

BADJI MOKHTAR –ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR ANNABA



جامعة باجي مختار عنابة

Année : 2018

Faculté des Sciences de l'Ingénieur
Département d'Électronique

Mémoire

Présenté en vue de l'obtention du diplôme de Master

Domaine : Sciences et Technologie

Filière : Automatique

Spécialité : Automatique et Systèmes

Intitulé

Multi-Classification des Défauts de Roulements d'une Machine Asynchrone par
Combinaison de l'Analyse en Composantes Principales et des Séparateurs à Vaste Marges

Présenté par :
Mlle. HAMIS Rabia

DEVANT Le JURY

Président :	H.A. ABBASSI	Professeur	UBMA
Directeur de Mémoire :	F. ARBAOUI	MCA	UBMA
Examineurs :	M.L. SAIDI	MCA	UBMA
	S. BENSAOULA	MCA	UBMA

Dédicaces

Je dédie ce modeste travail :

A mon père, qui m'encouragé tout le long de mes études, et qui m'a fourni tous les moyens nécessaires pour assurer la réussite dans mes études.

A mes cher frères, mes amis, pour leurs encouragements, leur amour, et leur soutien lors des moments difficiles.

Ma reconnaissance va aussi à tous ceux qui, de près ou de loin, ont contribué au succès de mon projet de fin d'étude.

Remerciements

En premier lieu, je remercie Allah le Tout-Puissant

الله سبحانه و تعالى

Qui m'a donné l'opportunité et la volonté de mener ce travail

الحمد لله

Le travail présenté dans ce mémoire a été effectué au Laboratoire d'Automatique et Signaux d'Annaba (LASA) du Département de L'électronique, Université Badji Mokhtar.

D'abord, mes propres remerciements pour mon encadreur **Dr. Arbaoui F. Maître de conférences** pour m'avoir accepté pour l'encadrement de ce sujet. Merci pour la confiance qu'il m'a accordé pour le traiter et l'accomplir. En effet, ce travail m'intéresse beaucoup, merci pour ses conseils précieux et la clarté de ses explications, et aussi pour sa grande contribution. Je tiens à lui exprimer toute mon respect et ma gratitude.

Mes remerciements s'adressent aussi aux membres du jury :

Mr. H.A. ABBASSI, Professeur de l'université de Badji Mokhtar Annaba.

Mr. M.L. SAIDI, MCA de l'université de Badji Mokhtar Annaba.

Mr. S. BENSAOULA, MCA de l'université de Badji Mokhtar Annaba.

Pour avoir accepté d'évaluer mon travail et pour avoir pris le temps de soigneusement étudier ce manuscrit. Leurs conseils ne seront qu'enrichissant et bénéfiques.

ملخص

العنوان: تصنيف عيوب المدرجة الكروية للآلة غير المتزامنة باستعمال تحليل المكونات الرئيسية وفاصل الهوامش الكبير

الخلل الناتج عن العيوب مثل تآكل أو تلف في المدرجة الكروية هو أحد الأسباب الرئيسية لسوء الأداء في الآلات غير المتزامنة، الذي يؤدي إلى خسائر مادية معتبرة.

الغرض من هذه المذكرة هو التشخيص الآلي لعيوب المدرجة الكروية على أساس التحليل الاهتزازي.

الشعاع المكون من المعلمات (الخصائص أو المؤشرات)، أو ما يسمى الشعاع الشكلي، تستخرج مؤشرات الإحصائية والطيفية لنتائج أمثل من الإشارات الاهتزازية الأصلية ومن الإشارات الاهتزازية التي تم معالجتها باستخدام تحليل المكونات الرئيسية.

أجريت الدراسة باستعمال قاعدة بيانات جامعة: *Case Western Reserve University* ، حيث تم الحصول على الإشارات الاهتزازية في أوضاع وظروف تشغيل مختلفة (في الحالة السليمة وفي حالة وجود العيب، تغيير في السرعة، تغيير في الحمولة وتغيير في شدة العيب).

علاوة على ذلك تمت إجراء دراسة مقارنة مع نوع آخر من المصنفات: الشبكة العصبية الاصطناعية. وقد أثبتت الطريقة المنتقاة أي فاصل الهوامش الكبير متعدد الأصناف، فعالية أفضل من حيث وقت ومعدل تصنيف بالمقارنة.

في مرحلة التصنيف، تم الاحتفاظ بفاصل الهوامش الكبير متعدد الأصناف.

أجريت مقارنة النتائج بين التحليل الزمني مع وبدون استخدام تحليل المكونات الرئيسية، وأظهرت النتائج فعالية هذا النهج.

الكلمات الدالة: فاصل الهوامش الكبير، المعلمات الإحصائية، تحليل المكونات الرئيسية، إشارة اهتزازية، عيوب المدرجة الكروية.

Résumé

Titre : Multi-Classification des Défaits de Roulements d'une Machine Asynchrone par Combinaison de l'Analyse en Composante Principale et des Séparateurs à Vaste Marges

Le défaut de roulement est l'une des causes dominantes de dysfonctionnements dans les machines asynchrones, entraînant une perte économique significative.

L'objectif de ce mémoire est le diagnostic automatisé des défauts de roulements basé sur l'analyse vibratoire.

Un vecteur de paramètres (Indicateurs ou caractéristiques), appelé aussi vecteur forme, est extrait de chacun des signaux de vibration bruts et spectres, il se compose d'indicateurs statistiques et spectraux pour une performance optimale dans un premier temps, ensuite les paramètres sont extraits des signaux traités en utilisant l'analyse en composantes principales (ACP).

L'étude a été menée sur la base de données de la Case Western Reserve University, où les signaux ont été acquis dans différents modes et conditions de fonctionnement (sain et avec défauts, variations de vitesse, variation de charges et variation de la sévérité du défaut).

Dans la phase de classification, les séparateurs à vaste marges (SVM) multi-classes ont été retenus.

Une comparaison des résultats dans le domaine temporel avec et sans analyse en composantes principales a été effectuée, les résultats obtenus montrent l'efficacité de cette approche.

Ensuite une étude comparative a été effectuée avec les classificateurs à base Réseaux de Neurones Artificiels (RNA). On a constaté une meilleure efficacité en termes de temps et taux de classification du multi-SVM par rapport aux RNA.

Mots clés : Machines à Vecteurs de Supports (SVM), Paramètres statistiques, analyse en composantes principales (ACP), défaut roulement, signal vibratoire.

Abstract

Title: Induction Machine bearings defects Multi-Classification by combination of Principal Component Analysis and Support Vector Machines

The rolling defect is one of the main causes of dysfunctions in induction machines, resulting in a significant economic loss.

The main purpose of this study is an automated diagnosis of bearing defects based on vibratory analysis.

A vector of parameters (Indicators or characteristics), also known as vector form and composed of statistical and spectral indicators for optimal performance, is extracted from the original vibration signals and from the signals processed using principal component analysis (PCA).

The study was carried out on the *Case Western Reserve University* database, where the signals were acquired in different modes and operating conditions (healthy and with defaults, speed variations, load variation and variation in defect severity).

In the classification phase, the Multi-Class Support Vector Machine (SVM) was chosen.

A comparison of the results of the temporal domain with and without principal component analysis was performed; the results obtained show the effectiveness of this approach.

As a result, a comparative study was achieved on an other type of classifiers: Artificial Neural Network (ANN). A better efficiency in terms of time and classification rate goes to the multi-class SVM compared to the ANN.

Key words: Support Vector Machines (SVM), Statistical parameters, principal component analysis (PCA), rolling failure, vibration signal.

Liste des figures

Chapitre 1

1.1	Constitution d'un roulement	4
1.2	(a) Ecaillage d'une bague intérieure, (b) Criques corrosion.....	5
1.3	Photos représentant les différents défauts de roulements. (a) Corrosion, (b) : Cratère bille, (c) : Cratère bague intérieure, (d) : Cratère bague extérieure, (e) : mécanismes généraux d'usure en présence de contaminant	5
1.4	Spectre théorique d'un défaut	6
1.5	Fréquence caractéristique d'un défaut de type écaillage affectant la bague extérieure	6
1.6	Fréquences caractéristiques du roulement à billes	7
1.7	Techniques d'extraction des caractéristiques	9

Chapitre 2

2.1	Représentation des étapes de la construction du modèle PCA.....	15
2.2	Illustration graphique de l'effet du centrage et de réduction d'une distribution bivariable, (a) Distribution d'origine, (b) Centrage et réduction de cette distribution	15

Chapitre 3

3.1	Illustration d'une classification binaire linéaire SVM	23
3.2	Illustration du choix de la marge avec un hyperplan optimal.....	23
3.3	Schéma de changement des représentations non linéaires vers un nouvel espace	26
3.4	Classification SVM binaire avec le noyau linéaire de Kernel	27
3.5	Classification SVM binaire avec le noyau poly 2 ^{ième} degré de Kernel	27
3.6	Classification SVM binaire avec le noyau Gaussien de Kernel	27
3.7	Classification SVM un contre tous (One Vs All)	29

3.8	Schéma des hyperplans linéaires pour la classification un contre tous (One vs all) / n=4	29
3.9	Classification SVM un contre un (One Vs One)	30

Chapitre 4

4.1	Le Banc d'essai de la CWRU.....	34
4.2	Signal original cas sain	35
4.3	Auto corrélation du signal original cas sain.....	35
	Figure 4.4 Variation de quelques.....	35
4.5	Signal original avec défaut bille.....	35
4.6	Auto corrélation du signal ave défaut bille.....	35
4.7	Variation de quelques paramètres temporels et fréquentiels du signal avec défaut bille.....	36
4.8	Signal original avec défaut bague interne.....	36
4.9	Auto corrélation du signal avec défaut bague interne.....	36
4.10	Variation de quelques paramètres temporels et fréquentiels du signal avec.....	37
4.11	Signal original avec défaut bague externe.....	37
4.12	Auto corrélation du signal avec défaut bague externe défaut bague interne.....	37
4.13	Variation de quelques paramètres temporels et fréquentiels du signal avec défaut bague externe.....	37
4.14	Schéma présentant les étapes de diagnostic.....	38

Liste des tableaux

4.1. Données vibratoires des défauts avec une fréquence d'échantillonnage de 12K HZ / CWRU.....	34
4.1. Composantes principales obtenues pour les deux cas.....	41
4.3 : Récapitulatif des résultats obtenus :.....	42

Table des matières

Dédicaces	
Remerciements	
Résumé / Abstract / ملخص	
Liste des figures	
Introduction générale :.....	1
Chapitre 1. Défauts Roulements et analyse vibratoires	
1.1. Roulement :	4
1.2. Types des défauts roulements :.....	4
1.2.1. Écaillage :.....	5
1.2.2. Grippage :.....	5
1.2.4. Détérioration des cages:.....	5
1.3. Fréquences caractéristiques et défauts :.....	5
1.3.1. Les paramètres caractéristiques d'un roulement :	7
1.4. Définition d'une vibration :.....	7
1.4.1 Analyse vibratoire :.....	8
1.4.2. Outils d'analyse des vibrations :.....	8
1.5. Les Méthodes d'analyse Vibratoire :.....	8
1.6. Extraction des caractéristiques :.....	8
1.7 Sélection et réduction des caractéristiques :.....	10
1.7.1 Caractéristiques ou paramètres temporelles :.....	10
1.7.2 Caractéristiques ou paramètres fréquentielles :	11
Chapitre 1 : L'analyse en Composantes Principales	
2.1. Introduction :.....	14
2.2. Principe de l'analyse en composantes principales :.....	14
2.3. Détermination d'une structure optimale du modèle ACP :.....	19
1. Le pourcentage cumulé de la variance totale (PCV) :.....	19
2. la méthode de validation cumulée (PRESS) :.....	19
3. le Scree test :	20

Chapitre 3 : Multi classification par les séparateurs à vaste marges

3.1. les algorithmes d'apprentissage automatique :	22
3.2. Séparateur à vaste marge (SVM) :	22
3.3. Hyperplan, marge et support vecteur :	22
3.4 Séparateur à vaste marge dans le cas linéaire :	23
3.4.1 Maximisation de marge :	23
3.4.2 Calcul de la marge :	24
3.4.3 Calcul de l'Hyperplan optimal :	24
3.5. Séparateur à vaste marge dans le cas non linéaire :	25
3.5.1 Les Noyaux :	26
3.6. Types multi-class des séparateurs à vaste marge :	27
3.6.1 Multi classification par l'approche un contre tous. :	28
3.6.2 Multi classification par l'approche un contre un :	30

Chapitre 4 : Résultats de simulation

4.1. Matériels et méthodes :	33
4.1.1. Présentation du Banc d'Essai de la CWRU :	33
4.1.2. Présentation des signaux et quelques-unes de leurs caractéristiques :	34
4.2. Méthodologie de diagnostic adopté :	39
4.2.1 .L'analyse en composantes principales :	39
4.2.2 .Multiclassification SVM :	39
4.3. Configuration de l'expérience :	40
4.3.1. Les données de défaut de roulement :	40
4.3.2. Sélection des caractéristiques en utilisant l'analyse en composantes principales :	40
4.3.3. Traitement des données et diagnostic utilisant l'approche multi-classes SVM :	41
4.4. Discussion des résultats :	43
Conclusion générale :	44

INTRODUCTION GENERALE

Introduction générale

Les machines industrielles automatisées sont composées de divers sous-systèmes, tels que le système de commande électrique, unités de commande(API), les actionneurs (vérins, moteurs asynchrones, distributeurs) ces composants sont impliqués dans l'exécution des opérations désirées des machines.

Les composants structurels d'une machine tels que les roulements à billes, sont soumis à l'usure au fil du temps, la surveillance de ces composants est donc nécessaire pour réduire régulièrement le risque de défaillances et de pannes par un diagnostic anticipatif ou précoce des défauts et assurer la disponibilité et l'utilisation efficace et optimale des machines qui sont les facteurs clés de l'économie de l'entreprise manufacturière.

L'objectif principal de ce mémoire est de mettre en œuvre un algorithme de surveillance de l'état des roulements à l'aide des machines à vecteurs de support.

Le succès de cette méthode est justifié par les solides bases théoriques qui la soutiennent. Il existe en effet un lien direct entre la théorie de l'apprentissage statistique et l'algorithme d'apprentissage du SVM.

La méthode SVM est une technique de classification particulièrement bien adaptée pour la reconnaissance des formes et traite des données très diversifiées ayant plusieurs dimensions telles que les textes, les images, la voix, les diagnostics médicales, la bio-informatique pour la reconnaissance de gènes ...etc. Le champ d'application des SVM est large et représente une méthode de classification intéressante.

Ce mémoire comporte principalement 4 chapitres. Le premier est consacré à la présentation des différents types de défauts des roulements ainsi que l'analyse vibratoire. Le deuxième chapitre, introduit la théorie de l'analyse en composante principale et son utilisation en diagnostic. Dans le troisième chapitre, nous détaillons la classification supervisée par les séparateurs à vaste marges.

Le dernier chapitre présente l'application expérimentale du système de diagnostic combinant l'analyse en composante principale avec les séparateurs à vastes marges sur les données du banc d'essai de la CWRU.

Le dernier chapitre présente l'application expérimentale du système de diagnostic sur les données du banc d'essai de la CWRU avec deux techniques de classification : les réseaux de neurones artificiels (RNA) et les séparateurs à vaste marge (SVM). Une comparaison des performances est donnée et les résultats obtenus sont interprétés et discutés.

On achèvera, enfin par une conclusion générale et des perspectives.

Chapitre

1

*Défauts Roulements & Analyse
Vibratoire*

Chapitre 1. Défauts Roulements & Analyse Vibratoire

1.1. Roulement :

Le roulement assure la liaison mobile en rotation avec un minimum de frottement entre deux éléments, tout en supportant des charges radiales ou axiales plus ou moins importantes [8].

De manière générale, les roulements sont constitués ; de deux bagues (externe et interne) l'une fixe et l'autre tournante avec le rotor. Elles sont séparées par un ensemble d'éléments roulants qui peuvent être des billes, des rouleaux ou des aiguilles selon l'application envisagée, une cage séparant à intervalles réguliers et guidant les éléments roulants.

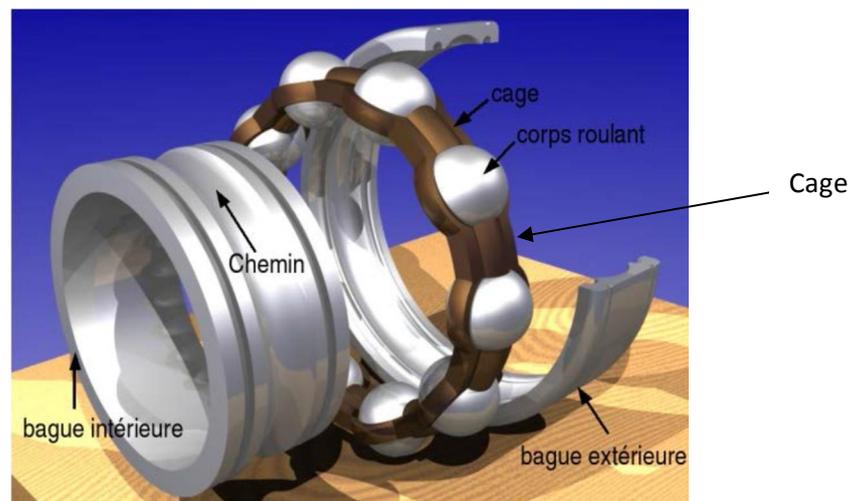


Figure1.1 : Constitution d'un roulement

1.2. Types des défauts roulements :

Selon leurs applications, et la criticité des machines sur lesquelles ils sont montés, les roulements nécessitent plus ou moins une surveillance et un entretien préventif. Pour cela le suivi des niveaux de vibration et de température, sont les sources d'information les plus utilisées.

Les aspects de ces détériorations se manifestant dans la plupart des cas par un arrachement de matière plus ou moins important, comme par exemple :

1.2.1. Écaillage :

Se caractérise par des traces de fissuration et d'arrachement de fragments de matière.

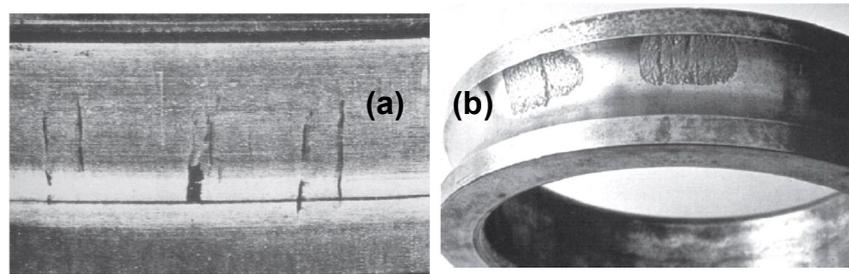


Figure 1.2 : (a) Ecaillage d'une bague intérieure, (b) Criques de corrosion

1.2.2. Grippage :

Des traces brunes dues à l'échauffement, des déformations de corps roulants, et des micro-fusions peuvent apparaître.

1.2.3. Coups, fissures, cassures :

Se manifestant avec des traces visibles d'outils, et des ruptures des bagues.

1.2.4. Détérioration des cages :

Se manifestant sous plusieurs formes : déformation ; usure ou rupture des cages.

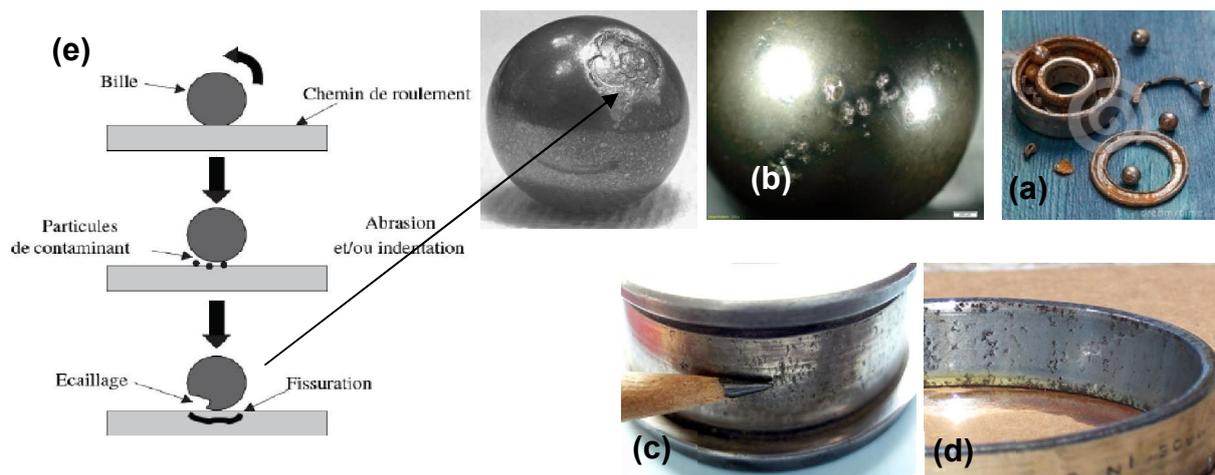


Figure 1.3 : photos représentant les différents défauts de roulements. (a) Corrosion, (b) : Cratère bille, (c) : Cratère bague intérieure, (d) : Cratère bague extérieure, (e) : mécanismes généraux d'usure en présence de contaminant.

1.3. Fréquences caractéristiques et défauts

La présence d'un défaut au niveau d'un roulement à un impact sur le spectre de son signal vibratoire tel qu'une augmentation de l'amplitude de la vibration aux fréquences du roulement,

mais aussi la présence des vibrations aux harmoniques de ces fréquences ainsi qu'à des fréquences de modulation, le spectre va faire apparaître un peigne de raie à la fréquence du défaut et des bandes latérales à la fréquence de rotation de l'arbre (F_0).

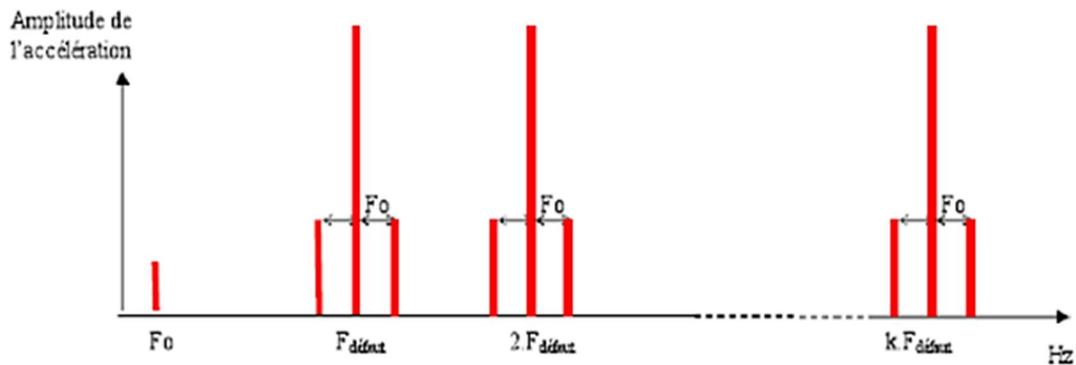


Figure 1.4 : Spectre théorique d'un défaut [3]

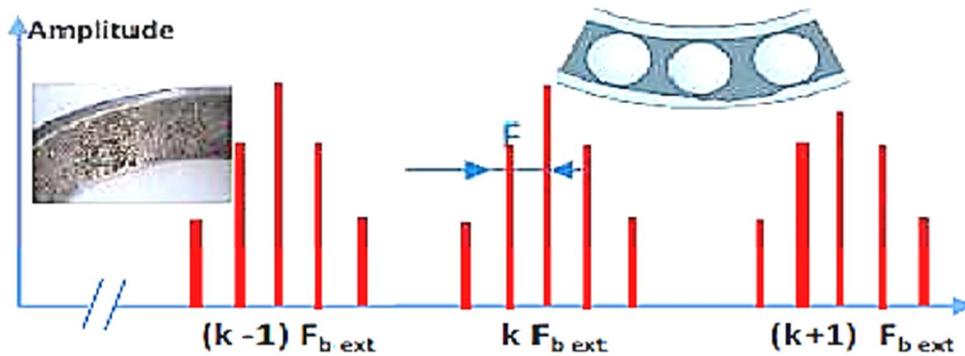


Figure 1.5 : Fréquence caractéristique d'un défaut de type écaillage affectant la bague extérieure [3]

Différentes fréquences sont obtenues pour des défauts, les quatre fréquences caractéristiques sont les suivantes [6]:

$$\left\{ \begin{array}{l} f_o(Hz) = f_r \left(\frac{N}{2} \right) \left(1 - \frac{D_b}{D_p} \cos\theta \right) \\ f_i(Hz) = f_r \left(\frac{N}{2} \right) \left(1 + \frac{D_b}{D_p} \cos\theta \right) \\ f_c(Hz) = f_r \left(\frac{1}{2} \right) \left(1 - \frac{D_b}{D_p} \cos\theta \right) \\ f_b(Hz) = f_r \left(\frac{D_b}{2D_p} \right) \left(1 - \frac{D_b^2}{D_p^2} \cos^2\theta \right) \end{array} \right. \quad (1.1)$$

f_b (Ball Frequency [BF]) : La fréquence de passage d'un défaut de bille ;
f_o (Ball Passing Frequency Outer Race [BPFO]) : La fréquence de passage d'un élément roulant sur un défaut de bague extérieure ;
f_c (Cage Frequency [CF]) : La fréquence de passage d'un défaut de cage ;
f_r la fréquence de rotation mécanique ;
(Ball Passing Frequency Inner Race [BPFI]) : La fréquence de passage d'un élément roulant sur un défaut de bague intérieure.

1.3.1. Les paramètres caractéristiques d'un roulement :

Sont les suivantes (figure 1.6) :

- N Le nombre de billes,
- D_b le diamètre de bille,
- D_p le diamètre moyen du roulement,
- θ l'angle de contact.

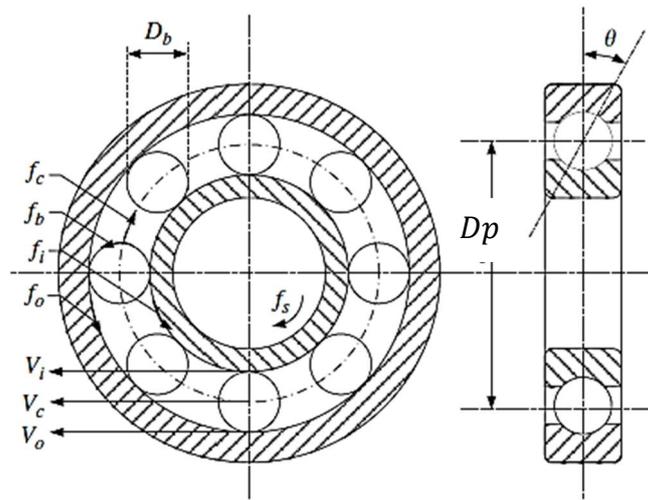


Figure 1.6 : Fréquences caractéristiques du roulement à billes [7]

1.4. Définition d'une vibration :

La vibration d'une machine soumise à une force périodique peut être décrite en termes de déplacement, de vitesse ou d'accélération. La vitesse du mouvement vibratoire correspond à la variation de son déplacement pour une unité de temps. L'accélération représente une variation de la vitesse par unité de temps.

Le déplacement exprimé en [μm], la vitesse en [mm/s] et l'accélération en [m/s^2].

A ces grandeurs sont associées une période temporelle et une phase. Ainsi, dans le cas d'une machine tournante, une vibration élémentaire peut se mettre sous la forme :

$$x(t) = X \cdot \sin(2\pi \cdot f \cdot t) \quad (1.2)$$

Une propriété fondamentale relie les trois représentations déplacement D , vitesse V et accélération a :

$$D = \frac{V}{2\pi f} = \frac{a}{(2\pi f)^2} \quad (1.3)$$

1.4. Analyse vibratoire

La modification de la vibration d'une machine constitue souvent la première manifestation physique d'une anomalie, cause potentielle de dégradations, voire de pannes par effet de fatigue au sein de la matière.

La surveillance des vibrations peut permettre de déceler l'apparition d'efforts perturbateurs à un stade précoce. La facilité relative de son installation rend l'analyse des signaux vibratoires une des techniques les plus utilisées pour la surveillance et le diagnostic des défauts des machines tournantes, cela dit il existe d'autres techniques de surveillance tel que :

- La thermographie infrarouge
- L'analyse des mesures ultrasonore

1.4.1. Outils d'analyse des vibrations

Les outils utilisés pour mesurer les vibrations se sont considérablement améliorés au cours des dernières années. Le capteur de choix pour la plupart des données de vibration sur les machines industrielles est un accéléromètre : c'est un capteur qui convertit le mouvement mécanique en un signal électrique.

1.5. Les Méthodes d'analyse Vibratoire :

Les techniques du domaine temps-fréquence sont principalement utilisées pour traiter des signaux vibratoires non stationnaires. On y distingue la transformée de Fourier courte, la distribution de Wigner-Ville, l'analyse en ondelettes notamment la décomposition en paquets d'ondelettes (Wavelet Packet Decomposition : WPD), la décomposition modale empirique et la transformée de Hilbert-Huang.

L'une des Méthodes la plus efficace pour l'analyse des signaux non stationnaires est l'analyse en composante principale (ACP ou PCA en anglais pour principal component analysis). C'est une méthode de la famille de l'analyse des données et plus généralement de la statistique multivariée, qui consiste à transformer des variables liées entre elles (dites « corrélées » en statistique) en nouvelles variables décorréelées les unes des autres. Ces nouvelles variables sont nommées « composantes principales », ou axes principaux. Elle permet de réduire le nombre de variables et de rendre l'information moins redondante.

1.6. Extraction des caractéristiques :

Les paramètres statistiques quantifient les informations sur la forme et l'énergie du signal, elles peuvent refléter l'état de santé mécanique sous différents aspects et peuvent se compenser mutuellement à différents niveaux de défaut.

L'extraction des caractéristiques repose essentiellement sur des techniques issues du traitement du signal. Ces techniques peuvent être regroupées dans deux catégories principales selon que le signal à traiter est stationnaire ou non stationnaire. La figure ci-dessous donne un aperçu des techniques utilisées pour l'extraction des caractéristiques.

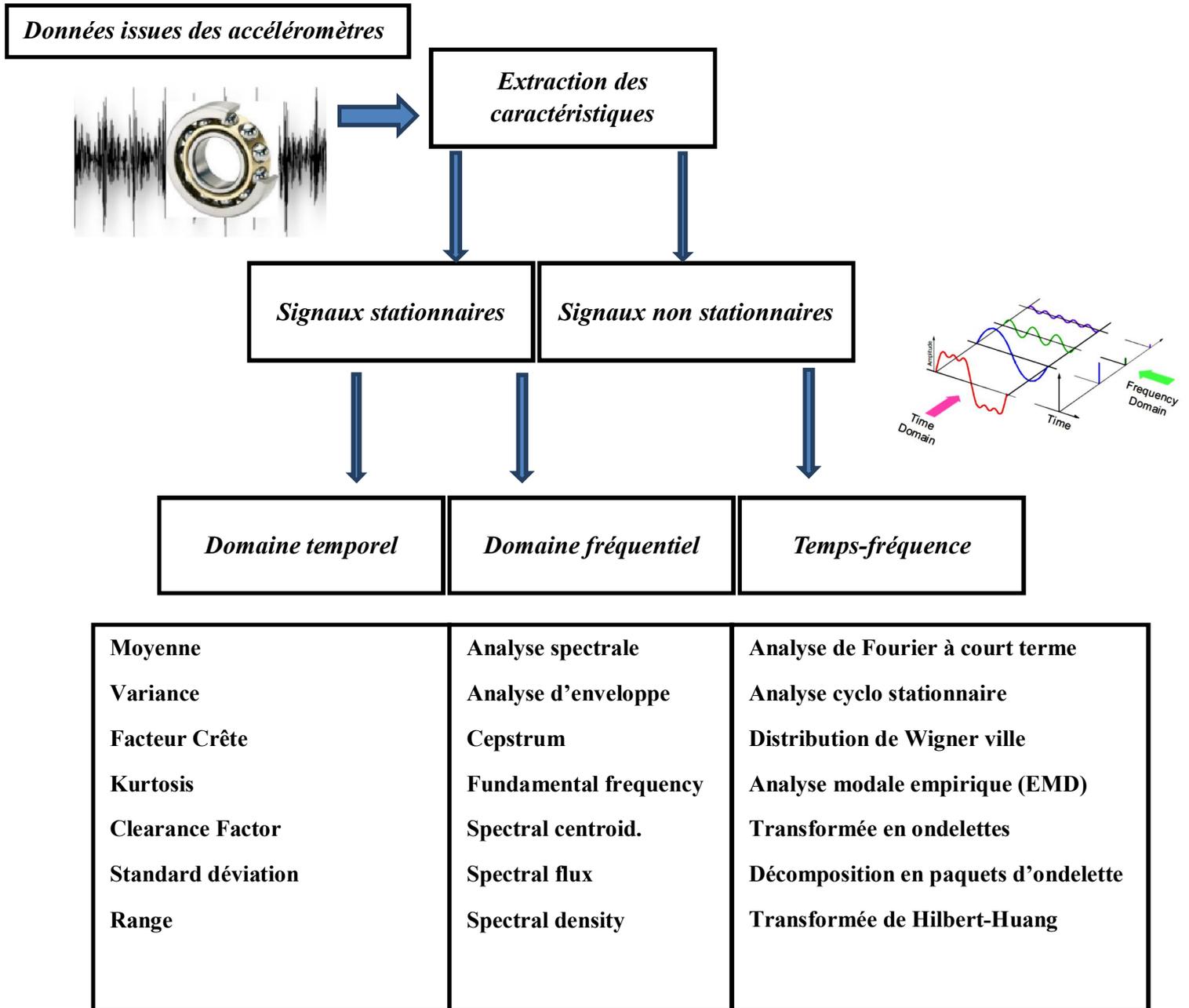


Figure 1.7 : Techniques d'extraction des caractéristiques [12]

1.7. Sélection et réduction des caractéristiques :

Après l'étape d'extraction des paramètres, l'étape de sélection est cruciale pour trouver quels sont les paramètres les plus importants et les plus fiables pour la détection du défaut. En effet, si les paramètres sont sélectionnés avec peu de pouvoir de discrimination, la conception ultérieure d'un classificateur conduirait à de mauvaises performances.

D'autre part, si les paramètres avec des informations riches sont sélectionnés, la précision de la classification augmente et le temps de calcul est réduit.

Afin d'avoir un système de diagnostic performant, il est nécessaire d'employer un processus de sélection des indicateurs les plus pertinents, permettant d'améliorer les performances de la classification

Les techniques de sélection d'indicateurs sont divisibles en trois catégories, selon la manière dont elles interagissent avec le classificateur [11] :

- Les approches avec filtres
- Les approches avec enveloppes (wrapper)
- Les approches intégrées (embedded)

Dans le cas de la classification par le séparateur à vaste marge on peut citer les travaux suivants :

- L'élimination récursive des indicateurs (Recursive Features Elimination (RFESVM)), présentée par Guyon [1].
- Dans [4], les auteurs proposent une approche SA-SVM basé sur le recuit simulé pour l'optimisation des paramètres des SVMs.
- Dans [10], les auteurs proposent un algorithme génétique AG pour la sélection d'un sous ensemble optimal d'indicateurs, dans le cadre de diagnostic automatisé des machines tournantes. Le sous ensemble sélectionné est utilisé pour entraîner deux classificateurs différents à savoir, les SVMs et les réseaux de neurones artificiels RNA

1.7.1. Caractéristiques ou paramètres temporelles

Elles concernent le calcul des paramètres statistiques du signal. Elles sont utilisées en classification, détection et diagnostic de défaillances où elles donnent des résultats acceptables.

Parmi les caractéristiques temporelles, on peut citer la moyenne, le carré de la moyenne quadratique (Root Mean Square : RMS) qui correspond à l'énergie moyenne du signal, la valeur de son pic qui peut renseigner sur la taille d'un défaut, le Kurtosis qui caractérise l'aspect impulsif du signal, le facteur de crête qui mesure les augmentations rapides des vibrations dans les roulements et le Skewness utilisé pour évaluer l'état de surface d'un composant [11].

$$\text{Moyenne : } x_m = \sum_{n=1}^N \frac{x(n)}{N} \quad (1.4)$$

$$\text{Variance : } x_{\text{var}} = \frac{\sum_{n=1}^N (x(n) - x_m)^2}{N} \quad (1.5)$$

$$\text{Kurtosis } x_{\text{kur}} = \frac{\sum_{n=1}^N (x(n) - x_m)^4}{(N - 1)\sigma^4} \quad (1.6)$$

$$\text{Clearance factor : CLF} = \frac{x_{\text{max}}}{\left(\frac{1}{N} \sum_{n=1}^N \sqrt{|x(n)|}\right)^2} \quad (1.7)$$

$$\text{Mean absolute deviation : MAD} = \sum_{n=1}^N \frac{|\sum_{n=1}^N (x(n) - x_m)|}{N} \quad (1.8)$$

$$\text{Range : } x_{\text{range}} = |x_{\text{max}} - x_{\text{min}}| \quad (1.9)$$

$$\text{Standarddeviation : STD} = \sqrt{\frac{\sum_{n=1}^N (x(n) - x_m)^2}{N}} \quad (1.10)$$

$$\text{PeakmagnitudetoRMSratio: Crest factor} = \frac{x_{\text{max}}}{\text{RMS}} \quad (1.11)$$

Où $x(n)$ / ($n= 1, \dots, N$) est l'amplitude au point d'échantillonnage n et N est le nombre de points d'échantillonnage. x_m est le moyen de x , σ est l'écart type de x .

1.7.2. Caractéristiques ou paramètres fréquentielles :

L'analyse fréquentielle consiste à visualiser l'énergie du signal à différentes fréquences. En connaissant les conditions critiques de fonctionnement du composant (vitesse, charge, géométrie des éléments constitutifs, ...etc.). On peut extraire des caractéristiques permettant d'isoler les défauts. Ces caractéristiques sont aussi appelées signatures fréquentielles.

La fréquence moyenne :

$$\text{Mean Frequency : } f_{\text{mean}} = \frac{\sum_{n=1}^N I_n \cdot f_n}{\sum_{n=1}^N I_n} \quad (1.11)$$

Où f_n et I_n sont les fréquences et les amplitudes au point n du spectre du signal

$$\text{Mean power frequency: MNF} = C_s = \frac{\sum_{n=1}^N P_n \cdot f_n}{\sum_{n=1}^N P_n} \quad (1.12)$$

Où P_n est la puissance spectrale du signal au point n [9].

$$\text{Densité spectrale de puissance DSP} = \frac{\sum_{n=1}^N P_n}{N} \quad (1.13)$$

$$\text{Spectral Entropy } E_{n_s} = -\sum_{n=1}^N P_n \log P_n \quad (1.14)$$

$$\text{Short time Energy} = \sum_{n=1}^N [x[m] W[n - m]]^2 \quad (1.15)$$

Références Bibliographiques

- [1] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- [2] Ilyes, KHELFI. (2014) *Diagnostic des machines tournantes par les techniques de l'intelligence artificielle*. Thèse de doctorat. Université de Badji Mokhtar Annaba.
- [3] Ishak T. (2012) *Extraction d'indicateurs robustes pour le diagnostic des défauts mécaniques : Comparaison de L'EMD et des ondelettes (WT)*. Thèse de doctorat. Université Ferhat Abbas de Sétif 1.
- [4] Lin, S. W., Lee, Z. J., Chen, S. C., & Tseng, T. Y. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied soft computing*, 8(4), 1505-1512.
- [6] McNerny S.A. and Dai Y. (2003), Basic vibration signal processing for bearing fault detection, *IEEE Transactions On Education*, Vol.46, No.1.
- [7] Ming, Y., Chen, J., & Dong, G. (2011). Weak fault feature extraction of rolling bearing based on cyclic Wiener filter and envelope spectrum. *Mechanical Systems and Signal Processing*, 25(5), 1773-1785.
- [8] Morel, J. (1992). *Vibrations des machines et diagnostic de leur état mécanique*. Collection de la Direction des études et recherches d'électricité de France.
- [9] Oskoei, M. A., & Hu, H. (2008). Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE transactions on biomedical engineering*, 55(8), 1956-1965.
- [10] Samanta, B., Al-Balushi, K. R., & Al-Araimi, S. A. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence*, 16(7), 657-665.
- [11] Tang, D., Wei, F., Qin, B., Liu, T., & Zhou, M. (2014, August). Coooolll: A deep learning system for twitter sentiment classification. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (pp. 208-212).
- [12] W. Yan, H. Qiu, and N. Iyer, (2008). Feature extraction for bearing prognostics and health

Chapitre

2

*L'analyse en Composantes
Principales*

Chapitre 2. L'analyse en Composantes Principales

2.1. Introduction :

L'analyse en composantes principales (ACP) est une méthode statistique d'analyse des données multivariée qui fait la projection orthogonale linéaire, elle projette des observations multidimensionnelles représentées dans un espace de dimension m (où m étant le nombre des variables observées) sur un sous espace de dimension inférieure ($\ell < m$) tout en minimisant la perte des informations, aussi en maximisant la variance des projections et la minimisation de la distance quadratique moyenne entre les points de données et leurs projections. On obtient, alors, un nouvel espace de variables appelées les composantes principales d'une variance maximale et non corrélées entre eux, ces derniers sont en combinaisons linéaires avec les variables originelles. Ces nouvelles variables sont ensuite utilisées à des fins de modélisation, d'estimation d'état, de diagnostic de fonctionnement de système.

La méthode ACP est fortement établie pour la réduction de dimensionnalité et l'analyse multivariée qui sont très importants dans des nombreux domaines comprennent la compression des données de grande dimension, le traitement d'image, la visualisation, l'analyse des données exploratoires, la reconnaissance des formes, et la prédiction de séries temporelles [1,2,5,6].

2.2. Principe de l'analyse en composantes principales :

L'ACP consiste à remplacer une famille de variables par de nouvelles variables appelées Composantes principales (CPs). Ces dernières sont de variances maximales et non corrélées deux à deux. Elles sont des combinaisons linéaires des variables originelles.

Les différentes étapes pour la construction du modèle PCA sont hiérarchisées comme suit :

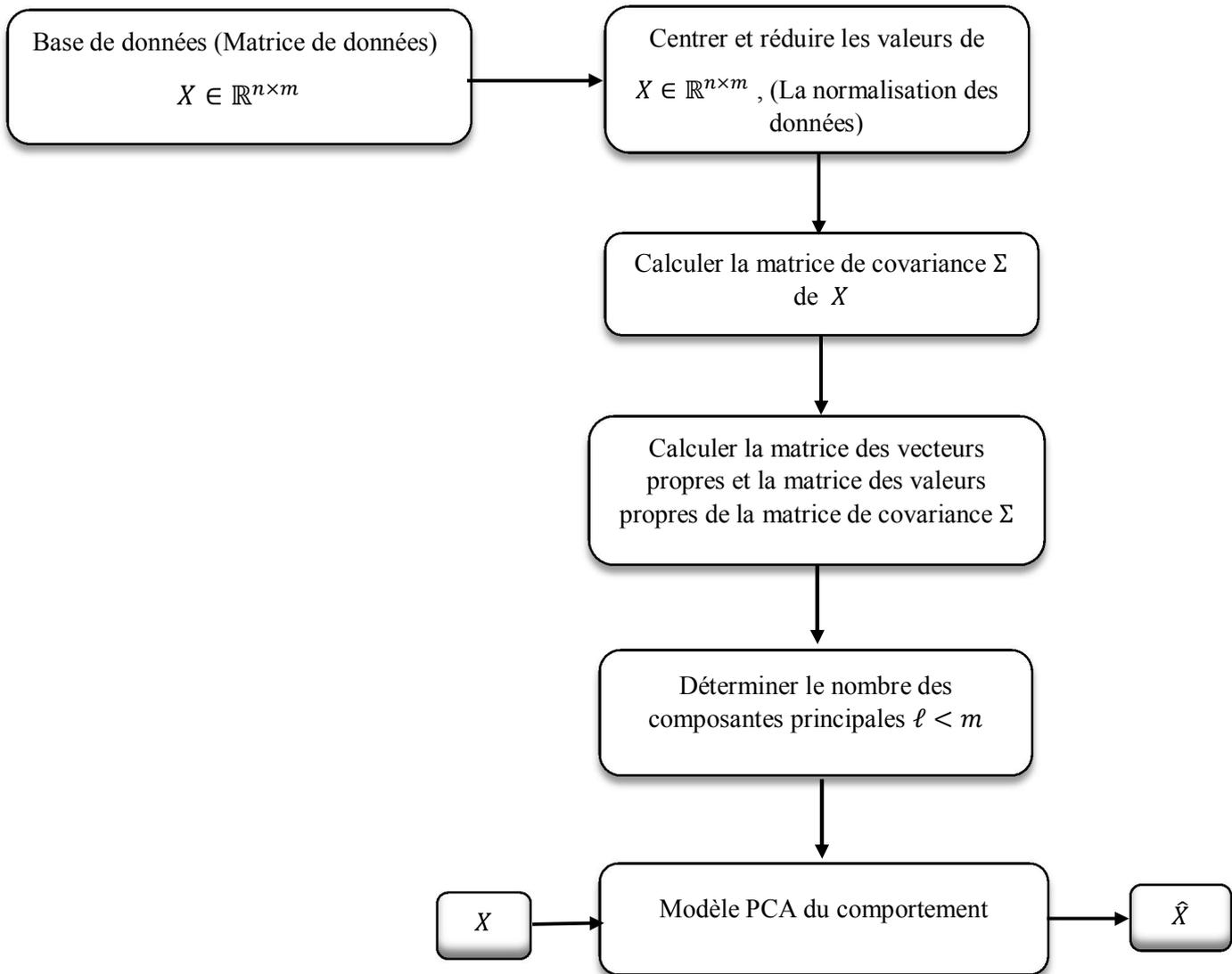


Figure 2.1 : Représentation des étapes de la construction du modèle PCA [2].

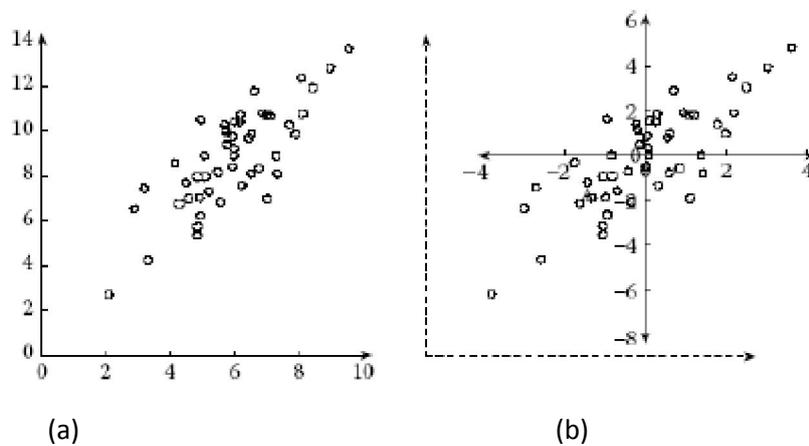


Figure 2.2 : Illustration graphique de l'effet du centrage et de réduction d'une distribution bivariable, (a) Distribution d'origine, (b) Centrage et réduction de cette distribution.

Considérons un ensemble de données collectées lors d'un fonctionnement normal du système étudié. Ces données peuvent être représentées par une matrice

$$X = [x(1), \dots, x(N)]^T \in \mathbb{R}^{N \times m} \quad (2.1)$$

où N représente le nombre des observations et m représente le nombre des variables mesurées. Chaque ligne de la matrice de données X représente une observation sous forme d'un vecteur de mesures collectées à un instant k , généralement centrées

$$x(k) = [x_1(k), \dots, x_m(k)]^T \in \mathbb{R}^m \quad (2.2)$$

où $x_j(k)$ ($j = \{1, \dots, m\}$) est la mesure de la variable j à l'instant k ,

Par définition, la matrice de covariance est donnée par :

$$\Sigma = E\{XX^T\} = \frac{1}{N}XX^T \in \mathbb{R}^{m \times m} \quad (2.3)$$

Selon le principe de l'ACP, on suppose qu'un vecteur de composantes $\hat{t} \in \mathbb{R}^\ell$ est associé à chaque vecteur d'observation dont il optimise la représentation au sens de la minimisation de l'erreur d'estimation de x ou la maximisation de la variance de \hat{t} . A chaque instant k , les vecteurs \hat{t} et x sont liés par une transformation linéaire de type $\hat{t}(k) = \hat{P}^T x(k)$ telle que la matrice de transformation $\hat{P} \in \mathbb{R}^{m \times \ell}$ vérifie la condition d'orthogonalité $\hat{P}^T \hat{P} = I_\ell \in \mathbb{R}^{\ell \times \ell}$.

L'optimisation de la représentation en se basant sur la matrice de projection \hat{P} est obtenue par la minimisation de l'erreur quadratique d'estimation de x . Notons par \hat{P} la matrice optimale de représentation, celle-ci peut être donnée par :

$$\hat{P} = \underbrace{\arg \min}_{\hat{P}} \{J_e(\hat{P})\} \quad (2.4)$$

Où J_e étant le critère d'erreur d'estimation de PCA qui devrait être minimisé.

Sous la contrainte d'orthogonalité de la matrice de projection, on peut écrire [1,2] :

$$\begin{aligned} J_e(\hat{P}) &= E\{\|x - \hat{x}\|^2\} = E\{\|x - \hat{P} \hat{P}^T x\|^2\} \\ &= E\{(x - \hat{P} \hat{P}^T x)(x - \hat{P} \hat{P}^T x)^T\} = E\{x^T x - \hat{t}^T \hat{t}\} \\ &= E\{\text{trace}(xx^T) - \hat{t}^T \hat{t}\} = \text{trace}\{\Sigma\} - E\{\hat{t}^T \hat{t}\} \\ &= \text{trace}\{\Sigma\} - J_v(\hat{P}) \end{aligned} \quad (2.5)$$

La minimisation du critère d'erreur d'estimation de PCA J_e est équivalent à la maximisation du J_v , alors [1] :

$$J_v(\hat{P}) = E\{\hat{t}^T \hat{t}\} = E\{\sum_{j=1}^{\ell} t_j^2\} = \sum_{j=1}^{\ell} E\{t_j^2\} = \sum_{j=1}^{\ell} \text{Var}\{t_j\} \quad (2.6)$$

D'après l'équation précédente, la maximisation du critère J_v est équivalente à une maximisation de la variance de la composante t_j . Le problème d'optimisation est donc reformulé comme suit :

$$\hat{P} = \underbrace{\arg \min}_{\hat{p}} \{J_e(\hat{P})\} = \underbrace{\arg \max}_{\hat{p}} \{J_v(\hat{P})\} \quad (2.7)$$

Soit $t \in \mathbb{R}$ la projection du vecteur de données x le long d'une direction représentée par un vecteur unitaire $p \in \mathbb{R}^m$. La composante t est donné par $t = x^T p = p^T x$, sous la contrainte $\|p\|^2 = p^T p = 1$. Elle représente une nouvelle variable ayant une moyenne et une variance qui dépendent des propriétés statistiques de x comme suit :

$$\begin{aligned} E\{t\} &= E\{p^T x\} = p^T E\{x\} = 0 \\ \text{Var}\{t\} &= E\{(t - E\{t\})^2\} = E\{t^2\} \\ E\{(p^T x)(x^T p)\} &= p^T E\{xx^T\}p \\ &= p^T \Sigma p \end{aligned} \quad (2.8)$$

L'expression précédente est un problème d'optimisation sous contrainte qui admet une solution par l'application de la méthode de multiplicateur de Lagrange, donc l'expression de la résolution est formulée par la fonction de Lagrange suivante [1,2] :

$$\mathcal{L}(p, \lambda) = J_v(p) - \lambda(p^T p - 1) = p^T \Sigma p - \lambda(p^T p - 1) \quad (2.9)$$

Où λ est le multiplicateur de Lagrange et le vecteur p maximisant le critère d'optimisation J_v , est solution du système d'équations suivant :

$$\begin{cases} \partial \mathcal{L}(p, \lambda) / \partial p = \Sigma p - \lambda p = 0 \\ \partial \mathcal{L}(p, \lambda) / \partial \lambda = p^T p - 1 = 0 \end{cases} \quad (2.10)$$

Un tel système d'équations admet des solutions réelles de la variable λ . Par la résolution directe de l'équation caractéristique [1,7] :

$$\text{Det}\{\Sigma - \lambda I_m\} = 0 \quad (2.11)$$

Ces solutions sont les m valeurs propres λ_i de la matrice de covariance Σ et qui sont associées au m vecteurs propres p_i [1,2] :

$$(\Sigma - \lambda_i I_m)p = 0 \implies \Sigma p_i = \lambda_i p_i \quad (2.12)$$

Elle est donnée aussi sous forme matricielle comme suit [1,2] : $\Sigma P = P \Lambda$

Avec $P = [P_1, \dots, P_m] \in \mathbb{R}^{m \times m}$: la matrice de projection des données, et

$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_m\} \in \mathbb{R}^{m \times m}$: La matrice diagonale des valeurs propres de Σ .

Alors nous avons les propriétés suivantes :

- la matrice de covariance Σ est symétrique,
- les m valeurs propres λ_i sont réelles et les m vecteurs propres p_i sont orthogonaux, alors les vecteurs propres p_i forment une base orthonormée [1,2] :

$$P^T P = P P^T = I_m \in \mathbb{R}^{m \times m} \quad (2.13)$$

Avec I_m la matrice d'identité d'ordre m .

Par conséquent :

$$p^T \Sigma p = \Lambda \quad (2.14)$$

Cela montre que les éléments diagonaux de Λ sont arrangés dans l'ordre décroissant [1]

$$\lambda_1 \geq \dots \geq \lambda_m$$

Le vecteur de données $x(k)$ peut se transformer sans aucune perte d'informations en un vecteur de composantes principales contenant de nouvelles variables non corrélées, qui sont des combinaisons linéaires des variables originelles. Ce nouveau vecteur des variables (vecteur des composantes principales) est exprimé par la relation suivante [1] :

$$t(k) = [t_1(k), \dots, t_m(k)]^T = P^T x(k) \in \mathbb{R}^m \quad (2.15)$$

Avec t_j ($j = \{1, \dots, m\}$) sont les composantes principales projeté dans le sous espace [1,7] :

$$t_j(k) = P_j^T x(k) = x^T P_j \quad (2.16)$$

Ces composantes non corrélées sont définies par [1] :

$$E\{t_i t_j\} = E\{p_i^T x x^T p_j\} = p_i^T \Sigma p_j = 0 \quad i \neq j \quad (2.17)$$

Aussi on peut représenter la matrice des composantes sous forme matricielle suivante [1] :

$$T = [t(1), \dots, t(N)]^T = X P \in \mathbb{R}^{N \times m} \quad (2.18)$$

Le vecteur de données $x(k)$ est défini par le vecteur de composantes principales associées avec le vecteur P [1] : $x(k) = P t(k) = \sum_{j=1}^m p_j t_j(k)$

L'estimation $\hat{x}(k)$ du vecteur de données $x(k)$ dans l'espace réduit qui est formé par les l premières valeurs propres est donnée par [1] :

$$\hat{x}(k) = \hat{P} \hat{t}(k) = \hat{P} \hat{P}^T x(k) = \hat{C} x(k) \quad (2.19)$$

Où $\hat{C} \in \mathbb{R}^{m \times m}$ est la matrice qui caractérise le modèle PCA et dont la matrice optimale de représentation est exprimée par [1] :

$$\hat{P} = [p_1, \dots, p_\ell] \in \mathbb{R}^{m \times \ell}$$

Ainsi le vecteur des l composantes est exprimé par [1] :

$$\hat{t}(k) = \hat{P}^T x(k) \in \mathbb{R}^\ell \quad (2.20)$$

Les pertes des informations pendant la réduction peuvent être récupérées à partir du vecteur résiduel comme suit [1] :

$$\tilde{x}(k) = \tilde{P} \tilde{t}(k) = \tilde{P} \tilde{P}^T x(k) = \tilde{C} x(k) \quad (2.21)$$

Avec $\tilde{P} = [p_{\ell+1}, \dots, p_m] \in \mathbb{R}^{m \times (m-\ell)}$

Et $\tilde{C} = \tilde{P}\tilde{P}^T = I_m - \hat{C} \in \mathbb{R}^{m \times m}$

La notion de la modélisation par PCA est divisée en deux parties, la réduction en un sous-espace principale, et un sous-espace résiduel, donc la décomposition du vecteur de mesures est donné par [1] :

$$x(k) = \hat{x}(k) + \tilde{x}(k) \quad (2.22)$$

2.3. Détermination d'une structure optimale du modèle ACP :

Le choix du nombre de CPs joue un rôle crucial dans la détermination des relations de redondance entre les variables. Un tel nombre est difficile à déterminer. Pour assurer une meilleure estimation du modèle PCA, des méthodes sont présentées pour aider dans son choix.

Pour cela, il existe des nombreux critères permettent de la détermination optimale des nombres des composantes retenues. Parmi ces critères, on peut citer la méthode du pourcentage cumulée de la variance totale, la méthode validation cumulée (PRESS), la méthode de la variance de l'erreur de reconstruction VER, et le Scree test [1,4].

1. Le pourcentage cumulé de la variance totale (PCV) :

Les valeurs propres de la matrice de corrélation représentent les variances des CPs. Par conséquent, elles ont été utilisées pour définir un critère PCV dans le but est la détermination de la structure optimale d'un modèle ACP. Ce critère représente le pourcentage de la variance expliquée par les CPs retenues comme suit :

$$PCV(\ell) = 100 \left(\frac{\sum_{a=1}^{\ell} \lambda_a}{\sum_{a=1}^m \lambda_a} \right) \% \quad (2.23)$$

Généralement, on essaie de retenir avec ce critère un nombre de CPs ayant un PCV compris entre 90% et 95% ou 99% de la variance totale.

2. la méthode de validation cumulée (PRESS) :

L'idée de la validation croisée est basée principalement sur la prédiction à travers un modèle ACP d'une mesure $x_i(k)$ de la matrice de données X . Le nombre des CPs retenues est optimal si la moyenne de la prédiction globale d'une telle mesure n'est plus significativement améliorée par l'ajout de CPs supplémentaires.

L'approche de la validation cumulée présente un critère qui est appelé PRESS, et le principe général de ce dernier est la minimisation de la somme des carrés des erreurs des prédictions, et qui est donné par la formule suivante :

$$PRESS(\ell) = \frac{1}{Nm} \sum_{k=1}^N \sum_{i=1}^m (\hat{x}_i^\ell(k) - x_i(k))^2 \quad (2.24)$$

Où $\hat{x}_i^\ell(k)$ est la prédiction de $x_i(k)$, N étant la taille du jeu de validation.

Il existe un algorithme simplifié de ce critère qui permet de calculer les composantes principales comme suit :

3. le Scree test :

Le Scree Test ou test du coude est une approche empirique proposée principalement pour la sélection du nombre optimal des CPs. Elle est basée sur un test du coude de la courbe du pourcentage de la variance résiduelle dont l'expression est la suivante :

$$PVR(\ell) = 100 \left(\frac{\sum_{a=\ell+1}^m \lambda_a}{\sum_{a=1}^m \lambda_a} \right) \% \quad (2.25)$$

Références Bibliographiques

- [1] Baligh MNASSRI .2012. Surveillance des processus industriels par analyse en composantes principales. Thèse de doctorat, Université d'Aix-Marseille.
- [2] Chaouch Alima. 2013. Surveillance de l'état de fonctionnement d'un procédé biologique de dépollution. Mémoire de magister, Université 20 Aout 1955 – Skikda
- [3] Anissa Ben Aicha. . Détection et localisation de défauts par Analyse en Composantes Principales certaine ou incertaine de type intervalle. Thèse de doctorat, l'Ecole Nationale d'Ingénieurs de Monastir.
- [4] Bergoug Aicha. 2016/2017. Détection et Localisation de Défauts par Analyse en Composantes Principale (ACP). Mémoire Master
- [5] R. Greiner, B. Póczos. Introduction to Machine Learning, University of Alberta <https://webdocs.cs.ualberta.ca/~greiner/C-466/>
- [6] Baligh Mnassri. Diagnostic de Défauts par Analyse en Composantes Principales, Laboratoire des Sciences de l'Information et des Systèmes – UMR CNRS 6168, Université de Saint-Jérôme.
- [7] Mohamed-Faouzi Harkat. 2003. Détection et Localisation de défauts par Analyse en Composantes Principales, thèse de doctorat, Institut National Polytechnique de Lorraine

Chapitre

3

Multiclassification par les
Séparateurs à Vastes Marges

Chapitre 3. Multi classification par les séparateurs à vaste marges

3.1. Les algorithmes d'apprentissage automatique :

Les algorithmes d'apprentissage automatique ont pour objet d'extraire de la connaissance à partir de données. Les algorithmes d'apprentissage automatique dit supervisés construisent un modèle de prédiction en utilisant un ensemble d'objets préalablement étiquetés, dit ensemble d'apprentissage.

À partir de l'ensemble d'apprentissage, ou base d'exemples, l'algorithme estime les paramètres du modèle de prédiction les plus performants possible, c'est-à-dire qui produisent le moins d'erreurs en prédiction.

À partir de ce modèle qui a été construit avec les paramètres optimaux, il sera possible de prédire la classe ou la catégorie d'un nouvel objet qui lui est soumis.

Un de ces algorithmes d'apprentissage automatique supervisé est le séparateur à vaste marge (SVM) ou machines à vecteurs de support.

3.2. Séparateur à vaste marge (SVM) :

Les séparateurs à vastes marges reposent sur deux idées clés : la notion de marge maximale et la notion de fonction noyau.

L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et A. Lerner, et en 1973 par Richard Duda et Peter Hart dans leur livre *Pattern Classification*.

Un SVM appartient à un ensemble de techniques d'apprentissage comme un perceptron, il trouve un séparateur linéaire entre les points de données de deux classes différentes,

En général, il peut y avoir plusieurs séparateurs possibles entre les classes (en supposant le problème linéairement séparable) et qu'un perceptron n'a pas de préférence parmi celles-ci, dans les SVMs, cependant, nous faisons un choix particulier parmi tous les séparateurs possibles : nous voulons celui avec la "marge" maximale.

3.3. Hyperplan, marge et support vecteur :

Le but du SVM-binaire est de trouver un classificateur qui va séparer les données et maximiser la distance entre deux classes. Dans l'exemple qui suit le classificateur est un classificateur linéaire appelé hyperplan qui sépare les deux ensembles de points.

Les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

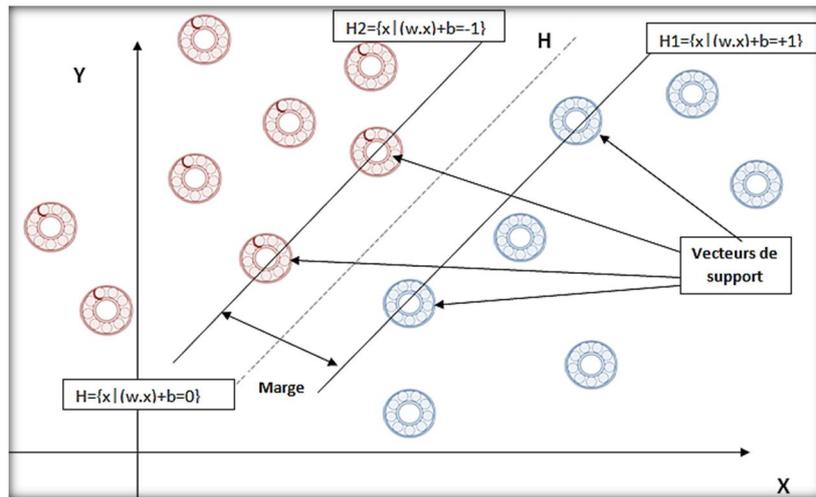


Figure 3.1 : Illustration d'une classification binaire linéaire SVM

Il est évident qu'il existe une multitude d'hyperplan valide mais la propriété remarquable des SVM est que cet hyperplan doit être optimal : cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale.

On appelle cette distance « marge » entre l'hyperplan et les exemples. L'hyperplan séparateur optimal est celui qui maximise la marge. Comme on cherche à maximiser cette marge, on parlera de séparateurs à vaste marge

3.4. Séparateur à vaste marge dans le cas linéaire :

3.4.1. Maximisation de marge :

Le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie (b) nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie (a) qu'avec une plus petite marge, l'exemple se voit mal classé.

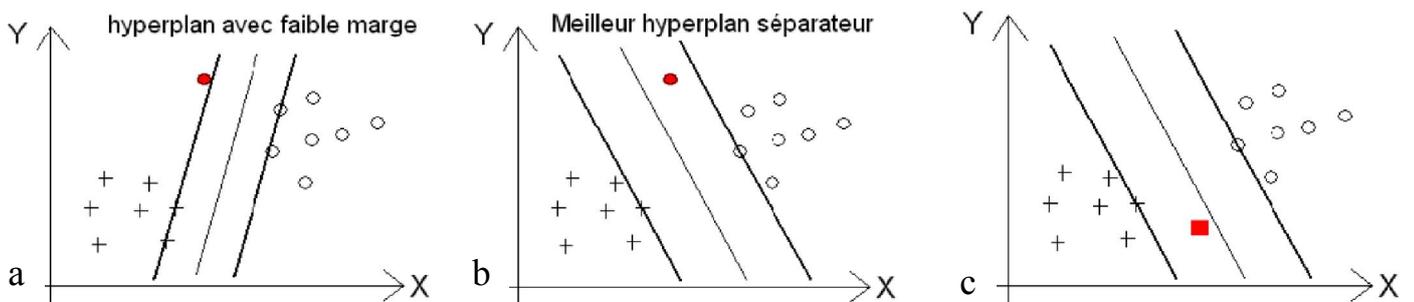


Figure 3.2 : Illustration du choix de la marge avec un hyperplan optimal

En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal. Dans le schéma (c), le nouvel élément sera classé dans la catégorie des « + ».

3.4.2. Calcul de la marge :

Rappelons qu'en deux dimensions la distance entre un point (x_0, y_0) et une droite.

$Ax + By + c = 0$, est donnée par la relation suivante :

$$\frac{|Ax_0 + By_0 + c|}{\sqrt{A^2 + B^2}} \quad (3.1)$$

De façon similaire, dans un modèle linéaire (voir figure (3.1)), la distance entre un point et le hyperplan séparateur qui a pour équation $wx + b = 0$ est donnée par :

$$d(x) = \frac{|wx + b|}{\|w\|} \quad (3.2)$$

3.4.3. Calcul de l'Hyperplan optimal :

L'hyperplan optimal est celui pour lequel la distance aux points les plus proches (Marge) est maximale. L'équation générale de l'hyperplan est $H : wx + b = 0$ où :

- w est le vecteur orthogonal à l'hyperplan.
- x un vecteur de l'hyperplan.
- $b/\|w\|$ est la distance entre l'hyperplan et l'origine.

Doit satisfaire les conditions suivantes pour une classification binaire :

$$\begin{cases} \text{Si } y_i = +1 ; w x_i + b \geq 1 \\ \text{Si } y_i = -1 ; w x_i + b \leq -1 \\ \text{Pour tous } \{x_i, y_i\} ; y_i(w x_i + b) \geq 1 \end{cases} \quad (3.3)$$

Soient $f(x) = w \cdot x + b$ tel que x_1 et x_2 deux points de classes différentes :

$$(f(x_1) = +1 \text{ et } f(x_2) = -1).$$

$$(w \cdot x_1) + b = +1 \text{ et } (w \cdot x_2) + b = -1 \text{ donc } (w \cdot (x_1 - x_2)) = 2$$

$$\text{D'où : } (w \cdot (x_1 - x_2)) / \|w\| = 2 / \|w\|.$$

On peut donc en déduire que maximiser la marge revient à minimiser son inverse $\|w\|/2$ sous la contrainte $y_i(w \cdot x + b) \geq +1$.

$$\min \frac{1}{2} \|w\| = \min \frac{1}{2} w^T w \quad (3.4)$$

$$\text{On a } \begin{cases} \min \frac{1}{2} w^T w \\ \text{sous la contrainte} \\ y_i(w \cdot x + b) \geq +1 \end{cases} \quad (3.5)$$

Ce genre de problème d'optimisation peut être résolu en associant un multiplicateur de Lagrange $\alpha_i \geq 0$.

Le Lagrangien est donné par :

$$L(w, s, \alpha) = \frac{1}{2} w^t w - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^N \alpha_i \quad (3.6)$$

Pour trouver la solution du problème primal, il faut minimiser le lagrangien par rapport à w et le maximiser par rapport aux α_i . Pour cette solution, les dérivées partielles du lagrangien s'annulent, selon les conditions de Kuhn-Tucker, on obtient :

$$\frac{\partial L(w,b,\alpha)}{\partial w} = 0, \frac{\partial L(w,b,\alpha)}{\partial b} = 0 \text{ et } \alpha_i \geq 0 \quad (3.7)$$

Cela revient à trouver les α_i et w qui annulent ses dérivées partielles :
On trouve :

$$w = \sum_{i=1}^N \alpha_i y_i x_i \text{ et } \sum_{i=1}^N \alpha_i y_i = 0 \quad (3.8)$$

Et en les réinjectant dans le Lagrangien on obtient le Lagrangien dual qui doit être maximisé

$$L(w, b, \alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i x_j \quad (3.9)$$

Si les α_i^* sont solutions de ce problème c.à.d. les multiplicateurs optimaux correspondant aux points les plus proches et sont non-nuls. Alors les points x_i sont appelés **vecteurs de supports**.

Et on a le w^* optimal : $w^* = \sum_{i=1}^n \alpha_i^* y_i x_i$ (3.10)

En le remplaçant dans l'équation de l'hyperplan séparateur on obtient l'hyperplan optimal

$$h(x) = \sum_{i=1}^n \alpha_i^* y_i x_i \cdot x + b \quad (3.11)$$

La fonction de décision associée est donc :

$$f(x) = \text{sign}(\sum_{i=1}^n \alpha_i^* y_i x_i \cdot x + b) \quad (3.12)$$

3.5. Séparateur à vaste marge dans le cas non linéaire

Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est d'utiliser des fonctions noyaux et de projeter les données de l'espace d'entrée (non linéairement séparables) dans un espace de plus grande dimension, de façon à ce que les données deviennent linéairement séparables.

Plus la dimension du nouvel espace est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les classes est élevée, ceci est illustré dans la figure (3.3).

En pratique, il revient à l'utilisateur du SVM d'effectuer des tests pour déterminer la fonction noyau qui convient le mieux pour son application.

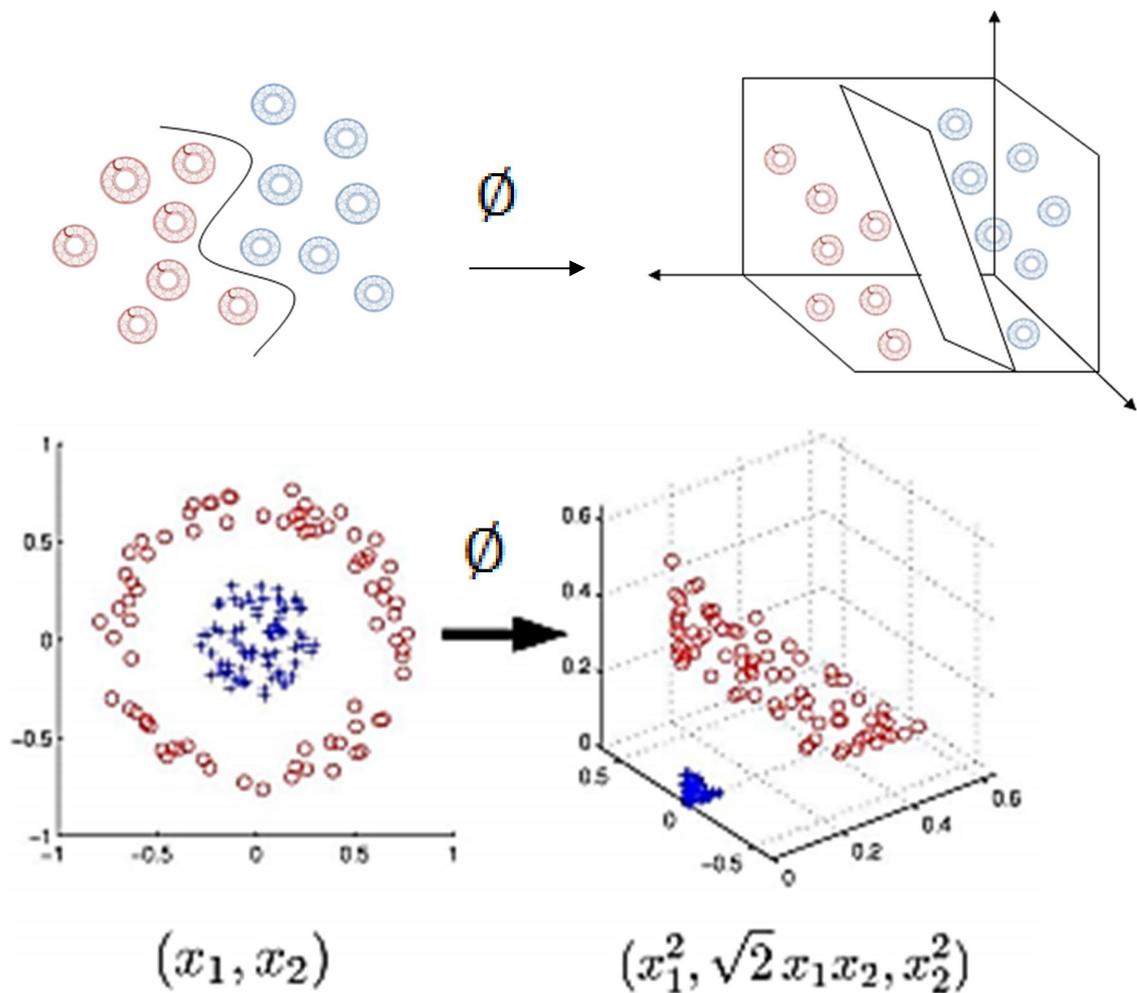


Figure 3.3 : Schéma de changement des représentations non linéaires vers un nouvel espace

3.5.1. Les Noyaux :

Le noyau permet de projeter les données dans un espace transformé *espace- φ* de grande dimension dans lequel s'opère plus facilement la séparation des classes.

Pour qu'il existe un *espace- φ* et une fonction φ correspondant à un noyau K , il suffit que K soit un "noyau de Mercer", K doit vérifier les propriétés suivantes :

K est une fonction continue, symétrique défini-positive qui associe à tout couple d'observations (x, x') une mesure de leur « influence réciproque » calculée à travers leur corrélation (les noyaux projectifs) ou leur distance (les noyaux radiaux)

On peut citer les exemples de noyaux : polynomiale, gaussien, sigmoïde et Laplacien.

Les différentes fonctions du Noyau (Kernel) : Les différentes fonctions du noyau Kernel sont les suivantes :

- Noyau linéaire : $K(x, z) = x \cdot x'$ (3.13)

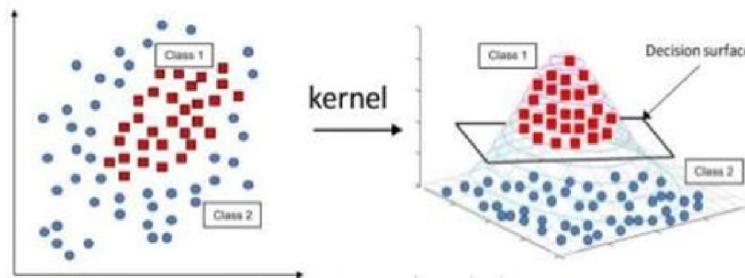


Figure 3.4 : Classification SVM binaire avec le noyau linéaire de Kernel.

- Noyau polynomial de degré d : $K(x, x') = (a * \langle x, x' \rangle + b)^d$ (3.14)

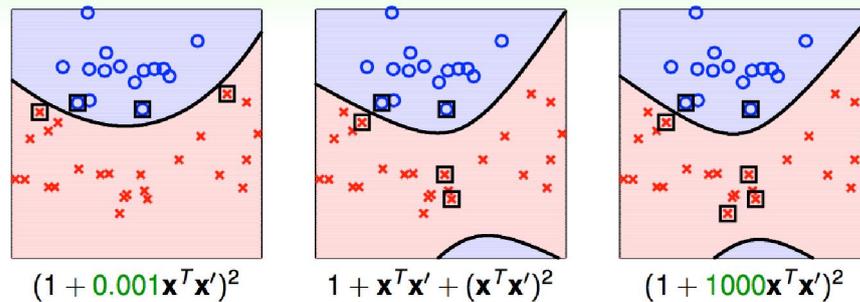


Figure 3.5 : Classification SVM binaire avec le noyau poly 2^{ième} degré de Kernel.

- Noyau Gaussien (Radial Basis Function, RBF) : $K(x, z) = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$ (3.15)

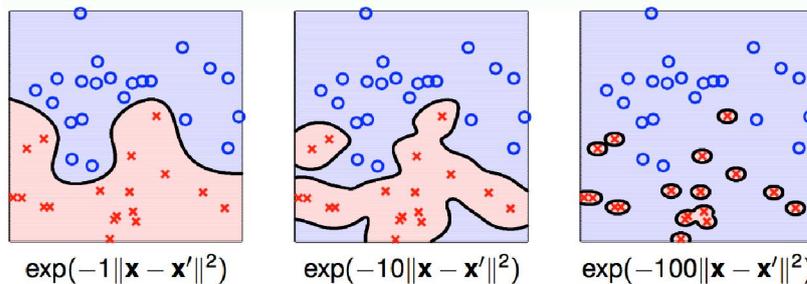


Figure 3.6 : Classification SVM binaire avec le noyau Gaussien de Kernel

3.6. Types multi-class des séparateurs à vaste marge

On discute dans la suite deux méthodes utilisées dans les vecteurs de support traitant des problèmes multi classes.

- un contre tous.
- un contre un.

Dans la méthode "un contre tous", un problème de classification de n classes est converti en n classificateurs binaires et pour le classificateur 'i' binaire, la classe i est séparé des autres classes.

L'approche "un contre un" transforme le problème de n classes en $n(n-1)/2$ classificateurs binaire qui couvrent tous les paires de classe.

3.6.1. Multi classification par l'approche un contre tous.

Pour un problème de N classes, le classificateur multiclassés1vsR établit d'abord N classificateurs bi-classe, chacun d'eux est responsable de chaque classe respectivement.

Pour le i ème classificateur bi-classe, 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 N classificateurs bi-classes, est formé pour trouver la fonction de décision, et alors toutes les fonctions de décision sont combinées pour obtenir la fonction finale de décision pour le problème de classification multi-classes figure (3.7).

Considérons un problème de n classes, pour l'approche "un contre tous", on détermine n fonctions de décision discrètes qui séparent une classe des autres classes avec une marge maximale :

$$D_{i(x)} = w_i^T \phi(x) + b_i \quad (3.16)$$

Si on a pour un vecteur d'entrée x :

$$D_{i(x)} > 0 \quad (3.17)$$

Les vecteurs de support doivent satisfaire :

$$y_i D_{i(x)} = 1$$

Le i ème SVM est formé avec tous les exemples dans la i ème classe avec des étiquettes positives et tous les autres exemples avec des étiquettes négatives, ainsi, l'hyperplan $D_i(x) = 0$ forme l'hyperplan de séparation optimal avec $D_i(x) \geq 1$ pour la i ème classe et ceux qui appartiennent aux classes restantes satisfont $D_i(x) \leq -1$.

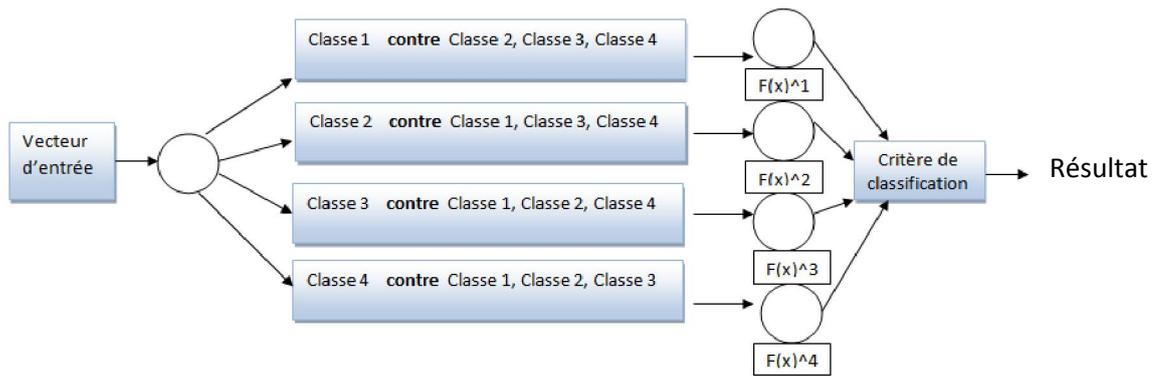


Figure 3.7 : Classification SVM un contre tous (One Vs All)

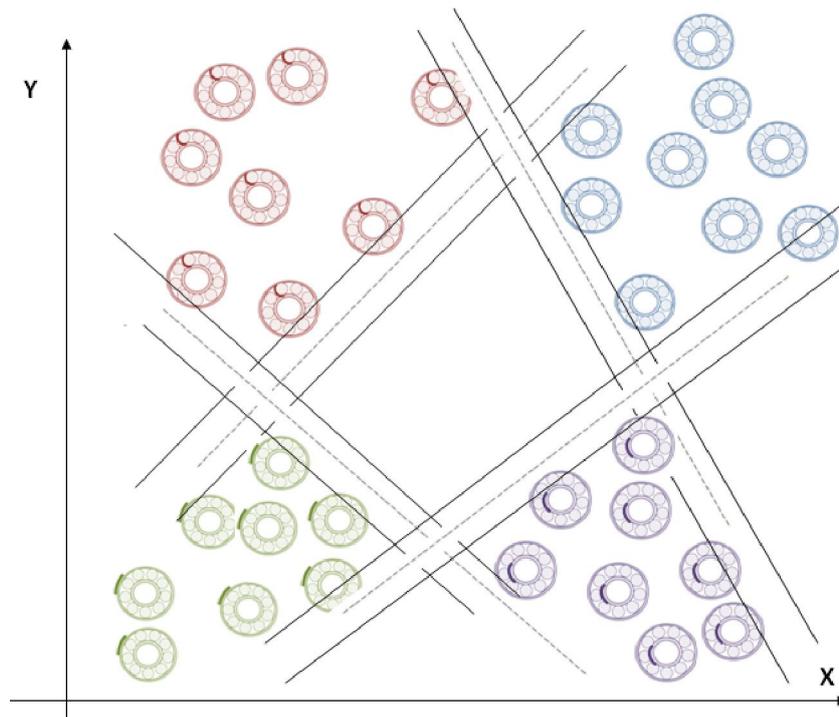


Figure 3.8 : Schéma des hyperplans linéaires pour la classification un contre tous (One vs all) / $n=4$

3.6.2. Multi classification par l'approche un contre un

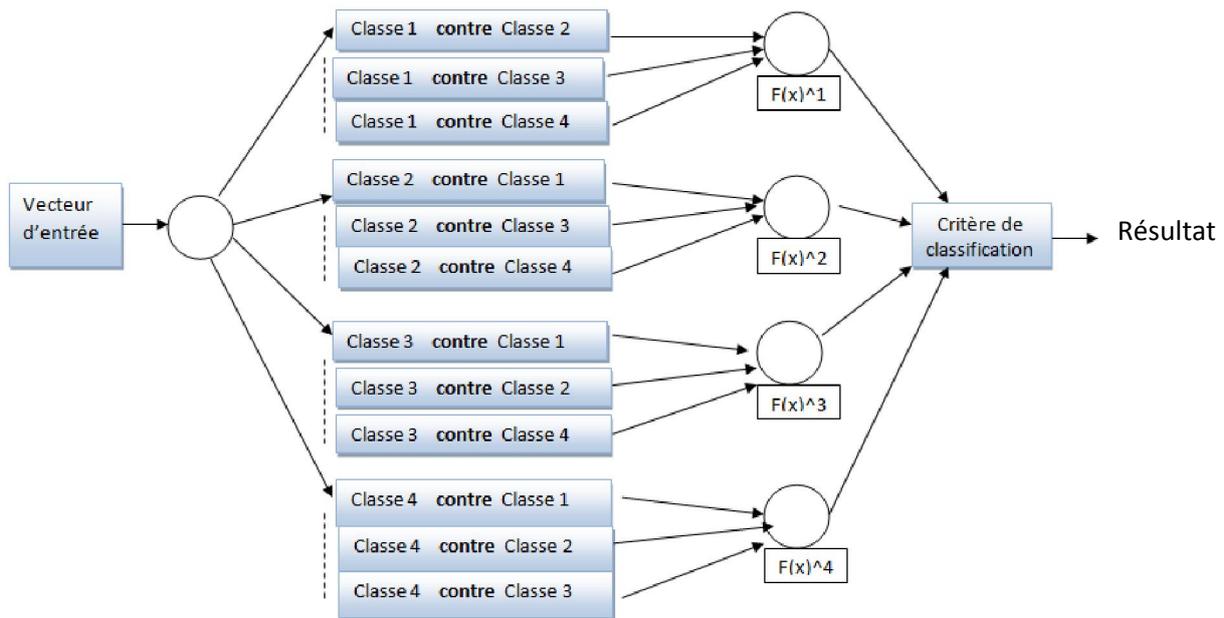


Figure 3.9 : Classification SVM un contre un (One Vs One)

Dans l'approche "un contre un", les fonctions de décision sont déterminées à partir des combinaisons des deux classes. Pour déterminer une fonction de décision pour une classe paire, nous utilisons les données d'apprentissage pour les deux classes correspondant. Ainsi, dans chaque session d'apprentissage, le nombre de données d'apprentissages est considérablement réduit par rapport à "un contre tous", qui utilisent tous les données d'apprentissage.

Mais le nombre de fonctions de décision est $n(n-1)/2$ comparativement avec n de "un contre tous", où n est le nombre de classes.

Soit la marge à maximiser pour la classe i contre la classe j est :

$$D_{ij}^{(x)} = w_{ij}^* \phi(x) - b_{ij}^* \quad (3.18)$$

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

Références Bibliographiques

- [1] De Souza, D. B., Chanussot, J., & Favre, A. C. (2014, May). On selecting relevant intrinsic mode functions in empirical mode decomposition: An energy-based approach. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 325-329). IEEE.
- [2] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- [3] Jean-Christophe Cexus, (2005) Analyse des signaux non-stationnaires par transformation de Huang, Operateur de Teager-Kaiser, et Transformation de Huang-Teager (THT). Acoustics. Université Rennes 1. French. <tel-00012141>.
- [4] M. G. Rilling. , (2007). Décompositions Modales Empiriques Contribution à la théorie, l'algorithme et l'analyse de performances. Thèse de doctorat. Université de Lyon.
- [5] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung et H.H. Liu, (1998): The empirical mode decomposition and the Hilbert spectrum for Nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London*, 454:903–995.

Chapitre

4

Résultats de simulation

Chapitre 4. Résultats de simulation

4.1. Matériels et méthodes :

4.1.1. Présentation du Banc d'Essai de la CWRU :

Comme représenté sur la figure 4.1, le banc d'essai se compose principalement d'un moteur (à gauche), d'un accouplement « transducteur/encodeur » (au centre), d'un dynamomètre (à droite) et d'un module d'acquisition des données.

Les signaux ont été enregistrés à l'aide des accéléromètres, qui ont été 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 dont des vitesses respectives sont 1797, 1772, 1750, 1730 tr/min.

Les capteurs ont été placés à la position 12 heures verticale des deux côtés accouplement et ventilateur de la cage du moteur en plus de deux autres capteurs à 6 et 3 heures.

La fréquence d'échantillonnage est $f_e = 12$ KHz et le nombre de points total est de

$N_p = 243938$ pts.

Les données ont été enregistrées pour :

- Les roulements sans défaut.
- Les défauts uniques du cotés accouplement (12.000 et 48.000 points/second).
- Les défauts uniques du cotés ventilateur (12.000 points/second.)

Dans ce travail, les roulements que nous avons considérés, sont le roulement du coté accouplement pour l'apprentissage et les tests. Les fréquences caractéristiques sont :

- Fréquence de la bague intérieure : $f_{bi} = 5.415 \times f_r$
- Fréquence de la bague extérieure : $f_{be} = 3.584 \times f_r$
- Fréquence de la cage : $f_c = 0.3982 \times f_r$
- Fréquence de la bille : $f_b = 4.7135 \times f_r$

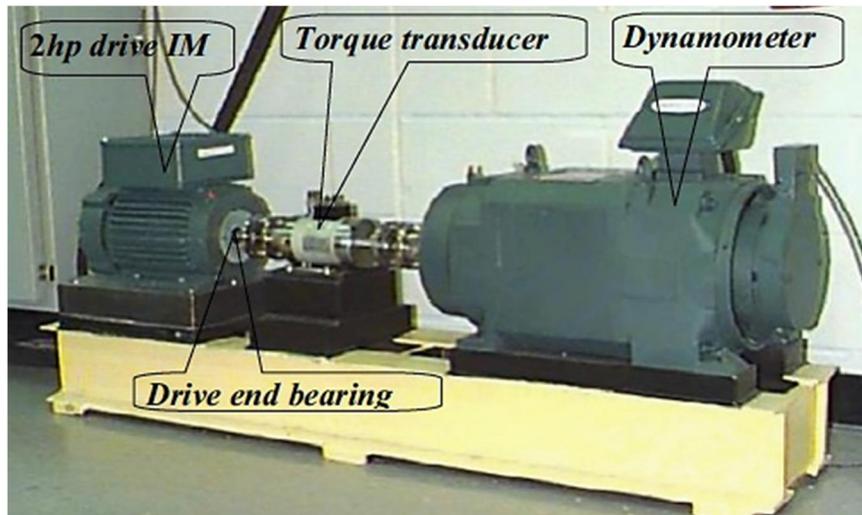


Figure 4.1 : Le Banc d'essai de la CWRU

Fault Diameter (inches)	Motor Speed (rpm)	Inner Race	Ball	Outer Race Position Relative to Load Zone (Load Zone Centered at 6:00)		
				Centered @6:00	Orthogonal @3:00	Opposite @12:00
0.007"	1797	I07_0	B07_0	O07@6_0	O07@3_0	O07@12_0
	1772	I07_1	B07_1	O07@6_1	O07@3_1	O07@12_1
	1750	I07_2	B07_2	O07@6_2	O07@3_2	O07@12_2
	1730	I07_3	B07_3	O07@6_3	O07@3_3	O07@12_3
0.014"	1797	I14_0	B14_0	O14@6_0		
	1772	I14_1	B14_1	O14@6_1		
	1750	I14_2	B14_2	O14@6_2		
	1730	I14_3	B14_3	O14@6_3		
0.021"	1797	I21_0	B21_0	O21@6_0	O21@3_0	O21@12_0
	1772	I21_1	B21_1	O21@6_1	O21@3_1	O21@12_1
	1750	I21_2	B21_2	O21@6_2	O21@3_2	O21@12_2
	1730	I21_3	B21_3	O21@6_3	O21@3_3	O21@12_3
0.028"	1797	I28_0	B28_0			
	1772	I28_1	B28_1			

Tableau 4.1 : Données vibratoires des défauts avec une fréquence d'échantillonnage de 12K HZ / CWRU

4.1.2. Présentation des signaux et quelques-unes de leurs caractéristiques :

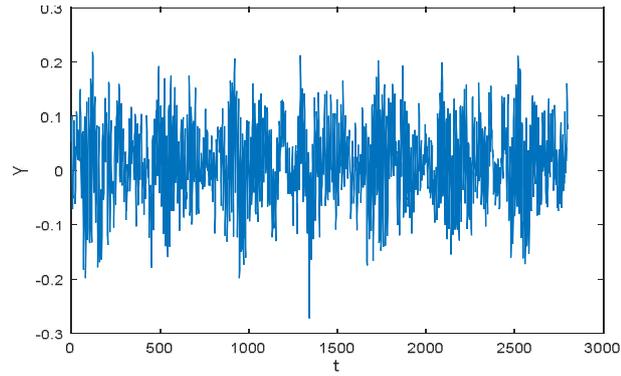


Figure 4.2 : Signal original cas sain

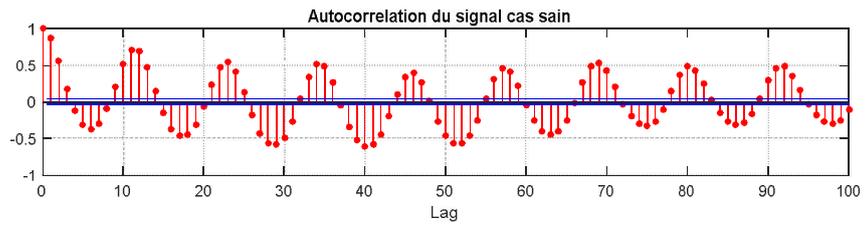


Figure 4.3 : Auto corrélation du signal original cas sain

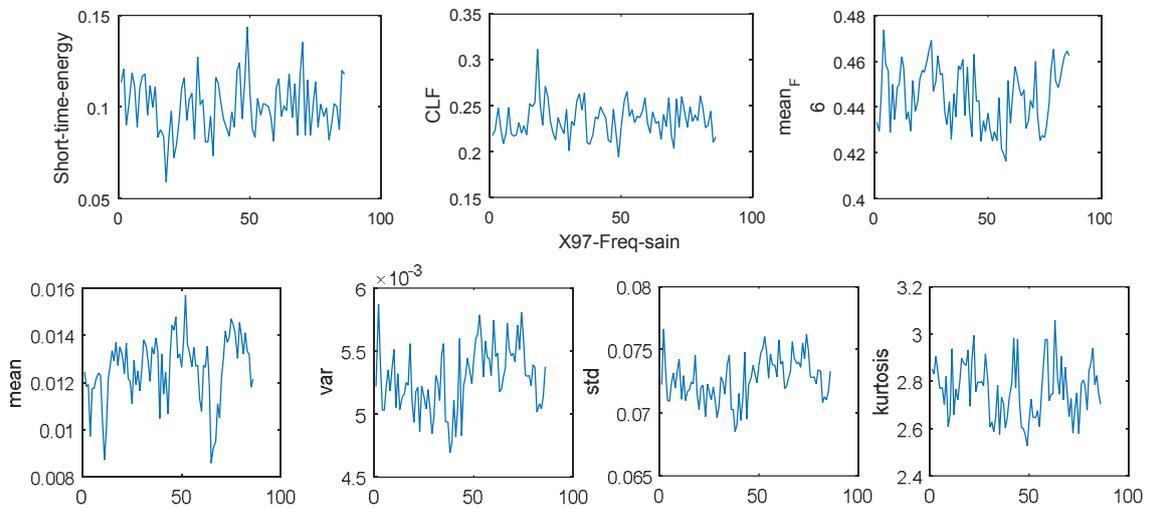


Figure 4.4 : Variation de quelques paramètres temporels et fréquentiels du signal sans défauts (sain)

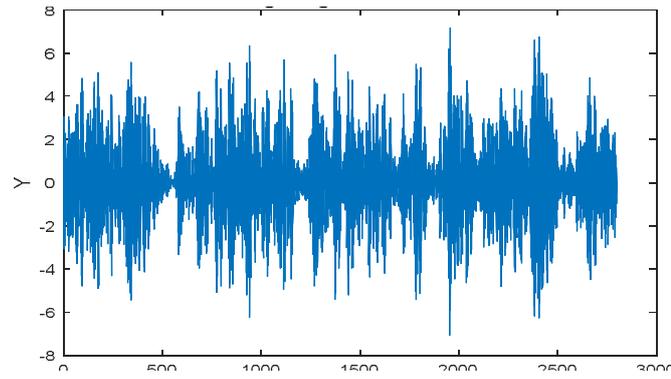


Figure 4.5 : Signal original avec défaut bille

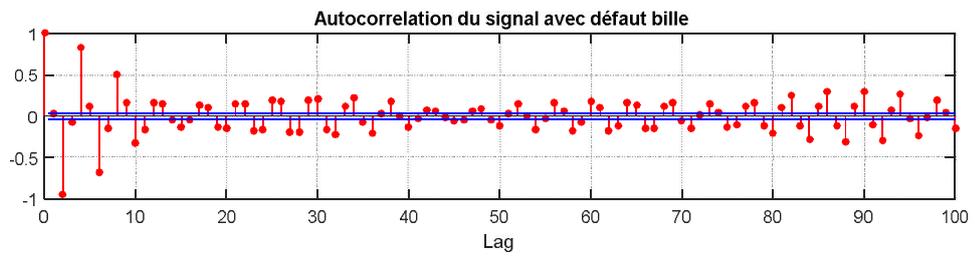


Figure 4.6 : Auto corrélation du signal avec défaut bille

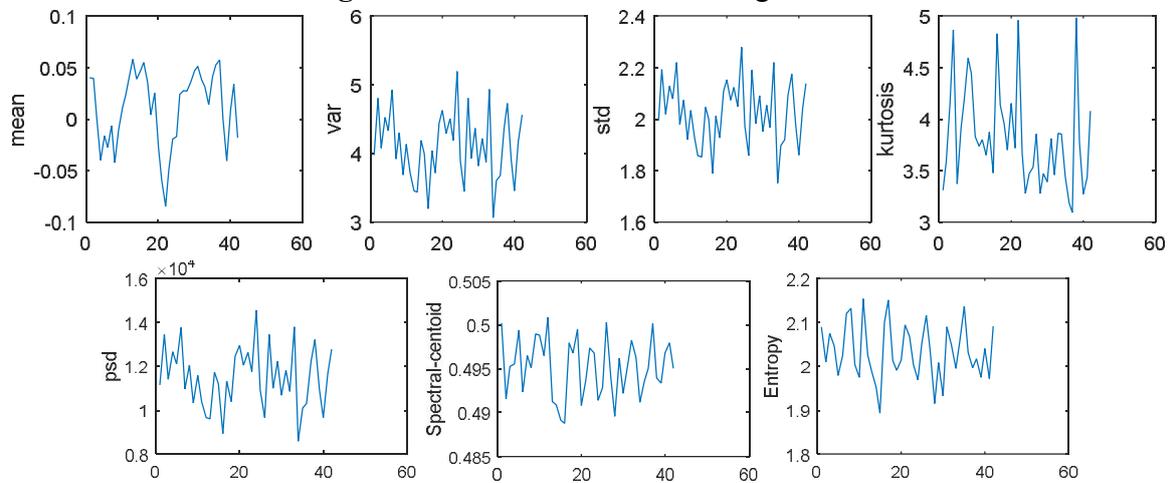


Figure 4.7 : Variation de quelques paramètres temporels et fréquentiels du signal avec défaut bille

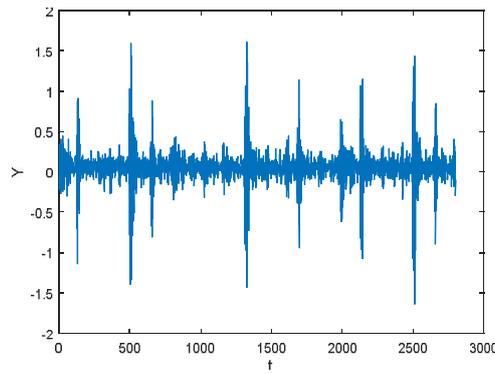


Figure 4.8 : Signal original avec défaut bague interne

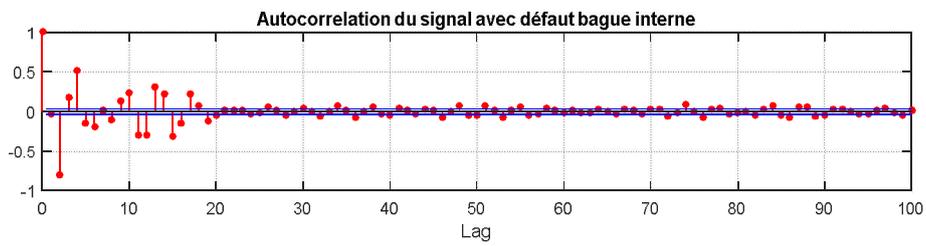


Figure 4.9 : Auto corrélation du signal avec défaut bague interne

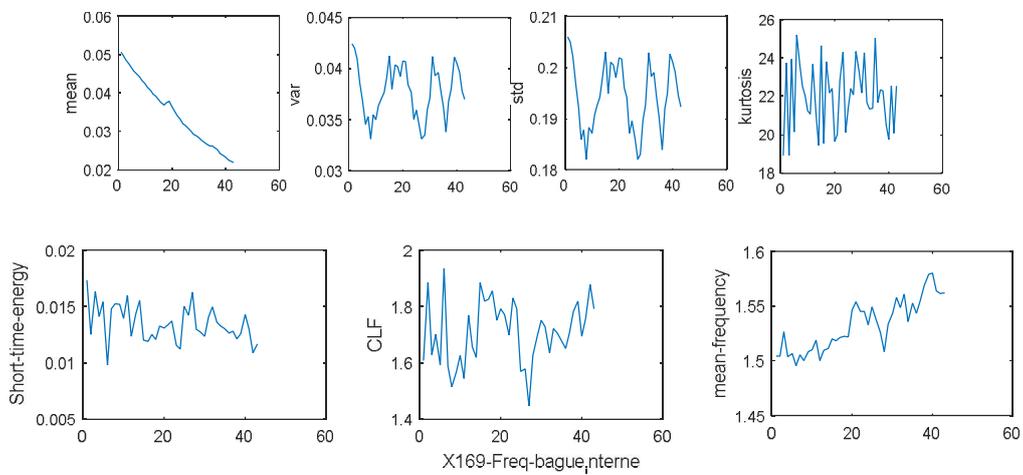


Figure 4.10 : Variation de quelques paramètres temporels et fréquentiels du signal avec défaut bague interne.

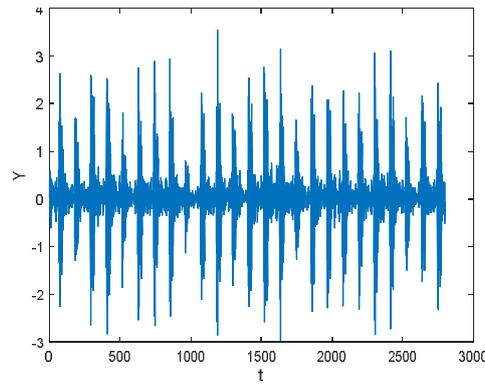


Figure 4.11 : Signal original avec défaut bague externe

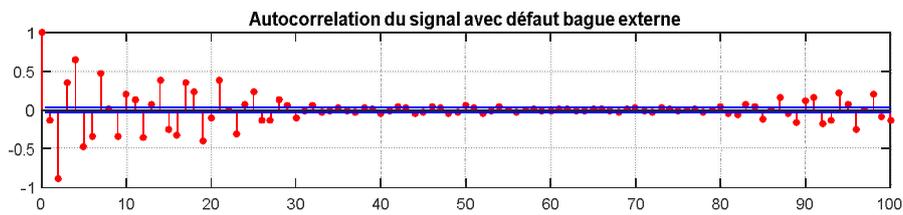


Figure 4.12 : Auto corrélation du signal avec défaut bague externe.

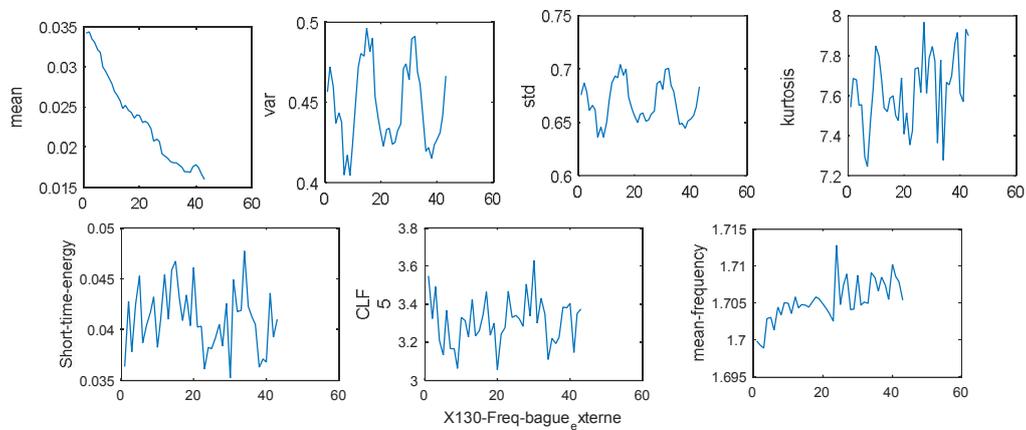


Figure 4.13 : Variation de quelques paramètres temporels et fréquentiels du signal avec défaut bague externe

4.2. Méthodologie de diagnostic adoptée :

La méthode utilisée dans notre travail peut être résumée comme le montre la figure 4.12 :

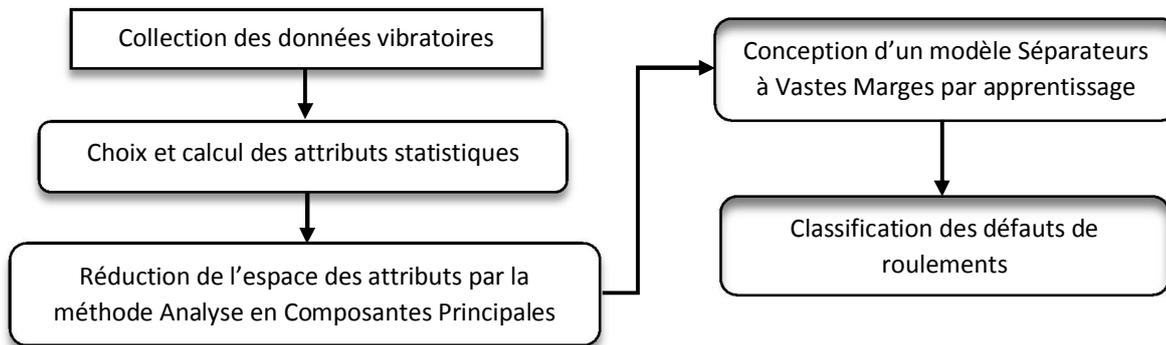


Figure 4.14 : Schéma présentant les étapes de diagnostic

4.2.1 L'analyse en composantes principales :

L'analyse en composantes principales (ACP) vise à transformer un certain nombre de variables corrélées en un plus petit nombre de variables décorrélées appelés composantes principales. Cette réduction de la dimensionnalité permet de réduire la complexité du traitement du signal. Les étapes de l'ACP s'opèrent comme suit :

- 1) Normalisation de la matrice d'observation X pour obtenir une matrice normalisée X^* .
- 2) Calcul la matrice de covariance R .
- 3) Calcul des vecteurs propres et des valeurs propres de la matrice de covariance R .
- 4) Choix des composants et la formation d'un vecteur caractéristique.
- 5) Obtenir le nouvel ensemble de données.

4.2.2 Multiclassification SVM

L'origine de l'algorithme SVM est une classification binaire où la classe étiquettes peut prendre deux valeurs : 1 et -1, mais en fait, les problèmes de classification comprennent plus de deux classes. Par exemple, l'implémentation de l'algorithme SVM binaire pour le diagnostic des défauts de roulement d'une machine asynchrone, où il y a le défaut de la bague intérieure, le défaut de la bague extérieure, et le défaut de billes ne peut être utilisée. Par conséquent, on adoptera la stratégie de classification par multi-classe SVM pour le diagnostic de ces défauts de roulements multiples. Plus précisément, on choisit la multiclassification SVM un contre tous (One Against All : OAA).

4.3. Configuration de l'expérience :

4.3.1. Les données de défaut de roulement :

La série temporelle des signaux de vibration ont été obtenues une expérience de l'université de Case Western Reserve Centre de données de roulement pour les conditions suivantes : le roulement normal, défaut de bague intérieure, défaut de bague extérieur, et le défaut de bille, et prélevés à 12 000 échantillons par seconde. Le type de roulement utilisé dans l'expérience a été 6205-2RS JEM, SKF roulement à billes à gorge profonde. Des exemples de signaux de chaque défaut ont été représentés dans les figures 4.5, 4.8 et 4.12 :

4.3.2. Sélection des caractéristiques en utilisant l'analyse en composantes principales :

Les caractéristiques statistiques originales choisies et calculées sont : kurtosis, skewness, valeur moyenne, écart-type, variance, max, et rms. Le schéma de l'expérience est décrit comme suit : extraire cinquante échantillons de chaque signal de défaut (y compris le cas sain) avec une longueur d'échantillon égale à 12000 (donc sur une durée de 1 seconde), puis égale à 1 200 points (à seulement 0,1 seconde). Par conséquent, on a obtenu 2000 échantillons au total. On a Calculé les paramètres (attributs statistiques) pour chaque échantillon à l'aide des formules des équations entre (1.4) et (1.15) pour réaliser enfin une matrice observée X. Cette la matrice observée X est transformée selon les étapes de l'analyse en composantes principales.

$$X^* = \frac{X - X_p}{S} \quad (4.1)$$

X^* étant la matrice normalisé, X_p étant la moyenne de signal de vibration et S est l'écart type.

Premièrement, la matrice X observés nécessite une normalisation selon la formule (4.1). Puis, on calcule le coefficient de corrélation R ainsi que ses valeurs propres et vecteurs propres. Les composantes principales sont présentées dans le tableau suivant :

Composantes Principales	Echantillons à 1200 points			Echantillons à 12000 points		
	P1	P2	P3	P1	P2	P3
	-1.5376	-0.8982	0.0848	-0.1978	2.2398	-0.1675
	-1.5073	-0.9052	0.1255	-0.3252	1.6003	0.7822
	-1.6283	-0.7230	-0.2430	-0.4151	1.6873	0.1664
	-1.7054	-0.5046	-0.8075	-0.5238	2.0233	-0.3065
	-1.4467	-0.9937	0.2589	-0.4749	1.1362	0.6737
	-1.6744	-0.7790	-0.2296	2.2039	-0.9757	1.2496
	-1.6228	-0.6769	-0.5235	2.6100	-0.3448	-0.4945
<i>Taux Contribution</i>	0.6548	0.3263	0.1079	0.6371	0.3049	0.0079
<i>Taux Contribution cumulatif</i>	0.6548	0.8811	0.989	0.6371	0.9420	0.0490

Tableau 4.2 : Composantes principales obtenues pour les deux cas

Selon le tableau précédent, on peut voir qu'il y a trois composantes principales qu'on nomme P1, P2, et P3 (cas 1200 pts et 12000 pts). Le rapport de contribution cumulatif des composantes principales a été 0,989(cas 12000 pts) et 0.991(cas 1200 pts) respectivement, pratiquement contenait l'information des caractéristiques d'origine.

La dernière étape est le calcul du résultat final, la matrice de la composante principale Z, qui contient un pourcentage de 98,8 et 99.1 des paramètres originels. Cette dernière a été calculée par multiplication de matrice de chargement gauche L selon la formule suivante :

$$Z = X * L$$

Enfin, l'analyse en composantes principales a été l'étape de prétraitement pour la multiclassification SVM.

4.3.3. Traitement des données et diagnostic utilisant l'approche multi-classes SVM :

Le diagnostic des défauts à l'aide de l'algorithme multi-classe SVM est basé sur la sélection des caractéristiques de trois types de défauts de roulement et du cas sain, de plus la méthode un contre tous (OAO), qui est une stratégie de l'algorithme SVM multi-classe a été utilisée dans cette expérience d'identification. Dans ce cadre, quatre paramètres sélectionnés par l'analyse en composantes principales, ont été utilisés pour la classe multi-SVM, et chaque type défaut contenait 50 échantillons. Par conséquent, il y avait 200 échantillons utilisés lors de l'étape de formation.

La matrice Z était de 200×7 , contient les informations pour l'apprentissage. Le tableau suivant présente l'établissement de formation SVM pour l'entraînement.

Le tableau suivant dresse le résultat de traitement par ACP multi-classification SVM en précisant le taux de bonne classification, le temps de classification ainsi que la matrice de confusion.

Le résultat a montré que l'OAO multi-classe SVM était approprié pour le diagnostic du défaut où les fonctions « Kernel » ou noyaux linéaire, quadratique et polynomiale ont donné les meilleurs résultats contrairement à la fonction RBF.

Le réseau de neurones utilisé pour la classification (pour des fins de comparaison avec la classification SVM) est un réseau multicouche possédant une couche cachée à 20 neurones et trois sorties binaires (0 0 0 pour reconnaître le cas sain, 1 0 0, 0 1 0 et 0 0 1 pour les défauts bague interne, bague externe et défauts billes respectivement).

	ACP + SVM			ACP + RNA		
	<i>Taux de classification</i>	<i>Temps de classification</i>	<i>Matrice de confusion</i>	<i>Taux de classification</i>	<i>Temps de classification</i>	<i>Matrice de confusion</i>
4X50 échantillons de 12000 pts (0.1s /échantillon)	99%	2.05 s	$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$	90%	4.3 s	$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}$
4X50 échantillons de 1200 pts (1s /échantillon)	95%	0.78 s	$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 1 \\ 1 & 0 & 4 \end{bmatrix}$	85%	3.1 s	$\begin{bmatrix} 5 & 0 & 0 \\ 1 & 4 & 1 \\ 0 & 0 & 5 \end{bmatrix}$

Tableau 4.3 : Récapitulatif des résultats obtenus

4.4 Discussion des résultats :

D'après le tableau 4.2, on voit bien que la méthode de diagnostic utilisant la classification par SVM avec un prétraitement par ACP donne un meilleur résultat que celle impliquant les réseaux de neurones. Ceci est évident en comparant les taux de bonne classification et les temps de classification obtenus.

Néanmoins, l'utilisation d'un nombre de points élevé dans un échantillon donne un meilleur taux de bonne classification au détriment d'un temps de classification plus lent. Donc, si on veut se rapprocher du diagnostic temps réel il faut opter pour les petits échantillons à condition de savoir bien choisir le vecteur de caractéristiques qui devrait comporter des détails temporels et fréquentiels sur le signal vibratoire originel. Le prétraitement et la réduction de la taille de la matrice d'observation devrait être aussi bien examiné dans le but d'avoir une espace de données pouvant donner une bonne classification.

CONCLUSION GENERALE

Conclusion Générale

En reconnaissance de formes, plusieurs méthodes ont été introduites pour le diagnostic mécanique des pannes, parmi lesquelles et les plus couramment utilisées sont les réseaux neuronaux artificiels, le séparateur à vaste marge et les réseaux de neurones flous.

Les séparateurs vastes marges qui sont basés sur la théorie de l'apprentissage statistique sont réputés pour l'utilisation d'un nombre d'échantillons réduit et ont une meilleure généralisation que les réseaux de neurones artificiels (par exemple) et c'est l'utilisation de noyaux qui permet cette généralisation. Cependant la précision d'une machine vectorielle de support est hautement déterminée par la sélection des paramètres optimaux.

La plus grande différence entre le SVM et le RNA est dans la méthode d'optimisation des paramètres du modèle, le RNA utilise la retro propagation du gradient qui ne garantit pas l'obtention d'une solution optimale par contre l'implémentation du SVM trouvera les paramètres optimaux.

Cependant, ils peuvent être complémentaire avec des combinaisons SVM et RNA. Par exemple, on peut utiliser plusieurs couches RNA et avoir la dernière classification via SVM à la couche de sortie.

L'algorithme développé dans cette étude peut être étendu à plusieurs types de défauts (composants, actionneurs, capteurs, ...etc.) sur les machines jusqu'à la conception d'un système d'avertissement adapté dans lequel l'apparition des défauts est rapportée à l'utilisateur sur l'écran de commande de la machine pour un diagnostic en temps réel.

Une autre idée est d'utiliser les informations existantes sur les machines et les exploiter avec cet algorithme pour en tirer un diagnostic fiable, sans pour autant déployer un matériel supplémentaire (capteurs et interfaces de calculs).

Comme perspectives, on voudrait explorer plusieurs méthodes de prétraitement en plus de l'ACP et plus de méthodes de classification dans un but de comparaison en premier, puis essayer de combiner celles ayant donné de meilleurs résultats jusqu'à obtenir une méthode présentant un temps de classification temps réel proche du 100%.