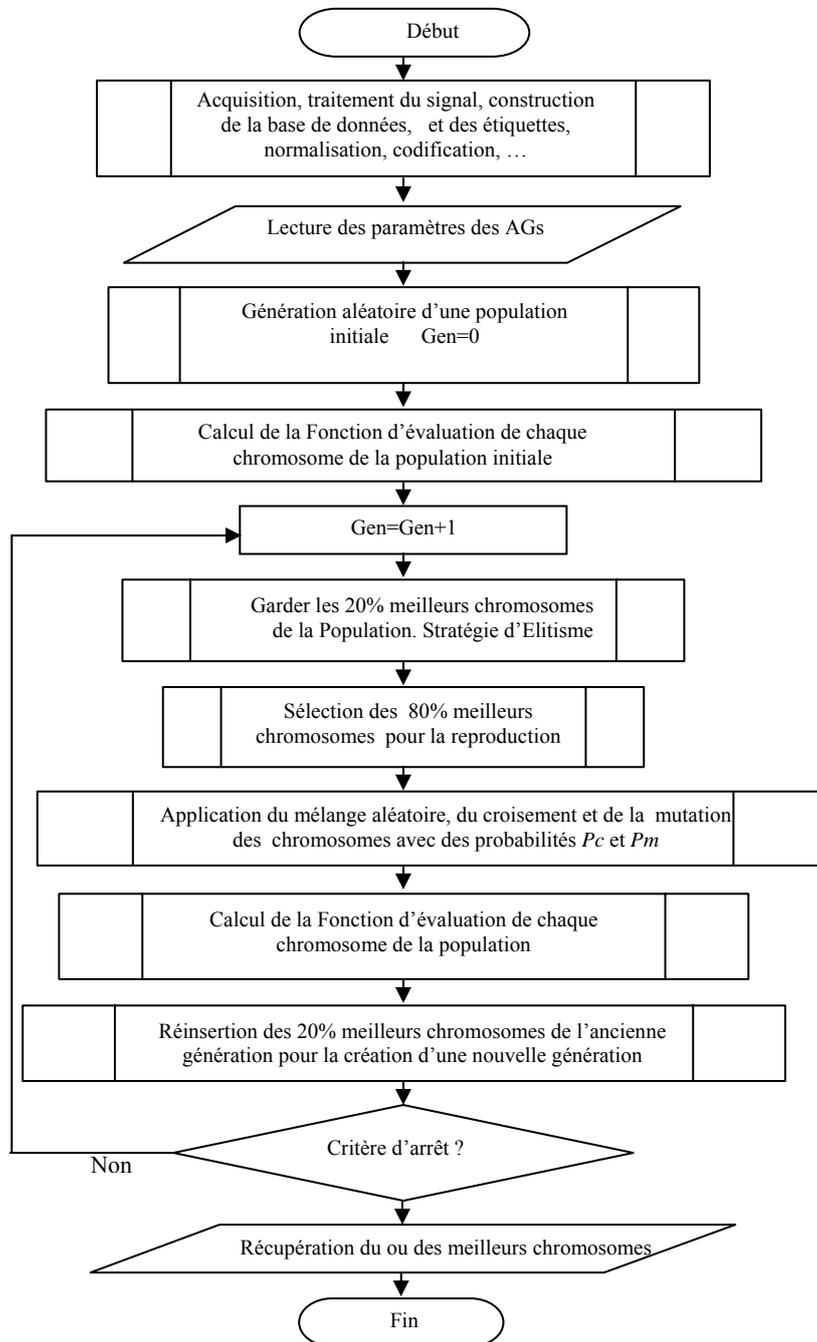


Figure 4.13 Organigramme de sélection du vecteur forme par Algorithmes Génétiques simples

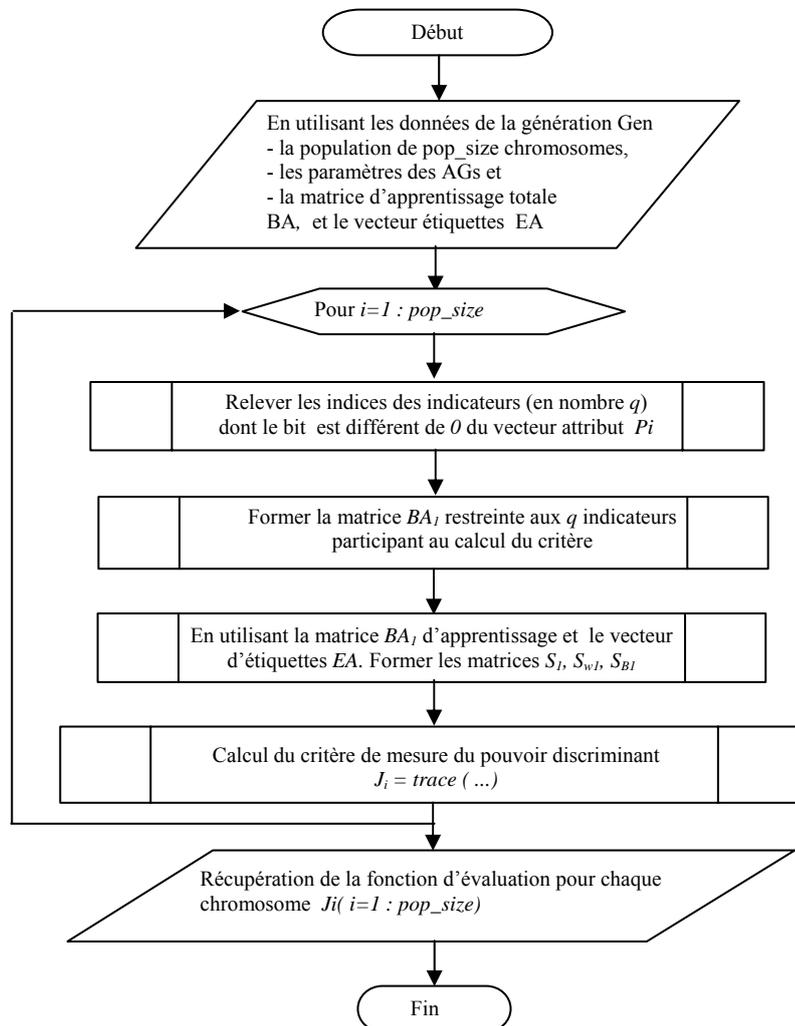


Figure 4.14 Organigramme de Calcul de la fonction d'évaluation « la trace »

4.5.4.1 Préparation des données

➤ Génération de caractéristiques

Les signaux bruts acquis par les capteurs sont bruités, ils contiennent des informations parasites et ils n'explicitent pas directement les informations utiles pour la décision. Ils sont donc difficilement exploitables. Il s'agit principalement dans cette étape **d'extraire et de mettre sous forme explicite l'information** contenue dans le signal. Cette phase est lourde en terme de calcul et nécessite toute la connaissance du dispositif étudié (chaîne cinématique, vitesse de rotation, couple, etc.) pour rechercher les paramètres adéquats par les méthodes de traitement adaptées. Il importe donc d'effectuer une première analyse des

paramètres afin de ne garder que les plus pertinents. Nous avons choisi d'utiliser les paramètres globaux d'un modèle AR de représentation comme indicateurs (Wang, 2003). Bien sûr, ces paramètres ne sont pas tous significatifs. Il s'agit dans notre cas de déterminer combien et quels paramètres il faut retenir. L'application de la technique de sélection passe par certaines préparations telles que la constitution de la matrice d'observations suivie d'une normalisation et d'un étiquetage des classes.

➤ Construction de la base de données, étiquetage et normalisation

- *Construction de la matrice* : Les données à traiter sont stockées dans un tableau BD de type observations/indicateurs de la forme :

$$BD = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad (4.12)$$

où $\begin{cases} N \text{ est le nombre d'observations} \\ M \text{ est le nombre d'indicateurs} \end{cases}$

Les N observations, représentées en lignes, seront divisées en N_c classes (w_1, w_2, \dots, w_{N_c}) ou modes de fonctionnement. Chaque classe w_i possède N_i observations.

Pour chaque observation, les valeurs prises par chaque indicateur sont notées :

$$(x_{ij})_{(1 \leq i \leq N, 1 \leq j \leq M)} \quad \text{avec} \quad \forall (i, j), x_{ij} \in \mathfrak{R}$$

- *L'étiquetage* : Il est réalisé par un expert. Il s'agit d'affecter pour chaque classe un code différent. Chaque classe w_i possède une étiquette $ET_i, i=1, \dots, N_c$. Nous avons un apprentissage supervisé, dans lequel nous imposons une réponse en sortie du classifieur pour une forme d'entrée donnée.

- *Normalisation* : Afin d'éviter qu'un indicateur ne prenne trop d'importance par rapport aux autres simplement parce qu'il varie dans de plus grandes proportions, nous procédons généralement à une normalisation des données en divisant les composantes de chaque indicateur par son écart-type et en les centrant.

$$x_{ij} = \frac{x_{ij} - m_j}{\sigma_j} \quad (4.13)$$

où σ_j représente l'écart-type du $j^{\text{ième}}$ indicateur et m_j sa moyenne.

➤ Codification du chromosome de contrôle des indicateurs

Le codage utilisé par les algorithmes génétiques est formé d'une chaîne de bits contenant toute l'information nécessaire à la description d'un point dans l'espace d'états. Il existe plusieurs types de codages tels que : le codage binaire, le codage réel et le codage Gray (Michalewicz, 1999).

Nous avons utilisé une codification binaire simple où le nombre de bits dans chaque chromosome correspond au nombre d'indicateurs possibles de l'ensemble de départ. Ainsi, un chromosome est codé par une chaîne de M bits. Si le $j^{\text{ème}}$ ($j=1, \dots, M$) bit a une valeur 1, alors l'indicateur correspondant doit être choisi et participe à la classification sinon l'indicateur correspondant doit être exclu. Ce type de codage a pour intérêt de permettre de créer des opérateurs de croisement et de mutation simples qui seront appliqués au cours des différentes générations du processus itératif.

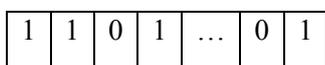


Figure 4.15 Codification représentant un chromosome

4.5.4.2 Génération de la population initiale

Le premier pas dans l'implantation des algorithmes génétiques est de créer une population d'individus initiaux. Chaque individu de la population est codé par un *chromosome* (Michalewicz, 1999). Une population est un ensemble de chromosomes. Si l'espace de recherche contient M indicateurs définis initialement, alors le nombre total de possibilités de candidats générés est de 2^M . Ce nombre est très élevé et le but est de trouver **une solution** parmi ce nombre de probabilités. Plusieurs chercheurs se sont penchés sur le problème qui consiste à déterminer la taille optimale de la population initiale pour atteindre la meilleure solution. La taille de la population doit être choisie de façon à réaliser un bon compromis entre le temps de calcul et la qualité du résultat (Eiben et al., 1999). Dans notre cas, cette taille est prise entre 20 et 50 de manière empirique.

La population initiale est constituée d'un ensemble de chromosomes générés aléatoirement. Un ensemble de solutions connues préalablement peut être injecté dans la population initiale.

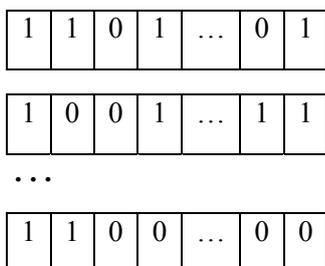


Figure 4.16 Codification représentant une population

4.5.4.3 Calcul de la fonction d'évaluation (Critère de la trace)

Dans notre cas d'une modélisation AR, nous avons remarqué que les matrices ne sont pas toujours de rang plein. Elles sont mal conditionnées (ill-conditionning). Par conséquent, nous avons cherché à minimiser la $trace(S_w)$ pour éviter d'inverser les matrices de variance, voir section 4.5.2.2 $V^{M'}$ sera donc la solution optimale au problème à résoudre si J_1 est minimale.

4.5.4.4 Fixation du nombre de générations

C'est un chiffre que l'expérimentateur doit fixer. Il est préférable qu'il soit assez grand afin de mieux visualiser la convergence de la solution. Certains auteurs (Haupt et al., 2004) utilisent un nombre de 1000 générations et d'autres 10 000. L'essentiel est de trouver des solutions en un nombre réduit de générations. Dans notre cas, nous avons trouvé qu'un nombre de 500 générations suffit largement.

4.5.4.5 Utilisation du critère d'arrêt (critère de convergence)

Le critère de convergence est la condition qui fait arrêter le déroulement des générations d'un AG. Dans notre cas, nous arrêtons la recherche lorsque la valeur absolue de la différence entre les valeurs minimums de la trace de deux générations successives est inférieure ou égale à ε (ε est une valeur faible qui dépend de la nature du problème, nous avons pris une valeur de 0.03). Si le critère d'arrêt n'est pas vérifié, nous prendrons le maximum pour le nombre de générations 500 comme critère d'arrêt.

$$\left| J_{Gen}^{\min} - J_{Gen+1}^{\min} \right| \leq 0.03 \quad (4.14)$$

D'autres critères peuvent être appliqués, nous trouvons entre autres dans (Haupt et al., 2004) : la différence entre le maximum et la moyenne de la fonction d'évaluation.

4.5.4.6 Sélection des individus pour la reproduction

Pour générer de nouveaux descendants (enfants), des parents sont sélectionnés en fonction de leur performance, puis des opérations génétiques leur sont appliquées. Pour sélectionner les meilleurs chromosomes dans une population, certaines techniques proposées par différents auteurs sont les suivantes : la sélection par roulette, par le reste stochastique, par le rang, par tournoi ou par troncature (Michalewicz, 1999).

Dans notre cas, nous avons choisi d'utiliser la sélection par troncature. Un taux de mortalité de 20 % est considéré, c'est-à-dire que 80 % des plus forts individus sont retenus pour la reproduction, alors que 20 % des faibles disparaissent sans avoir la possibilité de se reproduire c'est à dire subir un croisement et une mutation. Le croisement est mis en place pour que les nouveaux chromosomes gardent la meilleure partie des chromosomes anciens. Ceci dans le but d'obtenir de meilleurs chromosomes.

Néanmoins, il est quand même important qu'une partie de la population survive à la nouvelle génération car il est fort probable que les meilleurs chromosomes ne sont pas reproduits dans la prochaine génération. C'est pour cela que le modèle (stratégie) élitiste a été adopté.

4.5.4.7 Croisement (Hybridation)

Il existe là encore de nombreux algorithmes pour reproduire un enfant de 2 parents (2 parents $P1$ et $P2$ produisent toujours 2 enfants $E1$ et $E2$ par symétrie) : le croisement en 1 point, le croisement en 2 points, le croisement d'ordre linéaire. Le croisement en 1 point est le plus simple à mettre en œuvre. Il consiste à choisir une position p dans le chromosome, $E1$ contiendra la portion de $P1$ de 1 à p et la portion de $P2$ de $p+1$ à M (taille du chromosome).

La probabilité de croisement représente la fréquence à laquelle les croisements sont appliqués. Les probabilités de croisement les plus couramment utilisées appartiennent à l'intervalle $[0.6, 0.95]$. Eiben et al dans (Eiben et al., 1999) mentionnent que les valeurs inférieures à 0.6 sont rarement utilisées.

- si la probabilité est de 0%, la nouvelle génération est la copie de la précédente,
- si la probabilité est fixée à 100%, tous les descendants sont générés par croisement.

Nous avons choisi d'utiliser une probabilité de $P_c=95\%$.

4.5.4.8 Mutation

Comme pour le croisement, la mutation vise à modifier de façon aléatoire une partie de la population. C'est le second mécanisme d'innovation d'un AG. Ici, le principe est de choisir une valeur de remplacement aléatoire pour l'un des gènes des individus concernés de la population. A la différence du croisement, le taux de mutation est généralement faible et se situe entre 0.5% et 1% de la population totale (Michalewicz, 1999). Ce faible taux permet d'éviter une dispersion aléatoire de la population et n'entraîne que quelques modifications sur un nombre limité d'individus.

Le choix de la probabilité de mutation P_m et la probabilité du croisement P_c est un problème d'optimisation non linéaire complexe à résoudre. Pour les fixer, ces probabilités de mutation et de croisement dépendent de la nature de la fonction d'évaluation du problème à résoudre. Eiben et al (Eiben et al., 1999) rapportent une formule proposée par plusieurs auteurs pour déterminer la probabilité de mutation. Cette probabilité dépend de la taille du chromosome et est exprimée par la l'équation suivante : $P_m = 1/M$.

Il existe différentes sortes de mutations applicables. Nous trouvons entre autre : le déplacement, l'échange, la recherche locale et le remplacement (Michalewicz, 1999). Pour le remplacement : une position p du chromosome est choisie puis un nouveau

contenu est attribué de manière aléatoire ou non. C'est ce type de mutation que nous avons choisi avec une probabilité $P_m = 1/M$ (où M est la taille du chromosome).

4.5.4.9 Réinsertion des meilleurs chromosomes

Il existe plusieurs stratégies qui peuvent être adoptées dans la génération d'une nouvelle population. Dans un cycle de génération, un individu de la population courante est remplacé par un nouvel enfant. Il est fort probable que le meilleur individu de la génération courante ne sera pas reproduit dans la nouvelle génération. Alors il est intéressant d'introduire la stratégie de l'*élitisme* (Michalewicz, 1999). Elle consiste à copier le meilleur ou un sous-ensemble des meilleurs chromosomes de la génération courante à la prochaine génération. Cette stratégie d'élitisme permet d'augmenter la dominance des meilleurs individus dans une population, ce qui permet une amélioration des solutions à obtenir. C'est un bon moyen de conserver les meilleures solutions même s'il ne faut pas garder trop d'individus (sinon la convergence est trop rapide). Dans notre cas, nous conservons les 20 % meilleurs chromosomes pour les réinsérer dans la prochaine génération.

REMARQUE. — Cette section nous a permis d'avoir un aperçu global d'un système de sélection d'indicateurs basée sur les AGs. Dans le cadre de notre étude, nous nous sommes intéressés à l'ensemble de points de cette étape. Malheureusement, nous avons constaté qu'il n'existe pas de méthodologie formelle pour la conception du système. Tous les choix devront se faire par tâtonnement pour obtenir les performances recherchées et vérifier la robustesse des résultats.

4.5.5 Analyse des Résultats et interprétation

Les signaux du banc d'essai du CETIM (section 3.5) bien connu en littérature ont été utilisés pour cette étude. Nous présentons ci-dessous l'application des différentes étapes et les résultats obtenus de l'approche.

4.5.5.1 Préparation des observations

➤ Division des signaux

Comme nous disposons d'enregistrements d'une longueur de $N_s=60160$ points et pour **avoir suffisamment d'individus** dans les différents modes de fonctionnement, chaque signal est décomposé en K_1 tranches de N_M points avec éventuellement un recouvrement des tranches s_1 .

Le nombre d'échantillons pour chaque tranche doit être suffisamment grand pour couvrir un nombre suffisant du plus grand cycle correspondant à la plus grande période (la plus petite vitesse de rotation) donc à $T_2 = 0.063s$. En prenant un coefficient de 5, la longueur nécessaire de la tranche est de : $T_t = 0.315s$, soit un nombre d'échantillons

nécessaire par tranche de $N_t = 6300$ pts. En prenant pour longueur de la tranche N_M une valeur de puissance 2 supérieure à N_t alors $N_M = 8192$ échantillons et le nombre de tranches dans le cas d'un recouvrement 50 % est de :

$$K_t = \frac{N_s - s_1}{N_M - s_1} \cong 13 \text{ tranches} \quad (4.15)$$

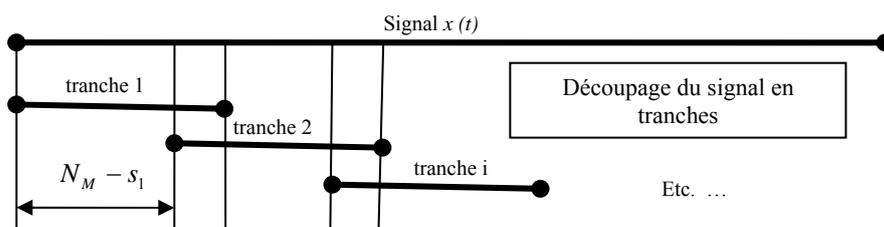


Figure 4.17 Découpage d'un signal en tranches ou sections

Les sept premières tranches de chaque signal décomposé sont utilisées pour l'apprentissage alors que les six tranches restantes sont considérées dans le test. Tous les prétraitements et traitements peuvent être alors effectués sur ces tranches.

➤ Prétraitement pour une analyse temporelle AR

Les prétraitements permettent d'éliminer les informations parasites et non porteuses d'informations pertinentes pour la reconnaissance. Pour ces prétraitements, le concepteur s'aide de connaissances qu'il possède sur les capteurs, les types de données, les caractéristiques propres des structures, les caractéristiques de fixation de la machine, la chaîne cinématique de la machine, etc.

Plusieurs opérations peuvent être envisagées. Dans notre cas, nous avons considéré le filtrage et la décimation.

Filtrage

Pour éviter l'effet de masque qui entache fortement la sensibilité des indicateurs "larges bandes", voir (Pachaud et Canetto, surveillance 4), (Boulenger et al. 2003). Les indicateurs doivent être calculés dans des bandes fréquentielles ni trop larges ni trop étroites et définies en fonction des caractéristiques cinématiques de la machine.

Nous avons choisi, après une analyse préliminaire, de calculer les indicateurs temporels dans différentes bandes latérales de largeur 167 Hz autour des fréquences d'engrènement, soit 10 raies latérales (dix fréquences de rotation) autour des 3 principales harmoniques (1.feng=333.33 Hz, 2.feng=666.66 Hz, et 3.feng=999.99 Hz). Les bandes de calcul représentant les fréquence de coupure sont alors les suivantes : [250-417 Hz], [583-750 Hz] et [916-1083 Hz]. Le signal de chaque tranche a donc été centré puis filtré par un filtre passe-bande. Le filtre utilisé est à réponse impulsionnelle finie à phase minimale

d'ordre 48 utilisant la méthode de la fenêtre de Hamming. Le gain est égale à zero à la fréquence de Nyquist.

Décimation

Pour chaque jour, nous disposons de 13 tranches de signaux de $N_M=8192$ points, échantillonnées à 20000 Hz. Par le filtrage ci-dessus, nous avons limité notre étude à une fréquence maximale de 1083 Hz. Les tranches de signaux sont donc sur-échantillonnées. Nous proposons de décimer ces signaux par un facteur 4. Nous obtenons ainsi des signaux de 2048 points échantillonnés à une fréquence $f_e=20000/4=5000$ Hz. Nous pouvons remarquer que la demi fréquence d'échantillonnage de Nyquist reste supérieure à la fréquence maximale du signal analysé. Le théorème de Shannon est donc bien vérifié. La fréquence d'échantillonnage $f_e > 4f_{\max}$. Le vecteur décimé est 4 fois plus court.

4.5.5.2 Constitution du vecteur forme

➤ Calcul des paramètres AR

Les paramètres AR d'un processus objet peuvent jouer le rôle d'identificateur de celui-ci. En particulier si le processus évolue dans le temps, soit par vieillissement soit par usure progressive. Les paramètres de son modèle comportemental évoluent également et peuvent inversement être un indicateur de "l'état de santé" du processus à surveiller.

Basés sur des travaux antérieurs (Felkaoui, 1995), nous avons développé un algorithme d'estimation des paramètres. La méthode d'identification du modèle AR utilisée est la procédure de la variable instrumentale utilisant la fonction Matlab « ivar ». Elle est suivie d'une autre fonction « polydata » pour récupérer ces paramètres. Les simulations numériques ont montré que l'algorithme proposé donne une bonne poursuite des variations de paramètres même en présence de bruit.

Nous avons calculé les indicateurs temporels pour chaque tranche du signal dans les différentes bandes, 20 paramètres par bande, ce qui donne un nombre initial de paramètres égal à 60 indicateurs par tranche de signal représentant une observation dans la matrice de reconnaissance de formes.

➤ Matrices d'observations

La base de données BD obtenue est formée d'un ensemble de 156 observations (13 tranches par 12 signaux) chacune de 60 indicateurs. Elle est partitionnée en deux sous ensembles distincts : une Base d'Apprentissage (BA) qui entraîne le classifieur et permet de construire la règle de décision, et une Base de Test (BT) des observations restants sur laquelle nous mesurons les performances et nous jugeons la validité et la fiabilité de la règle de décision dans la prochaine section 4.6.

➤ Choix des classes

Etant donné un vecteur étiquette ET contenant les modes de fonctionnement ou classes, chaque classe est identifiée par un numéro différent (0, 1, 2, ...). Afin de pouvoir étudier l'évolution d'un défaut dans le temps, nous avons considéré 2 possibilités : le cas de deux classes et le cas de trois classes.

4.5.5.3 Sélection d'indicateurs par AGs : Cas de deux classes

Le premier test consiste à répartir les signaux en 2 classes :

- Mode de fonctionnement sain,
- Mode de fonctionnement défectueux.

Nous présentons, ci-dessous, les résultats de sélection pour 2 variantes : la première variante considère le 12^{ième} et 13^{ième} jours comme faisant partie de la classe défaillante alors que la deuxième variante considère la classe défaillante dès le 9^{ième} jour.

1^{ière} variante : Pourquoi le 12^{ième} et 13^{ième} jours ? L'analyse des signaux temporels de la figure 3.13, nous permet de souligner l'apparition d'un motif lié au défaut avec l'apparence d'un choc de manière périodique. Par cette expérience, nous voulons voir les performances des paramètres AR et de la méthode de classification.

- La matrice d'apprentissage BA contient alors :
 - de la ligne 1 à 70, des observations de la classe 1
 - de la ligne 71 à 84, des observations de la classe 2
- La matrice de test BT contient alors :
 - de la ligne 1 à 60, des observations de la classe 1
 - de la ligne 61 à 72, des observations de la classe 2

2^{ième} variante : Pourquoi du 9^{ième} au 13^{ième} jour ? L'analyse du tableau 3.5 d'expertise nous permet de constater l'apparition d'un début d'écaillage sur la surface de la dent 15/16 dès le 9^{ième} jour. Par cette expérience, nous voulons voir s'il est possible de détecter d'une façon précoce l'apparition d'un défaut et bien sur tester les performances de la méthode de classification.

- La matrice d'apprentissage BA contient alors :
 - de la ligne 1 à 49, des observations de la classe 1
 - de la ligne 50 à 84, des observations de la classe 2
- La matrice de test BT contient alors :
 - de la ligne 1 à 42, des observations de la classe 1
 - de la ligne 43 à 72, des observations de la classe 2

Dans cette partie, nous avons essayé plusieurs variantes d'algorithmes génétiques afin d'effectuer des comparaisons. Les différents modes de réglages tels que la taille de la population, le nombre de générations, le critère d'arrêt, et les différents cas d'opérateurs

(de sélection, de croisement et de mutation) possibles ont été effectués. Dans cette expérience, nous avons testé également les deux variantes citées ci-dessous.

L'étude préliminaire théorique et expérimentale de la rapidité de convergence de l'algorithme de sélection nous a permis de choisir, pour le cas avec sélection, les paramètres des AG suivants :

- opérateur de sélection d'observations : sélection par troncature avec une stratégie généralisée du modèle élitisme,
- taille de la population : $N_p = pop_size = 20, 25 \text{ et } 30$
- taux de mortalité : 20%,
- nombre de génération maximale : $k_{max} = 500$,
- fonction d'évaluation : le critère de la $trace(S_w)$
- taille du chromosome : $N_b = M = 60$,
- opérateur de croisement : 1 point, probabilité $P_c = 95\%$,
- opérateur de mutation : probabilité $P_m = 1/N_b = 1/60$,
- critère d'arrêt : $\varepsilon \leq 0.03$ ou atteindre k_{max} .

➤ **Sélection d'indicateurs par AGs : Cas 12^{ème} et 13^{ème} jours défailants : 1^{ière} variante**

La figure 4.18 montre l'évolution de la valeur de la fonction d'adaptation du meilleur chromosome de la population courante au fur et à mesure des générations. Nous remarquons presque la même convergence. Pour les différentes populations initiales (essai1, essai2, essai3, ...), l'algorithme de sélection d'indicateurs proposé assure une convergence rapide vers la solution optimale. La position des indicateurs pertinents et non redondants (relativement au problème de séparabilité entre classes) ou le vecteur attribut optimal retenu pour différents essais et différentes tailles de la population est représenté sur le tableau 4.5.

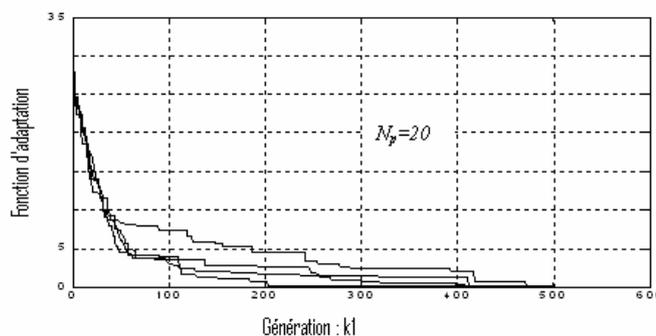


Figure 4.18 Effet de la population initiale aléatoire sur la convergence de la solution au fur et à mesure des générations et pour différents essais, cas 12^{ème} et 13^{ème} jours

Tableau 4.5 Vecteur attribut optimal x_{opt} et sa fonction d'évaluation f_{opt} pour différents essais et différentes tailles de la population N_p , cas 12^{ième} et 13^{ième} jour défailants

Taille de la popul.	1 ^{ier} essai		2 ^{ième} essai		3 ^{ième} essai		4 ^{ième} essai	
	x_{opt}	f_{opt}	x_{opt}	f_{opt}	x_{opt}	f_{opt}	x_{opt}	f_{opt}
$N_p=20$	20 40 60	0.0081	20 26 40 43 52 60	4.1915	20 40 44 60	0.6759	2 20 40 43 60	0.7438
$N_p=25$	40 41 42 60	0.3662	20 40	0.0054	20 40 56 60	1.6440	20 40 60	0.0081
$N_p=30$	41 60	0.0865	20 40 60	0.0081	20 60	0.0054	23 40 60	0.4642

Nous remarquons que pour les différents essais les résultats ne sont pas les mêmes. Ceci est tout à fait normal puisque la population initiale est générée d'une façon aléatoire. Toutefois, certains indicateurs se répètent : ce sont les composantes discriminantes **20, 40 et 60**. Ils seront retenus pour l'évaluation sur la base de test.

➤ **Sélection d'indicateurs par AGs : Cas du 9^{ième} au 13^{ième} jours défailants, 2^{ième} variante**

La figure 4.19 montre l'évolution de la valeur de la fonction d'adaptation du meilleur chromosome de la population courante au fur et à mesure des générations. La position des indicateurs pertinents et non redondants (relativement au problème de séparabilité entre classes) ou le vecteur attribut optimal retenu pour différents essais et différentes tailles de la population est représenté sur le tableau 4.6. Les mêmes remarques peuvent être faites pour ce qui est de la convergence. Les vecteurs formes optimaux retenus pour l'évaluation sur la base de test sont respectivement [1 20 40 60] et [20 40].

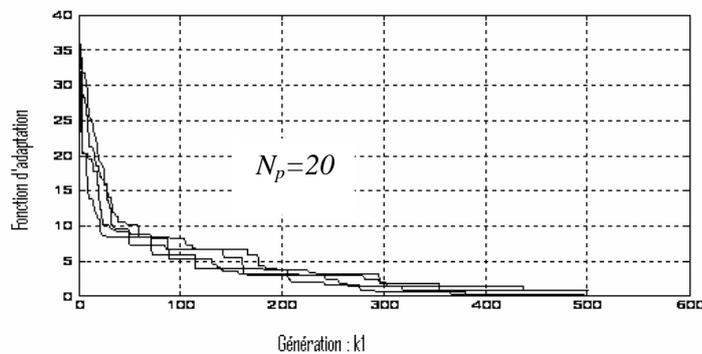


Figure 4.19. Effet de la population initiale aléatoire sur la convergence de la solution au fur et à mesure des génération, cas du 9^{ième} au 13^{ième} jour

Tableau 4.6 Vecteur attribut optimal et sa fonction d'évaluation pour différents essais et différentes tailles de la population : cas du 9^{ième} au 13^{ième} Jour défailants

Taille de la popul.	1 ^{er} essai		2 ^{ème} essai		3 ^{ème} essai		4 ^{ème} essai	
	x_{opt}	f_{opt}	x_{opt}	f_{opt}	x_{opt}	f_{opt}	x_{opt}	f_{opt}
$N_p=20$	20 26 40 60	1.4654	1 37 42 60	0.8633	37 60	0.4272	2 20 21 60	0.4438
$N_p=25$	1 20 40 60	0.1162	20 40 47 60	1.9509	20 40	0.0078	18 20 41 45 60	1.3487
$N_p=30$	20 40 59	0.0643	20 40	0.0078	20 60	0.0078	20 40 60	0.0117

4.5.5.4 Sélection d'indicateurs par AGs: Cas de trois classes

Le suivi de l'évolution ou de la transition entre les modes de fonctionnement est d'une grande importance pour le diagnostic précoce des phénomènes évolutifs. Ainsi, pour voir la **sensibilité** des paramètres AR, les **performances** de la méthode de classification et **valider** le comportement du système de surveillance et de diagnostic dans le cas évolutif, nous considérons 3 niveaux de sévérité.

L'analyse du tableau 3.7 d'expertise nous permet de constater une apparition d'un début d'écaillage sur la surface de la dent 1/2 au 7^{ième} jour et d'un autre sur la surface de la dent 15/16 au 9^{ième} Jour. Donc, nous avons choisi de répartir les signaux en 3 modes de fonctionnement : sain (**jour 2-6**), 1^{er} niveau de dégradation (**jour 7- 8**) et 2^{ième} niveau de dégradation (**jour 9 et 13**).

La matrice d'apprentissage BA contient alors :

- de la ligne 1 à 35 des observations de la classe 1
- de la ligne 36 à 49 des observations de la classe 2
- de la ligne 50 à 84 des observations de la classe 3

La matrice de test BT contient alors :

- de la ligne 1 à 30 des observations de la classe 1
- de la ligne 31 à 42 des observations de la classe 2
- de la ligne 43 à 72 des observations de la classe 3

L'évolution de la valeur de la fonction d'adaptation du meilleur chromosome pour les différentes populations initiales montre la même allure que dans le cas des figures 4.18 et 4.19 au fur et à mesure des générations. L'algorithme proposé assure toujours une convergence autour de l'itération 450. Le tableau 4.7 montre la position des indicateurs pertinents et non redondants pour les différents cas. Les deux vecteurs formes retenus sont respectivement [20 40 60] et [18 20 37 40 59 60]

Tableau 4.7 Le vecteur attribut optimal et sa fonction d'évaluation pour différents essais et différentes tailles de la population : cas de 3 classes

Taille de la popul.	1 ^{ier} essai		2 ^{ième} essai		3 ^{ième} essai		4 ^{ième} essai	
	x_{opt}	f_{opt}	x_{opt}	f_{opt}	x_{opt}	f_{opt}	x_{opt}	f_{opt}
$N_p=20$	5 20 22 40	2.1855	20 36 38 60	1.5188	18 20 22 40	0.8395	2 20 37 40 60	1.1714
$N_p=25$	20 60	0.0111	18 20 37 40 59 60	0.9877	20 40 60	0.0167	19 20 40 60	0.0961
$N_p=30$	20 40 60	0.0167	20 40	0.0111	20 38 40 60	0.2992	20 40 60	0.0167

4.5.6 Conclusion

L'algorithme de sélection d'indicateurs utilisé permet donc de choisir un sous-ensemble optimal d'indicateurs à partir d'un ensemble énorme de caractéristiques obtenues par l'estimation de paramètres autorégressifs (AR). Il s'agit d'une représentation parcimonieuse mais informative de quelques indicateurs sur les 60 initiaux du processus. Dans la prochaine section 4.6, nous allons étudier par le classifieur des k-ppv les performances des différents vecteurs formes (VF) sélectionnées.

4.6 Performance de l'algorithme des k-ppv

Les travaux présentés dans cette section portent sur la classification des défauts du réducteur de vitesses du CETIM (voir section 3.5). Nous avons pu montrer, que les performances des k-ppv sont étroitement liées à la pertinence des indicateurs. La procédure est validée sur des signaux sains et défectueux et en utilisant les paramètres d'un modèle AR

4.6.1 Principe des k-ppv

Le principe général de la méthode des k-ppv a été décrit brièvement dans la section 2.7.1 et consiste à rechercher parmi l'ensemble d'apprentissage EA, contenant l'ensemble des observations et leurs classes d'affectation (étiquettes ET_A), un nombre k d'observations parmi les plus proches possibles de l'observation à classer « x ». Puis, l'observation est affectée à la classe majoritaire parmi ces k individus trouvés, voir la figure 4.20 dans le cas de trois classes et la figure 4.21 pour l'organigramme. Le nombre k est fixé a priori par l'utilisateur. En d'autres termes, la manière la plus simple d'élaborer cette règle est de mesurer la distance entre le nouveau vecteur x et chacun des vecteurs de

l'ensemble d'apprentissage. Le vecteur x sera alors affecté à la classe la mieux représentée parmi ses k observations les plus proches.

L'introduction des rejets d'ambiguïté et de distance est nécessaire dans la règle des k – plus proches voisins afin de diminuer les erreurs de classifications et donc les erreurs de diagnostic.

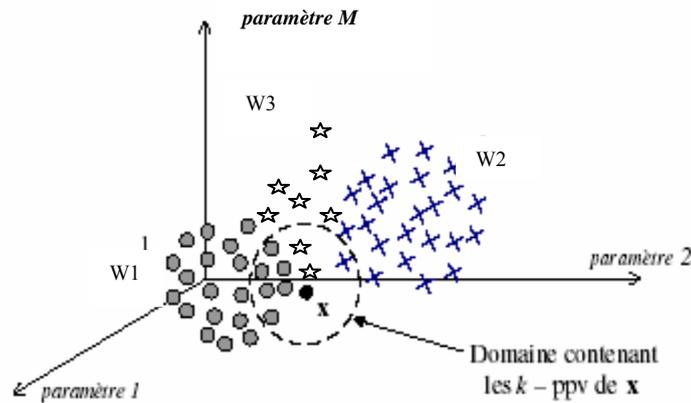


Figure 4.20 Illustration de l'affectation par la règle des k – ppv : cas de 3 classes

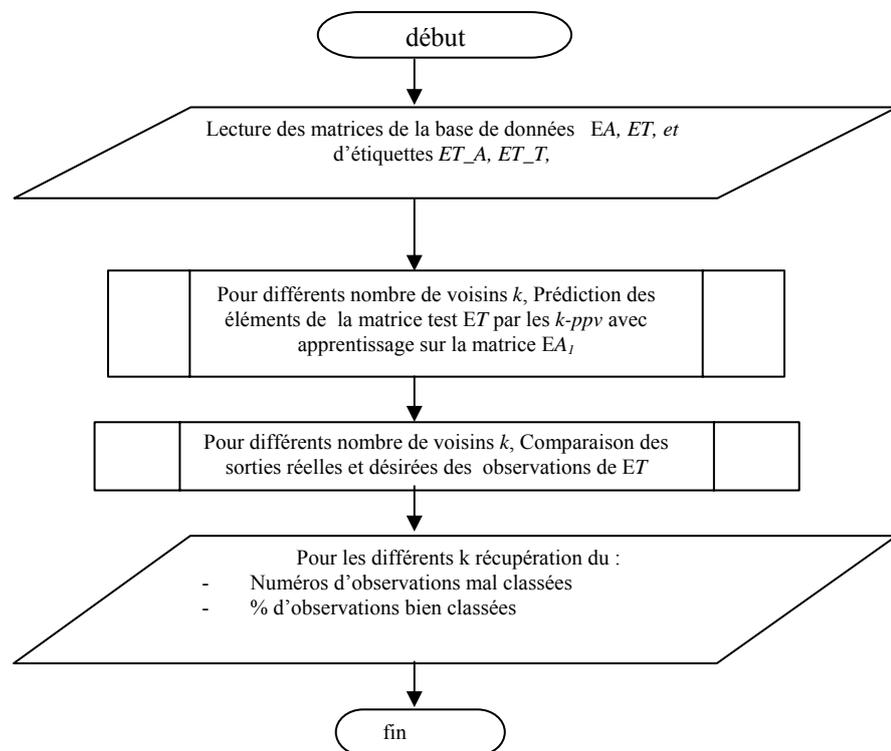


Figure 4.21 Organigramme de classification et de validation par les Kppv sans sélection du vecteur forme

La méthode des k -ppv a l'avantage d'être très simple à mettre en oeuvre et d'utiliser directement l'ensemble d'apprentissage. Elle ne fait aucune hypothèse a priori sur les données. La qualité de la discrimination par cette méthode dépend du choix du nombre k de voisins considérés. Il est cependant souvent nécessaire de faire varier ce nombre k pour obtenir les meilleurs résultats possibles (figures 4.22 et 4.23). Un autre problème important de la méthode des k -ppv est qu'elle nécessite un espace mémoire très important pour stocker les données et pour faire les différents calculs dans la phase de classification. De plus, elle a l'inconvénient d'utiliser les distances (Euclidienne, Manhattan, Max, Mahalanobis, etc. voir [annexe 2.3](#)) pour déterminer les voisins de l'individu à affecter, ce qui peut poser des problèmes si les dimensions à agréger ne sont pas homogènes ([MacKay, 2003](#)), ([Dubuisson, 1990](#))

Une première remarque peut être faite : pour un nombre de voisins $k=1$ ou 2 , l'algorithme de décision arrive toujours à séparer les classes quelle que soit leur nombre, la dimension du vecteur forme et la variante utilisée. Le résultat est parfait. Le pourcentage de réussite est de 100% malgré une diminution importante en nombre d'indicateurs. Nous remarquons également après une visualisation graphique de l'évolution du pourcentage de bien classés en fonction de k (figure 4.22) qu'il est préférable de choisir une valeur faible ($k=1$ à 4).

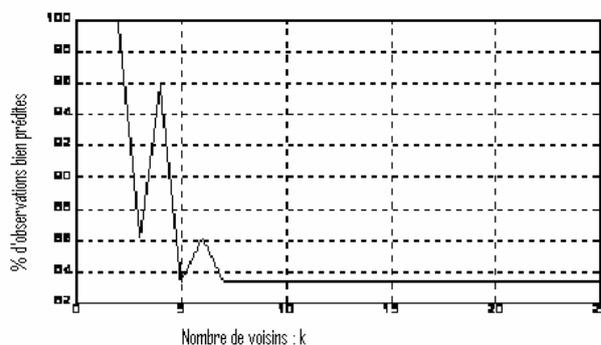


Figure 4.22 Allure de l'évolution du pourcentage d'observations bien prédites en fonction de k , cas de deux classes

4.6.2 Implémentation de l'algorithme

Le présent travail fait varier le nombre k des plus proches voisins (voir algorithme de la [figure \(4.21\)](#)). Puis valide expérimentalement l'algorithme avec et sans sélection d'indicateurs. La validation expérimentale de l'approche est réalisée sur un problème concret de la reconnaissance des défauts dans le cadre des signaux d'engrenage.

Nous présenterons une implémentation pour un nombre k et un élément x sous un environnement Matlab ci-dessous :

```
- lecture Arguments d'entrée
  X [n1] : élément (vecteur ligne) à classer de taille n1
  EA [n1,nc] : matrice des données d'apprentissages chaque
```

```

                                ligne représente un individu
    ET_A      :   vecteur   colonne   des   classes,   ensemble
d'apprentissage (étiquettes)
k : le nombre de voisins à considérer

% calcul des distances à toutes les classes sous Matlab
diff = x'*ones(1,nl)-EA';

dist = sqrt(diag(diff'*diff)); % distance euclidienne
[dd,ii] = sort(dist); % tri des distances, ii : indices
YY = ET_A(ii(1:k))'; % les classes des k premier voisins

% prendre la classe majoritaire de YY
[m,n] =size(YY);
for i=1:n
    z(i,1) = size(YY(find(YY==YY(i))),2);
    z(i,2) = YY(i);
end
[mx,i] = max(z(:,1));
class = z(i,2);
Affichage Argument de sortie
class : classe de l'élément fourni x
    
```

Nous soulignons que nous avons utilisé les matrices et vecteurs de données de la [section 4.5.5](#) pour les différents essais ci-dessous

4.6.3 Performance sans sélection

4.6.3.1 Cas 12^{ème} et 13^{ème} jours défailants : 1^{ière} variante

Le tableau 4.8 présente le pourcentage d'observations bien prédites et les numéros des observations mal prédites en fonction de k. Pour le nombre de voisins k= 7,8, 9, etc., toutes les observations mal prédites appartiennent à la 2^{ème} classe de 61 à 72. L'algorithme n'a pas pu séparer les deux classes. Il les a considéré toutes comme appartenant à la classe 1. Pour k=3, les observations mal prédites sont partagées entre les 2 classes. Nous trouvons respectivement 8, 14, 50, et 52 de la classe 1 et 61, 62, 65, 67, et 68 de la classe 2.

Tableau 4.8 Pourcentage de bien prédites et numéros des observations mal prédites en fonction de k, sans sélection, cas 12^{ième} et 13^{ième} jours

Le nombre de voisins k	Pourcentage de bien prédites	Numéros des observations mal prédites ou mal classées
k= 7,8, 9, ...	83.33	61 62 63 64 65 66 67 68 69 70 71 72
k=4	95.83	61 62 65
k=3	86.11	8 14 50 52 61 62 65 67 68 71
k=1 et 2	100	//

4.6.3.2 Cas du 9^{ème} au 13^{ème} jour défailants : 2^{ème} variante

L'évolution du pourcentage de bien classées ou du taux de succès de la classification en fonction de k-ppv ont toujours montré qu'il est préférable de choisir un faible k. Le pourcentage sur le tableau 4.9 est légèrement inférieur au précédent, 80.55 au lieu de 86.11 pour k=3 et 86.11 au lieu de 95.83 pour k=4. Il est tout à fait normal puisque c'est le début d'une dégradation qui n'est pas très avancée et qui n'est d'ailleurs pas bien marquée sur les signaux du jour 9 à 11. L'erreur de classification est plus probable. L'algorithme de décision n'arrive pas bien à séparer toutes les observations des deux classes. Les individus non prédits sont respectivement : 5, 8, 11, 14, 16, 17, 26, et 33 de la classe 1 et 33, 43, 45, 46, 47, 55, et 67 de la classe 2.

Tableau 4.9 Pourcentage de bien prédites et les numéros des observations mal prédites en fonction de k, cas du 9^{ème} au 13^{ème} jours

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal-classées
k=4	86.11	8 11 14 16 26 43 45 46 47 55
k=3	80.55	5 8 11 14 16 17 26 33 43 45 46 47 55 67
k=1 et 2	100	//

Des résultats meilleurs sont attendus par l'introduction du système de sélection d'indicateurs non redondants et pertinents car tous les indicateurs n'ont pas en général le même pouvoir discriminant. Certains peuvent même **apporter du bruit** au problème décisionnel posé.

4.6.3.3 Performance sans sélection d'indicateurs : cas de trois classes

L'évolution du pourcentage de bien classés ou du taux de succès de la classification en fonction de k (figure 4.23) montre toujours qu'il est préférable de choisir un faible k. Le pourcentage sur le tableau 4.10 est légèrement inférieur aux précédents. Ce qui implique une difficulté à séparer les trois modes de fonctionnement ou de dégradation sauf pour k=1 et 2.

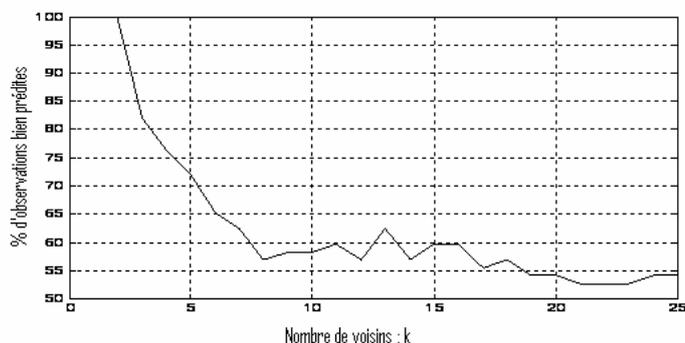


Figure 4.23 Evolution du pourcentage de bien prédites en fonction de k , cas de 3 classes sans sélection d'indicateurs

Tableau 4.10 Pourcentage de bien prédites et les numéros des observations mal prédites en fonction de k

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal prédites
K=4	76.38	5 8 11 14 16 19 24 26 27 30 33 34 35 43 45 46 55
K=3	81.94	3 5 8 11 14 16 17 26 32 33 35 55 67
k=1 et 2	100	//

4.6.4 Performance avec sélection

4.6.4.1 Cas 12^{ème} et 13^{ème} jours défaillants : 1^{ière} variante

— Performance du sous-ensemble d'indicateurs [20 40 60]

Les performances des k -ppv sur la base test, pour ce vecteur optimal, sont représentées sur le tableau 4.11. Les résultats obtenus montrent, par exemple, pour $k=4$ (4-ppv), une diminution du nombre d'indicateurs de 95 % (de 60 à 3) pour une perte « admissible » en performance de 5 % (de 95.83 % à 90.27 %) et une amélioration pour $k=3$ qui passe de 86.11% à 87.5 %. Nous remarquons également que pour $k=4$, toutes les observations sont correctement affectées aux classes correspondantes à l'exception de l'observation 38 de la classe 1. Les autres observations 62, 64, 65, 68, 70, et 71 sont toutes de la classe 2.

La séparation des deux classes est, donc, réalisée avec un taux d'erreur faible. Les résultats expérimentaux obtenus montrent de bonnes performances de la méthode de sélection des paramètres. En effet, le vecteur attribut optimal proposé par l'algorithme est suffisamment représentatif puisque ses composantes permettent une bonne séparation.

Tableau 4.11 *Pourcentage de bien prédites et numéros des observations mal prédites en fonction de pour le sous-ensemble d'indicateurs [20 40 60]*

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal prédites
K=4	90.27	38 62 64 65 68 70 71
K=3	87.5	9 38 62 64 65 67 68 70 71
k=1 et 2	100	//

— Performance du sous-ensemble d'indicateurs [20 26 37 40 43 52 60]

Une légère amélioration dans le taux de reconnaissance a été constatée pour ce vecteur forme. Le tableau 4.12 montre le pourcentage de bien prédites et les numéros des observations mal prédites en fonction de k. Pour k=4, par exemple, nous passons de 90.27 à 94.44 et pour k=3, nous passons de 87.5 à 90.27. Les individus mal prédits correctement sont respectivement pour k=3 : 26 de la classe 1 et 63, 64, 65, 67, 68, et 69 de la classe 2.

Tableau 4.12 *Pourcentage de bien prédites et les numéros des observations mal prédites en fonction de k pour le sous-ensemble d'indicateurs [20 26 37 40 43 52 60]*

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des Observations mal prédites
k=4	94.44	64 65 68 69
k=3	90.27	26 63 64 65 67 68 69
k=1 et 2	100	//

4.6.4.2 Cas du 9^{ième} au 13^{ième} jours défaillants : 2^{ième} variante

— Performance du sous-ensemble d'indicateurs [1 20 40 60]

Les performances du classifieur sur la base test pour ce vecteur optimal sont représentées sur le tableau 4.13. Les résultats obtenus pour les différents k montrent que l'algorithme n'a pas pu faire une séparation entre les 2 classes et que toutes les observations ont été affectées à la première classe. En effet, le vecteur attribut optimal proposé par l'algorithme ne permet pas une bonne séparabilité entre les classes.

Tableau 4.13 *Pourcentage de bien prédites et numéros des observations mal prédites en fonction de k pour le sous-ensemble d'indicateurs [1 20 40 60]*

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal prédites
k= 1, 2, 3, 4, etc.	58.33	de 43 à 72

— Performance du sous-ensemble d'indicateurs [20 40]

Une amélioration dans le taux de reconnaissance, par rapport au vecteur forme précédent, est à constater dans le tableau 4.14. Pour k=3, nous avons un pourcentage de

72.22 et pour $k=4$ un pourcentage de 83.33. Les individus non prédits correctement sont respectivement pour $k=3$: 4, 9, 12, 13, 16, 33, et 38 de la classe 1 et 43, 45, 47, 51, 52, 54, 55, 57, 59, 64, 65, 70, 71 de la classe 2. Nous pouvons conclure que contrairement à ce que l'intuition peut laisser penser, **la performance d'un classifieur n'augmente pas indéfiniment avec la taille du vecteur d'attributs**. Nous remarquons par exemple que pour l'ensemble des indicateurs, le pourcentage était de 86.11 pour 4 indicateurs, il est descendu à 58.33. Alors que pour les 2 indicateurs 20 et 40, le pourcentage a atteint 83.33 : c'est la "**malédiction de la dimensionnalité**".

Tableau 4.14 Pourcentage de bien prédites et les numéros des observations mal prédites en fonction de k pour le sous-ensemble d'indicateurs [20 40]

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal prédites
$k=4$	83.33	4 16 38 45 47 51 52 54 55 64 70 71
$k=3$	72.22	4 9 12 13 16 33 38 43 45 47 51 52 54 55 57 59 64 65 70 71
$k=1$ et 2	100	//

4.6.4.3 Performance des k -ppv : cas de trois classes

- Performance de deux sous ensembles choisis aléatoirement

L'utilisation dans les tableaux 4.15 et 4.16 de deux vecteurs formes respectivement [20 40 60] et [18 20 37 40 59 60] améliore nettement les résultats. Les résultats obtenus montrent par exemple pour $k=4$ (4 -ppv) une diminution du nombre d'indicateurs de 60 à 3 ou 6 pour une perte « admissible » en performance de 1 %. Une amélioration pour $k=3$ est constaté, le pourcentage passe de 81.94 % à 83.33 % pour un nombre d'indicateurs de 6, voir tableau 4.16. Les AGs ont donc permis de diminuer le nombre d'indicateurs nécessaire à une bonne classification sans trop perdre en performance.

Tableau 4.15 Pourcentage de bien prédites et les numéros des observations mal prédites en fonction de k pour le sous-ensemble d'indicateurs [20 40 60]

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal prédites
$k=4$	75	4 16 24 30 31 33 35 37 38 42 45 51 52 54 55 64 70 71
$k=3$	79.1667	4 9 12 13 16 20 33 37 38 41 43 45 55 64 70
$k=1$ et 2	100	//

Tableau 4.16 Pourcentage de bien prédites et numéros des observations mal prédites en fonction de k pour le sous-ensemble d'indicateurs [18 20 37 40 59 60]

Le nombre de voisins k	Pourcentage de bien Prédites	Numéros des observations mal prédites									
		10	11	13	16	19	31	32	33	35	36
$k=4$	75	36	37	38	40	41	48	62	65	70	
$k=3$	83.3333	11	16	18	19	30	31	32	35	36	48 65 71
$k=1$ et 2	100	//									

4.6.5 Conclusion

Le prétraitement des signaux et la recherche du vecteur forme constituent une étape très importante de la phase d'apprentissage du système de diagnostic par reconnaissance des formes.

Dans cette section 4.6, nous avons étudié les performances d'un des algorithmes de classification qui est « les k -ppv ». L'étude a été faite pour les cas **sans** et **avec** sélection d'indicateurs. Nous avons remarqué que la performance du classifieur n'augmente pas indéfiniment avec la taille du Vecteur Forme et qu'une diminution du nombre d'indicateurs, par exemple de 60 à 3, cause une perte en performance de 1 % seulement dans les observations bien classées. Pour un nombre de voisins $k=1$ ou 2, le résultat est toujours parfait c'est-à-dire de 100 %. Pour le cas de 3 classes et $k=3$ ou 4, le pourcentage de bien classées atteint au maximum les 83.33.

Comme le classifieur des k -ppv a l'inconvénient d'utiliser les distances (euclidienne, Manhattan, Mahalanobis, ...) pour déterminer les voisins de l'individu à affecter, ce qui peut poser des problèmes, si les dimensions ne sont pas homogènes. Nous voulons dans les sections 4.7 et 4.8 étudier les méthodes de calcul direct des frontières : le PMC et les SVM.

4.7 Performance de l'algorithme du PMC

On souligne dans (Bishop, 1995) que la reconnaissance de forme est un domaine d'application privilégié pour les techniques **connexionnistes des RNA** du fait de leurs propriétés d'apprentissage à partir d'exemples, de généralisation de leurs connaissances à des entrées inconnues, de traitement en parallèle et de résistance au bruit. Ils peuvent être également implémentés en circuits électroniques, offrant ainsi la possibilité d'un traitement en temps réel.

Il existe plusieurs types de réseaux de neurones qui se différencient par leurs méthodes à aborder l'espace de donnée et leurs façons de répartir les différentes classes. Les

architectures neuronales les plus utilisées en surveillance industrielle sont (Zemouri, 2003) (Jack, 2000) et (Samanta, 2003a) :

- le *Perceptron Multi Couches (PMC)*,
- les *Réseaux à base de Fonctions Radiales (RBFR)*,
- le *modèle de Hopfield*,
- les *réseaux de Kohonen. (LVQ supervisé, SOM non supervisé)*,
- les *réseaux ART* («Adaptive Resonance Theory»),
- les *réseaux de neurones probabilistiques PNN*.

Parmi les propriétés désirables d'un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage. L'apprentissage est un processus dynamique et itératif permettant de modifier les poids synaptiques (annexe 4.1) d'un réseau en réaction avec les stimuli qu'il reçoit de son environnement, afin d'adopter un comportement désiré. Le type d'apprentissage est déterminé par la manière dont les changements de paramètres surviennent. Plusieurs algorithmes d'apprentissage ont été développés depuis la première règle d'apprentissage de Hebb en 1949. Les méthodes les plus fréquemment sont (McCormick et al., 1997), (McCormick, 1998b) et (Jack, 2000) :

- l'algorithme de rétropropagation du gradient,
- la méthode de Newton,
- la méthode de Quasi-Newton,
- la règle de Hebb,
- l'algorithme Compétitif non supervisé, et
- l'algorithme par renforcement.

Chaque méthode d'optimisation possède des avantages et des inconvénients. De nos jours, les réseaux multicouches et la rétropropagation de gradient reste le modèle le plus étudié et le plus productif au niveau des applications. Nous en donnons une brève introduction dans l'annexe 4.1 De bonnes performances sont ainsi obtenues. Mais, un grand nombre d'architectures du réseau peut être utilisé. Les spécifications applicatives de chaque architecture de réseau ne reposent pas sur des résultats théoriques, le **recours à l'expérimentation** s'avère être le moyen le plus approprié pour choisir le type de réseau à adopter pour une application déterminée.

Dans le cadre de cette section, nous limiterons notre étude à l'utilisation du réseau de neurones multicouche (PMC). Ce type de réseau est dans la famille générale des réseaux à «propagation vers l'avant», c'est-à-dire qu'en mode normal d'utilisation, l'information se propage dans un sens unique, des entrées vers les sorties, sans aucune rétroaction. Son apprentissage est de type supervisé par correction des erreurs. Dans ce cas uniquement, le signal d'erreur est «rétropropagé» vers les entrées pour mettre à jour les poids des neurones. Comme il n'existe que des méthodes empiriques pour obtenir un réseau adapté

à la résolution de notre problème particulier. Cette étude fournit, dans la section 4.7.2 une méthodologie permettant à l'utilisateur/développeur de tester facilement les diverses possibilités neuromimétiques pour une application donnée sous environnement MATLAB.

4.7.1 Principe de la classification par RNA

Lorsqu'on cherche à résoudre un problème de classification, on a affaire à des formes ou à des entités, décrites par des nombres, susceptibles d'appartenir à des catégories, ou classes, différentes. La tâche d'un classifieur consiste à attribuer une classe à une forme inconnue qui lui est présentée ou – caractéristique très importante – à refuser de lui attribuer une classe si la forme inconnue est trop éloignée de celles qui sont utilisées pour l'apprentissage.

Pour jouer son rôle de classifieur, dans la phase d'apprentissage, le réseau de neurones doit approcher une fonction qui est le code de la classe de l'objet à la sortie dont les stimuli sont présentés à l'entrée du réseau. Ainsi, pour un problème à deux classes, la sortie désirée est égale à +1 si l'objet à classer appartient à l'une des classes et à 0 s'il appartient à l'autre classe. Pour un problème comportant N_c classes, on peut par exemple fixer les sorties désirées $d = [d_1, d_2, \dots, d_{N_c}]$ à l'aide de l'expression suivante :

$$d_i = \begin{cases} 1 & \text{si le stimulus (l'observation) appartient à la classe } i \\ 0 & \text{autrement} \end{cases}, \quad i = 1, 2, \dots, N_c$$

Par la suite, dans la phase d'exploitation, il suffira de présenter au réseau n'importe quel stimulus inconnu pour pouvoir procéder au classement de celui-ci dans l'une ou l'autre des classes. Dans cette phase les réseaux de neurones ne se contentent pas de prendre la décision d'attribuer un objet à une classe. Ils fournissent une information beaucoup plus riche, à partir de laquelle il est possible de prendre une décision nuancée (une probabilité d'appartenance); cela permet notamment de combiner cette probabilité avec des informations provenant, par exemple, d'autres systèmes de classification, utilisant d'autres algorithmes.

Pour plus de détails sur l'approche des réseaux de neurones et de leurs applications aux problèmes de classification, on peut se référer à [l'annexe 4.1](#) ou au livre de Bishop [\(1995\)](#).

Plusieurs applications mécaniques de l'approche des réseaux perceptron multicouches ont été publiées parmi lesquelles l'application de McCormick dans le diagnostic des défauts d'engrenages [\(McCormick, 1998b\)](#). Des applications des RNA dans le domaine de classifications des défauts mécaniques de roulement, des pompes, les processus industriels, etc. peuvent être trouvées dans [\(Jack, et al., 2000a\)](#), [\(Samanta et al., 2001\)](#), [\(McCormick et al., 1997\)](#) et [\(Sorsa et al., 1991\)](#).

Dans les prochaines sections sera discuté l'implémentation d'un classifieur basé sur les réseaux de neurones. Le choix du nombre des indicateurs « le nombre des neurones d'entrées » est optimisé en utilisant l'approche AG-ADL (section 4.5).

4.7.2 Implémentation des Réseaux de neurones sous matlab

Dans l'annexe 4.1, nous avons couvert un grand nombre de concepts et Il est utile de mettre quelques unes de ces idées ensemble avec un exemple de diagnostic. Il s'agit d'une structure hiérarchique relativement complexe, le but est de montrer comment construire, étape par étape, un réseau de neurones à propagation vers l'avant (feedforward).

Nous présentons l'organigramme d'optimisation d'un réseau de neurones sur la figure 4.24, la conduite détaillée des différentes étapes est décrite ci-dessous.

4.7.2.1 Génération d'indicateurs, construction de la base de données

a. Génération d'indicateurs

Il s'agit principalement, dans cette étape, **d'extraire et de mettre sous forme explicite l'information** contenue dans le signal. Cette phase est lourde en terme de calcul et nécessite toute la connaissance du dispositif étudié (chaîne cinématique, vitesse de rotation, couple, etc) pour rechercher les indicateurs adéquats par les méthodes de traitement du signal adaptées. Nous avons choisi d'utiliser des indicateurs temporels et fréquentiels comme paramètres représentant le signal.

b. Construction de la base de données, Etiquetage des classes

La performance d'un réseau lorsqu'elle est évaluée avec des données d'entraînement est presque toujours surestimée. Il faut bien comprendre que le réseau ne comporte aucune intelligence réelle, il ne fait qu'apprendre les associations qu'on lui fournit. A la limite, il peut les apprendre «par coeur».

Si l'on augmente la capacité de généralisation d'un réseau et l'on évalue correctement la performance d'un réseau, il faut le faire avec des données qui n'ont aucunement servi au processus d'apprentissage. En pratique, ceci implique de diviser la base de données BD que nous possédons en deux sous-ensembles distincts : les données d'entraînement EA et de test ET (voir section 4.5.4.1 et 4.5.5.2).

c. Normalisation

L'apprentissage du RNA peut devenir plus efficace si certaines étapes de pré-traitement sont appliquées sur les bases de données et les vecteurs des classes. Dans ce paragraphe, nous décrirons plusieurs routines de prétraitements que nous pouvons utiliser.

Il est souvent utile, avant l'apprentissage, de mettre à l'échelle les entrées (indicateurs) et les sorties (étiquettes). Une approche pour mettre à échelle les entrées et sorties du réseau est de normaliser les matrices et les vecteurs comme dans le cas de la section 4.5.4.1. Cette procédure est effectuée par la fonction « prestd ». Elle normalise les entrées

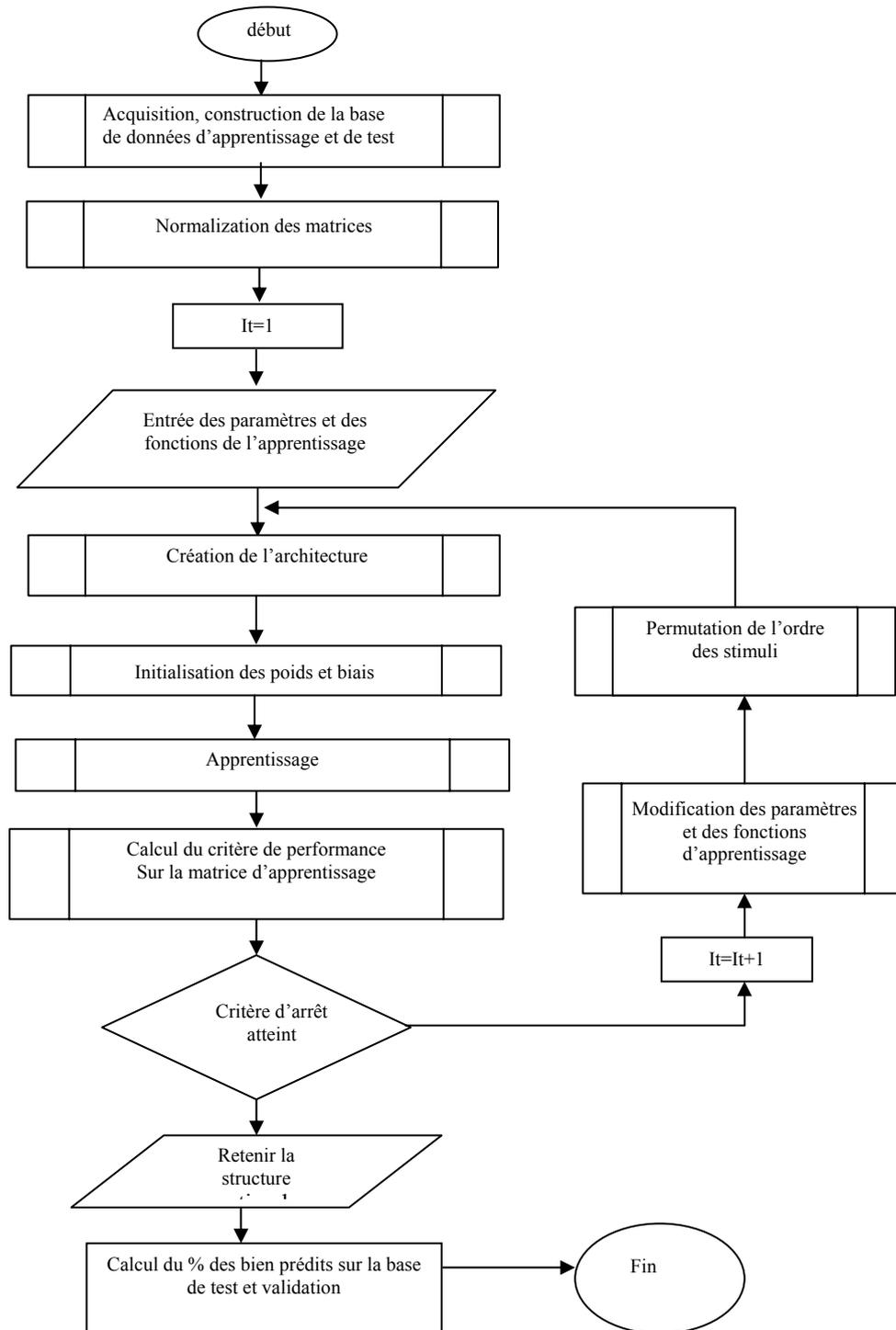


Figure 4.24. Algorithme pour l'évaluation de l'influence des différents paramètres et fonctions du réseau de neurones.

et sorties telles qu'elles auront une moyenne nulle et un écart type égal à 1. Le code suivant illustre l'utilisation de « `prestd` » :

$$[EAn, moyp, sigp, ET_An, moyt, sigt] = \mathbf{prestd}(EA, ET_A) \quad (4.16)$$

EA et ET_A : sont respectivement les données originales
des bases de l'entrée et de la sortie

EAn et ET_An : sont respectivement les données normalisées
des bases de l'entrée et de la sortie

Les Vecteurs Moyp et sigp : contiennent les vecteurs moyenne
et l'écart type des valeurs originales de l'entrée.

Les Vecteurs Moyt et sigt : contiennent les vecteurs moyenne
et l'écart type des valeurs originales de la sortie.

On peut également utiliser un autre prétraitement qui est la réduction du vecteur forme (nombre d'entrée du réseau de neurones) en utilisant l'analyse en composantes principales :

$$\begin{aligned} [EAtrans, transMat] &= \mathbf{prepca}(EAn, 0.001) \\ [R, Q] &= \mathbf{size}(EAtrans) \end{aligned} \quad (4.17)$$

4.7.2.2 Création d'un réseau

L'architecture d'un réseau consiste en une description du nombre de couches (N_1), du nombre de neurones (S_i) et de la fonction de transfert (TF_i) pour chaque couche. Une meilleure architecture à utiliser dépend du type de problème présenté.

La fonction de création d'un réseau est spécifique au modèle de réseau utilisé (`newc`, `newlvq`, etc). Pour les **réseaux multicouches**, la création du réseau est commandée par la fonction **newff** :

$$\mathbf{net} = \mathbf{newff} (PR, [S1 S2 \dots SN1], \{TF1 TF2 \dots TFN1\}, BTF, BLF, PF) \quad (4.18)$$

Avec :

PR : Plage des variations des entrées (affichage par `minmax(p)`) .

$$PR = \mathbf{minmax}(EAn)$$

Si : nombre des neurones dans la couche i , pour N_1 couches.

Les différentes fonctions qui doivent être, également, fixées dans le cas d'un réseau PMC pour la création d'un réseau sont la fonction d'activation, la règle d'apprentissage et la Fonction coût.

Les fonctions d'activation

TFi : fonction d'activation (ou Fonction de transfert) dans la couche i, elle peut être voir Annexe 4.1 :

- **Hardlim** : Fonctions Heaviside.
- **hardlims**: Fonctions signe.
- **logsig** : Fonction logarithme sigmoïde.
- **tansig** : Fonction tangente sigmoïde.
- **pureline** : fonction linéaire.
- **satlins** : Fonction linéaire a seuil.

Les plus utilisées sont logsig, tansig, purelin pour la rétropropagation du gradient ([toolbox nnet matlab 7.01 V4](#)). Nous avons considéré, dans notre cas, deux types de fonction la fonction sigmoïde et la fonction linéaire.

La règle d'apprentissage

2 types d'algorithmes d'apprentissage peuvent être utilisés Par « paquets= BTF » ou « incrémental=BLF » ; l'algorithme d'apprentissage à rétropropagation du gradient **par paquets BTF** (Batching mode) peut être un :

Algorithme lent

- Traingd : descente du gradient
- Traingdm : descente du gradient avec Momentum

Algorithme rapide (1 à 100 fois plus rapide que les algorithmes lent)

- Algorithmes du gradient à taux d'apprentissage
 - Taingda : Algorithme à taux d'apprentissage fixe
 - Traingdx : Algorithme à taux d'apprentissage variable
 - Resilient rétroPROPagation
 - Trainrp : apprentissage par l'algorithme RPROP.
 - Algorithmes du gradient conjugué
 - Traincgf : Gr. Conj. avec la mise à jour Fletcher-Reeves
 - traincgp : Gr.Conj. avec la mise à jour Polak-Ribière
 - traincgb : Gr. Conj. avec Powell-Beale Restarts
 - trainscg : Scaled Conjugate Gradient
 - Algorithmes de Quasi-Newton
 - trainbfg : Algorithme de BFGS (mise à jour de Broyden, Fletcher, Goldfarb, et Shanno)
 - trainoss : Algorithme One Step Secant
 - Algorithmes de Levenberg-Marquardt
 - trainlm : apprentissage par l'algorithme de Levenberg-Marquardt simple

- `trainbr` : version de `trainlm` avec modération automatique des poids.

L'apprentissage par paquets fonctionne par groupage c'est-à-dire que les mises à jour des poids sont accumulées sur l'ensemble des associations de la base d'apprentissage et appliquées une fois à la fin de chaque période d'itération.

Dans le cas où on choisit un apprentissage incrémental nous avons les fonctions BLF les poids sont modifiés à chaque présentation d'une entrée ou d'une observation.

La fonction BLF peut être, par exemple :

- **learngd - Gradient descent weight/bias learning function.**
(L'algorithme d'apprentissage sera la descente de gradient à taux d'apprentissage fixe.)
- **learngdm - Gradient descent w/momentum weight/bias learning function .** (version de `learngd` avec moment)
- **learnp - Perceptron weight/bias learning function.**

Nous avons considéré la fonction BTF avec comme algorithme (règle) d'apprentissage `trainrp` et `trainlm`.

La fonction coût

PF : fonction du coût. Ils en existent plusieurs voir annexe 4.1. Les deux les plus utilisées sont le « mse » et le « mae ». La déclaration se fait par exemple par le choix du paramètre : `net.pertformFcn='mae'`

`mae` pour mean absolute error,

`mse` pour mean square error

Exemple de création d'un réseau

Pour une seule couche cachée avec 10 neurones et une couche de sortie avec 2 neurones, nous aurons :

$$\text{net} = \text{newff}(\text{minmax}(\text{EAn}), [10, 2], \{\text{'tansig'}, \text{'tansig'}\}, \text{'trainrp'}, [], \text{'mae'}) \quad (4.19)$$

4.7.2.3 Initialisation des poids et biais

Avant l'apprentissage du réseau à rétropropagation du gradient, les poids et biais doivent être initialisés. La fonction « **newff** » initialisera automatiquement et par défauts les poids avec l'algorithme de Nguyen-Widrow. Mais, si on veut notre propre initialisation, ceci peut être fait par la commande « **init** ». Cette fonction prend un réseau objet comme entrée et retourne un réseau objet avec tous les poids et biais mis à jour ou réinitialisés.

Les deux options les plus pratiques de l'initialisation des poids sont :

- **l'algorithme de Nguyen-Widrow « `initnw` »** : cet algorithme est utilisé par défaut, il choisit des valeurs de sorte à distribuer la

région active de chaque neurone de la couche à travers l'espace de la couche de l'entrée, et

- **l'algorithme weight-and-bias initialization « initwb »** qui nous permet de choisir l'initialisation pour chaque série de poids et de biais séparément.

a. algorithme initwb

Quand on utilise `initwb`, nous avons à spécifier les fonctions pour chaque série de poids et biais séparément. L'option la plus communément utilisée pour les fonctions est « `rands` », qui met tous les poids et biais à un nombre aléatoire entre -1 et +1, d'autres fonction sont sur le tableau 4.17.

Dans le cas par exemple d'un réseau avec une couche cachée, plusieurs arguments doivent être définis :

```
net.initFcn = 'initlay'; % mettre net.initParam à des matrices vides;
net.layers{1}.initFcn = 'initwb'; % choisir l'algorithme d'initialisation de la couche cachée ;
net.layers{2}.initFcn = 'initwb'; % choisir l'algorithme d'initialisation de la couche de sortie ;
net.inputWeights{1,1}.initFcn = 'rands' % choisir la fonction d'initialisation des poids entre la couche d'entrée et la couche cachée
net.layerWeights{2,1}.initFcn = 'rands' % choisir la fonction d'initialisation des poids entre la couche cachée et la couche de sortie
net.biases{1}.initFcn='rands' % choisir la fonction d'initialisation des biais de la couche cachée
net.biases{2}.initFcn='rands' % choisir la fonction d'initialisation des biais de la couche de sortie
net=init(net) % permet d'initialiser les poids et biais.
```

Pour l'affichage des résultats nous avons à écrire :

```
W1=net.IW{1,1},
W2=net.LW{2,1},
b1=net.b{1},
b2=net.b{2},
et pour la sauvegarde l'instruction
save poids W1 W2 b1 b2
```

b. algorithme initnw

L'initialisation des poids et biais avec « `initnw` » est faite de la manière suivante :

```
net.initFcn='initlay'
net.layers{1}.initFcn='initnw'
```

```

net.layers{2}.initFcn='initnw'
net=init(net)

Affichage des résultats des poids et biais
W1=net.IW{1,1}
W2=net.LW{2,1}
b1=net.b{1},
b2=net.b{2},
save poids W1 W2 b1 b2

```

Tableau 4.17 Les différents cas de la fonction `net.layerWeights{i,j}` du toolbox Matlab

Fonctions d'initialisation des poids et biais	
Initzero	Zero-weight/bias initialization function
Midpoint	Midpoint-weight initialization function
Randnc	Normalized column-weight initialization function
Randnr	Normalized row-weight initialization function
Rands	Symmetric random-weight/bias initialization function

Dans notre cas, les poids sont initialisés aléatoirement à de valeurs faibles pour se situer en début d'apprentissage dans les parties non saturées des sigmoïdes. Nous avons utilisée la fonction par défaut « `initnw` ».

4.7.2.4 Apprentissage

Il existe de nombreuses variantes de l'algorithme d'apprentissage de rétropropagation, plusieurs ont été cités dans la section 4.7.2.2. Nous avons souligné, dans la même section que l'algorithme peut être implémenté de deux manière : le mode incremental et le mode par paquets. Pour une discussion détaillée de ces deux approches d'apprentissage, voir (Hagan et al., 1996).

En matlab, l'apprentissage par paquets peut être fait en utilisant soit « **adapt** » soit « **train** », mais « **train** » - formule 4.20- est généralement la meilleure option, puisqu'elle a l'avantage d'accéder à un nombre plus grand d'algorithmes efficaces dérivés de l'apprentissage par rétropropagation. L'apprentissage Incremental peut être seulement fait par le code « `adapt` ».

Le choix d'algorithme d'apprentissage nécessite la détermination des paramètres d'ajustement des poids à chaque itération et les arguments entrée/sortie. La détermination du critère d'arrêt et aussi cruciale dans la mesure ou la convergence peut passer par des minima locaux.

Les paramètres

Différentes paramètres d'apprentissage existe. Ils sont énumérées en [annexe 4.2](#) selon l'algorithme d'apprentissage. Nous avons cité le cas des algorithmes les plus utilisés tels que les algorithmes : de « **Levenberg-Marquardt backpropagation (LM)** », de « **Resilient backpropagation (RPROP)** » et de « **Scaled Conjugate Gradient back propagation (SCG)** »

Les paramètres sont, par exemple : le momentum, le taux d'apprentissage, le nombre d'itération maximal, etc.

Les arguments d'entrée et de sortie

Différents arguments doivent être définis ou donnés, nous trouvons :

Arguments d'entrée

- **net** – réseau créer par newff (réseau optimal).
- **EA** - matrice d'apprentissage, entrées du RNA.
- **ET_A** - matrice ou vecteur étiquettes cibles du RNA, default = zeros.
- **Pi** - Initial input delay conditions pour un réseau dynamique, default = zeros.
- **Ai** - Initial layer delay conditions pour un réseau dynamique, default = zeros.
- **VV** - matrice de validation, default = [].
- **ET** - matrice de test, default = [].

Arguments de sortie

- **net** – nouveaux réseau.
- **tr** - résultat de l'apprentissage (Itérations et fonction performance)
- **O** – matrice ou vecteur de sortie du RNA.
- **E** - les erreurs à la sortie du RNA.
- **Pf** - Final input delay conditions pour un réseau dynamique.
- **Af** - Final layer delay conditions pour un réseau dynamique.

La syntaxe de la fonction train est donnée par :

$$[\text{net}, \text{tr}, \text{O}, \text{E}, \text{Pf}, \text{Af}] = \mathbf{train}(\text{net}, \text{EA}, \text{ET_A}, \text{Pi}, \text{Ai}, \text{VV}, \text{TV}) \quad (4.20)$$

Nous avons, dans notre cas, utilisé un réseau statique à apprentissage par paquets les données de validation seront utilisées pour les tests, la fonction devient :

$$[\text{net}, \text{tr}, \text{O}, \text{E}] = \mathbf{train}(\text{net}, \text{EA}, \text{ET_A}, [], [], \text{VV}, \text{TV}) \quad (4.21)$$

Si on écrit VV et TV, elles doivent être vide ou une structure de la forme suivante :

VV.P, TV.P - les données de Validation et de test.

VV.T, TV.T - les étiquettes ou cibles de Validation et de test, par default = zeros.

VV.Pi, TV.Pi - Validation/test initial input delay conditions, par default = zeros.

VV.Ai, TV.Ai - Validation/test layer delay conditions, default = zeros.

La matrice de validation VV est utilisée pour arrêter l'entraînement d'une manière précoce afin d'avoir une bonne généralisation (voir annexe 4.1). La matrice test TV est utilisée pour mesurer les performances (% des bien prédits).

Si VV.T, VV.Pi, ou VV.Ai sont mises à de matrices vides alors des valeurs par défaut seront utilisées. Il est de même pour TV.T, TV.Pi, TV.Ai.

Remarques : Les fonctions d'apprentissage ne supportent pas toutes les matrices de validation et de test.

4.7.2.5 Calcul du pourcentage des bien prédits : recherche des performances en classification (Phase test)

La fonction "sim" active un réseau. Elle prend les entrées test normalisées d'un réseau ETn, et le réseau optimal « net », et retourne la sortie calculée de la base de test. Ceci est le mode par groupages de la simulation, dans laquelle toutes les observations sont placées dans une matrice.

Différents arguments doivent être définis (formule (4.22)), nous trouvons les :

Arguments d'entrée :

net - est le pointeur retourné par une fonction d'apprentissage « train ... »

ETn - entrée test du réseau

Pi - Initial input delay conditions pour des réseaux dynamiques, default = zeros.

Ai - Initial layer delay conditions pour des réseaux dynamiques, default = zeros.

ET_T - sorties cibles du réseau, default = zeros.

et retourne les arguments de sortie :

Y - sorties calculées du réseau.

Pf - Final input delay conditions pour des réseaux dynamiques.

Af - Final layer delay conditions pour des réseaux dynamiques.

E - erreurs du réseau.

perf - performance du réseau

La syntaxe de la fonction **sim** est donnée par :

$$[Y,Pf,Af,E,perf] = \mathbf{sim}(\text{net},ETn, Pi, Ai,ET_T) \quad (4.22)$$

Nous avons, dans notre cas, utilisé un réseau statique à apprentissage par paquets les données de validation seront utilisées pour les tests, la fonction devient :

$$[Y, [], [], E, perf] = \mathbf{sim}(\text{net}, \text{ETn}, [], [], \text{ET_T}) \quad (4.23)$$

Le réseau est entraîné pour des sorties binaires 1 ou 0. Toutefois, la fonction `sim` pour des vecteurs d'entrée bruités peut donner des valeurs différentes de 1 et 0. par exemple 0.6 au lieu de 1. L'utilisation de la fonction de post traitement « **compet** » permet de convertir les sorties calculées en sorties binaires.

$$\begin{aligned} Y &= \mathbf{compet}(Y) \\ Y &= \mathbf{find}(\mathbf{compet}(Y) == 1) \end{aligned} \quad (4.24)$$

Utilisation de Simulink : `gensim(net,st)` simule le réseau de neurones avec un block en simulink ,`st=-1` dans le cas statique.

4.7.2.6 Graphe

- `plotpv` - trace les vecteurs entrée/cible.
- `plotep` - trace la position des poids et biais sur une surface erreur.
- `plotperf` - trace la performance du réseau.

4.7.3 Performance de la classification

Les travaux présentés dans cette section portent sur la classification des défauts de roulements du moteur asynchrone (voir section 3.6). Nous avons pu montrer, que les performances du PMC sont étroitement liées à la pertinence des indicateurs. La procédure est validée sur des signaux roulements sains et défectueux et en utilisant des indicateurs temporels et fréquentiels.

4.7.3.1 Préparation des observations

Pour chaque mode de fonctionnement, nous disposons de 5 signaux enregistrés pour quatre couples résistants, donc quatre vitesses différentes, voir tableau 4.1 de la section 4.3. Ainsi, Nous avons 20 signaux échantillonnés à 12 kHz.

Pour avoir suffisamment d'individus dans les différents modes de fonctionnement (sans et avec défaut) et puisque nous disposons d'enregistrements d'une longueur de 243938 points, chaque signal est décomposé en 2 tranches, la première pour l'entraînement et la deuxième pour le test du réseau PMC.

Comme dans la section 4.3, nous avons choisi de calculer des indicateurs temporels et fréquentiels dans les quatre bandes adjacentes $[1-1500 \text{ Hz}]$, $[1500-3000 \text{ Hz}]$, $[3000-4500]$ et $[4500-6000 \text{ Hz}]$ et pour chaque tranche de signal.

Les indicateurs temporels sont les suivants : la valeur efficace, le facteur de crête, le skewness, le kurtosis, et les moments d'ordre 5 et 6. Les indicateurs fréquentiels sont les énergies du spectre dans chacune des quatre bandes.

Les deux ensembles d'indicateurs calculés forment donc une matrice d'apprentissage EA de (20individu \ 28 indicateurs), et une matrice de test ET de même taille formé en

utilisant la deuxième tranche de chaque signal. L'Ensemble d'Apprentissage (EA) entraîne le classifieur et permet de construire la règle de décision, et l'Ensemble de Test (ET) des observations mesure les performances et juge la validité et la fiabilité de la règle de décision. Par conséquent, nous avons deux vecteurs étiquettes, ET_A et ET_T, qui contiennent les modes de fonctionnement, chaque état (classe) est identifié par un suite de code binaire 0 et 1. Chaque classe contient plusieurs observations.

4.7.3.2 Choix des paramètres optimaux du réseau

L'étude théorique présentée en annexe 4.1 et quelques essais préliminaires, sur la vitesse de convergence des différents algorithmes d'apprentissage, nous ont permis de choisir pour notre cas les paramètres et fonctions suivants :

- a. **Classifieur** : perceptron multicouches (PMC) statique,
- b. **Nombre de couches** : 3 couches (1 couche d'entrée 1 couche cachée et une couche de sortie),
- c. **Nombre de neurones** : 28 à l'entrée, 5 à la sortie,
- d. **Fonction d'activation** : Sigmoidé pour les deux couches « cachée et de sortie » ou Sigmoidé pour la couche cachée et linéaire pour la couche de sortie,
- e. **Valeurs désirées à la sortie ET_A et ET_T** : « 0.05 » pour 0 et « 0.95 » pour 1,
- f. **Algorithme d'apprentissage** : Resilient backpropagation (Rprop), Levenberg-Marquardt backpropagation (trainlm) et scaled conjugate gradient backpropagation (trainscg)
- g. **Le type d'apprentissage** : par groupage (batching mode),
- h. **Fonction de performance** : l'erreur absolue moyenne « mae » et l'erreur quadratique moyenne « mse »,
- i. **Paramètres de la section** : prendre les valeurs par défaut sauf,
 - net.trainParam.epochs - Nombre maximal d'itérations (à définir)
 - net.trainParam.goal - Critère d'arrêt ou l'erreur minimale (à définir)
 - net.trainParam.lr - pas d'apprentissage (en anglais : Learning rate) (à définir)

Pour éviter les problèmes de saturation du réseau les données sont normalisées. Pour garder la même influence les **poids des connexions** sont initialisés au début d'une façon aléatoire puis réutilisés à chaque fois.

Pour ce travail, nous allons utiliser des données du banc d'essai du moteur asynchrone. Nous voulons optimiser :

- ❖ Le nombre d'itérations,
- ❖ Le nombre de neurones de la couche cachée,
- ❖ Le nombre d'observations,
- ❖ Les poids initiaux des connexions

qui permettent au système de reconnaissance de déterminer le niveau de dégradation des défauts de **roulement** à partir d'indicateurs vibratoires c'est-à-dire de s'approcher au maximum du minimum global de la fonction erreur. Tous les choix devront se faire par tâtonnement pour obtenir les performances recherchées et vérifier la robustesse des résultats. Les résultats présentés ici ne sont qu'un échantillon de la quantité énorme de travaux effectués dans le cadre de cette thèse.

Pour notre travail, nous avons considéré **5 classes** (tableau 4.1 page 101). Nous avons procédé à ce partage pour pouvoir étudier l'évolution d'un défaut dans le temps et les performances du perceptron multicouche « PMC ».

4.7.3.3 Performance du PMC sans sélection d'indicateurs

Le réseau est entraîné en utilisant la matrice d'apprentissage EA pour prédire les sorties de la matrice de test ET avec deux ensembles d'indicateurs (temporels et fréquentiels) et avec l'ensemble de combinaison. La performance de classification est calculée en utilisant la formule de la section 1.2.3.2. Le tableau 4.18 représente les performances de classification pour les trois ensembles d'indicateurs.

Tableau 4.18 Performance de la classification par les réseaux de neurones sans sélection d'indicateurs

Indicateurs	Performance apprentissage(%)	Performance test (%)	Mal prédit (en test)	Nbre d'itérations
temporels	100	90	5, 8	7
fréquentiels	100	100	-	100
combinés	100	85	5, 8, 9	11

Les résultats du tableau.4.8 montrent que le réseau est entraîné avec une performance de 100% pour les différents ensembles d'indicateurs. Nous avons obtenu une performance en test de 90% en utilisant les indicateurs temporels et une meilleure performance de 100%, avec les indicateurs fréquentiels, tandis que la combinaison des deux ensembles abaisse les performances à 85%. Cela peut être expliqué par la redondance de quelques indicateurs.

4.7.3.4 Performance du RNA avec sélection des indicateurs

La reconnaissance de formes (supervisée ou non supervisée) consiste à reconnaître (à classifier) des objets sur la base de leur vecteur forme (attributs). Contrairement à ce que l'intuition peut laisser penser, la performance des classifieurs n'augmente pas indéfiniment avec la taille du vecteur d'attributs. Après avoir augmenté en fonction du nombre de descripteurs, l'efficacité atteint un plateau et quelquefois se met à décroître au-delà d'un certain nombre d'attributs ("**malédiction de la dimensionnalité**"). Comme par ailleurs le coût (temps + mémoire) de la classification augmente avec la taille du vecteur

forme, on a tout intérêt donc à limiter le nombre d'indicateurs effectifs au nombre "optimal" (ni trop petit, ni trop grand). Ce nombre optimal est lié à la "dimensionnalité" du problème à traiter. La recherche de cette "**dimensionnalité intrinsèque**" (ou nombre de degrés de liberté du problème) est un sujet intéressant. Dans cette thèse, nous avons proposé, dans la section 4.5, une méthode de sélection indépendante du classifieur basée sur une hybridation de l'algorithme génétique et de l'analyse discriminante linéaire AG_ADL. Les résultats expérimentaux sur les données réelles de roulements ont montré l'efficacité de la méthode proposée.

Le réseau est entraîné et testé en utilisant le vecteur forme optimal. Plusieurs essais ont été réalisés afin de tester la convergence de notre algorithme. Le tableau 4.19 représente les performances de classification après sélection des indicateurs les plus pertinents.

Tableau 4.19 Performances de la classification avec sélection d'indicateurs

Indicateurs	Indicateurs sélectionnés	Performance apprentissage(%)	Performance test (%)	Nbre d'itérations
Essai 1	7 13 18 19	100	100	100
Essai 2	7 13 19 24	100	100	100
Essai 3	7 13 19 24	100	100	100
Essai 4	7 13	100	100	87

Les performances du RNA sur la base de test, en utilisant le vecteur forme optimal (tableau 4.19), montrent une amélioration de la performance de la classification pour les différents essais réalisés. Cette performance atteint 100%. (Toutes les observations sont correctement affectées aux classes correspondantes) avec un nombre d'indicateurs réduit. La séparation des cinq classes est, donc, réalisée avec un taux d'erreur nul. On peut aussi constater que pour les différents essais, le vecteur forme optimal n'est pas le même. Ceci est tout à fait normal puisque la population initiale est générée d'une façon aléatoire. Toutefois, certains indicateurs se répètent, ce sont : 7 13 19 et 24 qui correspondent respectivement à la valeur efficace calculée dans les bandes [1500-3000], [3000-4500],[4500-6000] et au moment d'ordre 6 calculé dans la bande [4500-6000]. On peut constater également que la classification est réalisée avec une performance de 100% en utilisant seulement les indicateurs 7 et 13.

4.7.4 Conclusion

L'utilisation des techniques d'optimisation par AG et de la classification par un réseau de neurone artificiel PMC à côté du traitement du signal a pu améliorer la surveillance des machines en détectant l'évolution d'un défaut dans le temps. C'est ce que nous avons pu mettre en évidence lors de cette section. Dans la prochaine section, nous considérons la nouvelle méthode des Supports vector machines (SVM).

4.8 Performance de l'algorithme des SVM

Les *Support vector machines*, souvent traduit par l'appellation de supports à *vaste marge* ou séparateurs à Vaste Marge (SVM) sont une classe d'algorithmes d'apprentissage initialement définis pour la discrimination c'est-à-dire la prédiction d'une variable qualitative initialement binaire. Ils ont été ensuite généralisés au cas de plusieurs classes et à celui de la prédiction d'une variable quantitative. Dans le cas de la discrimination d'une variable dichotomique, ils sont basés sur la recherche de l'*hyperplan de marge optimale* qui, lorsque c'est possible, classe ou sépare correctement les données tout en étant le plus éloigné possible de toutes les observations (figure 4.25). Le principe est donc de trouver un classifieur, ou une fonction de discrimination, dont la capacité de généralisation (qualité de prédiction) est la plus grande possible. Cette approche découle directement des travaux de Vapnik en *théorie de l'apprentissage* à partir de 1995 (Vapnik, 1998), (Vapnik, 2000). Elle s'est focalisée sur les propriétés de généralisation (ou prédiction) d'un modèle en contrôlant sa complexité.

L'annexe 4.3 présente une description des supports à vaste marge (SVM), en mentionnant les principes, et les bases théoriques de cette méthode, il précise les différents algorithmes d'apprentissage des SVM en détaillant beaucoup plus l'algorithme '*Sequential minimal optimisation*' (SMO), qui sera utilisé dans les sections suivantes pour étudier les performances des SVM dans les différents cas de classification : binaire, multi classes, avant et après sélection d'indicateurs, et avec les différents *noyaux*.

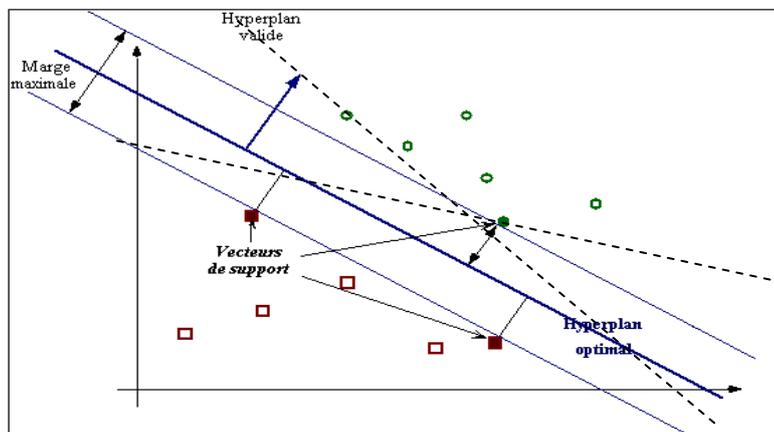


Figure 4.25 Exemple de séparation entre deux classes

(extraite de (Cornuéjol et al., 2002))

4.8.1 Principe des séparateurs à vaste marge (SVM)

Le principe de base des SVM consiste à ramener le problème de la discrimination à celui, linéaire, de la recherche d'un hyperplan optimal. Deux idées ou astuces permettent d'atteindre cet objectif (Bessel, 2005) :

- la première consiste à définir l'hyperplan comme solution d'un problème d'optimisation sous contraintes dont la fonction objectif ne s'exprime qu'à l'aide de produits scalaires entre vecteurs et dans lequel le nombre de contraintes "actives" ou vecteurs de supports contrôle la complexité du modèle.
- la deuxième est le passage à la recherche de surfaces séparatrices non linéaires, Elle est obtenu par l'introduction d'une fonction noyau (*kernel*) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (*feature space*) de plus grande dimension. D'où l'appellation couramment rencontrée de machines à noyau ou *kernel machines*.

L'introduction de noyaux, spécifiquement adaptés à une problématique données, lui confère une grande flexibilité pour s'adapter à des situations très diverses (reconnaissance de formes, de séquences génomiques, de caractères, détection de spams, diagnostics, etc.).

De nombreuses introductions et présentations des SVM sont accessibles sur des sites comme par exemple : www.kernel-machines.org. Le livre de référence sur ce sujet est celui de Vapnik (Vapnik, 1998). Pour plus de détails sur l'approche des SVM et de leurs applications aux problèmes de classification, on peut se référer à l'annexe 4.3. Les séparateurs à vaste marge ou SVM en abrégé ont été le sujet de plusieurs travaux (Samanta, 2004a, b), (Yang, 2005a), (Yang, 2005b), (Jack et al., 2002), (Kharoubi, 2002), les principes, et les bases théoriques de cette méthode ont été discutés dans plusieurs ouvrages, les plus connues sont (Burge, 1998) et (Vapnik, 2000).

Dans les prochaines sections sera discuté l'implémentation d'un classifieur basé sur les SVM. Le choix du nombre des indicateurs « à l'entrée du classifieur » est optimisé en utilisant l'approche AG-ADL discuté en (section 4.5).

4.8.2 Implémentation des SVM

Durant cette partie expérimentale, pour entraîner et tester les SVM, nous avons adapté à notre problème une boîte à outils MATLAB appelée : *Statistical Pattern Recognition Toolbox for Matlab*¹ (Vojtech et al., 2004), l'algorithme principal utilisé pour l'entraînement des SVM est celui de Platt : *sequential minimal optimisation* (SMO), cet algorithme est décrit dans la section 4.3 de l'annexe 4.3, mais pour des détails complémentaires voir (platt, 1998). Un pseudo code de cet algorithme est donné ci-dessous tiré de la *Boîte à outils « Statistical Pattern Recognition Toolbox for Matlab (STPRTool) »* :

Nous considérons 3 cas :

Example 1 : Classification binaire (deux classes) :

¹<http://cmp.felk.cvut.cz/pub/cmp/software/stprtool/index.html>
<ftp://cmp.felk.cvut.cz/pub/cmp/articles/Franc-TR-2004-08.pdf>

```

trn = load('Cetim_train');           %charger les données d'apprentissage
tst = load('Cetim_tst');             % charger les données de test
options.ker = 'rbf';                 %choix du noyau ( kernel)
options.arg = [0.1 0.5 1 5];         % paramètres du noyau
options.C = [1 10 100];              % constante de régularisation
options.solver = 'smo';               % Algorithme d'apprentissage
[model,Errors] = evalsvm(trn,options); % évaluation avec les
                                     différents paramètres du
                                     noyau et la constante C

figure; mesh(options.arg,options.C,Errors);
hold on; xlabel('arg'); ylabel('C');
ypred = svmclass(tst.X,model);       % la prédiction des sortie yi
ccolor(ypred,tst.y)                  % calcul de l'erreur de classification

```

Exemple 2 :Multiclass (un contre un)

```

data = load('pentagon');              %charger les données
options = struct('ker','rbf','arg',1,'C',10,'verb',1);%
                                     options du noyau et la constante C
model = oaosvm(data,options);         %fonction de décision 'un contre un'
figure;
ppatterns(data); ppatterns( model.sv.X);
pboundary( model );                  %visualisation des données

```

Exemple 3 :Multiclass (un contre tous)

```

data = load('pentagon');
options = struct('ker','rbf','arg',1,'C',10,'verb',1);
model = oaosvm(data,options);         %fonction de décision 'un contre tous'
figure;
ppatterns(data); ppatterns( model.sv.X);
pboundary( model );

```

4.8.3 Performance de la classification

Les travaux présentés dans cette section portent sur la classification des défauts de l'engrenage conique « 5 » **de la boîte de vitesse de l'hélicoptère (NAVY Westland data)** (section 3.7). Nous avons étudié les performances des SVM.

Dans la première partie, tous les ensembles d'indicateurs calculés participent au entraînement et au test des SVM mais, dans le deuxième partie, seulement, un sous

ensemble d'indicateurs pertinents est sélectionné par l'algorithme AGs_ADL (section 4.5) pour entraîner et tester les SVM. Nous avons pu montrer, que les performances du SVM sont étroitement liées à la pertinence des indicateurs. La procédure est validée sur les signaux avec trois niveaux (engrenage sain et 2 niveaux de dégradation) en utilisant des indicateurs temporels et fréquentiels.

➤ **Conduite des différentes étapes d'apprentissage et de test**

Voici donc un résumé de la démarche à suivre pour entraîner et tester les SVM :

1- préparer les données, calculer en traitement du signal, que l'on doit classer. Ces données seront divisées en deux sous-ensembles ; le premier forme la matrice d'apprentissage EA , le deuxième la matrice de test ET , ces deux ensembles sont stockés dans deux tableaux X de type observations/variables.

Nous avons alors N observation, représentées en lignes et M indicateurs, représentées en colonnes. Ces matrices seront divisés en N_c classes (w_1, w_2, \dots, w_{N_c}). Chaque classe w_i possède N_i observations,

2- former deux vecteurs d'étiquettes : ET_A pour l'apprentissage, et ET_T pour le test. Le vecteur ET_A est de taille N_a , égale au nombre d'observations de la matrice d'apprentissage, le vecteur ET_T est de taille N_t , égale au nombre d'observations de la matrice de test. Chaque élément de ces deux vecteurs prend la valeur de la classe de l'observation correspondante (1,2... N_c).

Pour le cas de la classification binaire (deux classes), nous attribuons les valeurs d'étiquettes suivantes : 1 pour les observations de la première classe, et 2 pour la les observations de la deuxième classe. Le problème qui se pose, c'est que pour la tâche de classification par les SVM, les étiquettes des observations doivent avoir seulement les valeurs (-1 ou 1). Pour résoudre ce problème nous utilisons la conversion suivante :

$$e' = e \times 2 - 3$$

Avec e : l'ancienne étiquette (1 ou 2)

e' : la nouvelle étiquette (-1 ou 1)

Dans le cas des SVM multi classes (*un contre un* où *un contre tous*), le vecteur d'étiquette peut prendre des valeurs supérieur à 2, et comme ces méthodes utilisent des combinaisons des SVM binaires, le problème est résolu de la même manière,

3- faire un choix du noyau et les valeurs de ses paramètres (exemple : le degré 'P' du noyau polynomial).

4- faire un choix des valeurs de la constante de régularisation 'C'.

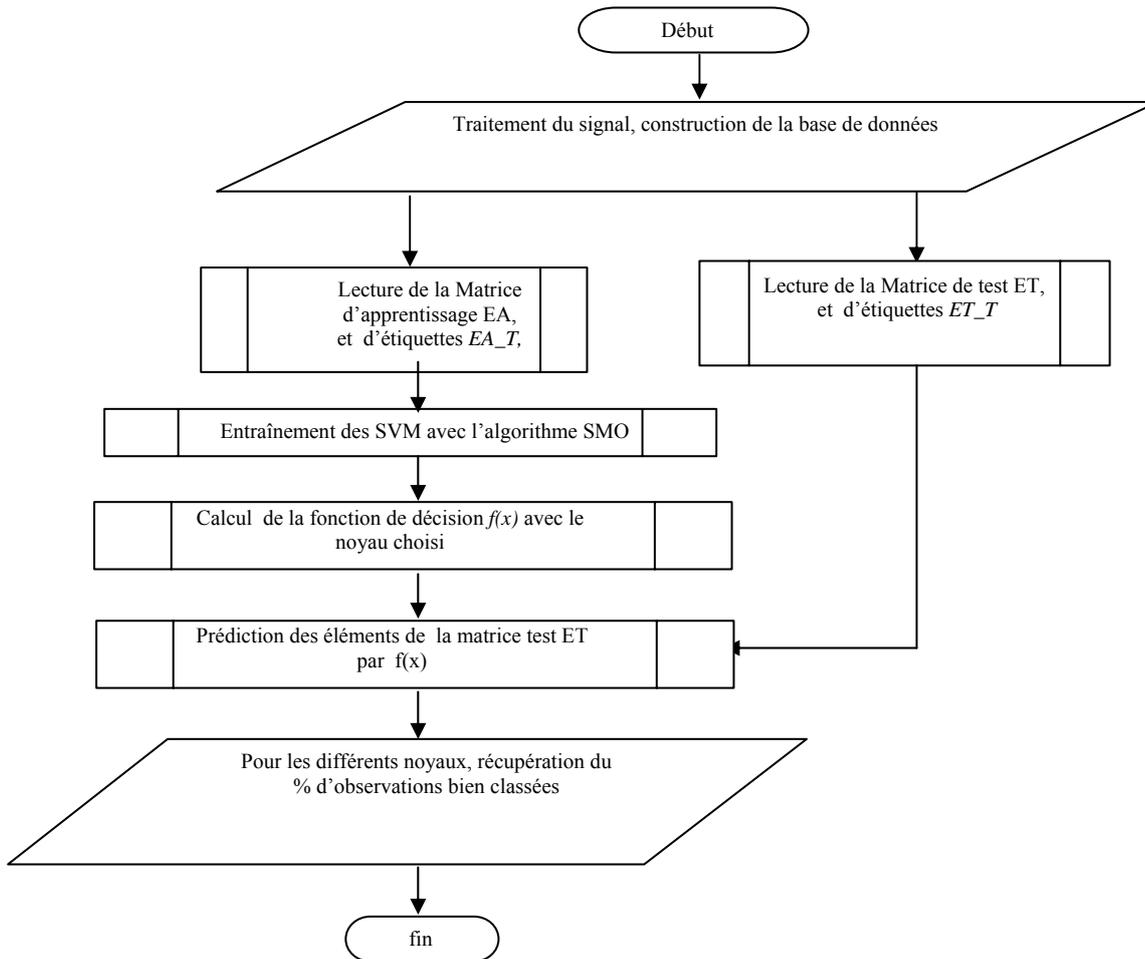


Figure 4.26 Organigramme de classification par les SVM sans sélection d'indicateurs. Utilisation du vecteur initial.

5- l'algorithme SMO utilise la matrice d'apprentissage EA et le vecteur d'étiquettes ET_A pour résoudre le problème d'optimisation appelé par l'équation (33) de l'annexe 4.3. La se problème, nous donne les valeurs optimales du vecteur α , les indices des observations appelés *vecteurs de supports*, ainsi que le biais 'b'.

6-Avec tous les paramètres choisis ou calculés précédemment, on peut établir la fonction de décision $f(x)$

7- la fonction de décision est maintenant établie, on utilise la matrice de test pour prédire les sorties y_i de chaque observation x_i . Si le signe de l'équation 21 de l'annexe 4.3 est positif, y (la sortie prédite) prend la valeur 1, dans le cas contraire, y prend la valeur 2.

8- pour voir quels sont les éléments mal prédits, on fait une comparaison entre le vecteur des sorties prédites et les sorties désirées du vecteur d'étiquettes (ET_T) .

9- On peut donc calculer la performance des SVM c'est-à-dire le nombre de sorties bien prédites.

L'organigramme ci-dessus donné par la figure 4.26 résume ces différentes étapes de classification sans sélection des indicateurs.

4.8.3.1 Préparation des observations

➤ Prétraitement des signaux

Dans cette partie les étapes de prétraitement restent les mêmes que celles des signaux de CETIM (section 4.5.5.1). Nous procédons au choix de nombre des tranches K_t et le nombre des points pour chaque tranche N_M . Ce dernier doit être suffisamment grand pour couvrir un nombre suffisant du plus grand cycle correspondant à la période de rotation $T = 0.023s$ (page 95). Si nous prenons un coefficient de 20, la longueur nécessaire de la tranche sera : $Ne = 2418 \times 20 = 48360$ points.

Si nous considérons pour N_M une valeur supérieure à Ne , $N_M = 2^{16} = 65536$ échantillons et puisque nous disposons d'un nombre d'échantillons N_s de 412464 points. Alors le nombre de tranches dans le cas d'un recouvrement 50 %, soit $s_1 = \frac{N_M}{2}$, est de :

$$K_t = \frac{N_s - s_1}{N_M - s_1} \cong 11 \text{ tranches} \quad (4.25)$$

Pour le défaut étudié, nous disposons de 9 signaux représentant l'état sain, 6 signaux pour le défaut niveau 1, et 9 signaux pour le défaut niveau 2 (figure 3.18). Après la division de chaque signal en 11 tranches nous obtenons **264 tranches** (24*11 observations) échantillonnées à 103166 Hz. Le signal de chaque tranche a donc été centré puis filtré par un filtre passe-bande. Ici, nous avons limité notre étude à une fréquence maximale de $5000\text{Hz} > 4 f_{eng1}$ (1108,9 Hz de engrenage 5).

Les tranches des signaux sont, donc, suréchantillonnées. Nous proposons de décimer ces signaux par un facteur 4. Nous obtenons ainsi des signaux de **16384 points** échantillonnés à une fréquence $f_e = 103116/4 = 25779$ Hz. On peut remarquer que le théorème de Shannon est vérifié.

➤ **Extraction des indicateurs**

Dans cette étape nous avons choisi d'extraire les trois ensembles d'indicateurs suivants :

1^{er} ensemble : composé des indicateurs temporels suivants : la valeur efficace, le facteur de crête, le skewness, le kurtosis, et les moments d'ordre (5 à 9). Ces indicateurs sont calculés dans différentes bandes latérales de largeur 426 Hz autour des fréquences d'engrènement, soit 10 raies latérales (dix fréquences de rotation fr_1) autour des 3 principales harmoniques (1. $f_{eng1} = 1108.9$ Hz, 2. $f_{eng1} = 2217.8$ Hz, et 3. $f_{eng1} = 3266.7$ Hz). Soit donc trois bande de calcul : [895-1321 Hz], [2004-2430 Hz] et [3113-3539 Hz]. Alors la matrice des données obtenue est de **264**individus / **27** indicateurs (27=9 indicateurs x.3 bandes)

2^{ème} ensemble : ici, nous avons choisi de calculer les coefficients du modèle AR dans les trois bandes indiquées ci-dessus, la méthode utilisée pour identifier l'ordre du modèle est la Décomposition en Valeurs Singulières (SVM) (Felkaoui, 1995), nous avons obtenu pour une bande fréquentielle de 426 Hz un ordre égal à 2. Puisque nous avons choisi de calculer ces coefficients dans trois bandes nous obtenons alors 6 coefficients pour chaque tranche, ce qui donne pour tous les signaux, une matrice de (**264**individus / **6** indicateurs)

3^{ème} ensemble : dans cette étape, nous avons choisi de calculer les énergies autours des fréquences caractéristiques dans les trois bandes précités, dans la 4^{ième} bande [4222,4648], et dans la bande totale [895-4648]. La matrice obtenue est, donc, composée de (**264** individus / **5** indicateurs).

Ces trois ensembles forment une matrice totale de (**264** individus / **38** indicateurs) ces individus appartiennent au trois classes suivantes :

1^{ière} classe (sans défaut) : composée de 99 individus (9x11).

2^{ième} classe (défaut niveau1) : composée de 66 individus (6x11).

3^{ième} classe (défaut niveau2) : composée de 99 individus(9x11).

Toutes les données ont été normalisées en accord avec l'équation (4.13). La matrice totale obtenue est divisée en deux sous ensembles : Le premier formé de 144 individus (6 tranches de chaque signal), utilisé pour l'entraînement des SVM, le deuxième composé de 120 (24 x 5 : les 5 tranches restantes de chaque signal), utilisé pour le test.

Nous avons constaté une variation des indicateurs qui caractérise bien la dégradation de l'état de l'engrenage pour les deux niveaux de l'écaillage sauf pour les coefficients qui ont été insensibles au défaut.

4.8.3.2 Performances des SVM sans sélection

Les SVM sont entraînés avec les trois ensembles d'indicateurs, et avec quatre noyaux différents : linéaire, polynomial, RBF, et sigmoïdal. Dans ce cas nous avons trois classes différentes ce qui implique l'utilisation des méthodes multi classes. Dans cette partie de travail nous avons choisi d'étudier les performances des deux méthodes: *un contre un* et *un contre tous*.

a. Un contre un

Cette méthode est basée sur des combinaisons de $q_j(x)$ classifieurs binaires équation 34 annexe 4.3, tel que :

$$q_j(x) = \begin{cases} 1 & \text{pour } f(x) \geq 0 \\ 2 & \text{pour } f(x) < 0 \end{cases} \quad (4.26)$$

Dans le cas étudié, nous avons trois classes, les classifieurs binaires que nous avons établi sont (1 contre 2), (2 contre 3), (1 contre 3). Les étapes de classification restent les mêmes que celle de la classification binaire, mais dans ce cas, chaque classifieur binaire ne peut affecter que les observation appartenant aux classes qu'il sépare. Donc pour avoir un modèle (fonction de décision) qui peut affecter n'importe quelle observation à l'une de ces trois classes nous avons utilisé une règle de vote majoritaire.

La stratégie de vote majoritaire pour le cas des SVM multi classes est utilisée pour l'affectation d'une observation x à une des N_c classes. Cette méthode affecte l'observation x à la classe $y \in \mathcal{Y}$ qui a la majorité des votes :

$$y = \arg \max_{y=1..g} v_y(x) \quad (4.27)$$

Le tableau 4.20 représente les performances de chaque classifieur binaire ainsi que la performance totale obtenue en utilisant la règle de vote majoritaire, pour les différents noyaux : linéaire polynomial, RBF, et sigmoïdal. Il compare les résultats obtenus avec les différents noyaux. Les symboles (perf 1-2), (perf 1-3) et (perf 2-3) désignent respectivement les performances des classifieurs binaires établis (observations bien prédites) pour les paires de classes (1 et 2), (1 et 3) et (2 et 3), le symbole (VS) désigne le nombre total des vecteurs de supports. C : est la meilleure valeur de la constante de régularisation obtenue par validation croisée

Les résultats montrent que les meilleures performances (91.11% à 100%) sont obtenues

Tableau 4.20 Performances, observations bien prédites, des SVM avec la méthode « un contre un » pour les différents noyaux sans sélection

Noyau (Kernel)	Nbre VS	Perf 1-2	Perf 1-3	perf 2-3	Performance totale (%)	C
Linéaire	81	88	100	80	90	100
Polynomial	116	72	100	80	76.66	1
RBF	34	89.33	91.11	78.66	92.5	100
Sigmoïdal	20	88	91.11	80	95.83	100

avec le classifieurs binaire établi entre la première est la troisième classe, cela est expliqué par le niveau développé du défaut et qui est traduit sur la figure 4.27 par l'éloignement des deux classes ; fonctionnent sain (sans défaut) et la classe (niveau 2). Pour le modèle global établi en utilisant la règle de vote majoritaire, on peut constater que la meilleure performance est obtenue avec le noyau sigmoïdal (95.83%), une performances de 92.5 avec le noyau RBF , et 90% avec le noyau linéaire. La figure 4.27 représente l'hyperplan optimal séparant les trois classes

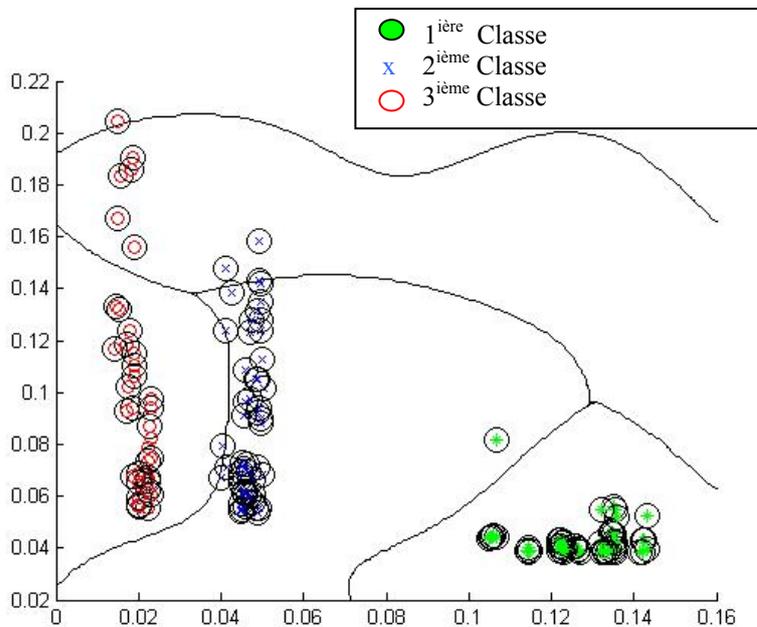


Figure 4.27 Séparation des trois classes (avec un noyau RBF)

Cette figure représente l'hyperplan séparant les trois classes (la classe du fonctionnement sain (sans défaut) en vert, la classe (niveau 1) en bleu, et la classe (niveau 2) en rouge), on peut constater que la distance entre les classe traduit l'état de dégradation de l'engrenage pour les deux niveau d'écaillage de l'engrenage.

b. Un contre tous

L'entrée du classifieur « un contre tous » est composée de la matrice d'apprentissage EA et de test ET, les vecteurs d'étiquettes : ET_A pour l'apprentissage, ET_T pour le test. Ces deux vecteurs sont composés des étiquettes $y_i \in \mathcal{Y} = \{1, 2, \dots, N_c\}$ avec N_c : le nombre des classes.

Le problème de multi classes est résolu par la méthode un contre tous, en transformant le problème en séries de N_c classifieurs binaires. Chaque classifieur binaire est entraîné et testé en suivant les mêmes étapes données par l'organigramme de la figure 4.26 chaque classifieur q_i est établi entre la classe i et les autres classes qui sont considérés comme une seule classe donc la sortie y_i' prédite par q_i a les valeur suivantes :

$$y_i' = \begin{cases} 1 & \text{pour } y = y_i \\ 2 & \text{pour } y \neq y_i \end{cases} \quad (4.28)$$

Pour les données que nous avons traité, trois classifieurs binaires ont été entraînés et testés (1 contre tous, 2 contre tous, 3 contre tous). Les résultats obtenus sont représentés sur le tableau 4.21.

Tableau 4.21 Performances des SVM avec la méthode « un contre tous » pour les différents noyaux (sans sélection)

Noyau (Kernel)	Nbre VS	Perf 1-tous	Perf 2-tous	perf 3-tous	Performance totale (%)	C
linéaire	26	100	100	100	100	100
Polynomial	96	100	100	97.5	98.33	100
RBF	32	100	100	100	100	200
Sigmoïdal	76	100	100	89.99	99.17	10

Les symboles (perf 1-tous) (perf 2-tous) (perf 3-tous) désignent respectivement les performances des classifieurs binaires en considérant :

Classe 1 (sans défaut) contre les autres (classe2 et3).

Classe 2 (défaut niveau1) contre les autres (classe1 et3).

Classe 3 (défaut niveau1) contre les autres (classe1 et2).

Les résultats obtenus montre que les meilleures performances sont obtenues avec le noyau linéaire (100%) et RBF(100%) pour les différents cas, on peut constater aussi que la séparation de la première et la deuxième classe des autres a été réalisée avec une performance de 100% et une performance de 89.99 à 100% pour la séparation de la troisième classe l'algorithme n'arrive pas à discriminer la 3^{ième} classe représentant le niveau 3. On peut remarquer également que le nombre des vecteurs de supports utilisés varie pour les différents noyaux. En effet nous avons eu une performance de 100% avec un nombre 26 vecteurs de supports c'est-à-dire 18% des vecteurs d'apprentissage sont utilisés pour le calcul de la fonction de décision avec le noyau linéaire, et 32 vecteurs de supports avec le noyau RBF.

4.8.3.3 Performances des SVM avec sélection

Nous avons déjà souligné que les techniques de la sélection pour la classification, sont proposées pour remplir plusieurs besoins simultanément :

- Réduire le coût d'extraction des indicateurs.
- Améliorer l'exactitude de classification.
- Améliorer la précision de l'évaluation de performance.

Les Algorithmes génétiques (AG) présentent un mécanisme idéal pour la sélection des indicateurs dans les hauts espaces dimensionnels. Utilisant les principes *Darwiniens* évolutionnaires, l'AG est capable de chercher à travers un grand nombre de solutions possibles rapidement, et arrive à une solution dans un temps beaucoup plus rapide que d'autres méthodes de sélection d'indicateurs comparables de la section 2.6. Les AGs ont été montrés pour être extrêmement robuste, et offrent une bonne exactitude sans exiger toute connaissance antérieure.

Pour évaluer les performances des SVM après sélection des indicateurs, la matrice de données présentée à l'entrée du classifieur est composé de $M' < M$ indicateurs. Où M est le nombre initial des indicateurs, et M' le nombre des indicateurs sélectionnés par le couplage AG_ADL (section 4.5)

La figure 5.21 représente l'organigramme de classification par les SVM utilisant le vecteur forme optimal

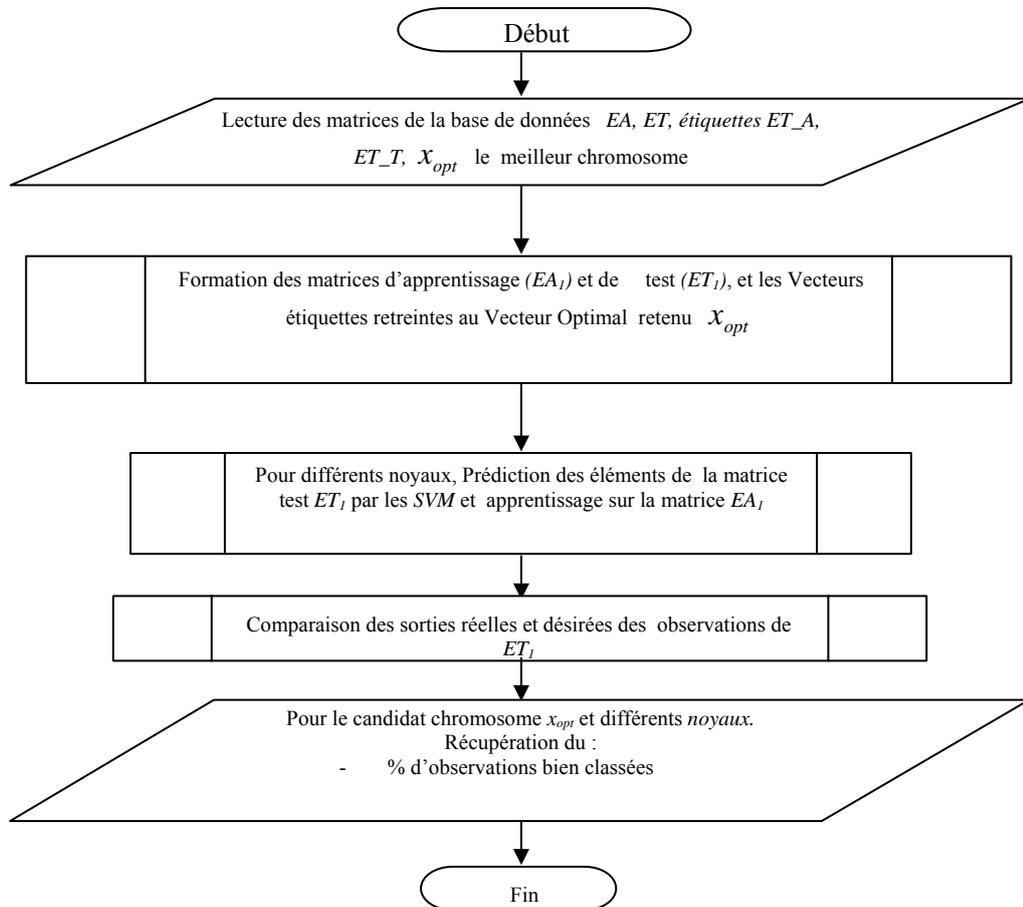


Figure 4.21 Organigramme de classification et de validation par les SVM utilisant le vecteur forme optimal

a. Performance de l'algorithme : Un contre un

En utilisant l'approche un contre un, et le vecteur forme optimal retenu par l'algorithme de sélection. Les SVM sont entraînés et testés en utilisant quatre noyau différents : linéaire, polynomial, RBF, et sigmoïdal.

Le tableau 4.22 représente les performances des SVM binaires établis pour chaque pair de classes et la performance totale en utilisant la règle de vote majoritaire

Tableau 4.22 Performances des SVM avec la méthode « un contre un » pour les différents noyaux, avec les indicateurs sélectionnés

Noyau (Kernel)	Vecteur forme optimal	Nbre 'SV'	1-2	1-3	2-3	Performance totale (%)	C
linéaire	8 9 14 17 23 25	59	89.33	100	80	86.83	100
Polynomial	6 7 8 9 15 17 23 24 25 34 38	100	66.66	100	80	86.66	100
RBF	6 7 8 9 15 17 23 24 25 38	26	86.66	100	80	100	100
Sigmoïdal	6 7 8 9 15 17 23 24 25	22	93.33	100	80	90.83	100

En comparaison avec les performances obtenues en utilisant le vecteur forme initiale (Tableau 4.20), les résultats du tableau 4.22 montrent que les performances des SVM sont améliorées pour les différents cas. Les SVM permettent donc, une bonne séparation entre les classes notamment la première et la troisième classe, où le taux de succès est maintenant de 100% pour les quatre noyaux. Cela est expliqué par le niveau d'écaillage développé, et le pouvoir discriminant des indicateurs retenus par l'algorithme de sélection du vecteur forme optimal. On constate aussi une amélioration de la performance totale pour tous les noyaux.

Les résultats obtenus montrent aussi, pour le noyau sigmoïdal, qu'une diminution du nombre d'indicateurs de (de 38 à 9) n'entraîne pas une perte en performance (comme les k-ppv). On peut constater également que la majorité des indicateurs sélectionnés sont les indicateurs temporels, qui ont donc, un pouvoir discriminant important.

b. Performance de l'algorithme : Un contre tous

Le tableau 4.23 représente les résultats obtenus avec la méthode « un contre tous », après sélection des indicateurs (en utilisant le vecteur forme optimal), et avec les quatre noyaux : linéaire, polynomial, RBF, et sigmoïdal.

Tableau 4.23 Performances des SVM avec la méthode « un contre tous » pour les différents noyaux, avec les indicateurs sélectionnés

Noyau (Kernel)	Vecteur forme optimal	Nbre 'SV'	Perf 1-tous	Perf 2-tous	Perf 3-tous	Performance totale (%)	C
linéaire	6 7 8 9 14 17 23 25	50	100	100	100	100	100
Polynomial	6 7 8 9 10 15 17 23 24 25 34 38	78	100	100	98.33	100	100
RBF	6 7 8 9 15 17 21 23 24 25	73	100	100	100	100	100
Sigmoïdal	6 7 8 9 15 17 23 24 25	25	100	100	98.33	99.83	100

Avec les différents noyaux utilisés, on peut constater que la performance des classifieurs binaires est améliorée en comparaison avec le tableau 5.21. Nous avons eu une performance de 100% avec le noyau linéaire en utilisant seulement 8 indicateurs et une performance de 100% avec le noyau RBF en utilisant 9 indicateurs. Le vecteur forme optimal n'est pas le même pour les différents essais. Ceci est toujours la cause de la génération aléatoire de la population initiale des AGs. On peut constater également que la majorité des indicateurs sélectionnés, sont les indicateurs temporels 6 7 8 9 15 17 23 24 et 25.

c. Les vecteurs de supports

L'examen du nombre des vecteurs supports utilisés dans les différents cas étudiés, peut donner beaucoup d'informations concernant les vecteurs qui sont utilisés réellement, et aussi, il donne une indication sur la complexité des données dans l'espace augmenté par les différents noyaux utilisés. Quand l'hyperplan est difficile à déterminer, beaucoup de vecteurs du support doivent être utilisés, alors que peu de vecteurs indiquent que l'hyperplan est relativement facile à identifier ; et par conséquent, la performance sera relativement bonne. Prenons comme exemple pour ces données de Westland, on peut constater que le nombre des vecteurs de supports varie en fonction des noyaux choisis, voir les résultats du tableau 4.20. Avec le noyau sigmoïdal, la fonction de décision est établi avec 20 vecteurs supports , et 34 en utilisant un noyau RBF, ce qui explique les performances totales obtenues avec ces deux noyaux en comparaison avec le noyau

linéaire et polynomial . On peut constater également que le noyau polynomial nécessite un nombre élevé de vecteurs supports pour établir la fonction de décision. En comparant les résultats des tableaux sans sélection et des tableaux avec sélection du vecteur forme, Nous pouvons constater que généralement après sélection des indicateurs, le nombre des vecteurs supports diminue pour les différents noyaux, ce qui indique que les classes sont bien séparées l'une à l'autre, et par conséquent le nombre nécessaire de vecteurs supports pour déterminer l'hyperplan optimal est faible.

4.8.4 Conclusion

Dans cette section 4.8, nous avons étudié les performances des SVM pour différentes applications. Deux cas principaux ont été étudiés : Dans le premier cas, tous les indicateurs calculés participent à la classification, et dans le deuxième cas, seulement un sous-ensemble d'indicateurs a été sélectionné pour effectuer la classification des données. Nous avons testé les performances des SVM binaires (séparation entre deux classes) et les performances des SVM multiclassés en utilisant les signaux de 'Westland'. Nous avons obtenu de bons résultats avec les deux approches utilisées. L'étude des performances des SVM, après sélection des indicateurs du vecteur forme optimal, montre une amélioration des performances dans la plupart des cas avec un nombre d'indicateurs très réduit.

Chapitre 5

Conclusion et perspectives

Les travaux présentés dans cette thèse ont abordés les différentes étapes nécessaires à la construction d'un système de reconnaissance de défauts des machines tournantes. Pour chacune de ces étapes à savoir : l'acquisition, les pré-traitements et traitements, la génération et la réduction des indicateurs, puis enfin la décision (classification), nous avons tenté de présenter le maximum d'information et les travaux bibliographiques récents..

Nous avons proposé une méthode d'optimisation pour la sélection d'indicateurs pertinents du système de reconnaissance. En mécanique, de nombreuses études ont été proposées dans la littérature (voir section 2.6). Notre orientation s'est focalisée sur un couplage algorithmes génétiques (AGs)/analyse discriminante linéaire (ADL). Ainsi, dès l'étape d'extraction des indicateurs, la sélection des paramètres pertinents et non redondants du système est effectuée. Cette sélection consiste à réduire les entrées du classifieur (k-ppv, réseau de neurones de type PMC et séparateurs à vaste marge SVM) tout en améliorant ou en maintenant le taux de reconnaissance de la classification.

Nos contributions principales sont :

- la mise en place de presque tous les programmes du traitement de signal sous un environnements Matlab avec des fenêtres interactives,
- l'étude et la mise en oeuvre d'une méthode classique de Fisher et d'une méthode hybride AGs-ADL pour sélectionner les indicateurs discriminants,
- l'étude et la mise en oeuvre des algorithmes de classification (cités ci-dessus) pour le cas des défauts des machines tournantes,

Une fois que nous avons déterminé les bons paramètres de l'AG tels que : la taille de la population, le nombre de génération, les probabilités de mutation et de croisement et les méthodes de reproduction. Nous avons effectué des tests de performances sur les bases des données réelles d'un réducteur à engrenages, des roulements d'un moteur asynchrone et des engrenages d'une boîte à vitesses d'un hélicoptère. Les meilleurs résultats ont été obtenus par l'utilisation des SVM.

Dans la section suivante nous présentons les principaux résultats de cette thèse..

5.1 Principaux résultats

5.1.1 Importance de la sélection d'indicateurs

Le traitement des signaux et la recherche du vecteur forme constituent une étape très significative de la phase de construction du système de diagnostic automatique des défauts par RdF.

Nous avons pu montrer dans la section 4.3 et 4.4 à travers la visualisation par ACP, que la séparabilité entre modes de fonctionnement est étroitement liée à la pertinence des indicateurs. La

procédure a été validée sur des signaux réels vibratoires de roulements (figure 4.5 et 4.8) et d'engrenage (figure 4.5 et 4.13). L'élimination, donc, des éléments redondants et inutiles est d'une grande importance dans la séparation des modes de fonctionnement, de plus, on peut constater visuellement sur les mêmes figures que la séparabilité des classes obtenue avec le VF réduit est meilleure que celle obtenue avec le VF initial.

5.1.2 Couplage AGs-ADL

L'hybridation de plusieurs méthodes afin d'exploiter les avantages de chacune, semble constituer une approche intéressante pour la robustesse des systèmes de reconnaissance automatique de défauts. Ainsi, nous avons couplé les AGs qui ont une grande capacité d'optimisation et l'ADL qui est une méthode efficace de discrimination.

L'algorithme proposé a permis de choisir un sous-ensemble optimal de 3 indicateurs sur les 60 initiaux. Il s'agit, bien sûr, d'une représentation parcimonieuse mais informative du processus. L'utilisation d'un classifieur comme les k-ppv permet de comparer les performances de la classification en utilisant le VF initial et le VF réduit.

5.1.3 Performances du classifieur k-ppv

On a remarqué après une visualisation graphique de l'évolution du pourcentage de bien classés en fonction des k-ppv et pour les différents cas qu'il est préférable de choisir une valeur faible de voisins (k=1 à 4).

Les résultats obtenus *avec sélection* sont satisfaisants. Ainsi, la réduction du nombre d'indicateurs a permis de réduire le temps de calcul requis pour l'apprentissage. Nous avons remarqué, également, que la performance du classifieur n'augmente pas indéfiniment avec la taille du Vecteur Forme et qu'une diminution du nombre d'indicateurs, par exemple de 60 à 3, cause une perte en performance de 1 % seulement dans les observations bien classées. Pour un nombre de voisins k=1 ou 2, le résultat est toujours parfait c'est-à-dire de 100 %. Pour le cas de 3 classes et k=3 ou 4, le pourcentage de bien classées atteint au maximum les 83.33.

Les résultats des deux autres méthodes basées sur un calcul direct des frontières : le PMC et les SVM seront vues dans les paragraphes suivants.

5.1.4 Performances du classifieur PMC

L'optimisation du réseau, au début, a permis de réduire le temps de calcul requis pour l'apprentissage du réseau et de réduire également la complexité du réseau de neurones.

Les performances du RNA sur la base de test, en utilisant le vecteur forme optimal (tableau 4.19), montrent une amélioration de la performance de classification pour les différents essais réalisés. Cette performance a atteint 100%. La séparation des cinq classes est, donc, réalisée avec un taux d'erreur nul. On peut aussi constater que pour les différents essais, le vecteur forme optimal n'est pas le même. Ceci est tout à fait normal puisque la population initiale est générée d'une façon aléatoire. L'utilisation des techniques d'optimisation par AGs et de la classification par un réseau de neurone artificiel PMC à coté du traitement du signal à pu, donc, améliorer la surveillance des machines

Dans la prochaine section, nous considérons la nouvelle méthode des Séparateurs à vaste marge (SVM : Support vector machines).

5.1.5 Performances du classifieur SVM

Nous avons testé les performances des SVM binaires (séparation entre deux classes) et les performances des SVM multiclassés en utilisant les signaux de 'Westland'. Nous avons obtenu de bons résultats avec les deux approches utilisées. L'étude des performances des SVM, après sélection des indicateurs du vecteur forme optimal, montre une amélioration des performances dans la plus part des cas et une diminution admissible dans certain cas mais avec un nombre d'indicateurs plus grand que dans les cas du PMC.

5.2 Perspectives

Si, de par l'extrême complexité de notre problème, les résultats que nous avons obtenus ont montré des caractéristiques intéressantes qui nous poussent à approfondir nos recherches dans plusieurs directions. Ainsi, plusieurs perspectives peuvent être envisagées, nous les présentons ci-dessous.

5.2.1 Notion de rejet

La mise en place de méthodes de diagnostic par RdF se heurte souvent au problème de la connaissance à priori du système étudié. En effet, il est rare de disposer d'un nombre suffisant d'observations correspondantes à tous les modes de fonctionnement possibles d'un système (qualifiés d'anormaux). Il est donc nécessaire de concevoir des systèmes de diagnostic adaptatif, dans lesquels la connaissance incomplète peut être enrichie en incluant de nouveaux modes de fonctionnement au fur et à mesure qu'ils apparaissent. Cela se traduit par la prise en compte possible du nombre de classes au niveau de la règle de décision. Les options de rejet pour ces règles de décision seront prises en compte pour résoudre ce problème.

Son intégration au sein d'un système de diagnostic par RdF permet de diminuer l'erreur de classification en utilisant deux types de rejet : le rejet en ambiguïté et le rejet en distance. Le rejet en ambiguïté, introduit par Chow (Chow, 1957), affecte une observation, se situant approximativement à égale distance entre plusieurs classes, dans une nouvelle classe appelée la classe de rejet en ambiguïté ω_0 . Cette classe rassemble les observations correspondant au cas d'indécision.

Lorsqu'une observation à classer se trouve dans une zone de l'espace éloignée de toute classe, elle n'est représentée par aucune classe connue. Pour cela, Dubuisson (Dubuisson, 1990) introduit une règle de rejet en distance. Les points situés loin des classes connues sont affectés dans une classe appelée la classe de rejet en distance ω_d .

La notion de rejet est la base de toute procédure de décision adaptative. Les rejets en ambiguïté peuvent amener à fusionner deux classes et les rejets en distance peuvent engendrer la création de nouvelles classes sous le contrôle du module d'adaptation.

5.2.2 Logique floue

L'introduction des ensembles flous dans le cadre de la RdF s'avère être un outil intéressant pour le suivi de l'évolution, ou de la transition entre modes de fonctionnement qui est d'importance pour le diagnostic de phénomènes évolutifs. Plusieurs cas ont été déjà envisagés, nous citons l'utilisation conjointe des RN et des règles floues pour le diagnostic de machines tournantes

(Baghli,1999) et le diagnostic des systèmes évolutifs en utilisant la Reconnaissance Floue des Formes (Mouchaweh, 2002). Nous pouvons imaginer plusieurs autres combinaisons.

5.2.3 Réseau de neurones dynamique

La surveillance dynamique et la prédiction d'une variable pour la maintenance prédictive nécessitent la prise en compte par le réseau de neurones de la dimension temporelle (Zemouri, 2003). Cette prise en compte du passé du signal n'est possible que par des architectures de réseaux de neurones dites temporelles. Les réseaux de neurones statiques (appliqués dans cette thèse) sont incapables d'assurer ce genre de traitement.

Une détection précoce d'un palier de dégradation ou même l'apprentissage d'une séquence d'évolution d'un système à événements discrets nécessite par conséquent, une architecture neuronale temporelle (dynamique). C'est dans ce contexte que de futurs travaux de recherches peuvent être envisagés

5.2.4 autres algorithmes d'optimisation

Nous comptons pour des travaux ultérieurs appliquer et comparer d'autres algorithmes d'optimisation et d'extraction de connaissances telles que (Jourdan, 2003) : Le recuit simulé, la recherche tabou, la recherche par dispersion, les colonies de fourmis, l'essaim de particules, les systèmes immunitaires artificiels et les algorithmes à estimation de distribution.

5.2.5 Autres méthodes hybrides pour la reconnaissance automatique des défauts dans les machines tournantes

L'hybridation de plusieurs méthodes afin d'exploiter les avantages de chacune, semble constituer une approche intéressante pour la robustesse des systèmes de reconnaissance automatique. La combinaison la plus courante de nos jours est l'hybridation des modèles de Markov cachés (HMM) et des réseaux connexionnistes (ou Neural Networks (NN)) (Haton, 1997). La principale raison est que ces deux systèmes présentent des avantages complémentaires. Les HMM ont une grande capacité à traiter les événements dans le temps, les réseaux de neurones sont experts dans la classification de formes statiques.

5.2.6 Autres

Application des 3 méthodes de reconnaissance pour une même base de données.

Références

- (Abe et al., 2006) N. Abe et M. Kudo, Non-parametric classifier-independent feature selection, *Pattern Recognition Vol. 39*, pp. 737 – 746, 2006
- (Addison et al., 2002) P.S. Addison, J.N. Watson et T. Feng, Low-oscillation complex wavelets, *Journal of Sound and Vibration* , 254, 733–762. (2002)
- (Afnor, 1984) AFNOR(1984) : Recueil de normes française des Chocs et vibrations. Tome 1, 2^{ème} édition 1984.
- (Alattass, 1994) M. Alattass, " Maintenance des machines tournantes: signature de défauts d'engrenages droits et hélicoïdaux," Thèse de l' I.N.S.A de Lyon, 1994, pp 198.
- (Almeida et al., 2002) R.G.T. De Almeida, S.A. Da Silva Vicente et L.R. Padovese, New technique for evaluation of global vibration levels in rolling bearings, *Shock and Vibration* 9 (2002) 225–234.
- (Andrade et al., 1999) F.A. Andrade, I. Esat et M.N.M. Badi, Gearbox fault detection using statistical methods, time–frequency methods (STFT and Wigner–Ville distribution) and harmonic wavelet—A comparative study, in: *Proceedings of the COMADEM '99, Chipping Norton, 1999*, pp. 77–85.
- (Angeli, 1995) C. Angeli and A. Chatzinikolaou. An expert system approach to fault diagnosis in hydraulic systems. *Expert Systems*, 12(4), 323-330, 1995.
- (Antoni et al. 2003) J. Antoni et R. Randall. A stochastic model for simulation and diagnostics of rolling element bearings with localized faults. *ASME Journal* , 2003
- (Antoni et al., 2002) J. Antoni et R. Randall Differential diagnosis of gear and bearing faults. *ASME Journal of Vibration and Acoustics*, 124(2), 165-171, 2002
- (Antoni, 2000) Antoni J. « *Apport de l'échantillonnage angulaire et de la cyclostationnarité au diagnostic par analyse vibratoire des moteurs thermiques* », Thèse de l'Institut National Polytechnique de Grenoble, 21 Février 2000, pp 229.
- (Antoni, 2007) J. Antoni Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions, *Journal of Sound and Vibration* 304 497–529 499, 2007
- (Aretakis et al., 1997) N. Aretakis et K. Mathioudakis, Wavelet analysis for gas turbine faultdiagnostics, *Journal of Engineering for Gas Turbines and Power* 119 870–876. (1997)
- (Arquès, 1996) P. Arquès « *Diagnostic prédictif de l'état des machines* », Edition Masson, Paris 1996, pp.269.
- (Arthur et al., 1997) N. Arthur et J. Penman, Inverter fed induction machine condition monitoring using the bispectrum, in: *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, Banff, Alta., Canada, , pp. 67–71. 1997*
- (Ayat, 2004) Ayat N. «Sélection de modèle automatique des machines à vecteurs de support: application à la reconnaissance d'images de chiffres manuscrits», thèse de doctorat, Montréal, le 20 janvier 2004.
- (Bachir, 2002) S. Bachir, Contribution au Diagnostic de la machine asynchrone par estimation paramétrique, Thèse de l' Université de Poitiers, France, 2002.
- (Baillie et al., 1996) D.C. Baillie et J. Mathew, A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings, *Mechanical Systems and Signal Processing* 10 (1996) 1–17.
- (Bardou, 1994) O. Bardou, " Sur des méthodes de surveillance et de diagnostic vibratoire de machines alternatives," Thèse de l' I. N. P. de Grenoble, 1994, pp 114.
- (Bautista et al., 1999) Martin-Bautista M. J. & Vila M.-A, «A survey of Genetic Feature Selection in Mining Issues». *Evolutionary Computation, CEC, Proceedings of the 1999 Congress on*, vol. 2, 1999, p. 1314-1321.

- (**Baydar et al., 2001a**) N. Baydar et A. Ball, A comparative study of acoustic and vibration signals in detection of gear failures using Wigner–Ville distribution, *Mechanical Systems and Signal Processing* 15 (2001) 1091–1107.
- (**Baydar et al., 2001b**) N. Baydar, Q. Chen, A. Ball et U. Kruger, Detection of incipient tooth defect in helical gears using multivariate statistics, *Mechanical Systems and Signal Processing* 15 (2001) 303–321.
- (**Baydar et al., 2003**) N. Baydar et A. Ball, Detection of gear failures via vibration and acoustic signals using wavelet transform, *Mechanical Systems and Signal Processing* 17 (2003) 787–804.
- (**Benahmed, 2002**) N. Benahmed, Optimisation de réseaux de neurones pour la reconnaissance de chiffres manuscrits isolés : sélection et pondération des primitives par algorithmes génétiques, Thèse de doctorat, Université de Québec, 2002
- (**Benoughram, 1999**) R. Benouaghran « *Contribution au diagnostic des machines tournantes complexes. Application à un laminoir* », Thèse de l’UTC de Compiègne, Avril 1999, pp.128.
- (**Bessel, 2005**) P. Bessel « *Data mining II. Modélisation Statistique & Apprentissage* » publications du laboratoire de statistique et probabilités, Université Paul Sabatier, Toulouse, Mars 2005.
- (**Bhanu et al., 2003**) B. Bhanu et Y. Lin «Genetic algorithm based feature selection for target detection in SAR images», *Image and Vision Computing*, vol. 21, 2003, p. 591–608
- (**Bishop, 1995**) C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK (1995).
- (**Blankenship et al., 1995**) G.W. Blankenship, R. Singh, Analytical solution for modulation sidebands associated with a class of mechanical oscillators, *Journal of Sound and Vibration* 179 (1995) 13–36.
- (**Bonato et al., 1997**) P. Bonato, R. Ceravolo, A. De Stefano, M. Knaflitz, Bilinear time–frequency transformations in the analysis of damaged structures, *Mechanical Systems and Signal Processing* 11 (1997) 509–527.
- (**Bonnardot, 2003**) F. Bonnardot « *Comparaison entre les analyses angulaire et temporelle des signaux vibratoires de machines tournantes. Etude du concept de cyclostationnarité floue* », thèse de doctorat .LAPSI, université Jean Monnet, Saint-Etienne janvier 2005
- (**Boudaoud, 1997**) N. Boudaouad, *Conception d’un système de diagnostic adaptatif en ligne pour la surveillance des systèmes évolutifs*, Thèse présentée devant l’Université de Technologie de Compiègne, France, 1997.
- (**Bouillaut et al. 2001**) L. Bouillaut, M. Sidahmed., « *Helicopter gearbox vibrations: Cyclostationary analysis or bilinear approach?* », International Symposium on Signal Processing and its application (ISSPA), Malaysia, 13-16 August 2001, IEEE 2001, pp 4.
- (**Bouillaut, 1999**) L. Bouillaut et M. Sidahmed, Approche cyclostationnaire et bilinéaire des signaux vibratoires d’engrenage, *Revue française de mécanique*, n°. 2, 1999, pp.87-93.
- (**Bouillaut, 2000**) L. Bouillaut, « *Approches cyclostationnaire et non-linéaire pour l’analyse vibratoire de machines tournantes : Aspect théoriques et applications au diagnostic* », Thèse de l’UTC de Compiègne, Novembre 2000, pp.183.
- (**Boulahbal et al., 1999**) D. Boulahbal, F. Golnaraghi, F. Ismail, Amplitude and phase wavelet maps for the detection of cracks in geared systems, *Mechanical System and Signal Processing* 13 (1999) 423–436.
- (**Boulenger et al.,1998**) A. Boulenger et C. Pachaud, *Surveillance des machines par analyse des vibrations : du dépistage au diagnostic*. 2^{ème} tirage Ed. AFNOR 1998. (bibliothèque cote S8/51294).
- (**Boulenger et al.,2003**) A. Boulenger, C. Pachaud, *Analyse vibratoire en maintenance. Surveillance et diagnostic des machines* (2ème édition,Dunod, 2003).

- (**Boumaza, 1999**) R. Boumaza, " Analyses factorielles des distributions marginales de processus, " Thèse de l'Université Joseph Fourier Grenoble, 1999, pp 128.
- (**Bourgain, 1988**) L. Bourgain, R. Dart et J. Bourgain. Machines tournantes et circuits pulsés : applications industrielles et médicales de l'analyse spectrale. Ed Dunod 1988
- (**Boustany, 2005**) R. Boustany, J. Antoni, A subspace method for the blind extraction of a cyclostationary source: Application to rolling element bearing diagnostics, *Mechanical System and Signal Processing* 19 (2005) 1245–1259.
- (**Braun, 1986**) S.G. Braun « Mechanical signature analysis ». Orlando, Florida, Academic press.
- (**Breneur, 2002**) C. Breneur, Eléments de maintenance préventive de machines tournantes dans le cas de défauts combinés d'engrenages et de roulements0 Thèse de L'école Doctorat de INSA, Lyon, France, 2002.
- (**Burge, 1998**) C. A Burge, Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, 2, 121–167 (1998)
- (**Canetto et al., 2001**) P. Canetto et C. Pachaud, Acceptante tests : a necessary step to be included in vibration monitoring and diagnosis, 4^{ième} Conférence Internationale des Méthodes de Surveillance et techniques de diagnostic acoustiques et vibratoires, compiégnne-France, 16-18 Octobre 2001, Vol. 2, pp. 707-714
- (**Capdessus et al., 1991**) Capdessus C., Sidahmed M., « *Analyse des vibrations d'un engrenage : cepstre, corrélation, spectre* », *Traitement de signal*, Volume 8, N° 5, 1991, pp.365-372.
- (**Capdessus et al., 1995**) Capdessus C., Sidahmed M., Lacoume J.L., « *Apport de la théorie des processus cyclostationnaires à l'analyse et au diagnostic des engrenages* », 2ème Conférence Internationale des Méthode de Surveillance, Senlis-France, 10-12 Octobre 1995, Vol. 1, pp. 391-401.
- (**Capdessus, 1992**) Capdessus C., « *Aide au diagnostic des machines tournantes par traitement du signal* », Thèse de Institut National Polytechnique de Grenoble, 1992, pp 130
- (**Casimir et al., 2006**) R. Casimir, E. Boutleux, G. Clerc et A. Yahoui "The use of features selection and nearest neighbors rule for faults diagnostic in induction motors," *Engineering Applications of Artificial Intelligence* 19, 2006, pp 169-177
- (**Casimir, 2003**) R. Casimir, « Diagnostic des défauts des machines asynchrones par reconnaissance des formes », Thèse de L'école Doctorale de Lyon, 2003, pp 172.
- (**Chandroth et al., 1999**) G.O. Chandroth, W.J. Staszewski, Fault detection in internal combustion engines using wavelet analysis, in: *Proceedings of the COMADEM '99*, Chipping Norton, 1999, pp. 7–15.
- (**Chandroth et al., 1999**) G.O. Chandroth, W.J. Staszewski, Fault detection in internal combustion engines using wavelet analysis, in: *Proceedings of the COMADEM '99*, Chipping Norton, 1999, pp. 7–15.
- (**Chen et al., 1998**) Y.D. Chen, R. Du, Diagnosing spindle defects using 4-D holospectrum, *Journal of Vibration and Control* 4 (1998) 717–732.
- (**Chen et al., 2003**) Chen Z., He Y, Chu F., Huang J., «Evolutionary strategy for classification problems and its application in fault diagnostics», *Engineering Applications of Artificial Intelligence*, vol. 16, 2003, p. 31–38.
- (**Chiollaz et al., 1987**) Chiollaz M., Flandrin P., Gache N., « *Utilisation de la représentation de Wigner-Ville comme outil de diagnostic des défauts de fonctionnement des moteurs thermiques* », 11ème Colloque GRETSI, Nice 1987, pp 579-582.
- (**Chiollaz et al., 1992**) Chiollaz M., Favre B., « *Caractérisation fine de bruits moteur par analyse temps-fréquence de Wigner-Ville* », *Traitement du Signal*, 1992, Vol. 8 n°5, pp.301-318.

- (Chow et al., 1995) T.W.S. Chow, G. Fei, Three phase induction machines asymmetrical faults identification using bispectrum, *IEEE Transactions on Energy Conversion* 10 (1995) 688–693.
- (Chow, 1957) CHOW C. K., An optimum character recognition system using decision, *R.E. Transaction Electronic Computer* (1957) 247-254.
- (Ciupke, 2005) Ciupke K., «A comparative study on methods of reduction and selection of information in technical diagnostic», *Mechanical Systems and Signal Processing*, vol. 19, n°5, 2005, p. 919-938.
- (Codrea Codrea C.M. et al., «Feature learning with a genetic algorithm for fluorescence fingerprinting of plant species», *Pattern Recognition Letters*, vol. 24, 2003, p. 2663–2673.
- (Cohen, 1989) L. Cohen, Time–frequency distribution—A review, *Proceedings of the IEEE* 77 (1989) 941–981.
- (Cornuéjol et al., 2002) Cornuéjol.A, Miclet.L «*Apprentissage artificiel ,concept et algorithmes*», Edition OYROLLES,2002.
- (Dalpiaz et al., 1997) G. Dalpiaz et A. Rivola, Condition monitoring and diagnostics in automatic machines: Comparison of vibration analysis techniques, *Mechanical Systems and Signal Processing* 11 (1997) 53–73.
- (Dalpiaz et al.2000) G. Dalpiaz, A. Rivola et R. Rubini, Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears, *Mechanical Systems and Signal Processing* 14 (2000) 387–412.
- (Dash et al., 1997) M. Dash et H. Liu, Feature Selection for Classification. *Intelligent Data Analysis 1*. 131-156.
- (Daubechies, 1992) I. Daubechies, "Ten lectures on wavelets", Philadelphia, society for industrial and applied Mathematics (SIAM) 1992.
- (Davies 1998) A. Davies, *Handbook of Condition Monitoring—Techniques and Methodology*, Chapman & Hall, UK, 1998.
- (Dron et al., 1998) J.P. Dron, L. Rasolofondraibe, C. Couet et A. Pavan, Fault detection and monitoring of a ball bearing benchtest and a production machine via autoregressive spectrum analysis, *Journal of Sound and Vibration* 218 (1998) 501–525.
- (Dubuisson, 1990) B. Dubuisson, « *Diagnostic et reconnaissance des formes* ». *Traité des nouvelles technologies. Série diagnostic et Maintenance*. Editions Hermès, Paris 1990, pp 320.
- (Dubuisson, 2001) B. Dubuisson, « *diagnostic, intelligence artificielle, et reconnaissance des formes* » Hermès, Paris, 2001.
- (Duda et al., 2001) R. O. Duda, P.E.Hart, D.G.Stork: *Pattern Classification*, 2nd edition, NewYork: wiley Interscience. (2001).
- (Duda et al.,1973) R.O. Duda et P.E. Hart, *Pattern classification and scene analysis*, John Wiley & Sons, New York, 1973.
- (Eiben, 1999) Eiben A.E., Hinterding R., & Michalewicz Z., « Parameter Control in Evolutionary Algorithms», *Evolutionary Computation, IEEE Transactions on*, vol. 3, n°2, 1999, p. 124-141.
- (Eisenmann et al., 1997) R.C. Eisenmann, Sr. and R.C. Eisenmann, Jr., *Machinery Malfunction Diagnosis and Correction: Vibration Analysis and Troubleshooting for the Process Industries* , Hewlett- Packard Professional books, Prentice-Hall, Upper Saddle River, NJ (1997).
- (Elbadaoui et al.1999) Elbadaoui M, Guillet F., Danière J., Surveillance des systèmes complexes à engrenages par analyse cepstrale du signal, *traitement du signal*, vol.16, n°.5, 1999, pp.37-391.
- (Elbadaoui, 1999a) M. Elbadaoui, « *Contribution au diagnostic vibratoire des réducteurs complexes à engrenages par l'analyse cepstrale* », Thèse de l'Université Jean Monnet, 9 Juillet 1999, pp 141.

- (**Felkaoui et al. 1994**) A. Felkaoui, B.Fortas, A.Apostoliouk: Sur la sélection de l'ordre dans l'analyse spectrale moderne des processus linéaires, ICSS'94 (1994). Alger
- (**Felkaoui, 1995**) Felkaoui A., Contribution à l'identification des signaux et des systèmes dynamiques linéaires, Thèse de magister, Université de Sétif, Algérie, 1995.
- (**Flandrin, 1993**) Flandrin P., « *Temps-fréquence* », Traité des Nouvelles Technologies, Coll. Hermes 1993.
- (**Fukunaga , 1990**) K. Fukunaga, Introduction to Statistical Pattern Recognition (Second Edition), Academic Press, New York, 1990.
- (**Gardner et al., 1994**) Gardner W., The cumulant theory of cyclostationary times series, Part II : development and applications, IEEE transactions on Signal Processing, Vol.42, n°12, , pp.3409-3429, 1994.
- (**Gardner, 1986**) Gardner W., Introduction to random process with applications to signals and
- (**Garga et al., 1997**) A.K. Garga, B.T. Elverson, D.C. Lang, AR modeling with dimension reduction for machinery fault classification, in: Critical Link: Diagnosis to Prognosis, Haymarket, 1997, pp. 299–308.
- (**Gelman et al., 2000**) Gelman L. M., Kripak D.A., Fedorov V.V. and Udovenko L.N., « *condition monitoring diagnosis methods of helicopter units* », Mechanical system and signal processing (2000), pp 613-624.
- (**Gemmel et al., 1994**) B. D. Gemmel, J. R. McDonald, R. W. Stewart, R. N. T. Brooke, and B. J. Weir. A consultative expert system for fault diagnosis of turbinegenerator plant. Proceedings of the Institution of Mechanical Engineers: Part A,208:257{266, 1994.
- (**Goldman, 1999**) S. Goldman, Vibration Spectrum Analysis: A Practical Approach, Industrial Press, New York, 1999.
- (**Golub et al., 1996**) G. H. Golub et C. F. Van Loan, *Matrix Computations*, Third Edition, Johns Hopkins University Press, Baltimore, Maryland, USA,1996
- (**Gu et al., 2002**) S. Gu, J. Ni, J. Yuan, Non-stationary signal analysis and transient machining process condition monitoring, International Journal of Machine Tools and Manufacture 42 (2002) 41–51.
- (**Guo et al., 2005**) H.Guo, L. B. Jack, and A. K Nandi, « Feature Generation Using Genetic Programming With Application to Fault Classification», *IEEE Transactions On Systems, Man, and Cybernetics—PART B: Cybernetics*, Vol. 35, n° 1, 2005, p. 89-99.
- (**Hagan et al., 1996**) M.T. Hagan, H.B. Demuth, M.H. Beale, Neural Network Design, Chapters 11 and 12, PWS Publishing Company, Boston, MA 1996.
- (**Hallinan, 2001**) J. Hallinan, The Practical Handbook of Genetic Algorithms Applications : Chapter 5 Feature Selection and Classification in the Diagnosis of Cervical Cancer, 2nd Edition, Edited by Lance D. Chambers, Chapman & Hall/CRC, Boca Raton, Florida, USA, 2001.
- (**Hardle, 2003**) W. Hardle ET L. Simar, Applied Multivariate Statistical Analysis, Berlin, 2003
- (**Harris et al., 2002**) C.M. Harris, A.G. Piersol, Harris' Shock and Vibration Handbook, McGraw-Hill, New York, 2002.
- (**Haupt et al., 2004**) Haupt R.L. et Haupt S. E. Practical Genetic Algorithms, , (2nd edition). John Wiley & Sons, Inc Publication, 2004.
- (**Hayes, 1996**) M.H. Hayes, Statistical Digital Signal Processing and Modeling, Wiley, New York, 1996.
- (**Haykin 1994**) S. Haykin, Neural Networks, a Comprehensive Foundation. (Macmillan, New York, NY., 1994)
- (**Ho et al., 1997**) D. Ho, R.B. Randall, Effects of time delay, order of FIR filter and convergence factor on SANC, Fifth International Congress on Sound and Vibration, Adelaide, 1997.

- (**Ho et al., 1998**) D. Ho, R.B. Randall, Improving the efficiency of SANC in its application to bearing diagnostics, Comadem'98, Launceston, AU, 1998.
- (**Ho et al., 2000**) D. Ho, R.B. Randall, Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals, *Mechanical Systems and Signal Processing*, vol. n° (2000) pp.
- (**Hong et al., 1991**) Q.Y. Hong et S.A.Kwong, «Genetic classification method for speaker recognition Engineering», *Applications of Artificial Intelligence*, vol. 18, 2005, p. 13–19.
- (**Hurd et al., 1991**) H.L Hurd et N.L. Gerr, “Graphical Methods for determining the presence of periodic correlation”, *Journal of time series analysis*, Vol.42, no 4, p.337-350, 1991.
- (**Jack et al., 2000a**) L.B. Jack, A.K. Nandi, “Comparison of neural networks and support vector machines in condition monitoring applications,” *Proceedings of 13th Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM)*, Houston, TX, USA, 2000, pp. 721–730.
- (**Jack et al., 2000b**) L.B. Jack, A.K. Nandi, “Support vector machines for detection and characterisation of rolling element bearing faults,” *Proceedings of Institution of Mechanical Engineers. Part C: Journal of Mechanical Engineering Science* 215 (2000) 1065–1074.
- (**Jack et al., 2000c**) L.B. Jack, A.K. Nandi, “Genetic algorithms for feature extraction in machine condition monitoring with vibration signals,” *IEE Proceedings of Vision & Image Signal Processing* 147 (2000) 205–212.
- (**Jack et al., 2002**) L.B. Jack, A.K. Nandi, “Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms,” *Mechanical Systems and Signal Processing* 16 (2002) 373–390.
- (**Jack, 2000**) L.B.Jack: Application of artificial intelligence in machine condition monitoring, thèse de doctorat, DEEAE, université de Liverpool (2000).
- (**Jacobs, 1988**) R. A. Jacobs, Increased rates of convergence through learning rate adaption *Neural Networks* 1:295-307, 1988.
- (**Jourdan, 2003**) L. Jourdan Métaheuristiques pour l'extraction de connaissances : Application à la génomique, thèse de Doctorat DE L'U.S.T.L. Lille présentée et soutenue le 26 Novembre 2003.
- (**Key et al, 1981**) M.K. Key et JR S. L Marple., *Spectrum Analysis -A modern perspective*. *Proceeding of the IEEE*, Vol. 69, No. 11, pp.1380-1419 (November 1981).
- (**Kharoubi, 2002**) Kharoubi J. « *Etude de Techniques de Classement - Machines à Vecteurs Supports - pour la Vérification Automatique du Locuteur* »Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 2002.
- (**Kiral et al.2006**) Z. Kiral et H. Karagülle (2006), Vibration analysis of rolling element bearings with various defects under the action of an unbalanced force, *Mechanical Systems and Signal Processing* 20 (2006) 1967–1991
- (**Kirkham et al., 1999**) C. Kirkham, A. Long, O. Taylor, and C. Isbell. Adaptive online systems for condition monitoring: The NEURAL-MAINE project. In *Proceedings of COMADEM '99*, pp. 317{325, Kingham, Oxford, 1999. Coxmoor Publishing Ltd.
- (**Kohavi, 1995**) R.Kohavi “A study of cross-validation and bootstrap for accuracy estimation and model selection”. In: Cs Mellosh (ed.) *Proceedings of the 14th International Joint Conf of Artificial Intelligence*, Morgan Kaufmann publishers, Inc: 1137-1143, (1995)
- (**Koizumi et al. 2000**) T. Koizumi et N. Tsujiuchi, Y. Matsumura, Diagnosis with the correlation integral in time domain, *Mechanical Systems and Signal Processing* 14 (2000) 1003–1010.
- (**Koo et al., 2000**) I.S. Koo, W.W. Kim, The development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant, *ISA Transactions* 39 (2000) 309–316.

- (**Krishnappa,1999**) G. Krishnappa et M. Donovan, Railway bearing diagnostic: Laboratory Data Analysis, report TP. Prepared for Transportation Development CentreTransport by National Research Council Canada Vancouver, Canada, September 1999 (53 pages)
- (**Kudo et al., 2000**) M. Kudo, J. Sklansky, Comparison of algorithms that select features for pattern classifiers, *Pattern Recognition, Vol. 33*, pp. 25-41, 2000.
- (**Lacoume & al. 97**) J.-L. Lacoume, P.-O. Amblard and P. Comon, “ Statistiques d’ordre supérieur pour le traitement du signal”, édition Masson, France, Oct.1997.
- (**Lacoume, 2001**) Lacoume J.L., Les statistiques d’ordre supérieur pour la surveillance acoustique et vibratoire. 4^{ième} Conférence Internationale des Méthodes de Surveillance et techniques de diagnostic acoustiques et vibratoires, compiègne-France, 16-18 Octobre 2001, Vol. 1, pp. 49-72
- (**Lau et al., 2000**) Soon-Seng Lau, Peter J. Sherman, « *The influence of period variation on Time-Frequency analysis of the Westland helicopter data* », Mechanical system and signal processing (2000) 14(4). pp. 571-677.
- (**Laurentiu, 2004**) Laurentiu T. A., Neural networks in multiphase reactors data mining : feature selection, prior knowledge, and model design, Thèse Ph.D., Faculté des études supérieures, Université Laval, QUÉBEC, 2004.
- (**Li et al., 2003**) W. Li, G. Zhang, T. Shi et S. Yang, Gear crack early diagnosis using bispectrum diagonal slice, Chinese Journal of Mechanical Engineering 16 (2003) 193–196 (English Edition).
- (**Lin et al., 2004**) J. Lin, M.J. Zuo et K.R. Fyfe, Mechanical fault detection based on the wavelet denoising technique, Journal of Vibration and Acoustics 126 (2004) 9–16.
- (**Liu et al, 1996**) Liu T. I., Singonahalli J. H. et Iyer R. N., « Detection of roller bearing defects using expert system and fuzzy logic», *Mechanical Systems and Signal Processing*, vol. 10, n°5, 1996, p. 595-614.
- (**Liu et al., 1992**) T. I. Liu et J. M. Mengel. Intelligent monitoring of ball bearing conditions. Mechanical Systems and Signal Processing, 6(5):419{431, 1992.
- (**Liu et al., 2004**) Z. Liu, X. Yin, Z. Zhang, D. Chen et W. Chen, Online rotor mixed fault diagnosis way based on spectrum analysis of instantaneous power in squirrel cage induction motors, IEEE Transactions on Energy Conversion 19 (2004) 485–490.
- (**Loughlin et al., 2000**) P. Loughlin et F. Cakrak, « *Conditional moments analysis of transients with application to the helicopter fault data* », Mechanical system and signal processing (2000) 14(4), pp 515-522.
- (**Luo et al., 2003**) G.Y. Luo, D. Osypiw et M. Irle, On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing, Journal of vibration and Control 9 (2003) 931–947.
- (**MacKay, 2003**) MacKay David J.C., « *Information Theory, Inference, and Learning Algorithms*», Cambridge University Press 2003. pp 640.
- (**Malhi et al., 2004**) Malhi A., and Gao R. X., «PCA-Based Feature Selection Scheme for Machine Defect Classification», *IEEE Trans. On Instrumentation and measurement* , VOL. 53, n°6, 2004, p. 1517-1525.
- (**McCormick et al., 1997**) A.C. McCormick et A.K. Nandi, Real time classification of rotating shaft loading conditions using artificial neural networks, IEEE Transactions on Neural Networks 8 (3) (1997) 748–757.
- (**McCormick et al., 1999**) A.C. McCormick et A.K.Nandi (1999), Bispectral and trispectral features for machine condition diagnosis condition diagnosis, IEE Proceedings -Vision, Image and Signal Processing, - Vol.146, n.5, 1999.
- (**McCormick, 1998a**) A.C. McCormick, 'cyclostationarity in rotating machine vibration' Mechanical systems and signal processing (1998) 12(2).225-242

- (**McCormick, 1998b**) A. C. McCormick, “Cyclostationary and higher order statistical signal processing algorithms for machine condition monitoring” a dissertation for the degree of doctor of philosophy, University of Strathclyde, September 1998
- (**McFadden et al., 1986**) McFadden P.D., Smith J.D., Effect of Transmission Path on Measured Gear Vibration. *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, 1986, pp. 377-378.
- (**McFadden, 1986**) P.D. McFadden, Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration, *Transactions of the ASME Journal of Vibration, Acoustics, Stress, and Reliability in Design* 108 (1986) pp.165–170.
- (**Mechefske et al., 1992**) C.K. Mechefske et J. Mathew, Fault detection and diagnosis in low speed rolling element bearing. Part I: The use of parametric spectra, *Mechanical Systems and Signal Processing* 6 (1992) 297–307.
- (**Meltzer et al., 2004**) G. Meltzer et N.P. Dien, Fault diagnosis in gears operating under non-stationary rotational speed using polar wavelet amplitude maps, *Mechanical Systems and Signal Processing* 18 (2004) 985–992.
- (**Mendel, 1991**) J. Mendel, “Tutorial on higher-order statistics (spectra) in signal processing and system theory : theoretical results and some applications”, *Proceedings of the IEEE*, vol. 79, no. 3, March 1991, pp. 278-305
- (**Meng et al., 1991**) Q. Meng, L. Qu, Rotating machinery fault diagnosis using Wigner distribution, *Mechanical Systems and Signal Processing* 5 (1991) 155–166.
- (**Michalewicz, 1999**) Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. Springer Verlag, collection *Artificial Intelligence*, 3rd Edition, New York. 1999.
- (**Miller et al., 1999a**) A.J. Miller et K.M. Reichard, A new wavelet basis for automated fault diagnostics of gear teeth, in: *Inter-Noise 99: Proceedings of the 1999 International Congress on Noise Control Engineering*, vols. 1–3, Poughkeepsie, 1999, pp. 1597–1602.
- (**Miller, 1999b**) A.J. Miller, A new wavelet basis for the decomposition of gear motion error signals and its application to gearbox diagnostics, M.Sc. Thesis, Graduate Program in Acoustics, The Pennsylvania State University, State College, PA, 1999.
- (**Minnicino et al., 2004**) M.A. Minnicino et H.J. Sommer, Detecting and quantifying friction nonlinearity using the Hilbert transform, in: *Health Monitoring and Smart Nondestructive Evaluation of Structural and Biological System III*, vol. 5394, Bellingham, 2004, pp. 419–427.
- (**Mitchell, 1993**) J. S. Mitchell, *Introduction to Machinery Analysis and Monitoring*, PenWel Books, Tulsa (1993).
- (**Morel & al.1992**) J. Morel, *Vibrations des machines et diagnostic de leur état mécanique*. Ed . EYROLLES 1992.
- (**Mouchaweh, 2002**) M. S. Mouchawek, Conception d’un système de diagnostic adaptatif et prédictif basé sur la méthode Fuzzy Pattern Matching pour la surveillance en ligne des systèmes évolutifs Application à la supervision et au diagnostic d’une ligne de peinture au trempé. Thèse de Doctorat, Université de Technologie de Reims Champagne-Ardenne, France, 2002.
- (**Nandi et al., 1999**) Nandi S., Toliyat H., Condition monitoring and diagnosis of electrical machines- a review, *Proceedings of the IEEE -IAS Annual Meeting Conference*, 8 pages, 1999
- (**Nandi et al., 2005**) Nandi. A. K. Zhang, k, L. et L. B. Jack, Fault detection using genetic programming, *Mechanical Systems and Signal Processing* 19 (2005) 271–289.
- (**Nandi, 1999**) A. K. Nandi, “Modified self-organising map for automated novelty detection applied to vibration signal monitoring” *Mechanical Systems and Signal Processing*, accepted January 2005.

- (Oehlmann, 1995)** Oehlmann H., R.Brie, Y.Begotto, M.Tomziack, "Time frequency analysis of gear spalling in a gearbox", *2ème Conférence Internationale sur les Méthodes de Surveillance Acoustique et Vibratoire et Techniques de diagnostic*, Senlis, Octobre 1995
- (Osuna et al., 1997)** Osuna E, Freund R., Girosi, F., Improved training algorithm for support vector machines, Proceedings of the IEEE, Neural Networks for Signal Processing '97, 1997
- (Oukhellou, 1997)** L. Oukhellou, "Paramétrisation et Classification de Signaux en Contrôle Non Destructif. Application à la Reconnaissance des Défauts de Rails par Courants de Foucault," Thèse de l'Université de Paris-Sud, 1997, pp 211.
- (Pachaud et al., 2001)** C. Pachaud et P. Canetto, Machines diagnosis : which criteria for which indicators, 4^{ième} Conférence Internationale des Méthodes de Surveillance et techniques de diagnostic acoustiques et vibratoires, compiègne-France, 16-18 Octobre 2001, Vol. 2, pp. 539-552
- (Pan et al., 1998)** M.-C. Pan, H. Van Brussel, P. Sas et B. Verbeure, Fault diagnosis of joint backlash, Journal of Vibration and Acoustics, Transactions of the ASME 120 (1998) 13–24.
- (Parizeau, 2004)** Parizeau M., Réseaux de neurones, édition université Laval, automne 2004.
- (Parker et al., 2000)** B.E. Parker, H.A. Ware, D.P. Wipf, W.R. Tompkins, B.R. Clark, E.C. Larson et H.V. Poor, Fault diagnostics using statistical change detection in the bispectral domain, Mechanical Systems and Signal Processing 14 (2000) 561–570.
- (Pau, 1986)** L. F. Pau. Survey of expert systems for fault detection, test generation and maintenance. Expert Systems, 3(2): pp.100-111, 1986.
- (Pen et al., 2004)** Z.K. Peng et F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography, Mechanical Systems and Signal Processing 18 (2004) 199–221
- (Platt, 1998)** Platt.J.C «*Sequential Minimal Optimization:A Fast Algorithm for Training Support Vector Machines*» Technical Report MSR-TR-98-14 April 21, 1998.
- (Polycarpou et al., 1995)** P. Polycarpou et A. Helmicki. Automated fault detection and accommodation: A learning systems approach. IEEE Transactions on Systems, Man, and Cybernetics, 25(11):1447-1458, 1995.
- (Poyhonen, 2004)** S. Poyhonen, P. Jover et H. Hyotyniemi, Signal processing of vibrations for condition monitoring of an induction motor, in: ISCCSP: 2004 First International Symposium on Control, Communications and Signal Processing, New York, 2004, pp. 499–502.
- (Prieur et al., 1995)** P. Prieur et G. Durso, "Des indices de cyclostationnarite pour la surveillance des engrenages", *15ème colloque GRETSI sur le traitement du signal et des images*, Juan-Les-Pins, Septembre 1995, p.1241-1244.
- (Qu et al., 1998)** L. Qu, X. Liu, G. Peyronne et Y. Chen, The holospectrum: A new method for rotor surveillance and diagnosis, Mechanical Systems and Signal Processing 3 (1989) 255–267.
- (Raad, 2003)** A. Raad, Contributions aux statistiques cycliques d'ordre supérieur : applications au diagnostic des défauts d'engrenage, Thèse de doctorat, Université Technologie de Compiègne, (2003).
- (Ramesh, 1999)** Ramesh V.E., & Murty N., «Off-Lin signature Verification Using Genetically Optimized Weighted Features», *Pattern Recognition*. Vol. 32, n° 2, 1999, p.217-233.
- (Randall , 2004)** R.B. Randall, Detection and diagnosis of incipient bearing failure in helicopter gearboxes, Engineering Failure Analysis, Vol. 11 (2004) 177–190
- (Randall et al., 2001)** R.B. Randall, J. Antoni et S. Chobsaard, The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals, Mechanical Systems and Signal Processing 15 (2001) 945–962.
- (Randall, 1987)** R.B. Randall, Frequency Analysis, K. Larsen and Son, Denmark, 1987.

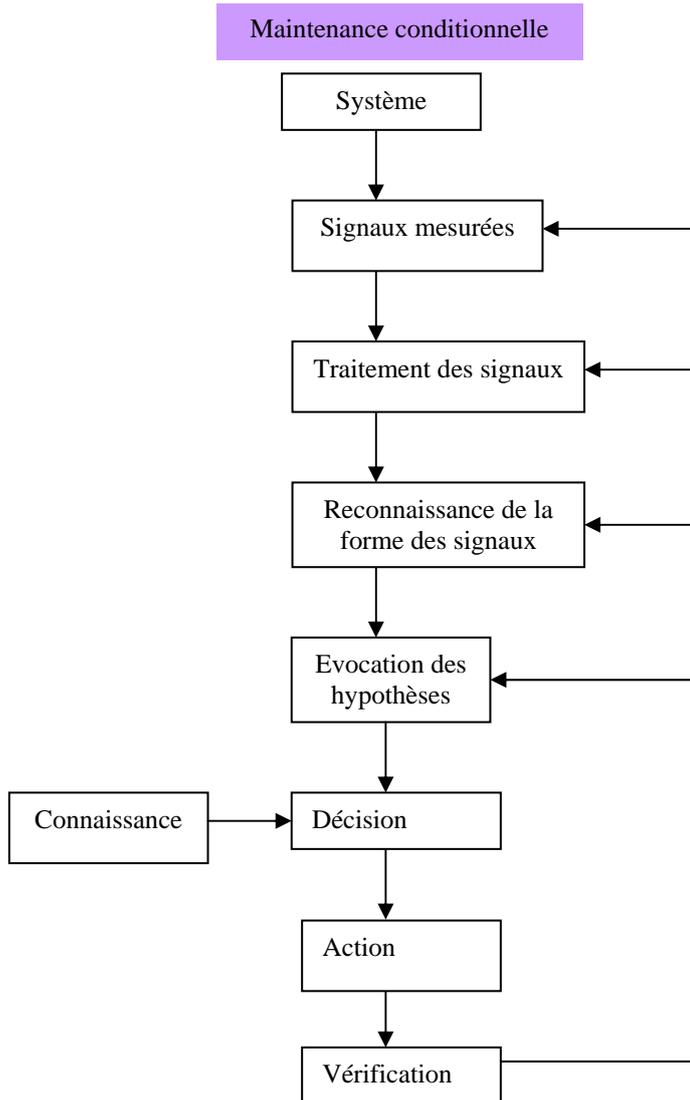
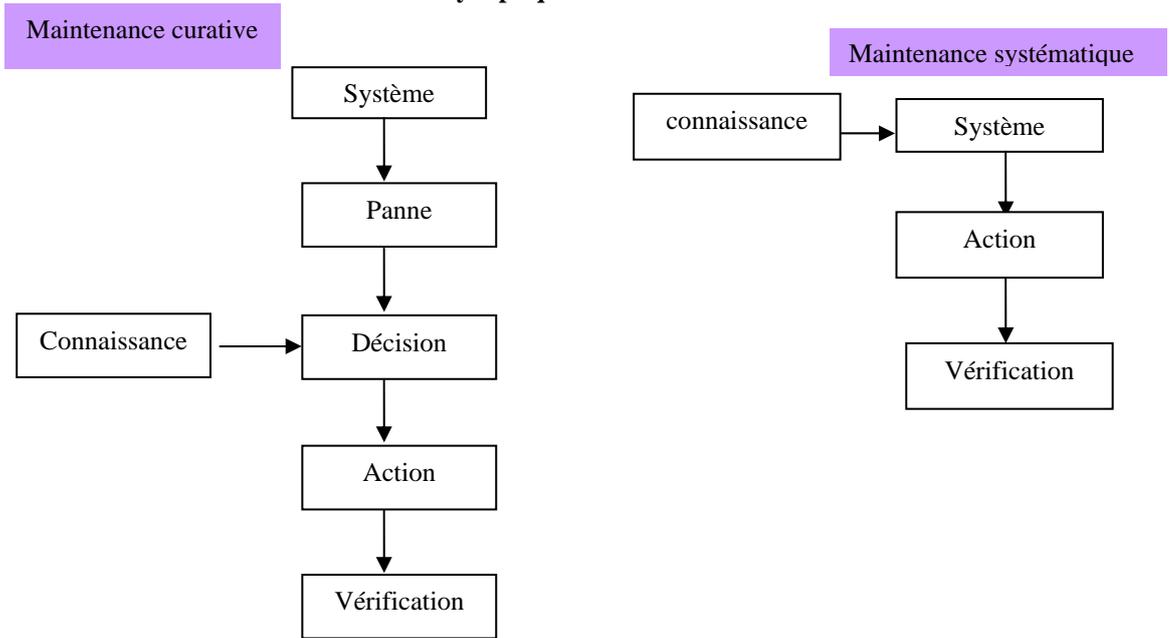
- (Rojas et al., 2006)** Rojas A. et Nandi A. K. (2006), “Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines .” *Mechanical Systems and Signal Processing* 20 (2006) 1523–15
- (Roy et al., 1990)** S. Roy and J. J. Shynk., Analysis of the momentum lms algorithm. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38: 2088-2098. 1990.
- (Rubini et al., 2001)** R. Rubini et U. Meneghetti, Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings, *Mechanical Systems and Signal Processing* 15 (2001) 287–302.
- (Salami et al., 2001)** M.J.E. Salami, A. Gani et T. Pervez, Machine condition monitoring and fault diagnosis using spectral analysis techniques, in: *Proceedings of the First International Conference on Mechatronics (ICOM '01)*, vol. 2, Kuala Lumpur, Malaysia, 2001, pp. 690–700.
- (Samanta et al., 2001)** B. Samanta, K.R. Al-Balushi et S.A. Al-Araimi, “Use of genetic algorithm and artificial neural network for gear condition diagnostics,” *Proceedings of COMADEM 2001*, University of Manchester, UK, 2001, pp. 449–456.
- (Samanta et al., 2003a)** B. Samanta et K. R. Al-Balushi “ Artificial Neural network based fault diagnostics of rolling element bearings using time domain features” . *Mechanical Systems and Signal Processing*, Volume 17, Issue 2, March 2003, Pages 317-328
- (Samanta et al., 2003b)** B. Samanta, K. R. Al-Balushi et S. A. Al-Araimi, “Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection” . *Engineering Applications of Artificial Intelligence*, Volume 16, Issues 7-8, October-December 2003, Pages 657-665
- (Samanta, 2003c)** B. Samanta, Artificial neural networks and genetic algorithms for gear fault detection, *Mechanical Systems and Signal Processing* 18 (2004) 1273–1282.
- (Samanta, 2004a)** B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing*, Volume 18, Issue 3, May 2004, Pages 625-644(2004)
- (Samanta, 2004b)** B. Samanta, “Artificial neural networks and genetic algorithms for gear fault detection”. *Mechanical Systems and Signal Processing*, Volume 18, Issue 5, September 2004, Pages 1273-128, (2004)
- (Sarkar, 1995)** Sarkar D., « méthodes to speed up error back propagation learning algorithm », *ACM computing surveys*, vol.27, n°4, p.519-542, 1995.
- (Schenck, 1994)** Schenck, *Vibrations Equilibrage sur Site : Application à la maintenance industrielle*. Ed. Schenck S.A. (1994).
- (Schoen et al., 1995)** R.R. Schoen, T.G. Habetler, Effects of time-varying loads on rotor fault detection in induction machines, *IEEE Transactions on Industry Applications* 31 (1995) 900–906.
- (Sidahmed et al., 1990)** M. Sidahmed et Y. Garnier “ surveillance et diagnostic de réducteurs a engrenages” le traitement de signal en mécanique Journées CETIM-GRECO TdSI sentis 7-8 mars 1990 pp 29-41.
- (Singh et al., 2003)** G. K. Singh, S. A. S. Al Kazzaz, Induction machine drive condition monitoring and diagnostic research : a survey, *Electric Power Systems Research*, Vol. 64, pp.145-158, 2003
- (Sorsa et al., 1991)** T. Sorsa, H. Koivo et H. Koivisto. Neural networks in process fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(4):815{ 825, 1991.
- (Stack et al., 2004)** J.R. Stack, R.G. Harley et T.G. Habetler, An amplitude modulation detector for fault diagnosis in rolling element bearings, *IEEE Transactions on Industrial Electronics* 51 (2004) 1097–1102.
- (Stack et al., 2004)** J.R. Stack, T.G. Habetler et R.G. Harley, Bearing fault detection via autoregressive stator current modeling, *IEEE Transactions on Industry Applications* 40 (2004) 740–747.

- (Starr et al., 1999) A. Starr, M. Desforges et J. Esteban. Strategies in data fusion - a condition monitoring approach. In Proceedings of COMADEM '99, University of Sunderland, pp. 399-408, Kingham, Oxford, 1999. Coxmoor Publishing.
- (Staszewski et al., 1994) W.J. Staszewski, G.R. Tomlinson, Application of the wavelet transform to fault detection in a spur gear, *Mechanical Systems and Signal Processing* 8 (1994) 289-307.
- (Swami et al., 1997) Swami A., Giannakis G.B., Zhou G., Bibliography on higher-order statistics, *Signal Processing*, 6 (1997) 65-126.
- (Swami, 1998) Swami A., Higher-Order Spectral Analysis Toolbox (HOSA), edited by United Signals & Systems, Inc. and The Mathworks, Inc., ver.2 Matlab 5.3, 1998. systems, McMillan, (1986),
- (Taylor, 1997) J. I. Taylor, Back to the Basics of Rotating Machinery Vibration Analysis, Vibration Consultants, Inc., Tampa Bay, FL (1994).
- (Théodoridis et al., 2003) S. Théodoridis et K. Koutroumbas, Pattern recognition, 2nd edition, Academic Press, Elsevier (USA). 2003
- (Thomas, 2003) Thomas M., 2003, *Fiabilité, maintenance prédictive et vibration de machines*, Ecole de Technologie Supérieure, Montréal, Québec, Canada, p.616
- (Thorsen et al., 1995), O.V. Thorsen, M. Dalva, A survey of faults on induction motors in offshore oil industry, petrochemical industry, gas terminal, and oil refineries. *IEEE Transaction on Industry Applications (IAS)*, Vol. 31, n.5, pp. 1186-1196, 1995.
- (Toliyat et al., 2003) H.A. Toliyat, K. Abbaszadeh, M.M. Rahimian et L.E. Olson, Rail defect diagnosis using wavelet packet decomposition, *IEEE Transactions on Industry Applications* 39 (2003) 1454-1461.
- (Tonshoff et al., 2003) H.K. Tonshoff, X. Li et C. Lapp, Application of fast Haar transform and concurrent learning to tool-breakage detection in milling, *IEEE/ASME Transactions on Mechatronics* 8 (2003) 414-417.
- (Torkkola, 2003) K. Torkkola «Feature Extraction by Non-Parametric Mutual Information Maximization», *Journal of Machine Learning Research*, vol. 3., 2003, p.1415-1438.
- (Van der Merwe et al., 2002) N.T. Van der Merwe et A.J. Hoffman, A modified cepstrum analysis applied to vibrational signals, in: Proceedings of 14th International Conference on Digital Signal Processing (DSP2002), vol. 2, Santorini, Greece, 2002, pp. 873-876.
- (Vapnik, 1998) V N. Vapnik, « *Statistical Learning Theory*»: Wiley Interscience publication, NewYork, 1998.
- (Vapnik, 2000) V. N. Vapnik, *The nature of statistical theory. Collection Statistical for Engineering and Information science*. 2nd edition, Springer 2000
- (Vojtech et al., 2004) F. Vojtech et Hlaváč V.«Statistical Pattern Recognition Toolbox for Matlab», CZECH TECHNICAL UNIVERSITY; juin 2004.
- (Walker et al. 1986) T. C. Walker et R. K. Miller. Expert Systems 1986. SEAI Technical Publications, Madison, GA, 1986.
- (Wang et al., 1993) W.J. Wang et P.D. McFadden, Early detection of gear failure by vibration analysis I. Calculation of the time-frequency distribution,
- (Wang et al., 1993) W.J.Wang, and P.D. McFadden (1993). Early detection of gear failure by vibration analysis : calculation of the time-frequency distribution. *Mechanical Systems and Signal Processing*, 7, 193-203.
- (Wang et al., 1996) W.J. Wang et P.D McFadden, Application of wavelets to gearbox vibration signals for fault detection, *Journal of Sound and Vibration* 192 (1996) 927-939.

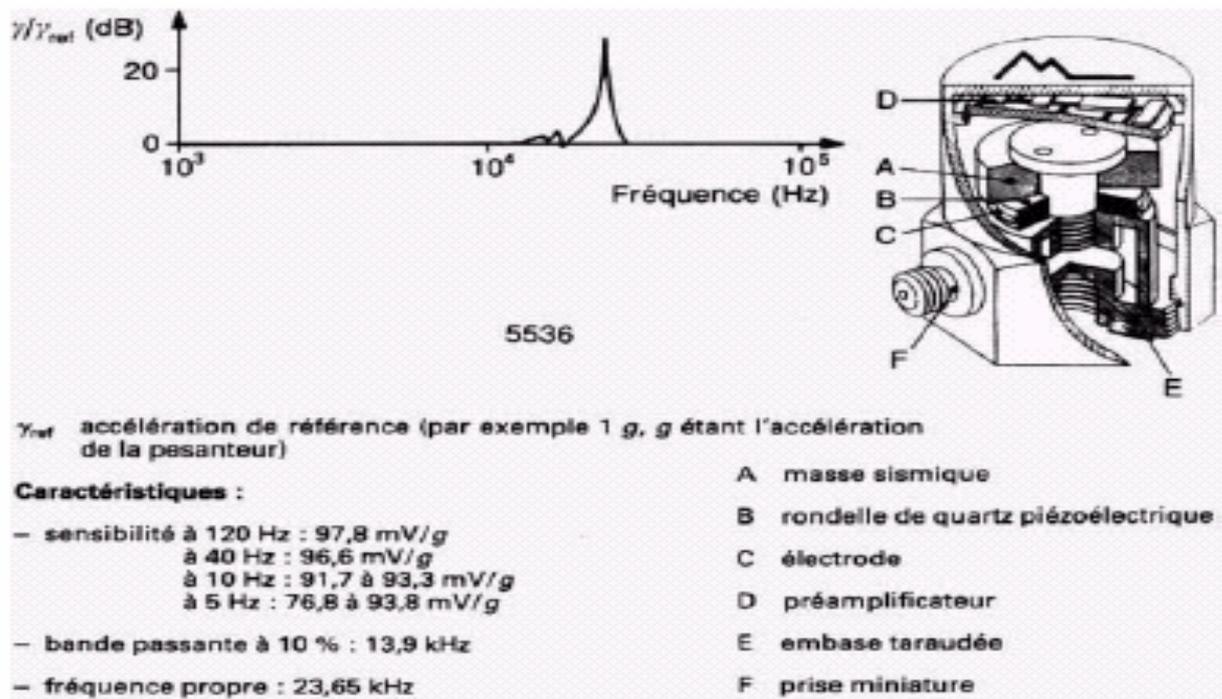
- (Wang et al., 2001a) W.J. Wang, J. Chen, X.K. Wu et Z.T. Wu, The application of some non-linear methods in rotating machinery fault diagnosis, *Mechanical Systems and Signal Processing* 15 (2001) 697–705.
- (Wang et al., 2001b) W.J. Wang, Z.T. Wu et J. Chen, Fault identification in rotating machinery using the correlation dimension and bispectra, *Nonlinear Dynamics* 25 (2001) 383–393.
- (Wang et al., 2002a) C.-C. Wang et G.-P.J. Too, Rotating machine fault detection based on HOS and artificial neural networks, *Journal of Intelligent Manufacturing* 13 (2002) 283–293.
- (Wang et al., 2002b) W. Wang et Wong A.K. , Autoregressive model-based gear fault diagnosis, *Trans. Of the ASME. J.Vib. & Acoustics*, Vol.124, Apr2002, pp.172-179.
- (Wang et al., 2003a) C.Wang et R.X. Gao, Wavelet transform with spectral post-processing for enhanced feature extraction, *IEEE Transactions on Instrumentation and Measurement* 52 (2003) 1296–1301.
- (Wang et al., 2003b) W.J. Wang et R.M. Lin, The application of pseudo-phase portrait in machine condition monitoring, *Journal of Sound and Vibration* 259 (2003) 1–16.
- (Wang, 2003) W. Wang, An evaluation of some emerging techniques for gear fault detection. *Structural Health Monitorig*, Vol. 2 n°. 3 pp.225-242, 2003 Sage publications.
- (Weiss et al., 1991) S. Weiss et C.A. Kulikowski Computer systems that learn, classification and prediction methods from Statistics, Neural Nets, Machine learning and Expert systems. San Mateo, California Morgan Kaufmann Publishers. (1991)
- (Williams et al.,2000) W.J. Williams et E.J. Zalubas. « *Helicopter transmission fault detection via Time-frequency, scale and spectral methods* ». *Mechanical system and signal processing* (2000) 14(4). pp 545-559.
- (Wong et al., 20005) M.L.D. Wong, L.B. Jack, and A.K. Nandi, « *Modified self-organising map for automated novelty detection applied to vibration signal monitoring* », *Mechanical Systems and Signal Processing*, accepted 25 January 2005.18 p.
- (Xiong et al., 2002) L. Xiong, T. Shi, S. Yang et R.B.K.N. Rao, A novel application of wavelet-based bispectrum analysis to diagnose faults in gears, *International Journal of COMADEM* 5, 2002,31–38.
- (Xu et al., 1991) Y.-G. Xu et Y.-L. Yan, Research on Haar spectrum in fault diagnosis of rotating machinery, *Applied Mathematics and Mechanics* 12 (1991) 61–66 (English Edition).
- (Yang et al., 1997) J.Yang et V. Honavar Feature Subset Selection Using A Genetic Algorithm. *Genetic Programming. A data mining perspective*. New York: Kluwer Academic. (1997).
- (Yang et al., 2002) D.-M. Yang, A.F. Stronach, P. Macconnell et J. Penman, Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks, *Mechanical Systems and Signal Processing* 16 (2002) 391–411.
- (Yang et al., 2005a) B.-S Yang et W.W. Hwang: Cavitation detection of butterfly valve using support vector machines, *Journal of sound and vibration* 287 (2005) 25–43.
- (Yang et al., 2005b) B.-S. Yang et W.W. Hwang: Condition classification of small reciprocating compressor for refrigerator using artificial neural network and support vector machines, *Mechanical Systems and Signal Processing* 19 (2005) 371–390.
- (Yang et al., 2005c) H. Yang, J. Mathew et L. Ma, Fault diagnosis of rolling element bearings using basis pursuit, *Mechanical Systems and Signal Processing* 19 (2005) 341–356.
- (Yangping et al., 2000) Z. Yangping, Z. Bingquan et W. DongXin. Application of genetic algorithms to fault diagnosis in nuclear power plants. *Reliability Engineering and Safety Systems*, 67:153{160, 2000.
- (Yann et al., 1997) Yann Guermeur et H P. Moisy « Théorie de l'apprentissage de Vapnik et SVM, Support Vector Machines » Article de l'école Normale Supérieure de Lyon. 1997.

- (Yen et al., 2000)** G.G. Yen et K.-C. Lin, Wavelet packet feature extraction for vibration monitoring, IEEE Transactions on Industrial Electronics 47 (2000) 650–667.
- (Young et al., 1993)** R.K. Young, Wavelets Theory and Its Applications, Kluwer Academic Publishers, Boston, 1993.
- (Ypma, 2001)** Ypma Alexander « *Learning methods for machine vibration analysis and health monitoring* », Thèse de l'université technique de Delft -pays bas, 12 Novembre 2001, pp221.
- (Yu et al., 1997)** C.B. Yu, H.B. He, Y. Xu et F.L. Chen, Identification method of acoustic information flow of bearing state, in: Condition Monitoring '97, 1997, pp. 311–315.
- (Zacksenhouse et al., 2000)** M. Zacksenhouse, S. Braun, M. Feldman et M. Sidahmed, « *Toward helicopter gearbox diagnostics from a small number of examples* », Mechanical Systems and Signal Processing (2000), pp 21.
- (Zadeh, 1987)** L.A. Zadeh, "Fuzzy Sets and Applications: Selected Papers ", ed. R.R. Yager et al. (John Wiley, New York, 1987).
- (Zemouri, 2003)** Zemouri M. R. « *Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques : Application à la e-maintenance* », Thèse de l'Université de Franche-Comté, 28 Novembre 2003, pp 277.
- (Zhan et al., 2003)** Y. Zhan, V. Makis et A.K.S. Jardine, Adaptive model for vibration monitoring of rotating machinery subject to random deterioration, Journal of Quality in Maintenance Engineering 9 (2003) 351–375.
- (Zhan et al., 2006)** Y. Zhan, V. Makis et A.K. Jardine, Adaptive state detection of gearboxes under varying load conditions based on parametric modelling . Mechanical Systems and Signal Processing 20 (2006) 188-221.
- (Zhang et al., 2005)** S. Zhang, J. Mathew, L. Ma et Y. Sun, Best basis-based intelligent machine fault diagnosis, Mechanical Systems and Signal Processing 19 (2005) 357–370.
- (Zhang, 2005)** L. Zhang, L. B. Jack et A. K. Nandi «Fault detection using genetic programming», *Mechanical Systems and Signal Processing*, vol. 19, 2005, p. 271–289.
- (Zhou, 2006)** Zhou R «Feature extraction of musical content for automatic music transcription», These de doctorat , Chinese Academy of Science, Beijing, Chine.2006.
- (Zhuge et al., 1991)** Q. Zhuge et Y. Lu, Signature analysis for reciprocating machinery with adaptive signal-processing, Proceedings of the Institution of Mechanical Engineers Part C—Journal of Mechanical Engineering Science 205 (1991) 305–310.
- (Zwingestein, 1994)** G. Zwingestein, Diagnostic des défaillances : théorie et pratique pour les systèmes industriels. Ed. HERMES 1994. (bibliothèque cote S8/51136).

Annexe 1.1
Schéma synoptique de la maintenance



Annexe 2.1

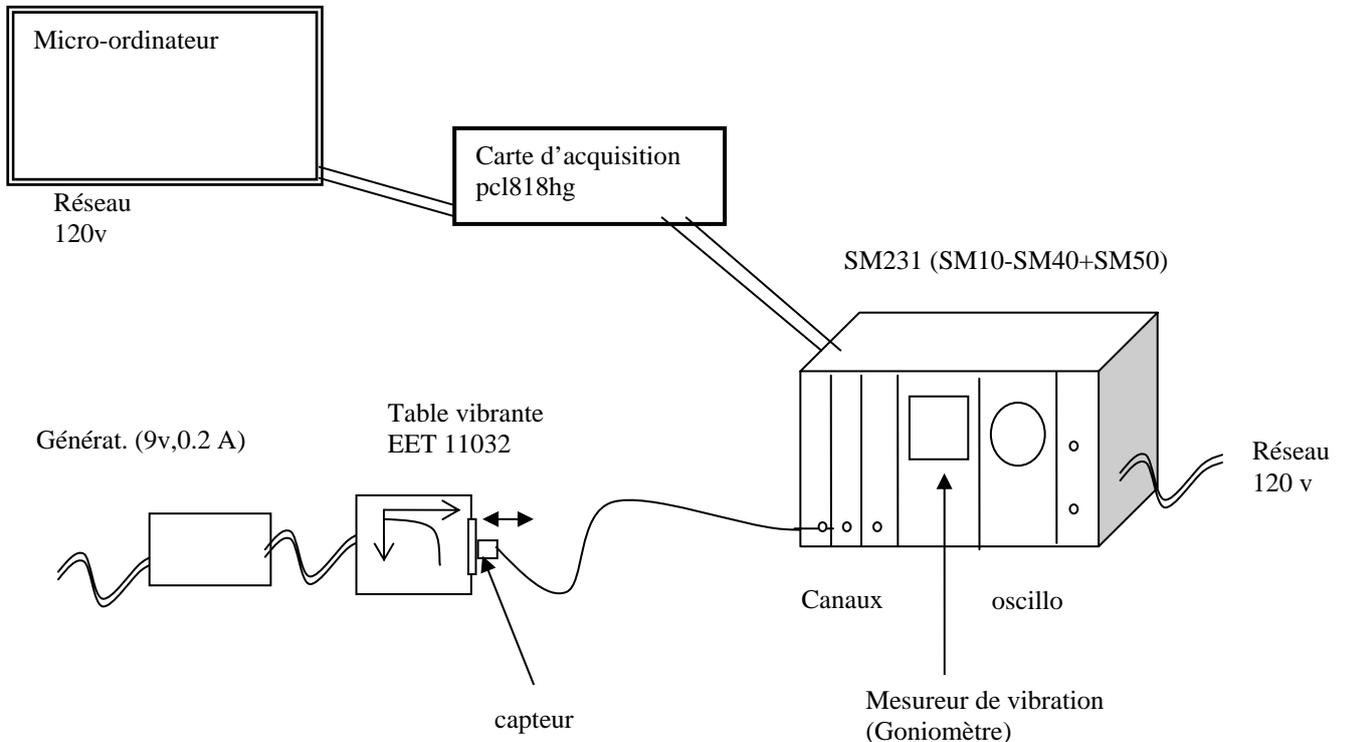
Schéma d'un accéléromètre
et sa fonction de transfert

Annexe 2.2

Signaux de Sétif

Un capteur piézo-électrique transforme l'excitation vibratoire en signal électrique qui, amplifié et éventuellement intégré à l'aide du SM231, est transmis à l'ordinateur à travers une carte d'acquisition. La figure 1, Annexe 4, présente le schéma synoptique général du système.

On présente ci-dessous les différentes parties du circuit de calibration



Les signaux sont visualisés sur un oscilloscope pour un contrôle de forme ils sont ensuite digitalisés par l'intermédiaire de la carte d'acquisition de façon à être stockés sur une mémoire de masse interne ou externe.

La fréquence d'échantillonnage est imposée par le théorème de Shannon qui stipule que le signal à étudier ne doit pas contenir des composantes fréquentielles dont la fréquence est supérieure à la moitié de la fréquence d'échantillonnage. Si cette condition n'est pas remplie, un phénomène de repliement vient perturber les analyses fréquentielles voir paragraphe 2.5.1 partie sur le filtre anti-repliement .

a- Table vibrante :

La table vibrante (table électrodynamique EET 11032) est un instrument qui a pour but de calibrer le mesureur de vibration (SM231). L'excitation de cette table avec le générateur de tension (9V, 0.2A) crée des vibrations sur le noyau de celle-ci. la fixation du capteur sur l'arbre se fait à l'aide d'un goujon, celui-ci est relié à l'entrée d'un canal (le SM10) sur le couvercle de la table, nous pouvons trouver la courbe d'étalonnage.

b- Le capteur piézo-électrique : est un KD 35 qui présente une réponse linéaire jusqu'à 5000 Hz avec un facteur de transmission de 5.00mV/ms^{-2} . Sa fréquence de résonance est de 25 kHz. Voir Annexe 0. Deux capteurs sont utilisés suivant les deux directions : axiale, radiale.

Caractéristique des capteur :

- Plage d'utilisation (bande passante) : 2Hz à 15 HZ, à l'intérieur de laquelle l'amplitude mesurée par le capteur ne dépasse pas une marge d'erreur fixée par le constructeur (par exemple l'amplitude mesurée à $\pm 10\%$ ou $\pm 3\text{ dB}$). C'est la plage de linéarité de la réponse voir figure.

- Gamme de mesure (Gamme dynamique) : de la plus petite à la plus des valeurs acceptées par le capteur.
- Sensibilité : 10 pC/g=10 millivolts/micromètre.
- Tenue en température : au delà d'une certaine température qui est fonction des matériaux constituant le capteur, la température peut faire chuter rapidement la précision de la mesure et même endommager définitivement le capteur.

c- L'amplificateur intégrateur et mesureur des descripteurs (SM231) : à la place d'un module de conditionnement annexe 6, celui-ci permet d'amplifier le signal d'accélération, de l'intégrer une fois pour mesurer la vitesse et deux fois pour obtenir l'élongation. il sert également, avec sa sortie oscilloscope, comme adaptateur du signal et fournit une impédance de sortie de 600 Ohm.

Cet instrument de mesure est très sensible. Il possède essentiellement trois canaux, un goniomètre (SM40) et un oscilloscope (SM50). Grâce à cet instrument, on peut évaluer l'accélération, la vitesse et le déplacement. Certaines caractéristiques techniques de l'instrument sont représentées dans le tableau suivant :

Domaine de travail fréquentiel. Pour mesurer les valeurs efficaces et crête à crête	2 Hz au 15 kHz
Durée de choc	1ms à 16 ms
Accélération maximale mesurée	a < 32 000 ms ⁻²
Accélération minimale mesurée	a < 0.032 ms ⁻²

d- La carte d'acquisition de données PCL-818HG : La PCL 818 HG est une carte haute performance de multi-fonctions pour l'acquisition de données en utilisant un micro-ordinateur.

Elle offre les cinq caractéristiques les plus demandées dans la mesure et le contrôle telle que :

- Résolution 12 bits
- Conversions analogique-numérique et numérique-analogique, (la fréquence d'échantillonnage maximale est de 100 MHz).
- 16 entrées et 16 sorties numériques, TTL/ DTL compatible.
- 16 entrées analogiques mono (ou 8 différentielles) et 1 sortie analogique,
- 1 compteur à usage externe,
- Un amplificateur d'instrumentation programmable qui permet la conversion des entrées de faible niveau : le Gain peut être de x 0.5, 1, 5, 10, 50, 100, 500, 1000
- Un multiplexeur intégré permet d'acquérir plusieurs voies.
- Des gammes d'entrée analogique sélectionnées par programme : ±0.005V, ±0.01V, ±0.05V, ±0.1V, ±0.5V, ±1V, ±5V, ±10V, 0-0.01V, 0-0.1V, 0-1V, 0-10V.
- 3 modes de déclenchement : * par programme, ** par trigger programmable, *** par impulsion externe.
- Transfert de données par un programme de contrôle, interruption et DMA (Direct Memory Acces).

e- Module d'acquisition de données (software) :

C'est un module qui gère cette carte en exploitant toutes ses fonctionnalités ce module offre 4 modes d'acquisition est une fonction de sauvegarde ainsi que la visualisation.

d. Calibration : la première opération à effectuer avant chaque série d'essais est une calibration de la chaîne de mesure. Cette calibration est effectuée grâce à la table vibrante qui délivre un signal calibré sinusoïdal d'amplitude 9.8m/s⁻² est de fréquence 78 Hz. Cette calibration permet de tenir déterminer le coefficient multiplicateur qui permettra d'assurer une analyse exacte de chaque série d'essais.

Annexe 2.3

Métriques

Il existe plusieurs types de métriques. Parmi les plus connus, dans le cas des données numériques, la métrique de Minkowski. Elle est très utilisée. Elle fournit un indice de dissimilation entre deux vecteurs \mathbf{x} et \mathbf{y} :

$$d(\mathbf{x}, \mathbf{y}) = \left[\sum_{j=1}^d |x_j - y_j|^r \right]^{1/r} \quad r \geq 1 \quad (1)$$

Où d désigne la dimension des vecteurs.

La métrique de Minkowski possède les propriétés d'une mesure de distance car elle vérifie :

$$d(\mathbf{x}, \mathbf{y}) = 0 \text{ si et seulement si } \mathbf{x} = \mathbf{y} \quad (2)$$

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{y}, \mathbf{z})$$

Les distances les plus utilisées pour certaines valeurs de r sont :

- La distance euclidienne classique pour $r = 2$:

$$d(\mathbf{x}, \mathbf{y}) = \left[\sum_{j=1}^d |x_j - y_j|^2 \right]^{1/2} = \left[(\mathbf{x} - \mathbf{y})^t (\mathbf{x} - \mathbf{y}) \right]^{1/2} \quad (3)$$

- La distance de Manhattan ou city – block pour $r = 1$:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^d |x_j - y_j| \quad (4)$$

- Ou encore, la distance du "Max" pour $r \rightarrow \infty$:

$$d(\mathbf{x}, \mathbf{y}) = \text{Max}_{1 \leq j \leq d} |x_j - y_j| \quad (5)$$

Enfin, il existe une mesure particulière de distance appelée distance de Mahalanobis. Elle est définie par :

$$d^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^t \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y}) \quad (6)$$

L'introduction de l'inverse de la matrice de variance – covariance \mathbf{S} dans l'expression (6) signifie que cette mesure de distance prend en compte la dispersion de l'ensemble d'apprentissage, dans chaque dimension.

Annexe 3.1

Types de défauts sur les roulements

Nous distinguons quatre types de dégradations sur les roulements (Nandi, 1999), (Ypma, 2001) :

a- Défaut sur la bague interne :

Il est caractérisé par la présence d'une raie à la fréquence caractéristique du défaut ($f_{d,bi}$). Cette fréquence est modulée par la fréquence de rotation (bandes latérales autour de la raie de défaut).

$$f_{d,bi} = \frac{N}{2} \left(1 + \frac{d_b}{d_m} \cos(\alpha) \right) (f_{rot} - f_{ext}) \quad (1)$$

Avec : f_{rot} : fréquence de rotation de la bague interne

f_{ext} : fréquence de rotation de la bague externe, cette fréquence de rotation de la bague externe est nulle si la bague est emmanchée dans le carter.

d_b : diamètre de la bille

d_m : diamètre moyen du roulement.

N : nombre de billes du roulement

α : angle de contact

b- Défaut sur la bague externe :

Ce défaut est caractérisé par la présence d'une raie à la fréquence ($f_{d,be}$). Bien que la charge appliquée sur la bague externe soit constante, on peut remarquer une modulation d'amplitude à la fréquence de rotation de l'arbre autour de la fréquence de défaut.

$$f_{d,be} = \frac{N}{2} \left(1 - \frac{d_b}{d_m} \cos(\alpha) \right) (f_{rot} - f_{ext}) \quad (2)$$

c- Défaut sur les éléments roulants :

La première fréquence caractéristique de défaut correspond à la fréquence de rotation de l'élément roulant sur lui-même. De plus, cet élément roulant rencontre une fois la bague interne et une fois la bague externe par tour, il génère donc des chocs à 2 fois cette fréquence.

$$f_{d,er} = \frac{1}{2} \frac{d_m}{d_b} \left(1 - \left(\frac{d_b}{d_m} \cos(\alpha) \right)^2 \right) (f_{rot} - f_{ext}) \quad (3)$$

d- Défaut sur la cage : Ce défaut se manifeste par la présence de raies à la fréquence $f_{d,ca}$ et ses harmoniques. A l'instar de la bague interne, on pourra constater des bandes latérales autour de $f_{d,er}$ (fréquence de défaut d'éléments roulants) et $2.f_{d,er}$, ces bandes seront distantes d'une fréquence égale à la fréquence de défaut de la cage $f_{d,ca}$.

$$f_{d,ca} = \frac{1}{2} \left(1 - \frac{d_b}{d_m} \cos(\alpha) \right) f_{rot} + \frac{1}{2} \left(1 - \frac{d_b}{d_m} \cos(\alpha) \right) f_{ext} \quad (4)$$

Annexe 4.1

Réseau de Neurones artificiels

Les réseaux les plus appropriés pour la classification sont les réseaux multicouches qu'on appelle communément perceptron multicouches et l'algorithme d'apprentissage couramment utilisé est l'algorithme de rétro-propagation du gradient. Nous en donnons une brève introduction dans les paragraphes qui suivent.

1. Le neurone formel

L'unité de base des réseaux de neurones est le neurone formel (ou cellule) représenté sur la Figure 1. Un neurone formel est une unité de traitement transformant un signal, généralement, multidimensionnel Z (stimulus) en un signal de sortie unidimensionnel ou multidimensionnel O par la suite d'opérations élémentaires suivante (Dubuisson 2001 p.182) :

$$y = \Phi(Z, W) = W \cdot Z^T = \sum_{i=1}^{Ne} w_i z_i - b \quad (1)$$

$$O = f(y) = f(W \cdot Z^T)$$

Ce neurone formel est un processeur élémentaire qui réalise une somme pondérée des signaux qui lui parviennent. La valeur de cette sommation est comparée à un seuil « b » et la sortie du neurone est une fonction **non linéaire** du résultat:

En prenant la convention de noter par $Z = [-1 \ z_1 \ z_2 \ \dots \ z_{Ne}]$ le vecteur d'entrées augmenté, et par $W = [b \ w_1 \ \dots \ w_{Ne}]$ le vecteur de poids augmenté. Il vient que la non-linéarité est assurée par la fonction de transfert (ou d'activation) f .

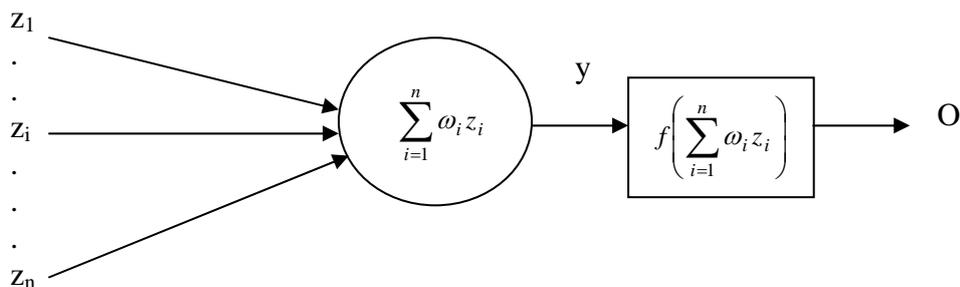
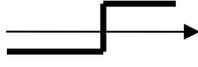
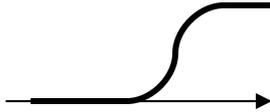
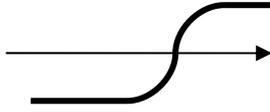


Figure 1 Modèle d'un neurone formel

Différentes fonctions de transfert pouvant être utilisées comme fonction d'activation du neurone. Nous les énumérons quelques une au dessous (Marc Parizeau, ...):

Tableau. 1 Fonction d'activation $O = f(y)$.

Nom de la fonction	Relation d'entrée/sortie	Forme de la courbe
seuil	$O = 0$ si $y < 0$ $O = 1$ si $y \geq 0$	
Seuil symétrique	$O = -1$ si $y < 0$ $O = 1$ si $y \geq 0$	
linéaire	$O = y$	
Linéaire saturée	$O = 0$ si $y < 0$ $O = y$ si $0 \leq y \leq 1$ $O = 1$ si $y > 1$	
Linéaire saturée symétrique	$O = -1$ si $y < -1$ $O = y$ si $-1 \leq y \leq 1$ $O = 1$ si $y > 1$	
Linéaire positive	$O = 0$ si $y < 0$ $O = y$ si $y \geq 0$	
Logarithme sigmoïde	$O = \frac{1}{1 + e^{-y}}$	
Tangente sigmoïde	$O = \frac{e^y - e^{-y}}{e^y + e^{-y}}$	
compétitive	$O = 1$ si $y_{\max imum}$ $O = 0$ autrement	

Les trois les plus utilisées sont les fonctions «seuil» (en anglais «hard limit»), «linéaire» et « sigmoïde».

2. Perceptron Multicouche

Un réseau de neurones est un maillage de plusieurs neurones, généralement organisé en couches. Un neurone réalise une fonction non linéaire de ses entrées. Ces dernières peuvent être les sorties d'autres neurones : on réalise ainsi un réseau de neurones, qui n'est donc rien d'autre que la composition de fonctions non linéaires élémentaires. La manière dont on «relie» entre elles les sorties des neurones aux entrées d'autres neurones définit "l'architecture" du réseau. Les neurones sont arrangés par couche. Pour le réseaux perceptron multicouches (PMC) introduit par Rosenbalt. le PMC se place comme un réseau non bouclé, Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avales. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. La Figure 2. présente un réseau perceptron multicouche (3 couches).

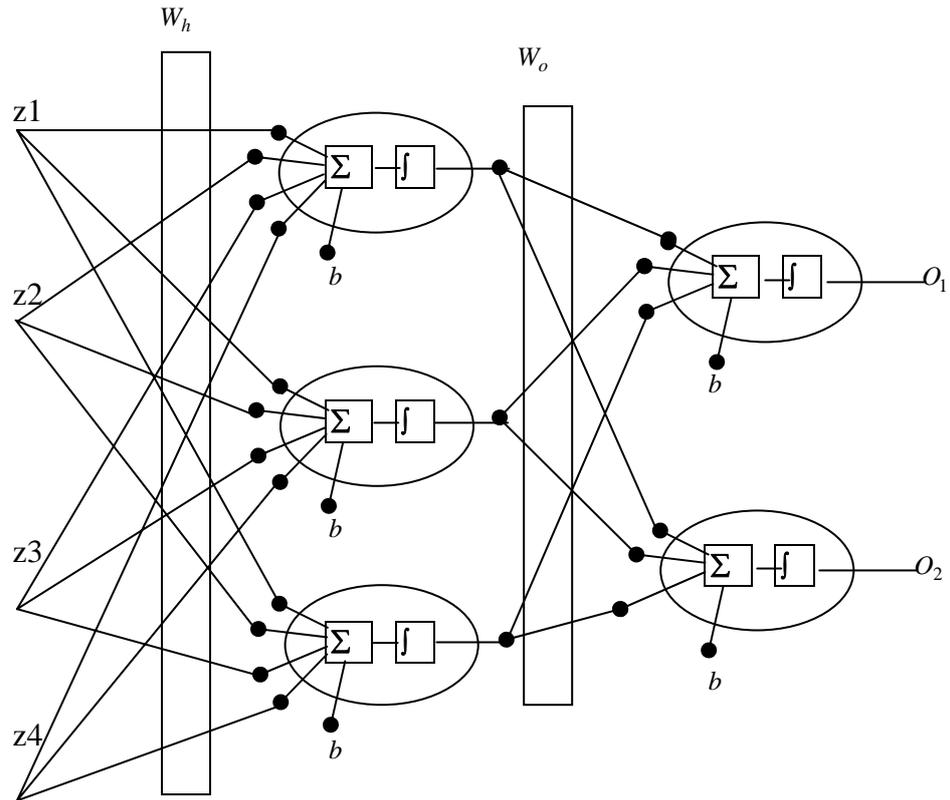


Figure 2. Structure d'un réseau perceptron multicouche (3 couches).

Par définition on appelle couche d'entrée (la première couche) l'ensemble des neurones d'entrées, couche de sortie (la dernière couche) l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelées couches cachées.

- \mathbf{Z} : est le vecteur d'entrée du réseau.
- \mathbf{O} : est le vecteur de sortie du réseau.
- w_{ji} : poids synaptique de la couche cachée.
- w_{kj} : poids synaptique de la couche de sortie
- i = indice des neurones de la couches d'entrée $i=1, Ne$
- j = indice des neurones de la couches cachée $j=1, Ncc$
- k = indice des neurones de la couche de sortie $k=1, Ns$

la matrice des poids de la couche cachée sera :

$$W_h = \begin{bmatrix} b_1 & w_{11} & \cdots & w_{1j} & \cdots & w_{1Ncc} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_i & w_{i1} & \cdots & w_{ij} & \cdots & w_{iNcc} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{Ne} & w_{Ne1} & \cdots & w_{Nej} & \cdots & w_{NeNcc} \end{bmatrix} \quad (2)$$

la matrice des poids de la couche de sortie sera :

$$W_o = \begin{bmatrix} b_1 & w_{11} & \cdots & w_{1k} & \cdots & w_{1Ns} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_j & w_{j1} & \cdots & w_{jk} & \cdots & w_{jNs} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{Ncc} & w_{Ncc1} & \cdots & w_{Ncck} & \cdots & w_{NccNs} \end{bmatrix} \quad (3)$$

Les Réseaux PMC sont très puissants et peuvent résoudre aisément des problèmes non linéairement séparables grâce à leurs propriétés fondamentales qui sont **l'apprentissage et la généralisation** des propriétés que nous discuterons plus tard.

Notre choix pour ce modèle était motivé par le fait, qu'il a montré dans de nombreux travaux antérieurs son efficacité en tant qu'outil de modélisation appliqué aux données empiriques. En plus, il s'est avéré que c'est un réseau bien adapté à réaliser une séparation entre deux ou plusieurs classes.

Dans notre étude nous considérons la situation la plus simple où le **réseau est statique** c'est-à-dire n'a ni un retour ni des retards (has no feedback or delays).

3. Propriétés du PMC

Vingt ans après la publication de l'ouvrage où **Minsky et Papert** exposaient les limitations de Perceptron simple, **Cybenko et al. (1989)** et **Hornik et al. (1989)** établissent les réseaux de neurones comme une classe d'**approximateurs universels** (aptitude à la séparation des classes non linéairement séparables et à l'approximation d'aussi près que l'on veut d'une fonction non linéaire). Il a été ainsi démontré qu'un perceptron multicouche avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée. Cette propriété de parcimonie est précieuse dans les applications industrielles. Néanmoins, cette propriété ne permet pas de choisir, pour un type de fonction donné, le nombre de neurones optimal dans la couche cachée.

Une autre propriété intéressante des réseaux neuronaux est sa capacité d'apprendre par des exemples. **L'apprentissage** est une phase du développement du RNA durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. Nous essayons dans le paragraphe 3.4. de donner des détails sur son déroulement. Cet apprentissage est réalisé directement à partir des données numériques et se concrétise, du point de vue interne, par une adaptation des seuls éléments susceptibles de varier dans un réseau de neurones : les poids synaptiques.

D'autres propriétés attractives des réseaux de neurones, incluant les PMC, peuvent être résumées comme :

- **la généralisation** : Bien performer avec des données qu'il n'a jamais vu auparavant.
- **La robustesse** : Possibilité de traiter des données bruitées.
- **l'aspect parallèle de l'évolution de la procédure à travers l'architecture** : la plupart des applications existantes n'exploitent pas encore aujourd'hui complètement ce parallélisme, mais déjà quelques exemples d'**implantation hardware** existent. L'un des avantages de ce parallélisme est la tolérance aux pannes. En effet, l'ensemble des poids synaptiques constitue la mémoire du réseau neuronal. Cela signifie, que l'information est distribuée dans ces poids. Il en résulte qu'une cellule défectueuse conduira à un nombre de connexions détruites, ce qui, **si le réseau est suffisamment volumineux**, représente un pourcentage faible de la mémoire. En d'autres termes, les performances du réseau neuronal sont amoindries, mais le système continue de fonctionner. (**Une grande tolérance sur les erreurs** : La défaillance d'un petit nombre de neurones n'affecte pas beaucoup la performance totale du système.)

4. Algorithme d'apprentissage du réseau multicouche

La méthode classique pour l'apprentissage du perceptron multicouche est l'algorithme de rétro-propagation supervisé. L'apprentissage dit supervisé par la présence d'un « professeur » qui possède une connaissance approfondie de l'environnement dans lequel évolue le réseau de neurones Figure 3. En pratique, les connaissances de ce professeur prennent la forme d'un ensemble de N_s couples de vecteurs d'entrée et de sortie que nous noterons $\{(Z_1, d_1), (Z_2, d_2), \dots, (Z_{N_c}, d_{N_c})\}$, où Z_i désigne un stimulus (entrée) et d_i la cible pour ce stimulus, c'est-à-dire les sorties désirées du réseau. Chaque couple (Z_i, d_i) correspond donc à un cas d'espèce de ce que le réseau devrait produire (la cible) pour un stimulus donné. Pour cette raison, l'apprentissage supervisé est aussi qualifié d'apprentissage par des exemples.

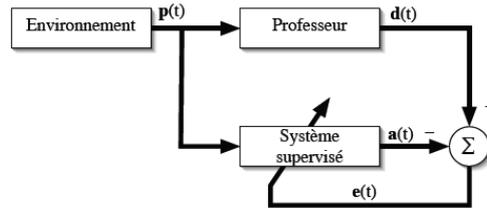


Figure 3 – Schéma bloc de l'apprentissage supervisé.

Cet apprentissage exploite une idée simple : la minimisation itérative d'un critère de l'erreur en sortie du réseau, on initialise les matrices de connexion au hasard, puis l'on fait évoluer ces matrices de manière à ce qu'elles autorisent l'association souhaitée, c'est-à-dire jusqu'à ce qu'un critère d'erreur (entre les sorties réellement obtenues et les sorties souhaitées) soit minimal « quasi nul »

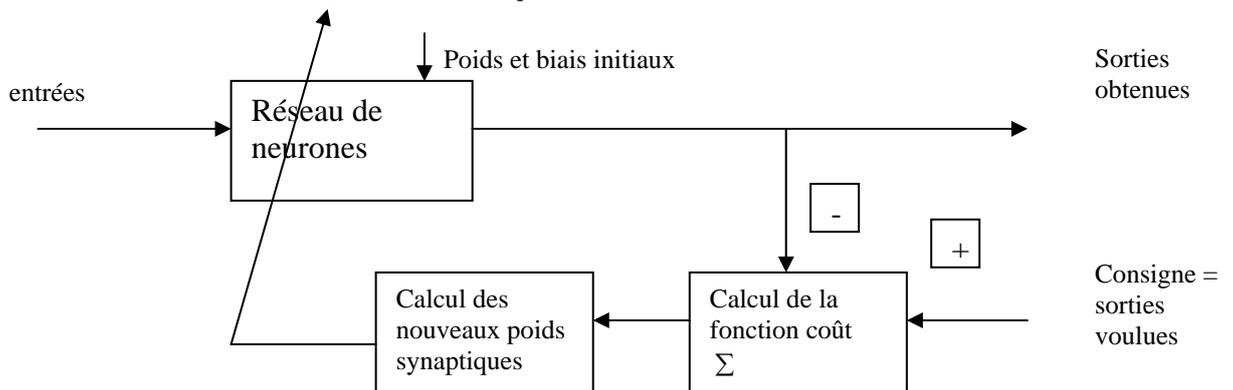


Figure 4. Schéma bloc de l'apprentissage supervisé

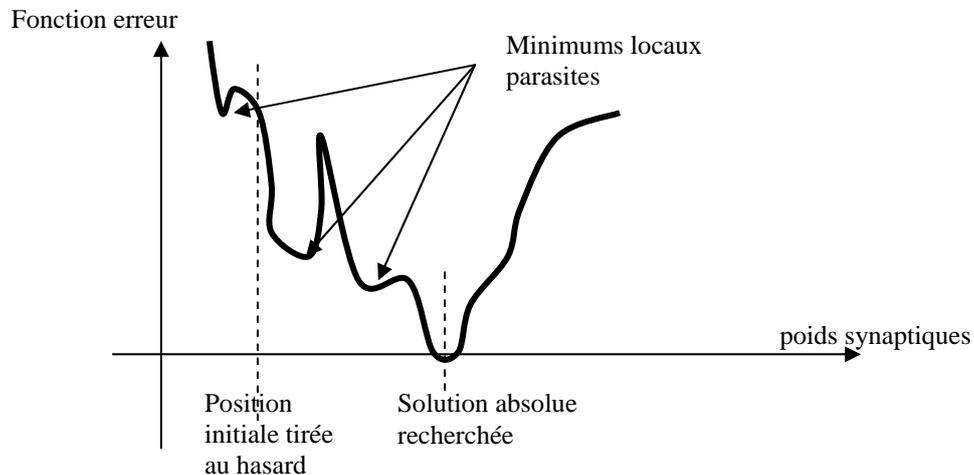


Figure 5. Illustre l'apprentissage qui consiste à chercher la configuration de poids synaptiques minimisant la fonction erreur

Résumé de l'algorithme (heuristique ou règle) d'apprentissage de rétropropagation du gradient (Dubuisson, 2001)

Cet algorithme repose sur la minimisation de l'erreur quadratique entre les sorties calculées et celles souhaitées. Le terme rétro-propagation du gradient provient du fait que l'erreur calculée en sortie est transmise en sens inverse vers l'entrée. Nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. De nos jours, les réseaux multicouches et la rétropropagation de gradient avec un apprentissage supervisé reste le modèle le plus étudié et le plus productif au niveau des applications.

1. Présentation d'un couple d'exemple entrée/sortie : appliquer un vecteur d'entrée $Z = [z_0 \quad z_1 \quad \dots \quad z_{Ne}]^T$ aux nœuds d'entrée puis initialiser les poids du réseau,

2. Propagation : Exécuter l'échantillon d'apprentissage à travers le réseau :

$$Y = W_h^T Z, \quad X = f_1(Y), \quad (4)$$

$$P = W_o^T X, \quad O = f_2(P), \quad (5)$$

3. Calcul des termes d'erreur du signal de la couches de sortie et de la couches cachée :

$$\text{Vecteur d'erreur de la couche de sortie } e_o = (O - d) f_2'(P) \quad (6)$$

$$\text{Vecteur d'erreur de la couche d'entrée } e_h = f_1'(Y) W_o^T e_o \quad (7)$$

4. Rétropropagation : Mise à jour des poids de la couche de sortie et la couche cachée respectivement avec un taux d'apprentissage $0 < \mu < 1$

$$W_o(i+1) = W_o(i) + \mu e_o O \quad (8)$$

$$W_h(i+1) = W_h(i) + \mu e_h X$$

5. Répéter ce processus jusqu'à ce que l'erreur moyenne quadratique (MSE) pour tous les stimuli (N

observations) d'entraînement à l'itération k devienne acceptable : $MSE(k) = \frac{1}{N} \sum_{i=1}^N E_p(i)$

(9)

Avec $E_p(i) = (O - d)_i^T (O - d)_i$ l'erreur entre la sortie réelle et la sortie désirée pour un stimulus

i (pour une observation i)

O : est le vecteur sortie calculée du réseau

d : est le vecteur sortie désirée du réseau

Remarque :

Cette approche suppose que l'apprentissage fonctionne par **groupage** (Par **paquets, hors ligne, batch mode**), c'est-à-dire que les mises à jour des poids sont accumulées sur l'ensemble des associations de la base d'apprentissage et appliquées une fois à la fin de chaque période (itération).

L'apprentissage incrémental (on ligne ou adaptative) : les poids sont modifiés à chaque présentation d'une entrée.

Pour notre étude nous considérons un réseau statique avec un apprentissage supervisé par groupage (Batch Training with Static Networks).

Un réseau est dynamique : When a network contains delays, the input to the network would normally be a sequence of input vectors that occur in a certain time order.

5. Considérations pratiques

Malgré les nombreuses réussites de l'application de l'algorithme du rétro-propagation du gradient, il a ses limitations. Les plus ennuyeux sont (Dubuisson 2001) :

- Une extrême lenteur de convergence,
- Une grande sensibilité aux conditions initiales,
- Une tendance aux oscillations,
- Une propension à rester dans les minimums locaux,

De nombreux travaux ont été consacrés à l'amélioration des performances de l'algorithme d'optimisation de rétropropagation du gradient. Les méthodes les plus utilisées sont citées à la section 3.6. Nous pouvons trouver à titre d'exemple des références tutoriales telles que (Sarkar, 1995) (Roy et al., 1990) et (Jacobs, 1988). L'efficacité et la convergence de l'algorithme dépendent, en grande partie de paramètres qu'un utilisateur doit fixer nous présentons ci-dessous quelques uns.

5.1. conditions initiales des poids et biais

L'initialisation des poids avant l'application de l'algorithme d'apprentissage par rétro-propagation du gradient est importante. Cette initialisation influe sur la vitesse de convergence du réseau mais aussi sur la qualité du réseau obtenu (Kolen et Pollack 90).

a. vitesse de convergence de l'apprentissage

Le temps de convergence d'un réseau dépend de l'espace initial de représentation des poids et de l'espace final après convergence. En effet, plus les poids initiaux sont proches de leur valeur finale et plus la convergence est rapide. On peut distinguer deux types de méthodes dans la littérature (Thimm et Fiesler 94) : les méthodes d'initialisation aléatoire dans un intervalle choisi de manière adéquate et les méthodes basées sur des techniques non aléatoires (Burel, 1991) (Fahlman, 1988). (Lee et al., 1993) ont montré théoriquement que la **saturation prématurée des neurones** (variations des poids trop petites pour avoir un effet sur la sortie des neurones) augmentait avec les valeurs maximales des poids. Ils en concluent que des valeurs initiales petites accroissent la vitesse de convergence mais aussi que des valeurs trop petites dégradent cette vitesse.

D'autres travaux sont basés sur des schémas d'initialisation et de méthode pseudo-inverse (Chen et Nutter, 1991) où seule la première couche est initialisée aléatoirement, les autres couches étant initialisées en fonction des sorties produites par la couche qui les précède.

b. Qualité du réseau résultant

L'apprentissage par rétro-propagation du gradient peut être vu comme l'optimisation d'une fonction ayant les poids du réseau pour paramètres (Figure 5). Ceci mène à une convergence de la fonction vers un minimum local qui peut être aussi global. Lorsque c'est le cas, on peut considérer que l'apprentissage a été fait correctement. Lorsque ce n'est pas le cas, le réseau obtenu n'est pas optimal et ne permet pas généralement une bonne classification. Si dans la plupart des cas, la solution trouvée est très proche de la solution optimale et donc acceptable, il reste des cas où la solution est médiocre.

Des études ont été faites (Wessel et Barnard, 1992) pour initialiser les poids de manière adéquate aux données à apprendre. (Denoeux et Lengellé, 1993) montrent des résultats plus robustes lorsque l'apprentissage est effectué avec des prototypes, ils permettent aussi d'accélérer la vitesse de convergence du réseau.

D'autres travaux, issus du domaine des algorithmes évolutionnaires ont été menés concernant l'influence de l'initialisation des poids.

Les poids sont initialisés aléatoirement à de valeurs faibles pour se situer en début d'apprentissage dans les parties non saturées des sigmoïdes.[...]

5.2. Moment (Momentum) (α)

L'algorithme de rétropropagation du gradient peut s'arrêter sur un minimum local. Une technique pour réduire cette possibilité est de rajouter un terme d'inertie dont le rôle est de **filtrer les oscillations** dans la trajectoire de la descente du gradient (Roy et al., 1990). Le terme du momentum produit deux effets distincts selon la situation. Premièrement, lorsque la trajectoire du gradient a tendance à osciller, il contribue à la stabiliser en ralentissant les changements de direction. Par exemple, avec $\alpha = 0.8$, cela correspond d'emblée à ajouter 80% du changement précédent au changement courant. Deuxièmement, lorsque le gradient courant pointe dans la même direction que le gradient précédent, le terme d'inertie contribue à augmenter l'ampleur du pas dans cette direction et donc à accélérer la convergence.

5.3. Taux d'apprentissage variable (μ)

Pour assurer la convergence, le pas doit être lentement réduit, au cours de l'apprentissage. La valeur de μ a un effet significatif sur les performances du réseau, si μ est petit nous avons une lenteur de convergence. Pour μ grand nous avons des oscillations (Panzani, 2004). Le pas d'apprentissage peut être également adaptatif fournissant un apprentissage plus rapide.

La variation dynamique du taux d'apprentissage peut également améliorer la convergence. Plusieurs approches peuvent être considérées. Par exemple, on peut adopter la stratégie suivante présentée par Panzani dans (Panzani, 2004) :

1. Si l'erreur totale, calculée pour toutes les associations de la base d'apprentissage, augmente d'une période à l'autre par plus d'un certain pourcentage β (typiquement de 1 à 5%) à la suite d'une mise à jour des poids, alors cette mise à jour doit être abandonnée et le taux d'apprentissage doit être multiplié par un facteur $0 < \rho < 1$, et le momentum doit être fixé à zéro ;
2. Si l'erreur totale diminue à la suite d'une mise à jour des poids, alors celle-ci est conservée et le taux d'apprentissage est multiplié par un facteur $\gamma > 1$; si le momentum avait précédemment été fixé à zéro, alors on lui redonne sa valeur originale ;
3. Si l'erreur totale augmente par moins de β , alors la mise à jour des poids est acceptée et le taux d'apprentissage reste inchangé ; Si le momentum avait précédemment été fixé à zéro, alors on lui redonne sa valeur originale ;

Cette approche suppose que l'apprentissage fonctionne par groupage, c'est-à-dire que les mises à jour des poids sont accumulées sur l'ensemble des associations de la base d'apprentissage et appliquées une fois à la

fin de chaque période. Dans certains cas cela peut accélérer grandement la convergence. Dans d'autres cas, cette approche peut aussi nuire à la convergence.

Il faut comprendre que ce genre de technique ajoute des paramètres (α, β, γ) et qu'il faut fixer a priori. Pour un problème donné, certaines combinaisons de paramètres peuvent être bénéfiques et d'autres non. Parfois, l'emploi d'une telle méthode peut même entraîner une divergence rapide là où la rétropropagation des erreurs avec momentum produisait une convergence lente

5.4. Les fonctions de transfert (d'activation)

Cette fonction, qui a été présentée au paragraphe 3.1. tableau 1., est responsable du comportement du neurone. Les fonctions d'activation utilisées sont variées, certains modèles emploient même plusieurs fonctions différentes dans le même réseau.

Généralement le perceptron multi couche utilise des fonctions d'activation sigmoïde pour les couches cachées et des fonctions linéaires pour la couche de sortie dans le cas d'une approximation de fonction et les fonctions sigmoïdes dans toutes les couches dans le cas de la reconnaissance de formes [Matlab toolbox du nnet v.4]. Ces fonctions d'activation permettent de borner les valeurs entre 0 et 1.

Nous avons dans notre cas utilisé les fonctions sigmoïdes logsig et tangsig.

5.5. La Complexité du modèle

Il a été démontré qu'un perceptron multicouches avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée. [Cybenko et al., 1989]. On peut constater que le nombre de neurones dans les différentes couches est le suivant :

Couche d'entrée : La première couche est appelée couche d'entrée. Elle recevra les données source que l'on veut utiliser pour l'analyse. Dans le cas de l'aide au diagnostic industriel, cette couche recevra les indicateurs de surveillance. Sa taille est donc directement déterminée par le nombre de variables d'entrées ou la dimension du VF.

Couche cachée : Le choix de sa taille n'est pas **implicite et doit être ajusté**. La taille est très importante pour le fonctionnement du réseau, et particulièrement les propriétés de généralisation. Si la couche cachée est trop petite, alors le réseau trouvera extrêmement difficile la répartition des données en différentes classes par contre si la couche cachée est trop grande, alors le réseau va apprendre "par cœur" les détails spécifiques de l'ensemble des données d'apprentissage et là le réseau va souffrir de ce qu'on appelle le "sur-apprentissage" ou le « surajustement », dans ce cas la généralisation du réseau va être extrêmement pauvre (Jack, 2000). En général, on peut commencer par une taille moyenne mais ce n'est pas toujours le meilleur choix. Il sera souvent préférable pour obtenir de bons résultats, d'essayer le plus de tailles possibles.

Différentes méthodes de sélection de modèle peuvent être utilisées [Dubuisson 2001]. Les techniques de **rééchantillonnage** (validation simple, validation croisée, ...) sont les plus utilisées avec de tels outils, les RNA se révèlent être performants. Nous les discuterons au **paragraphe 3.6**

Il existe également de nombreuses règles empiriques et heuristiques pour déterminer le nombre de neurones dans une couche cachée des détails sont dans les références (), ().

Enfin, nous disons que pour définir le nombre de neurones pour la couche cachée, dans notre travail, nous avons effectué un grand nombre d'expériences. Par exemple, on a varié la taille du réseau puis on a effectué un apprentissage complet pour chaque taille et puis on a choisi la structure qui conduit aux meilleurs résultats.

Couche de sortie : la taille de la couche de sortie est déterminée par le nombre de classes exigées (ou mode de fonctionnement). Deux stratégies peuvent être adaptées, dans la première, on utilise un réseau avec une C.S. ayant n neurones pour une classification pouvant atteindre jusqu'à 2^n défauts, un neurone de sortie mis à 0 ou à 1 selon que le défaut est absent ou présent. La seconde utilise un réseau de neurones propre à chaque défaut, c'est-à-dire qu'il faut utiliser un nombre de neurones égale aux nombres de défauts à détecter (codage disjonctif de 1-sur-M), à l'ordre d'une sortie pour chaque classe.

En dépit qu'elle demande plus de temps d'implémentation, cette dernière stratégie s'est avérée plus performante au niveau des résultats () (), expliquant ainsi notre choix initial d'appliquer cette seconde stratégie à notre étude.

5.6. Phénomène de saturation

Une autre considération pratique, soulignée par Parizeau dans (Parizeau,), dont on doit tenir compte lorsqu'on entraîne un PMC concerne le phénomène de saturation des neurones où, sous certaines conditions, les neurones peuvent à toute fin pratique cesser d'apprendre tellement leur convergence devient lente.

Par conséquent, à cause de ce phénomène de saturation, il importe soit de **normaliser** les données à l'entrée d'un PMC, c'est-à-dire de les transformer de manière à éviter tout risque de saturation, soit d'initialiser les poids sur la première couche en choisissant un intervalle de valeurs aléatoires ajusté aux stimuli d'apprentissage. Une autre alternative serait de fixer tous les poids à zéro. Bien que ceci réglerait certes le problème de la saturation des neurones, ce n'est malheureusement pas une alternative viable. En effet, il se trouve que l'origine de l'espace des poids correspond souvent à un lieu d'instabilité de la fonction d'erreur du réseau. Et ceci peut facilement entraîner la divergence de l'algorithme de rétropropagation.

5.7. Critère d'arrêt

Plusieurs critères d'arrêts peuvent être utilisés avec l'algorithme de rétropropagation des erreurs. Le plus commun consiste à fixer un nombre maximum de périodes d'entraînement, ce qui fixe effectivement une limite supérieure sur la durée de l'apprentissage. Ce critère est important car la rétropropagation des erreurs n'offre aucune garantie quant à la convergence de l'algorithme (). Il peut arriver, par exemple, que le processus d'optimisation reste pris dans un minimum local. Sans un tel critère, l'algorithme pourrait ne jamais se terminer. Un deuxième critère commun consiste à fixer une borne inférieure sur l'erreur ou encore sur la racine carrée de cette erreur. Pour les différents cas possibles, dépendant de l'application, il est parfois possible de fixer a priori un objectif à atteindre. Lorsque l'indice de performance choisi diminue en dessous de cet objectif, on considère simplement que le PMC a suffisamment bien appris ses données et on arrête l'apprentissage.

Les deux critères précédents sont utiles mais ils comportent aussi des limitations. Le critère relatif au nombre maximum de périodes d'entraînement n'est aucunement lié à la performance du réseau. Le critère relatif à l'erreur minimale obtenue mesure quant à lui un indice de performance mais ce dernier peut engendrer un phénomène dit de **sur-apprentissage** qui n'est pas désirable dans la pratique, surtout si l'on ne possède pas une grande quantité de données d'apprentissage, ou si ces dernières ne sont pas de bonne qualité.

Un processus d'apprentissage par correction des erreurs, comme celui de la rétropropagation, vise à réduire autant que possible l'erreur que commet le réseau. Mais cette erreur est mesurée sur un ensemble de données d'apprentissage. Si les données sont bonnes, c'est-à-dire qu'elles représentent bien le processus physique sous-jacent que l'on tente d'apprendre ou de modéliser, et que l'algorithme a convergé sur un optimum global, alors il devrait bien performer sur d'autres données issues du même processus physique. Cependant, si les données d'apprentissage sont partiellement corrompues par du bruit ou par des erreurs de mesure, alors il n'est pas évident que la performance optimale du réseau sera atteinte en minimisant l'erreur, lorsqu'on la testera sur un jeu de données différent de celui qui a servi à l'entraînement. On parle alors de la capacité du réseau à **généraliser**, c'est-à-dire de bien performer avec des données qu'il n'a jamais vues auparavant. Panzani dans (Panzani, 2004) a bien expliqué le phénomène. Par exemple, la figure 4 illustre le problème du **sur-apprentissage** dans le contexte d'une tâche d'approximation de fonction. La droite en pointillés montre une fonction linéaire que l'on voudrait approximer en ne connaissant que les points noirs. La courbe en trait plein montre ce qu'un réseau hypothétique pourrait apprendre. On constate que la courbe passe par tous les points d'entraînement et donc que l'erreur est nulle. De toute évidence, ce réseau ne généralisera pas bien si l'on échantillonne d'autres points sur la droite !

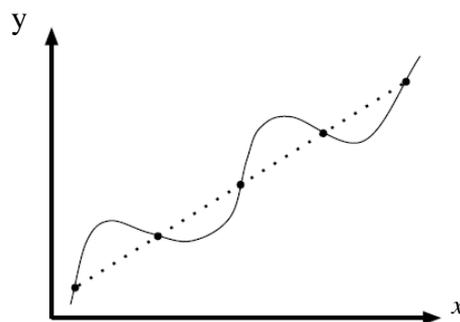


Figure 6. Illustration du phénomène de sur apprentissage pour le cas simple d'une approximation de fonction.

Une solution à ce problème consiste à utiliser un autre critère d'arrêt basé sur une technique dite de validation croisée (en anglais «cross-validation») paragraphe 3.6. Cette technique consiste à utiliser deux ensembles indépendants de données pour entraîner notre réseau : un pour l'apprentissage (l'ajustement des poids) et l'autre pour la validation, c'est-à-dire le calcul d'un indice de performance (une erreur, un taux de reconnaissance ou tout autre mesure pertinente à l'application). Le critère d'arrêt consiste alors à stopper l'apprentissage lorsque l'indice de performance calculé sur les données de validation cesse de s'améliorer pendant plusieurs périodes

d'entraînement. La Figure 7 illustre clairement ce compromis entre **sur-apprentissage**, **sous-apprentissage** et **bonne généralisation** pour une complexité donnée (soit le nombre d'itérations ou le nombre de neurones).

Dans notre cas, pour trouver un compromis entre erreur d'apprentissage et erreur de test, on compare pour différentes complexités. L'erreur moyenne sur la base d'apprentissage et sur la base de test indépendante. Les valeurs optimales correspondant aux valeurs pour lesquelles l'erreur sur la base de test est minimale. Nous avons également utiliser le pourcentage des bien prédits comme critère de performance (Figure 8.).

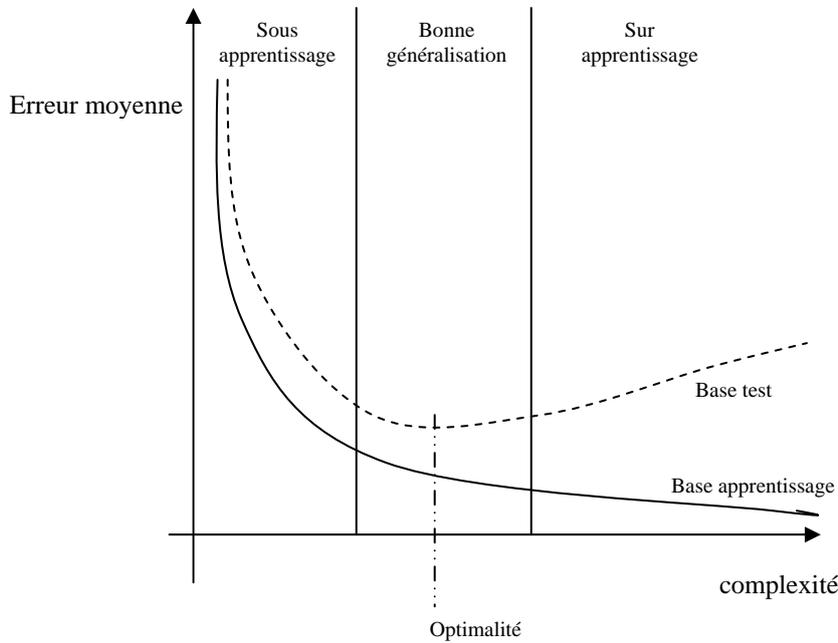


Figure 7. Complexité optimale du réseau représentée par l'erreur

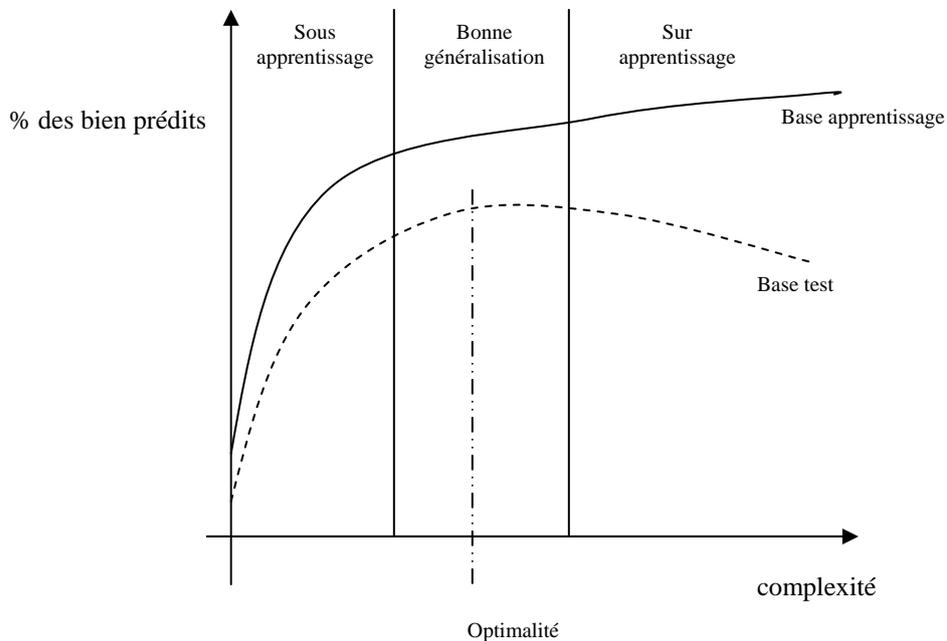


Figure 8. Complexité optimale du réseau représentée par le % des bien prédits

5.8. Fonction de Performance

Pour déterminer le critère d'arrêt et la complexité du réseau (modèle). Différentes fonctions de performance pouvant être utilisées :

- erreur absolue moyenne (mae),
- erreur quadratique moyenne (mse),
- erreur quadratique moyenne avec régularisation (msereg),
- somme des erreurs quadratiques (sse) ;

Les deux fonction les plus utilisées sont : le « mae » et «mse». La fonction « mae » est d'habitude utilisée en classification tandis que la fonction « mse » est plus utilisée en approximation de fonctions.

Dans notre cas, comme nous allons le constater à la section 6 la fonction mae n'a pas donnée de bon résultat, notre **critère de minimisation** sera donc le « mse ». Il nous permettra de contrôler le déroulement itératif de notre algorithme. L'apprentissage peut s'achever alors, soit par convergence de cette erreur à une valeur inférieure à certaine tolérance ou en arrivant à l'itération maximale. Les valeurs doivent être fixée a priori.

Bien sur nous pouvons utiliser d'autres distances telles que : Euclidienne, Manhattan, Mahalanobis , etc. qui sont très employées en discrimination (Duda & al., 2001), (Dubuisson 2001).

5.9. Influence du nombre d'observations sur les performances du RNA

Comme la sortie est non linéaire par rapport aux paramètres, La fonction de coût peut présenter des minima locaux, et les algorithmes d'apprentissage ne donnent aucune garantie de trouver le minimum global. De plus, le nombre d'exemples d'apprentissage joue un rôle fondamental dans l'existence des minima locaux, Car :

- Si l'on dispose d'un nombre suffisant d'exemples (beaucoup d'exemples par rapport au nombre de paramètres), le problème de minima locaux ne se pose pratiquement pas : il suffit, par prudence, d'effectuer quelques apprentissages avec des initialisations différentes des paramètres ;
- Si le nombre d'exemples est insuffisant, non seulement des minima locaux apparaissent, mais, de surcroît, le minimum global de la fonction de coût ne correspond pas forcément aux valeurs des paramètres recherchées ; il est donc inutile dans ce cas de mettre en œuvre des algorithmes coûteux pour chercher le minimum global.

5.10. Autres considérations pratiques

Nous énumérons ci-dessous d'autres considérations pratiques pour l'entraînement du PMC citées par Marc Parizeau dans [Parizeau ...]. Selon les circonstances, celles-ci peuvent aussi avoir un effet appréciable sur la performance de l'algorithme de rétropropagation des erreurs.

1. Lorsqu'on utilise une couche de sortie non-linéaire, c'est-à-dire une couche dont les neurones possèdent des fonctions d'activation non linéaires telles que la sigmoïde ou la tangente hyperbolique, il importe de ne pas chercher à saturer les neurones en fixant des sorties désirées qui tendent vers l'asymptote de la fonction. Dans le cas de la sigmoïde, par exemple, au lieu de fixer des sorties désirées à 0 ou à 1, on peut les fixer à 0.05 et 0.95. Ainsi, la rétropropagation des erreurs ne cherchera pas à entraîner les poids dans une direction qui pourrait rendre le neurone incapable de s'adapter. Autrement, la fonction logsig « sigmoïde » a une pente très petite quand on opère très loin du point centre. Pour la reconnaissance de forme et dans la couche de sortie nous utiliserons la fonction sigmoïde et nous voulons que le réseau opère aux queues de la fonction sigmoïde.

2. À chaque période d'entraînement (itération), il importe de permuter l'ordre de présentation des stimuli pour réduire la probabilité qu'une séquence de données pathologique nous garde prisonnier d'un piètre minimum local. En effet, la performance de la méthode de la descente du gradient peut dépendre grandement de cet ordre de présentation qui engendre des trajectoires différentes dans l'espace des paramètres, et des trajectoires différentes peuvent nous amener à des minimums locaux différents. Même s'il existe des séquences pathologiques, le fait de permuter les données à chaque période nous garantit que l'on ne tombera pas systématiquement sur les mêmes.

3. Dans le contexte d'un problème de classification à N_c classes, on associe généralement un neurone de sortie distinct à chacune d'elles. Ainsi, on interprétera chaque neurone sur la couche de sortie comme indiquant si oui ou non le stimulus d'entrée appartient à la classe correspondante. On construira les vecteurs « d » de sorties désirées avec deux valeurs possibles pour chaque composante : une valeur pour le «oui» et une valeur pour le «non». Si l'on choisit la fonction d'activation logistique, on pourra coder le «oui» avec une valeur proche de 1 et le «non» avec une valeur proche de 0 (voir item 1 ci-dessus). En mode de reconnaissance, on pourra classer un stimulus inconnu dans la catégorie associée au neurone ayant produit la sortie maximale.

4. Effectuer l'apprentissage d'un réseau quelconque revient à estimer les bonnes valeurs pour chacun de ses poids. Pour pouvoir estimer les paramètres d'un système quelconque possédant un certain nombre de degrés

de liberté (paramètres indépendants), il est nécessaire de posséder au moins un nombre équivalent de données. Mais dans la pratique il en faut bien plus ! Une règle heuristique nous indique que pour pouvoir espérer estimer correctement les Np poids d'un réseau de neurones, $10 Np$ données d'entraînement sont requises.

5. La performance d'un réseau lorsqu'évaluée avec ses données d'entraînement est presque toujours sur-estimée. Il faut bien comprendre que le réseau ne comporte aucune intelligence réelle. Il ne fait qu'apprendre les associations qu'on lui fournit. A la limite, il peut les apprendre «par coeur». Certains auteurs (Parizeau, ...), ..., estiment que si l'on veut évaluer correctement la performance d'un réseau, il faut le faire avec des données qui n'ont aucunement servi au processus d'apprentissage, ni pour la rétropropagation des erreurs, ni pour la validation croisée. En pratique, ceci implique de diviser les données d'entraînement en trois sous-ensembles distincts : les données d'entraînement, de validation et de test. La proportion relative de ces ensembles peut évidemment varier selon l'application, mais une bonne proportion se situe aux alentours de 50-20-30%, respectivement.

6. Quelques remarques sur l'application des algorithmes d'apprentissage

Ces remarques bibliographiques concernent une comparaison en vitesse de convergence et en mémoire de 6 différents algorithmes d'apprentissage.

La rétropropagation de gradient est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches. Nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. Plusieurs algorithmes d'amélioration de la convergence voir section 3.7.2. pouvant être utilisée, nous trouvons :

- les méthodes à pas μ adaptatifs telles que : Bold Driver (BD), Self-adaptative Back-Propagation (SAB), RProp, algorithmes du gradient conjugué, etc. Ils permettent d'accélérer la convergence en augmentant le pas d'adaptation tant que l'erreur décroît et en revenant à des valeurs faibles de pas en cas d'augmentation du critère d'erreur (Bishop 1995).
- Les méthodes d'optimisation du second ordre fondées sur le calcul ou l'approximation de la matrice Hessienne telles que : La méthode de Newton, La méthode de Levenberg-Marquardt, etc.

Une présentation détaillée de quelques unes de ces méthodes est donnée dans (Sarkar, 1995)

Les Algorithmes du gradient du 1^{er} ordre peuvent suivre un chemin en Zig-Zag pour converger. Les méthodes basées sur la dérivée du 2^{ème} ordre telles que Gauss Newton based Levenberg-Marquardt optimization algorithm, peut donner une convergence en un petit nombre d'itération.

Les algorithmes du gradient conjugué, en particulier SCG, performant sur des variétés des problèmes, spécifiquement dans le cas où la taille du réseau est grande. L'algorithme SCG est aussi rapide que l'algorithme LM dans les problèmes d'approximation des fonctions (plus rapide pour les réseaux de grandes tailles), et il est aussi rapide que l'algorithme RPROP pour les problèmes de classification. L'algorithme SCG nécessite relativement une place mémoire très modeste.

La méthode de Levenberg-Marquardt présente un intérêt pratique car elle peut être utilisée sans avoir à choisir le pas. L'algorithme de Levenberg-Marquardt est la plus rapide et assure la meilleure convergence vers un minimum de l'erreur quadratique, pour les problèmes d'approximation des fonctions où le nombre des poids du réseau est inférieur à cents. Quand le nombre de poids augmente l'efficacité de l'algorithme LM diminue, car la taille du Hessien augmente et nécessite une très grande place dans la mémoire, cet algorithme est pauvre pour les problèmes de classification. The original description of the Levenberg-Marquardt algorithm is given in (Marquardt, 1963).

L'algorithme RPROP est la plus rapide et assure la meilleure convergence vers un minimum de l'erreur quadratique, pour les problèmes de classification, mais cet algorithme est pauvre pour les problèmes d'approximations des fonctions et il nécessite une place mémoire très modeste. **Rprop** is generally much faster than the standard steepest descent algorithm. It also has the nice property that it requires only a modest increase in memory requirements.

La performance de l'algorithme trainbfg est similaire à celle de trainlm. Il n'exige pas un espace mémoire comme celui de trainlm, mais les calculs augmentent en fonction de la taille du réseau puisque il faut inverser une matrice à chaque itération.

L'algorithme traingdx est plus lent que les autres méthodes, et a exige le même espace mémoire que le trainrp, mais il peut être très utile pour certaines situations dans lesquelles il est mieux utile d'avoir une convergence plus lente. Par exemple, quand nous utilisons l'arrêt précoce «early stoping» on aura peut être un résultat inconsistant (non robuste) si on utilise un algorithme qui converge rapidement. On dépasse le point pour lequel l'erreur sur le test de validation est minimale.

Annexe 4.2

Paramètres d'apprentissage

Différentes paramètres d'apprentissage doivent être choisis. Ils sont énumérées ci-dessous selon l'algorithme d'apprentissage du paragraphe .., nous énumérons le cas des plus utilisés qui sont le « **Levenberg-Marquardt backpropagation** (LM) », l'algorithme « **Rprop** » et le « **Scaled Conjugate Gradient backpropagation** (SCG)»

- Déroulement de l'apprentissage selon les paramètres de l'algorithme « **Levenberg-Marquardt backpropagation** » trainlm montrés ici avec leurs valeurs par défaut

```
net.trainParam.epochs 100 Maximum number of epochs to train
net.trainParam.goal    0 Performance goal
net.trainParam.max_fail 5 Maximum validation failures
net.trainParam.mem_reduc 1 Factor to use for memory/speed trade off.
```

When using the training function trainlm, the parameter mem_reduc is used to determine how many rows of the Jacobian are to be computed in each submatrix. If mem_reduc is set to 1, then the full Jacobian is computed, and no memory reduction is achieved. If mem_reduc is set to 2, then only half of the Jacobian will be computed at one time. This saves half of the memory used by the calculation of the full Jacobian.

```
net.trainParam.min_grad 1e-10 Minimum performance gradient
net.trainParam.mu       0.001 Initial Mu
net.trainParam.mu_dec   0.1 Mu decrease factor
net.trainParam.mu_inc   10 Mu increase factor
net.trainParam.mu_max   1e10 Maximum Mu
net.trainParam.show     25 Epochs between displays (NaN for no displays)
net.trainParam.time     inf Maximum time to train in seconds
```

- Déroulement de l'apprentissage selon les paramètres de l'algorithme « **Resilient backpropagation** : **Rprop** » trainrp montrés ici avec leurs valeurs par défaut

```
net.trainParam.epochs 100 Maximum number of epochs to train
net.trainParam.show    25 Epochs between displays (NaN for no displays)
net.trainParam.goal    0 Performance goal
net.trainParam.time    inf Maximum time to train in seconds
net.trainParam.min_grad 1e-6 Minimum performance gradient
net.trainParam.max_fail 5 Maximum validation failures
net.trainParam.lr      0.01 Learning rate
net.trainParam.delt_inc 1.2 Increment to weight change
net.trainParam.delt_dec 0.5 Decrement to weight change
net.trainParam.delta0  0.07 Initial weight change
net.trainParam.deltamax 50.0 Maximum weight change
```

- Déroulement de l'apprentissage selon les paramètres de l'algorithme «**Scaled Conjugate Gradient backpropagation**» `trainscg` montrés ici avec leurs valeurs par défaut

<code>net.trainParam.epochs</code>	100	Maximum number of epochs to train
<code>net.trainParam.show</code>	25	Epochs between displays (NaN for no displays)
<code>net.trainParam.goal</code>	0	Performance goal
<code>net.trainParam.time</code>	inf	Maximum time to train in seconds
<code>net.trainParam.min_grad</code>	1e-6	Minimum performance gradient
<code>net.trainParam.max_fail</code>	5	Maximum validation failures
<code>net.trainParam.sigma</code>	5.0e-5	Determines change in weight for second derivative approximation.
<code>net.trainParam.lambda</code>	5.0e-7	Parameter for regulating the indefiniteness of the Hessian.

Nous avons choisis dans les deux cas :

- `net.trainParam.epochs` - Nombre maximal d'itérations 10000,
- `net.trainParam.show` - Pas d'itération pour affichage des résultats 50,
- `net.trainParam.goal` - Critère d'arrêt ou l'erreur quadratique minimale $mse=0$,
- `net.trainParam.lr` - pas d'apprentissage (en anglais : Learning rate) toujours compris en 0 et 1.

Annexe 4.3

Support vector Machines

1. Théorie de l'apprentissage de Vapnik

Dans cette section, nous rappelons quelques éléments essentiels de la théorie de l'apprentissage statistique. Nous introduisons par la même occasion le principe du risque structurel que minimise les SVM.

a) Risque structurel

Considérons un problème de classification à deux classes et soit les paires de données étiquetées :

$$(x_1, y_1), \dots, (x_l, y_l) \in R^N \times \{\pm 1\} \quad (1)$$

Où x_i représente la $i^{\text{ème}}$ observation de l'ensemble d'apprentissage et y_i son étiquette. Et soit l'ensemble de fonctions \square_α défini :

$$\{f_\alpha : \alpha \in A\}, f_\alpha : R^N \rightarrow \{\pm 1\} \quad (2)$$

α : Vecteur de paramètres.

A : l'ensemble des paramètres.

Supposons aussi qu'il existe une distribution $P(x,y)$ des données dont on ignore le modèle. La tâche d'apprentissage de \square_α qui approxime au mieux cette distribution consiste à minimiser le risque réel donné par :

$$R(\alpha) = \int \frac{1}{2} |f_\alpha(x) - y| dP(x, y) \quad (3)$$

b) Risque empirique

Dans l'équation (3) ne connaissant pas $P(x, y)$, il est difficile d'estimer le risque $R(\alpha)$. Il est possible toutefois de considérer une fonction de risque empirique de la forme :

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^l \frac{1}{2} |f_\alpha(x) - y| \quad (4)$$

Si l'ensemble d'apprentissage est fini, la minimisation du risque empirique donné par l'équation (4) ne garantit pas un minimum pour le risque réel (Ayat, 2004).

En 1979, Vapnik a mis au point le principe de la minimisation du risque structurel qui évite le sur-apprentissage des données à la convergence de la procédure d'apprentissage. Le risque empirique, par ailleurs, représente une estimation assez optimiste du risque réel dont la minimisation ne garantit pas la convergence vers une solution acceptable.

Le risque structurel constitue une borne supérieure de l'erreur de généralisation qui s'écrit :

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \phi\left(\frac{h}{l}, \frac{\log(\eta)}{l}\right) \quad (5)$$

Pour $\alpha \in A$ et $l > h$ avec une probabilité d'au moins $1-\eta$ et où ϕ est un terme de confiance donnée par :

$$\phi\left(\frac{h}{l}, \frac{\log(\eta)}{l}\right) = \sqrt{\frac{h(\log \frac{2l}{h}) - \log(\frac{\eta}{4})}{l}} \quad (6)$$

Le paramètre h est appelé VC ou dimension de Vapnik-Chervonenkis. Il décrit la capacité de l'ensemble de fonctions solutions. Pour un problème de classification binaire, h est le nombre maximum de points séparables selon 2^h configurations par l'ensemble de fonctions solutions. La minimisation de cette borne consiste à opérer une régularisation de l'erreur empirique $R_{emp}(\alpha)$ via le terme de pénalité $\phi\left(\frac{h}{l}, \frac{\log(\eta)}{l}\right)$. Donc un minimum est garanti en minimisant à la fois l'erreur empirique et le terme de régularisation.

La limite supérieure du risque réel (erreur de généralisation) définie dans l'équation (5) constitue un principe essentiel de la théorie des supports à vaste marge qui nécessite encore quelques remarques pertinentes.

Selon l'équation (5), étant donné un ensemble d'apprentissage de taille l , il est possible de borner $R(\alpha)$ en minimisant la somme des quantités $R_{emp}(\alpha)$ et $h(\{f_\alpha: \alpha \in A'\})$, tel que A' représente un sous-ensemble de A . Par ailleurs, l'erreur empirique dépend de la solution particulière trouvée par la machine d'apprentissage, à savoir f_α , et peut être réduite en choisissant des valeurs appropriées pour les paramètres α_i composant α .

La capacité h (dimension VC) quant à elle, dépend de l'ensemble de fonctions $\{f_\alpha: \alpha \in A'\}$ que la machine d'apprentissage peut inférer. Dans le but de contrôler h , il est possible de considérer des structures $S_n = \{f_\alpha: \alpha \in A_n\}$ de plus en plus complexes, telles que :

$$S_1 \subset S_2 \subset \dots \subset S_n \subset \dots, \quad (7)$$

et dont les capacités respectives vérifient l'inégalité :

$$h_1 \leq h_2 \leq \dots \leq h_n \leq \dots, \quad (8)$$

Pour un ensemble observations $(x_1, y_1), \dots, (x_i, y_i)$, la minimisation du risque structurel vise à choisir f_α parmi l'ensemble de fonctions $\{\square_\alpha: \square \in A_n\}$ possibles, qui minimise le risque garanti pour désigner la limite supérieure de l'erreur de généralisation donnée dans l'équation (5). Le processus de choisir le bon sous-ensemble de fonctions solutions revient à contrôler la complexité du classificateur en cherchant le meilleur compromis entre une faible erreur empirique et une complexité moindre

2. Les bases mathématiques des SVM

Les systèmes d'apprentissage appelés supports à vaste marge (*Support vector machines*) ou SVM en abrégé sont des algorithmes basés sur les trois principes mathématiques suivants (Yann, 1997) :

a) Le principe de Fermat (1638):

Les points qui minimisent ou maximisent une fonction dérivable annulent sa dérivée. Ils sont appelés points stationnaires

b) Le principe de Lagrange (1788):

Pour résoudre un problème d'optimisation sous contraintes il suffit de rechercher un point stationnaire x du lagrangien L de la fonction f à optimiser,

$$L(x, \alpha) = f(x) + \sum_{i=1}^k \alpha_i f_i(x) \quad (9)$$

Où $f(x)$ est la fonction à optimiser (fonction objective), les α_i sont des constantes appelées coefficients de Lagrange ou *multiplicateurs de Lagrange*, les f_i exprimant les contraintes

c) Le principe de Kuhn-Tucker (1951) :

Les relations de Kuhn-Tucker peuvent s'appliquer au cas qui nous intéresse. Avec des fonctions f et f_i convexes, il est même toujours possible de trouver un point-selle (x_0, α^*) qui vérifie :

$$\min_x L(x, \alpha^*) = L(x_0, \alpha^*) = \max_{\alpha \geq 0} L(x_0, \alpha) \quad (10)$$

Nous allons maintenant montrer comment ces principes peuvent être appliqués à la recherche d'un hyperplan séparateur optimal dans le cadre de la classification. Cette notion est claire dans le cas où les données sont linéairement séparables et elle permet alors de définir la notion de *vecteurs de supports*. On verra ensuite comment elle peut s'étendre au cas non linéairement séparable puis comment on peut définir des supports à vaste marge non linéaires.

3. Principes des supports à vaste marge (SVM)

Les supports à vaste marge ou SVM en abrégé ont été le sujet de plusieurs travaux (Samanta ; 2004b), [Yang, 2005a], (Yang, 2005b), (Jack et al., 2001) et (Kharoubi, 2002), les principes, et les bases théoriques de cette méthode ont été discutés dans plusieurs ouvrages, les plus connues sont (Vapnik, 1998), (Vapnik, 2000), (Burge, 1998).

3.1 Construction de l'hyperplan optimal

Pour bien décrire la technique de construction de l'hyperplan optimal séparant des données appartenant à deux classes différentes dans deux cas différents : Le cas des données linéairement séparables et le cas des données non linéairement séparables, supposons nous avons les données empiriques :

$$(x_1, y_1), \dots, (x_i, y_i) \in \mathbb{R}^N \times \{\pm 1\} \quad (11)$$

5.1.1 Cas des données linéairement séparables

Dans ce paragraphe nous présentons la méthode générale de construction de l'hyperplan optimal qui sépare les données appartenant à deux classes différentes, linéairement séparables.

La figure 1 donne une représentation visuelle de l'hyperplan optimal dans le cas des données linéairement séparables.

Soit $H : (w \cdot x) + b$ l'hyperplan qui satisfait les conditions suivantes :

$$\begin{cases} w \cdot x_i + b \geq 1 & \text{si } y_i = 1 \\ w \cdot x_i + b \leq -1 & \text{si } y_i = -1 \end{cases} \quad (12)$$

Ceci est équivalent à :

$$y_i(w \cdot x_i + b) \geq 1 \quad \text{Pour } i=1, \dots, m \quad (13)$$

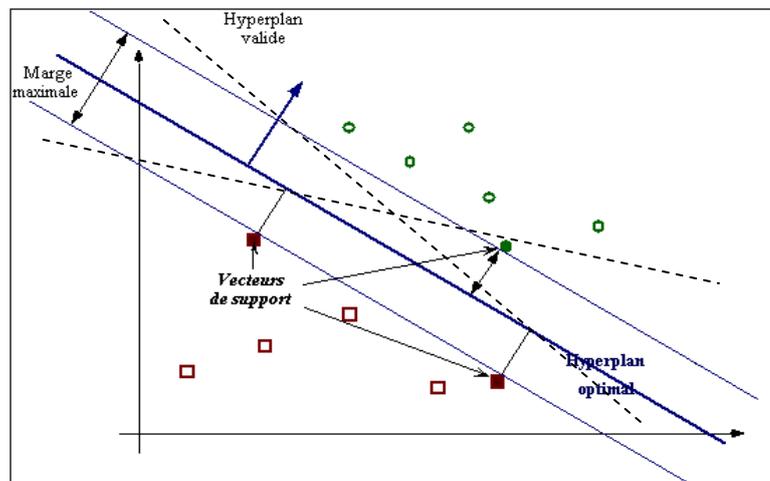


Figure 1 Exemple de séparation entre deux classes par les SVM (Cornuéjol, 2002)

Comme nous l'avons déjà mentionné, un hyperplan optimal (HO) est un hyperplan qui maximise la marge M qui représente la plus petite distance entre les différentes données des deux classes et l'hyperplan. Maximiser la marge M est équivalent à maximiser la somme des distances des deux classes par rapport à l'hyperplan.

La distance d'un point à l'hyperplan est :

$$d(x) = \frac{|w \cdot x + b|}{\|w\|} \quad (14)$$

L'hyperplan optimal est celui pour lequel la distance aux points les plus proches (*marge*) est maximale. Cette distance vaut :

$$\begin{aligned} M &= \min_{x_i / y_i} \frac{w \cdot x + b}{\|w\|} - \max_{x_i / y_i} \frac{w \cdot x + b}{\|w\|} \\ &= \frac{1}{\|w\|} - \frac{-1}{\|w\|} \\ &= \frac{2}{\|w\|} \end{aligned} \quad (15)$$

Maximiser la marge revient donc à maximiser $\frac{2}{\|w\|}$ ce qui est équivalent à minimiser $\frac{\|w\|^2}{2}$ sous la

contrainte (13) Ceci est un problème de minimisation d'une fonction objective quadratique avec contraintes linéaires [Vapn.98] :

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \forall i \quad y_i (w \cdot x_i + b) \geq 1 \end{cases} \quad (16)$$

L'introduction des multiplicateurs de Lagrange donne le lagrangien primaire qui s'écrit :

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [y_i (w^T x_i + b) - 1] \quad (17)$$

α_i : Les multiplicateurs de Lagrange.

Le lagrangien doit être minimal par rapport à w et b et maximal par rapport à α . Le point optimal est un point selle qui vérifie :

$$\begin{aligned} \frac{\partial L(w, b, \alpha)}{\partial w} &= 0 \\ \frac{\partial L(w, b, \alpha)}{\partial b} &= 0 \end{aligned} \quad (18)$$

Résoudre l'équation (18) donne :

$$\begin{aligned}
 w &= \sum_{i=1}^m \alpha_i y_i x_i \\
 \sum_{i=1}^m \alpha_i y_i &= 0
 \end{aligned} \tag{19}$$

m : nombre des points pour lesquels $\alpha_i > 0$.

En substituant l'expression de w dans l'équation (17) on peut écrire le problème dual sous la forme :

$$\begin{cases}
 \max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\
 \forall i \alpha_i \geq 0 \\
 \sum_{i=1}^l \alpha_i y_i = 0
 \end{cases} \tag{20}$$

Donc on peut écrire la fonction de décision des SVM linéaires sous la forme suivante :

$$f(x) = \text{sgn}\left(\sum_{i=1}^m \alpha_i^* y_i (x \cdot x_i) + b^*\right) \tag{21}$$

Où les coefficients α_i^* et b^* sont obtenus par résolution de l'équation (20).

5.1.2 Cas des données non linéairement séparables

Dans ce cas où les données sont non-linéairement séparables (figure 2), l'hyperplan optimal est celui qui satisfait les deux conditions suivantes :

-La distance entre les vecteurs bien classés et l'hyperplan optimal doit être maximale.

-la distance entre les vecteurs mal classés et l'hyperplan optimal doit être minimale.

Pour formaliser tout cela, on introduit des variables de pénalité non-négatives ζ_i pour $i=1,..,m$ appelées variables d'écart. Ces variables transforment l'inégalité (16) comme suit (Kharoubi, 2002) :

$$\begin{cases}
 \min \frac{1}{2} \|w\|^2 + C \sum \zeta_i \\
 \forall i y_i (w \cdot x_i + b) \geq 1 - \zeta_i
 \end{cases} \tag{22}$$

Où C est un paramètre de régularisation. Elle permet de concéder moins d'importance aux erreurs. Cela mène à un problème dual légèrement différent de celui du cas des données linéairement séparables. Donc maximiser le lagrangien donné par l'équation (20) par rapport à α_i sous les contraintes suivantes :

$$\begin{cases}
 \max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\
 \forall i 0 \leq \alpha_i \leq C \\
 \sum_{i=1}^l \alpha_i y_i = 0
 \end{cases} \tag{23}$$

Le calcul de α_i^* et b^* , et de la fonction de décision $f(x)$ reste exactement le même que pour le cas des données linéairement séparable.

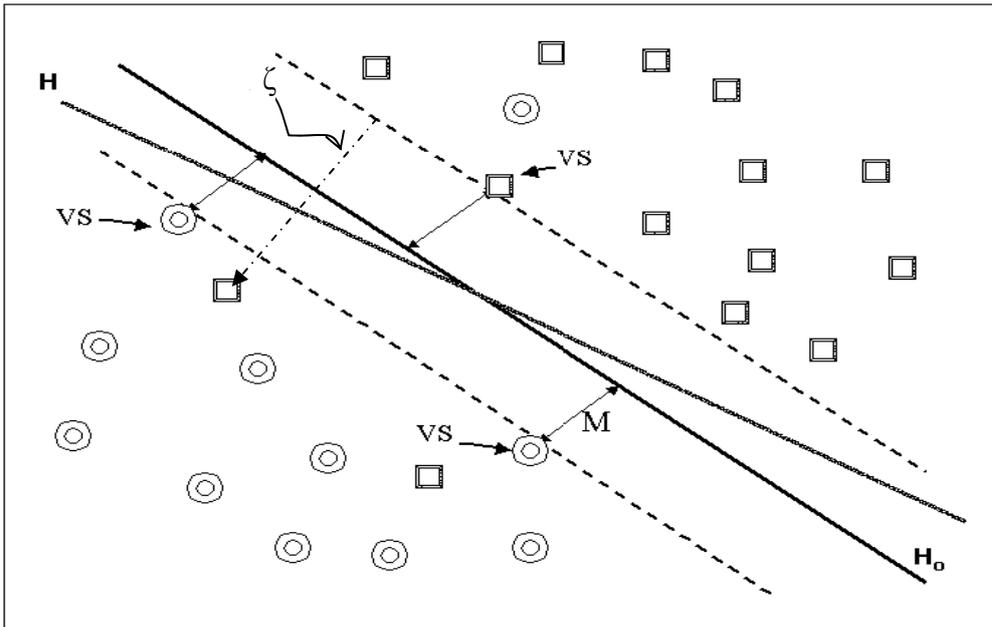
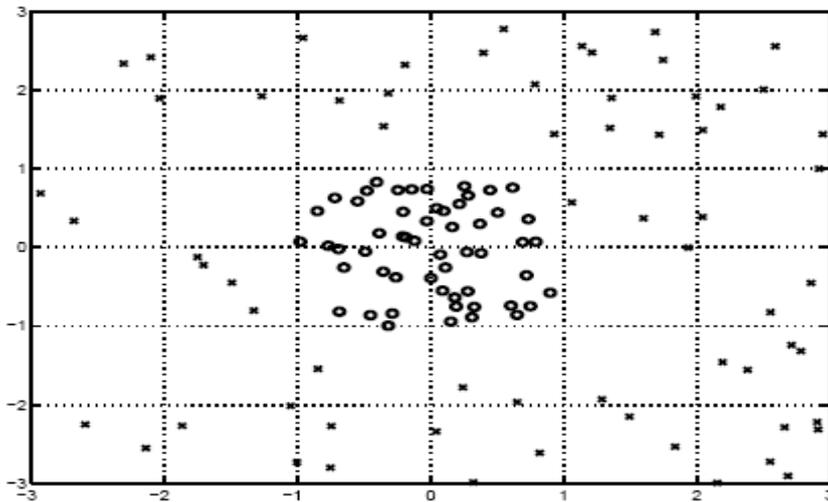


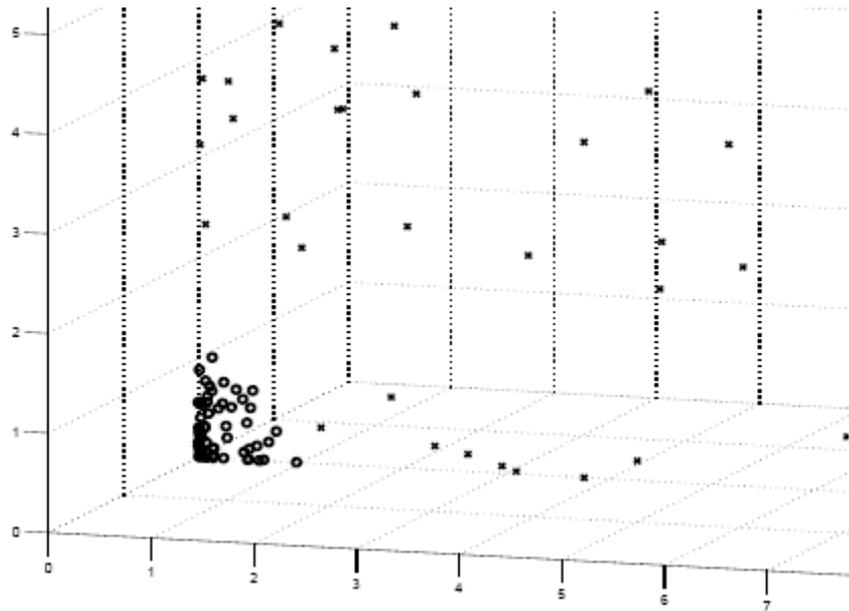
Figure 2 Hyperplans séparateurs dans le cas de données non-linéairement séparables (VS sont les Vecteurs de Supports). (Kharoubi, 2002)

5.2 Les SVM non linéaires

Plus la dimension de l'espace de description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les classes est élevée. En transformant l'espace d'entrée en un espace de redescription de très grande dimension, éventuellement infinie, il devient donc possible d'envisager d'utiliser la méthode des SVM. (Burge, 1998).



a) Espace de départ : exemples tirés aléatoirement dans R^2 appartenant à deux classes non-linéairement séparables.



b : Espace augmenté : l'image des exemples de (a) dans R^3 en utilisant la transformation Φ

Figure 3 Illustration de l'effet du changement d'espace par une fonction noyau. Les données non linéairement séparables dans l'espace de départ X sont à présent séparables dans l'espace augmenté (Kharoubi, 2002).

Notons Φ une transformation non linéaire de l'espace d'entrée X en un espace de *redescription* $\Phi(x)$:

$$X = (x_1, \dots, x_d)^T \text{ à } \Phi(X) = (\Phi(x_1), \dots, \Phi(x_d), \dots)^T \quad (24)$$

Généralement, le vecteur image $\Phi(x)$ est de dimension supérieure à la dimension de l'espace d'origine d . Les données sont projetées via la fonction Φ telle que :

$$\Phi(x_i) \cdot \Phi(x_j) = k(x_i, x_j) \quad (25)$$

Où k est appelée fonction noyau ou *kernel*. C'est à dire le produit scalaire dans l'espace des caractéristiques va être représentable comme un noyau de l'espace d'entrée. Le classificateur est donc construit sans utiliser explicitement la fonction Φ , donc sans nécessairement connaître cette fonction. Une famille de fonctions noyaux qui permet cette représentation et qui sont très appropriées aux besoins des SVM peut être définie comme l'ensemble des fonctions symétriques qui satisfont la condition suivante (Kharoubi, 2002) :

▪ **Théorème de Mercer :**

Pour être sûr qu'une fonction symétrique $k(x, x')$ admet un développement de la forme suivante :

$$k(x, x') = \sum_{k=1}^{+\infty} \beta_k \Phi_k(x) \cdot \Phi_k(x') \quad (26)$$

tel que les $\beta_k > 0$ décrit un produit interne dans l'espace des caractéristiques, il est nécessaire et suffisant que la condition suivante soit satisfaite :

$$\iint k(x, x') g(x) g(x') dx dx' \geq 0 \quad (27)$$

pour toute fonction $g \neq 0$ avec :

$$\int g^2(x) dx \geq 0 \quad (28)$$

Parmi les noyaux couramment utilisés par les chercheurs, on peut citer :

- Le noyau Polynomial : $k(x, x') = (x \cdot x' + 1)^q$ (29)

Où q est le degré du polynôme à déterminer par l'utilisateur.

- Le noyau RBF (*radial basis function*) : $k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$ (30)

Où σ est à déterminer par l'utilisateur.

- Le noyau Sigmoidal : $k(x, x') = \tanh(ax \cdot x' - b)$ (31)

Prenons l'exemple suivant [Ayat04]. Donc on a la transformation Φ tel que :

$$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$X = (x_1, x_2) \rightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$k(u, v) = \Phi(u) \cdot \Phi(v)$$

$$\begin{aligned} &= (u_1^2, \sqrt{2}u_1u_2, u_2^2) \begin{pmatrix} v_1^2 \\ \sqrt{2}v_1v_2 \\ v_2^2 \end{pmatrix} \\ &= (u_1^2v_1^2 + 2u_1v_1u_2v_2 + u_2^2v_2^2) \\ &= (u_1v_1 + u_2v_2)^2 \\ &= \left[(u_1, u_2) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right]^2 = (u \cdot v)^2 \end{aligned}$$

Le Lagrangien se transcrit dans ce cas par :

$$L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (32)$$

l : taille de la base d'apprentissage.

Donc le problème d'optimisation (23) devient :

$$\begin{cases} \max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \forall i \quad 0 \leq \alpha_i \leq C \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \quad (33)$$

La résolution de ce problème d'optimisation conduit à la fonction de décision suivante :

$$f(x) = \text{sgn} \left(\sum_{i=1}^m \alpha_i^* y_i k(x_i, x) + b^* \right) \quad (34)$$

Où les coefficients α_i^* et b^* sont obtenus comme précédemment par résolution de l'équation (33).

5.2.1 Conditions de Karush-Kuhn-Tucker (KKT)

La résolution de l'optimisation quadratique de l'équation (33) est basée sur les conditions de convergence dites de «Karush-Kuhn-Tucker» (KKT) qui établissent les conditions nécessaires (mais parfois suffisantes) de convergence de la fonction objective duale. Ces conditions sont relativement simples et s'écrivent :

$$\begin{aligned} \alpha_i = 0 &\Rightarrow y_i f(x_i) \geq 1 && \text{et } \zeta_i = 0 \\ 0 < \alpha_i < C &\Rightarrow y_i f(x_i) = 1 && \text{et } \zeta_i = 0 \\ \alpha_i = C &\Rightarrow y_i f(x_i) \leq 1 && \text{et } \zeta_i \geq 0 \end{aligned} \quad (35)$$

Les équations de 35 reflètent une propriété importante des SVM stipulant qu'une grande proportion des exemples d'apprentissage sont situés en dehors de la marge et ne sont pas retenus par le modèle. Par conséquent, leurs multiplicateurs α_i sont nuls.

Les conditions de KKT traduisent le fait que seulement les variables α_i des points situés sur la frontière de la marge ($0 < \alpha_i < C$) ou à l'intérieure de celle-ci ($\alpha_i = C$) sont non nulles. Ces points sont les **vecteurs de supports** du classificateur (Ayat, 2004).

Les SVM produisent alors une solution en n'utilisant qu'un sous ensemble réduit des données d'apprentissage. Sans cette propriété, l'entraînement des SVM sur de gros ensembles de données ainsi que son stockage deviennent extrêmement prohibitifs.

5.2.2 Calcul du biais

Le paramètre de biais b permet de définir des surfaces de séparation ne passant pas par l'origine. Son calcul exploite les vecteurs de supports respectant l'inégalité $0 \leq \alpha_i < C$ dont les ζ_i correspondants sont nuls. L'égalité suivante est alors vérifiée :

$$y_i \left(b + \sum_{i=1}^l \alpha_i y_i k(x_i, x_j) \right) = 1 \quad (36)$$

En considérant la moyenne calculée sur cet ensemble des vecteurs de supports, une valeur stable de b peut s'écrire :

$$b = \frac{1}{N_{SV}} \left(y_i - \sum_{i=1}^l \alpha_i y_i k(x_i, x_j) \right) \quad (\text{III.37})$$

Où N_{SV} représente le nombre de vecteurs de support considérés.

4. Algorithmes d'apprentissage des SVM

Afin de trouver les paramètres du SVM, il est nécessaire de résoudre le problème d'optimisation quadratique convexe donné par l'équation (33) dont la formulation matricielle s'écrit encore :

$$L_D = -\frac{1}{2} \alpha^T K \alpha + 1^T \alpha \quad (38)$$

Où K est une matrice semi-définie positive dont les composantes $k_{ij} = y_i y_j k(x_i, x_j)$ et 1 est le vecteur unité de même taille que α . Comme la fonction objective est convexe, tout maximum local

est aussi un maximum global. Toutefois, il peut y avoir des solutions optimales différentes en terme de α_i donnant lieu à des performances différentes.

Il existe une grande variété de méthodes et de logiciels traitant de la résolution de problèmes quadratiques. Cependant, quelques méthodes seulement sont capables de traiter un grand nombre d'exemples. Dans le cas contraire, l'apprentissage des SVM de quelques centaines d'exemples prendrait énormément de temps de calcul et assez de ressources mémoire. Seulement, il est possible de dériver des algorithmes qui exploitent la forme particulière de la fonction objective duale des SVM. Dans cette section, nous allons présenter trois approches différentes pour la résolution du problème quadratique des SVM. Ensuite, une attention particulière sera portée à l'algorithme SMO de Platt (Platt, 1998) que nous utilisons pour l'entraînement du classificateur dans la partie expérimentale.

4.1 Méthode de chunking

La résolution de la fonction objective duale de l'équation (32) avec un très grand nombre d'exemples donne lieu à un vecteur α creux. Selon les données, plusieurs des paramètres α_i sont soit nuls ou égales à C. S'il y a moyen de savoir a priori lesquels α_i seront nuls, il est possible de réduire la taille de la matrice K sans altérer la valeur de la fonction objective. Aussi, une solution α est valide si et seulement si elle respecte les conditions de KKT. Vapnik (Vapnik, 1998) était le premier à décrire une méthode qui exploite cette propriété en prenant en compte seulement les α_i non nuls ou ceux violant les conditions de Karush Kuhn Tucker. La taille de ce sous ensemble dépend du nombre de vecteurs de supports, de la taille des données et de la complexité du problème de classification. Cette méthode se comporte assez bien sur des problèmes de quelques centaines de vecteurs de supports. Des tâches plus complexes requièrent un schéma de décomposition de l'objective en sous problèmes plus facile à résoudre. Cette technique est décrite ci-dessous.

4.2 Méthode de décomposition successive

Cette méthode est similaire à celle du «Chunking» dans la mesure où elle considère aussi une succession de sous problèmes quadratiques à résoudre. La différence est que la taille des sous problèmes retenus est fixe. Cette méthode est basée sur la constatation qu'une succession de sous-problèmes quadratiques ayant au moins un exemple qui ne vérifie pas les conditions de KKT converge toujours vers une solution optimale (Ayat, 2004). Osuna et al., (Osuna et al., 1997) suggèrent de conserver la taille du sous-problème fixe et d'ajouter ou d'enlever un exemple à la fois. Ceci permet d'entraîner de gros ensembles de données. En pratique, cette stratégie peut converger lentement si diverses heuristiques ne sont pas prises en compte. En effet, il est possible d'adopter des stratégies sophistiquées afin d'inclure ou d'exclure quelques exemples de la fonction objective. Différentes stratégies peuvent aussi accélérer la convergence même avec quelques milliers de vecteurs de supports. L'algorithme de SVMlight est une implémentation de cette méthode.

5.3 Méthode de minimisation séquentielle : SMO

La méthode d'optimisation par minimisation séquentielle (de l'anglais «Sequential Minimal Optimization») proposée par Platt (Platt, 1998), peut être perçue comme le cas extrême des méthode de décomposition successive (Ayat, 2004). A chaque itération, elle résout un problème quadratique de taille égale à deux. La résolution de ce dernier est analytique et donc nul besoin de recourir à un module d'optimisation quadratique. Encore faut-il choisir le bon couple de variables (α_i, α_j) à optimiser durant chaque itération. Les heuristiques que l'auteur utilise sont basées sur les conditions de KKT. Son implémentation est relativement simple. Ici nous donnons les grandes lignes de cet algorithme :

Nous choisissons (au hasard pour l'instant) deux multiplicateurs (appelons les α_1 et α_2) et nous modifions leurs valeurs de manière à augmenter le plus possible L_D (33) tout en satisfaisant les contraintes de la boîte ($0 \leq \alpha_i \leq C$), et la contrainte d'égalité $\sum_{i=1}^l \alpha_i y_i = 0$

Nous nous assurons de satisfaire la contrainte d'égalité lorsqu'à chaque itération, les nouvelles valeurs α'_1 et α'_2 sont choisies de manière à satisfaire :

$$y_1 \alpha'_1 + y_2 \alpha'_2 = y_1 \alpha_1 + y_2 \alpha_2 \quad (39)$$

Il est à noter qu'il est impossible de satisfaire la contrainte d'égalité si un seul multiplicateur est modifié à chaque itération. C'est pour cette raison que deux multiplicateurs sont modifiés à chaque itération.

Trouvons maintenant les valeurs de α'_1 et α'_2 qui maximisent L_D . Ensuite, nous verrons ce qu'il faut faire pour s'assurer d'avoir $0 \leq \alpha'_1, \alpha'_2 \leq C$

Pour cela, réécrivons L_D uniquement en termes de α_1 et α_2 :

$$L_D(\alpha_1 + \alpha_2) = \alpha_1 + \alpha_2 - \frac{1}{2} G_{11} \alpha_1^2 - \frac{1}{2} G_{22} \alpha_2^2 - \frac{1}{2} G_{12} \alpha_1 \alpha_2 - \alpha_1 v_1 - \alpha_2 v_2 + c \quad (40)$$

Où c est une constante qui ne dépend pas de α_1 et α_2 et où :

$$v_i = \sum_{j=3}^m \alpha_j G_{ij} \quad \text{pour } i = 1, 2 \quad (41)$$

Puisque α_1 et α_2 seront modifiés en gardant constant $y_1 \alpha_1 + y_2 \alpha_2$, exprimons α_1 en fonction de α_2 :

$$\alpha_1 = d - s \alpha_2 \quad \text{avec } s = y_1 y_2 \text{ et } d = \text{constante}$$

Et substituons dans L_D pour obtenir :

$$L_D(\alpha_2) = d - s \alpha_2 + \alpha_2 - \frac{1}{2} G_{11} (d - s \alpha_2)^2 - \frac{1}{2} G_{22} \alpha_2^2 - G_{12} (d - s \alpha_2) \alpha_2 - (d - s \alpha_2) v_1 - \alpha_2 v_2 + c \quad (42)$$

Au maximum de $L_D(\alpha_2)$, nous devons satisfaire :

$$0 = \frac{\partial L_D}{\partial \alpha_2} = 1 - s + s G_{11} (d - s \alpha_2) - G_{22} \alpha_2 - d G_{12} + 2s G_{12} \alpha_2 + s v_1 - v_2 \quad (43)$$

Donc, pour se rendre au maximum de L_D , nous devons choisir pour α_2 la valeur α_2^u satisfaisant (u signifie «unconstrained») :

$$(G_{11} + G_{22} - 2s G_{12}) \alpha_2^u = d(s G_{11} - G_{12}) + s v_1 - v_2 + 1 - s \quad (44)$$

En utilisant $d = \alpha_1 + s \alpha_2$, $s = y_1 y_2$, $G_{ij} = y_i y_j k_{ij}$ et :

$$K_i = \sum_{j=1}^m \alpha_j y_j k_{ji} \quad (45)$$

$$\eta = 2k_{12} - k_{11} - k_{22}$$

nous obtenons, après quelques lignes d'algèbre :

$$\alpha_2'' = \alpha_2 - y_2 \frac{(K_1 - y_1) - (K_2 - y_2)}{\eta} \quad (46)$$

Or, lorsque nous aurons trouvé α_2' , la valeur de α_1' sera donnée par la contrainte :

$$\alpha_1' = \alpha_1 + s(\alpha_2 - \alpha_2') \quad (47)$$

Ainsi on aura $0 \leq \alpha_1' \leq C$ SSI :

$$0 < \alpha_1 + s(\alpha_2 - \alpha_2') < C \quad (48)$$

Cas : $S = 1$ c à d ($y_1 = y_2$) :

Dans ce cas on doit satisfaire :

$$0 < \alpha_1 + (\alpha_2 - \alpha_2') < C \quad (49)$$

ce qui entraîne

$$(\alpha_1 + \alpha_2) - C \leq \alpha_2' \leq (\alpha_1 + \alpha_2) \quad (50)$$

Or, puisqu'il faut également avoir $0 \leq \alpha_2' \leq C$ on a finalement que :

$$\max[0, (\alpha_1 + \alpha_2) - C] \leq \alpha_2' \leq \min[C, (\alpha_1 + \alpha_2)] \quad (51)$$

Cas : $S = -1$ c à d ($y_1 \neq y_2$) :

de même nous obtenons :

$$\max[0, (\alpha_2 - \alpha_1)] \leq \alpha_2' \leq \min[C, C + (\alpha_2 - \alpha_1)] \quad (52)$$

Dans tous les cas, nous devons avoir $U \leq \alpha_2' \leq V$ avec :

$$U = \begin{cases} \max[0, (\alpha_1 + \alpha_2) - C] & \text{si } y_1 = y_2 \\ \max[0, (\alpha_2 - \alpha_1)] & \text{si } y_1 \neq y_2 \end{cases} \quad (53)$$

$$V = \begin{cases} \min[C, (\alpha_1 + \alpha_2)] & \text{si } y_1 = y_2 \\ \min[C, C + (\alpha_2 - \alpha_1)] & \text{si } y_1 \neq y_2 \end{cases} \quad (54)$$

Considérant que α_2'' est donné par l'équation (46), la nouvelle valeur α_2' est donnée par :

$$\alpha_2' = \begin{cases} V & \text{si } \alpha_2'' > V \\ \alpha_2'' & \text{si } U \leq \alpha_2'' \leq V \\ U & \text{si } \alpha_2'' < U \end{cases} \quad (55)$$

Et, ayant obtenu α'_2 , la nouvelle valeur α'_1 est obtenu de l'équation (47). Typiquement l'algorithme débute avec $\alpha = 0$. Une paire de multiplicateurs est choisie et les nouvelles valeurs sont obtenues selon les équations (55) et (47), et on recommence avec une autre paire de multiplicateurs.

À chaque itération, la valeur de L_D augmente et α satisfait toutes les contraintes. Éventuellement α sera très près de l'optimal α^* . Donc, avant la mise à jour de (α_1, α_2) , l'équation (45) nous donne pour K_i :

$$K_i = \alpha_1 y_1 k_{1i} + \alpha_2 y_2 k_{2i} + \sum_{j=3}^m \alpha_j y_j k_{ji} \quad (56)$$

et après la mise à jour de (α_1, α_2) , nous avons :

$$K'_i = \alpha'_1 y_1 k_{1i} + \alpha'_2 y_2 k_{2i} + \sum_{j=3}^m \alpha_j y_j k_{ji} \quad (57)$$

Le nouveau K'_i est donc simplement reliée à l'ancien K_i par :

$$K'_i = K_i + (\alpha'_1 - \alpha_1) y_1 k_{1i} + (\alpha'_2 - \alpha_2) y_2 k_{2i} \quad (58)$$

ce qui complète la description de l'algorithme SMO.

Notons ici, qu'en trouvant K_i qui maximise L_D et en calculant le biais b en accord avec (37) on peut établir la fonction de décision 34.

5. Les supports à vaste marge multiclassés

Pour un problème à deux classes, l'hyperplan (w, b) du SVM délimite les deux partitions selon le signe de $f(x)$ où $f(x) = w^* x + b$. Par ailleurs, bien que le SVM soit un classificateur binaire, il peut facilement être étendu pour décider de l'appartenance de données multiclassés. En particulier, on trouve deux schémas de classification [Ayat.04] :

1. *Un-contre-Tous*. Cette méthode est simple d'usage et donne des résultats raisonnables. Elle consiste à entraîner K classificateurs différents, séparant chaque classe des $K-1$ restantes. Ainsi, pour chaque exemple de test, K valeurs de sortie $f(x)$ sont disponibles.

Une façon naïve mais simple de classification consiste à attribuer l'exemple à la sortie de plus grande amplitude.

2. *Un-contre-Un*. Cette méthode requiert l'apprentissage de $K(K-1)/2$ classificateurs pour tous les couples de classes possibles. Durant le test, la méthode requiert la combinaison de toutes les sorties de classifieurs pour qu'une décision soit émise.

5.1 Un –contre -tous (One-against-all Multi class Classifier)

La première stratégie est le prétendu de classificateur un-contre-tous, qui est intuitif et simple. Pour un problème de K classes, le classificateur de multi-classes d'un-contre-tous établit d'abord K classificateurs binaire, chacun d'eux est responsable de chaque classe respectivement. Pour le $n^{\text{ième}}$ classificateur binaire, toutes les données d'entraînement ont besoin d'être impliqué ; le résultat positif est le point de données dans la classe n , et le résultat négatif est le point de données dans les autres classes $n-1$. Chacun des K classificateurs binaires est formé pour trouver la fonction de décision par les SVM binaires, alors toutes les fonctions de décision sont combinées pour obtenir la fonction finale de décision pour le problème de classification multi-classes (Zhou, 2006) :

$$f(x) = \arg \max_{i \in \mathcal{V}S} \sum y_i \alpha_i^n k(x_i, x) + b^n \quad (59)$$

Avec : $\sum_{i \in \mathcal{V}S} y_i \alpha_i^n k(x_i, x) + b^n$ est la fonction de décision de l' $n^{\text{ième}}$ classificateur binaire.

La figure 4 illustre un exemple de classification de trois classes par l'approche un contre tous

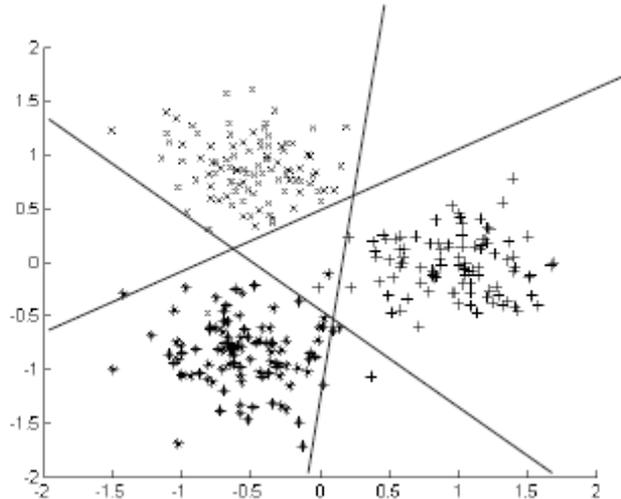


Figure 4 Problème à trois classes : frontières de décision linéaires dans la stratégie Un-contre-Tous

5.2 Un –contre- un (One-against-one Multi-class Classifier)

La deuxième stratégie emploie quelques arrangements et combine des paires de classificateurs binaires pour résoudre le problème de classification de multi-classes ; c'est le classificateur un-contre-un. Dans cette stratégie, un classificateur binaire doit être établi pour chaque paire possible de classes et le nombre total de ces classificateurs binaires est égale à $K(K-1)/2$. Les données d'entraînement de chaque classificateur binaire inclut seulement les données d'entraînement des deux classes impliquées. Un classificateur de multi-classes un-contre-un peut produire seulement la fonction de décision de classe mais pas la probabilité de classe. L'évaluation de probabilité est importante pour quelques applications pratiques, dans lesquelles le classificateur n'est pas employé en isolation mais il est intégré avec certaines autres connaissances de niveau plus élevé (Zhou, 2006).

Le problème de probabilité de classe pour la classification de K-classe ($K > 2$) peut être exprimé plus clairement et formellement comme suit :

Etant donné un ensemble d'entraînement de classification de K-classes, l'observation x , et une étiquette y , le premier problème l'estimation de la probabilité postérieure (μ) à partir de $\mu_{i,j} = p(y = i | y = i \text{ ou } j, x)$ de chaque classificateur binaire, et le deuxième problème est l'estimation de la probabilité $P(y = i | x)$ $i = 1 \dots K$ basé sur l'évaluation par paires du $\mu_{i,j}$ de probabilité pour les nouvelles données. Comme il montré dans la section précédente, la sortie de la fonction de décision des SVM n'est pas une probabilité. Car l'entraînement direct d'un classificateur binaire des SVM avec une sortie de probabilité est très difficile dans la pratique. Généralement, la manière utilisé pour résoudre le problème se déroule en deux phases ; d'abord un classificateur binaire standard de SVM est formée pour obtenir la fonction de décision, et alors l'évaluation du $\mu_{i,j}$ peut être convertée pour estimer la relation entre la probabilité $\mu_{i,j}$ et la sortie de la fonction de décision. La probabilité $\mu_{i,j}$ peut être décrite comme suit :

$$u_{i,j} = F(f_{i,j}(x)) \quad (60)$$

Où $f_{i,j}(x)$ est la fonction de décision du classificateur binaire impliquant la $i^{\text{ième}}$ classe et la $j^{\text{ième}}$ classe. L'estimation directe de la fonction F dans (60) est toujours un facteur difficile ; habituellement avec l'acceptation de F appartenant à un ensemble paramétrisé de fonction, l'évaluation de F est réduite à estimer les paramètres. Pour un exemple, dans [Platt.99] la fonction sigmoïde est suggéré comme modèle de F , et l'équation (60) peut être écrite comme suit :

$$u_{i,j}(x) = \frac{1}{1 + \exp(Af_{i,j}(x) + B)} \quad (61)$$

où A et B sont les paramètres et doivent être estimés par quelques manières. Avec la prétention que toutes les probabilités $\mu_{i,j}$ sont connues pour chaque classificateur binaire, plusieurs méthodes sont suggérées pour obtenir la probabilité $P(y=i|x)$ basée sur la combinaison des évaluations de toutes les probabilités $\mu_{i,j}$.

Par exemple : une règle intuitive de la combinaison des tous les classificateurs binaires doit employer un arrangement de vote et l'assigner à la classe qui gagne le plus paires des comparaisons comme : soit $r_{i,j}$ l'évaluation du $\mu_{i,j}$, et la règle de vote est :

$$f_{\max}(x) = \arg \max_i \sum_{j:j \neq i} I(r_{i,j} > r_{j,i}) \quad (62)$$

là où I est une fonction *indicateur* et $I(x) = 1$, si x est vrai, autrement $I(x) = 0$. Une estimation simple des probabilités des classes peut être calculée selon l'équation (63) :

$$P(y=1|x) = \frac{2 \arg \max_i \sum_{j:j \neq i} I(r_{i,j} > r_{j,i})}{K(K-1)} \quad (63)$$

6. Conclusion

Ce chapitre est organisé selon trois parties majeures. Après un bref historique sur les SVM, la première partie introduit la théorie de l'apprentissage statistique et le principe du risque structurel. La notion de régularisation par la maximisation de la marge y est expliquée.

Ensuite, nous avons analysé le principe de la transformation de données dans l'espace augmenté à l'aide des noyaux de Mercer. Dans la deuxième partie nous avons discuté des stratégies d'apprentissage des SVM sur de gros ensembles de données. En particulier, nous avons détaillé l'algorithme SMO. Enfin, dans la dernière partie, nous avons abordé les différentes stratégies de combinaison pour la classification de données multiclasse. En particulier, nous avons expliqué les approches un contre un et un contre tous, que nous utiliserons dans la partie expérimentale, et nous avons discuté les différents schémas de vote afin de fournir une classe d'appartenance.

ملخص

التعرف الآلي على العيوب الميكانيكية يحضنا حاليا باهتمام كبير. خاصة معرفة مستويات التآكل التي تشكل فرعا هاما نظرا لما تقتضيه الحاجة للتنبؤ بمدى حياة عناصر الآلة للقيام بصيانة مشروطة. العمل المقدم خلال هذه المذكرة يندرج في إطار التشخيص الآلي للأخطاء بواسطة التعرف على الأشكال و الذكاء الاصطناعي. أولا، سنعرض المنهجية المتبعة لتصميم نظام التعرف الآلي على العيوب الميكانيكية. ثم نعرض بالتفصيل مختلف المراحل : من مرحلة اقتناء الإشارة إلى مرحلة التصنيف. بعد ذلك نهتم باختيار المؤشرات التي لها القدرة على التمييز و الغير مشوشة، هذه المرحلة تعتبر جد هامة لتصميم النظام، سنبين خلالها ايجابية تقليص شعاع شكل (VF) الإشارات الحقيقية باستعمال طريقة فيشر. ثم مراقبة قاعدة المعطيات بواسطة طريقة التحليل في المركبات الأساسية (ACP) . أيضا نقترح طريقة المزوجة بين الخوارزميات الجينية (AGs) التحليل و الفاصل الخطي (ADL). و في النهاية نستعمل ثلاثة خوارزميات للتصنيف : الجار الأقرب (k-ppv), برسيبترون متعدد الطبقات (MLP) و الحوامل ذات الهامش الواسع (SVM). بالتالي، سنقدم مشكل التشخيص كمشكل تصنيف. سنستعمل اختبارات القدرة على التصنيف باختيار المؤشرات و بدون اختيارها. تقييم كل النتائج يكون استنادا إلى نسبة الملاحظات المصنفة جيدا للإشارات الحقيقية : للمدرجات و مسننات طائرة هلكوبتر. أفضل النتائج كانت باستعمال طريقة الحوامل ذات الهامش الواسع. المزوجة بين (AGs-ADL) مكنتنا أيضا من الحصول على أداء جيد.

كلمات المفاتيح : الآلات الدوارة, التشخيص الاهتزازي, التعرف على الأشكال, الذكاء الاصطناعي, التصنيف, معالجة الإشارة, تحليل المعلومات.

Abstract

The automatic recognition of mechanical defects and particularly the detection of the levels of degradation constitute a great interest nowadays. Indeed, one would have to gain by predicting the lifetime of machine elements for a condition-based maintenance.

The work we present in this thesis is part of the diagnosis automation by the methods of pattern recognition and artificial intelligence. Initially, we expose the methodology adopted for the system design of a mechanical defects automatic recognition. Then, we describe in detail the various stages while going from the acquisition to the decision. In second part, we are interested in the selection of the most relevant and non-redundant indicators which constitute a significant stage of the system development. Moreover, we show the advantage that there is to use the reduced form vector (FV) by the means of real signals and visualization by principal components analysis (PCA). We propose then, a method based on a hybridization of the genetic algorithms (GAs) and linear discriminant analysis (LDA). Lastly, three algorithms of classification are considered: the k nearest neighbours (knn), the multilayer perceptron (MLP) and the support vector machines (SVM). These choices are dictated by the fact that the diagnosis problem constitutes a perfect example of a classification problem.

Performance tests for the cases with and without selection of indicators are carried out. The evaluation of the results are conducted on the basis of the rate of well classified observations and the real signals of bearings and gears. The best results are obtained by the use of SVM. The GAs-LDA hybridization, also, allowed us to improve and achieve a good performance.

Keywords: rotating machines, vibration diagnosis, pattern recognition, artificial intelligence, classification, signal processing, data mining