

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE FERHAT ABBAS – SETIF
UFAS (ALGERIE)

MEMOIRE

Présenté à la faculté des sciences de l'ingénieur
Département d'optique et mécanique de précision
Pour l'obtention du diplôme de

MAGISTER

Option : Optique et mécanique de précision

Par

Mr. ZIANI Ridha

THEME

**Etude des performances de la classification par les
Supports à Vaste Marge (SVM): application au
diagnostic vibratoire**

Soutenu le :

devant la commission d'examen :

BOUZID Djamel	Prof. à l'université de Sétif	Président
ZEGADI Rabah	Prof. à l'université de Sétif	Rapporteur
BOUZID Saïd	M.C. à l'université de Sétif	Examineur
BELHATTAB Nacereddine	M.C. à l'université de Sétif	Examineur
KHEMAL Yamina	M.C. à l'université de Sétif	Examineur

REMERCIEMENTS

Ce travail a été réalisé au Laboratoire de Mécanique de Précision Appliquée. A l'institut d'Optique et de Mécanique de Précision (UFAS - SETIF).

Je tiens à remercier très vivement Monsieur le Professeur ZEGADI Rabah et Monsieur FELKAOUI Ahmed qui ont dirigé cette thèse, pour l'aide précieuse qu'ils m'ont apportée dans l'élaboration de ces travaux. Je leur remercie pour leurs conseils, leurs soutiens, leurs disponibilités et surtout leurs grandes qualités humaines.

Je tiens à remercier très vivement Monsieur le Professeur BOUZID Djamel de m'avoir fait l'honneur d'accepter d'être le président de jury.

Je remercie Monsieur BOUZID Saïd, Maître de Conférence au département d'optique et Mécanique de Précision (UFAS Sétif), D'avoir bien voulu participer au jury.

Je tiens également à remercier Monsieur BELHATTAB Nacereddine, Maître de Conférence au département d'optique et Mécanique de Précision (UFAS Sétif), D'avoir bien voulu participer au jury.

Je tiens également à adresser mes remerciements à Madame KHEMAL Yamina, Maître de Conférence au département de mathématiques, d'avoir acceptée de participer au jury.

Mes sincères remerciements s'adressent spécialement à tous les membres du Laboratoire LMPA et à l'ensemble du personnel administratifs et techniques du département, pour l'aide que chacun d'eux a pu m'apporter et pour les bons moments passés ensemble.

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

SOMMAIRE

Introduction générale	1
CHAPITRE I : Maintenance et analyse vibratoire	3
I.1 Introduction	3
I.2 Approche de maintenance	3
I.2.1 La maintenance corrective	3
I.2.2 La maintenance préventive	4
I.2.3 Conclusion sur la maintenance	4
I.3 Causes de défaillances et localisation des défauts dans les machines tournantes	5
I.3.1 Défauts d'engrenages	5
I.3.1.1. Défauts uniformes	5
I.3.1.2. Défauts localisés	5
I.3.2 Défauts liés aux arbres	7
I.3.2.1 Le balourd	7
I.3.2.2 Le désalignement	8
I.3.3 Défauts de roulements	8
I.3.3.1. Défauts indépendants de la fatigue	8
I.3.3.2. Défauts dépendants de la fatigue	9
I.3.3.3. Types de défauts sur les roulements	9
I.4 Vibration	11
I.5 Les méthodes d'analyse des signaux vibratoires	11
I.5.1. Analyse temporelle	11
I.5.2 Analyse fréquentielle	15
I.5.2.1L' analyse de l'enveloppe	16
I.5.2.2 l'analyse cepstrale	17
I.5.2.3 Exemples d'applications	17
I.5.3. Analyse temps -fréquence	18
I.5.4. Analyse temps –échelle (ondelette)	19
I.3.5 Analyse cyclostationnaire	20
I.6 Conclusion	20
CHAPITRE 2 : Diagnostic et reconnaissance de formes	21
II.1 Introduction	21
II.2 Principe de diagnostic par reconnaissance de formes	21
II.3 Panorama des méthodes de classification	23
II.3.1 Méthodes de classification non supervisées	24
II.3.1.1 Méthodes non hiérarchiques	24
II.3.1.1.1 Méthode de leader	24
II.3.1.1.2 Méthode de k -means	24
II.3.1.1.3 Méthode des nuées dynamiques	25
II.3.1.2. Méthodes hiérarchiques	25
II.3.2 Méthodes de classification supervisées	25
II.3.2.1 Méthodes d'apprentissage inductif	25
II.3.2.1.1 Méthode des k plus proches voisins (k -ppv)	26
II.3.2.1.2. Méthode bayésienne	26
II.3.2.1.3. Méthodes d'analyse discriminante	27
II.3.2.1.3.1 Discrimination linéaire	27
II.3.2.1.3.2 Discrimination quadratique	30
II.3.2.1.4 Les supports à vaste marge (SVM)	30

II.3.2.1.5 L'approche des réseaux de neurones	30
II.3.2.1.6 Affectation par arbres de décision	31
II.3.2.2 Méthodes d'apprentissage déductif : Affectation par système expert	32
II.4 Performance des méthodes de classification	33
II.5 Conclusion	33
CHAPITRE III : Les supports à vaste marge (SVM)	34
III.1 Introduction	34
III.2 Théorie de l'apprentissage de Vapnik	35
III.3 Les bases mathématiques des SVM	36
III.4 Principes des support à vaste marge (SVM)	37
III.4.1 Construction de l'hyperplan optimal	37
III.4.1.1 Cas des données linéairement séparables	37
III.4.1.2 Cas des données non linéairement séparables	39
III.4.2 Les SVM non linéaires	40
III.4.2.1 Conditions de Karush-Kuhn-Tucker (KKT)	43
III.4.2.2 Calcul du biais b	44
III.5 Algorithmes d'apprentissage des SVM	44
III.5.1 Méthode de Chunking	45
III.5.2 Méthode de décomposition successive	45
III.5.3 Méthode de minimisation séquentielle : SMO	45
III.6 les supports à vaste marge multiclassés	48
III.6.1 Un –contre –tous (One-against-rest Multi-class Classifier)	48
III.6.2 Un –contre- un (One-against-one Multi-class Classifier)	49
III.7 Conclusion	50
CHAPITRE IV : Sélection des paramètres du vecteur forme	51
IV.1 Introduction	51
IV.2 Sélection des paramètres par des méthodes séquentielles "constructive" et "destructive" (SFS & SBS)	51
IV.2.1 Principe des algorithmes d'ordonnancement	51
IV.2.1.1 Procédure séquentielle constructive (SFS)	52
IV.2.1.2 Procédure séquentielle destructive (SBS)	52
IV.3 Les algorithmes génétiques	53
IV.3.1 Principe	53
IV.3.1.1 Codification des données	53
IV.3.1.1.1 Types de codage	53
IV.3.1.2 Génération aléatoire de la population	54
IV.3.1.3 Fonction d'évaluation (fonction <i>fitness</i>)	55
IV.3.1.4 La sélection	56
IV.3.1.5 Le croisement	59
IV.3.1.6 la mutation	60
IV.3.1.7 Critère d'arrêt	60
IV.4 Conclusion	61
CHAPITRE V : Performances des SVM	62
V.1 Introduction	62
V.2 Applications sur des signaux aléatoires	62
V.2.1 L'erreur d'estimation des paramètres du modèle en fonction du rapport signal sur bruit (SNR)	64
V.2.2 Effet du choix de l'ordre sur l'estimation des coefficients du modèle AR	65
V.3 Conduite des différentes étapes d'apprentissage et de test des SVM.....	65

V.3.1 Classification des signaux aléatoires	68
V.3.2 Classification des signaux d'engrenages simulés	68
V.3.2.1 Extraction des indicateurs	69
V.3.2.2 Performances des SVM en fonction du SNR pour différents ordres	69
V.4 Conclusion	70
V.5. Application sur des signaux réels	70
V.5.1. Banc d'essai CETIM	70
V.5.1.1 Préparation Des Observations	72
V.5.1.1.1 Division Des Signaux	72
V.5.1.1.2 Filtrage.....	72
V.5.1.1.3 Décimation	73
V.5.1.1.4 Extraction Des Indicateurs	73
V.5.1.1.5 Normalisation	74
V.5.1.2 Performances des SVM (sans sélection)	76
V.5.2 Boite de vitesse d'un hélicoptère « NAVY Westland data »	77
V.5.2.1 Préparation des observations	80
V.5.2.1.1 Prétraitement des signaux	80
V.5.2.1.2 Extraction des indicateurs	81
V.5.2.3 Performances des SVM (sans sélection)	83
V.5.2.3.1 Un contre un	83
V.5.2.3.1 Un contre tous	85
V.6 Sélection des paramètres des paramètres du vecteur forme (indicateurs)	86
V.6.1 Sélection par AG	87
V.6.1.1 Architecture proposée du système de sélection par AG	87
V.6.1.2 Conduite des différentes étapes du système de sélection	87
V.6.1.2.1 Préparation des données	87
V.6.1.2.2 Codification des chromosomes de contrôle des indicateurs	89
V.6.1.2.3 Génération de la population initiale	89
V.6.1.2.4 Calcul de la fonction d'évaluation (fitness)	89
V.6.1.2.5 Nombre de génération	91
V.6.1.2.6 Le critère d'arrêt	91
V.6.1.2.7 Sélection des individus pour la reproduction	91
V.6.1.2.8 Le croisement	91
V.6.1.2.9 La mutation	92
V.7 Performances des SVM avec les indicateurs sélectionnés	92
V.7.1 Résultat et discussions	93
V.7.1.1 Le banc d'essai de CETIM (avec sélection)	93
V.7.1.2 NAVY Westland data (avec selection)	95
V.7.1.2.1 Un contre un	95
V.7.1.2.2 Un contre tous	96
V.7.1.3 Les vecteurs de supports.....	96
V.8 conclusion	97
Conclusion générale et perspectives	98
Annexes	99
Références bibliographiques	

LISTE DES ABRÉVIATIONS ET SIGLES

RDF : Reconnaissance de formes

SVM : Supports à Vaste Marge.

RBF : Noyau à base radiale.

VC : Dimension de Vapnik-Chernovenkis

VS : Vecteurs de supports

AG : Algorithmes génétiques.

N_b : Taille du chromosome

N_p : Taille de population

P_m : Probabilité de mutation

P_c : Probabilité de croisement

x_{opt} : Vecteur forme optimal

k_{max} : Nombre de génération maximal

BA : Matrice d'apprentissage

BT : Matrice de test

EA : Vecteur d'étiquettes d'apprentissage

ET : Vecteur d'étiquettes de test

S_w : Matrice de dispersion intra-classe

S_B : Matrice de dispersion inter-classe

tr : La trace de la matrice

x : Vecteur d'entrée arbitraire.

y : Sortie désirée bipolaire associée au vecteur x .

$P(x, y)$: Probabilité d'observation du couple (x, y) .

l : Nombre d'exemples d'apprentissage.

$f(x)$: Fonction de décision du SVM.

x_i : Vecteur d'entrée d'indice i .

INTRODUCTION GENERALE

Le premier rôle de la surveillance industrielle est d'augmenter la disponibilité des installations industrielles afin de réduire les coûts directs et indirects de la maintenance des équipements de production. Les coûts directs de cette maintenance sont ceux relatifs aux diverses pièces de rechange, main d'oeuvre, etc. Par contre, les coûts indirects sont essentiellement dus au manque à gagner engendré par un arrêt de production. On comprend alors que l'enjeu d'une bonne politique de surveillance est très important pour les entreprises soucieuses d'avoir une meilleure maîtrise des coûts de maintenance.

On conçoit donc, dans ce contexte, la nécessité, pour les responsables de maintenance, de chercher à n'intervenir que lorsque c'est indispensable. Ajoutons aussi que, les facultés humaines n'étant pas extensibles à l'infini, la complexité des installations actuelles imposent le recourt à de nouveaux outils, capable d'apporter une aide efficace à la décision. Ces outils doivent permettre évaluer rapidement l'état de santé d'une machine sans arrêter ou perturber son fonctionnement.

Pour établir ce diagnostic, il faut s'appuyer sur le fait que la machine émet vers l'extérieur de nombreux signaux qui sont symptomatiques de son fonctionnement, tel que la chaleur dégagée, la puissance absorbée, le bruit, les vibrations...etc. Parmi les signaux fournis, les vibrations présentent des avantages considérables dus à la richesse des informations qu'elles véhiculent [Alatt.94], de leur transmission quasi instantanée et de leur rapport direct avec le mouvement d'un mécanisme.

Une machine en bon état produit un niveau donné de vibrations. L'apparition d'un défaut ou d'une anomalie se traduit par une modification caractéristique de cet état de référence. C'est pourquoi, la plus part des méthodes modernes de maintenance sont basées sur le passage vers une maintenance prédictive (conditionnelle) susceptible de prévoir avec suffisamment de précision, l'évolution de l'état d'une machine.

Les différentes méthodes de diagnostic se regroupent en trois catégories principales [Lure03] : les méthodes symboliques, les méthodes internes, et les méthodes externes.

Les **méthodes symboliques** constituent d'avantage des outils d'aide au diagnostic que des méthodes propres au développement d'un système industriel de diagnostic. En effet, il s'agit plutôt d'outils d'analyse des relations entre les causes (défauts) et les effets (symptômes) utilisables après la détection de la présence d'une défaillance sur le procédé.

Les **méthodes internes** de diagnostic sont très intéressantes, lorsque l'on dispose d'un modèle interne du procédé ou que celui-ci est facilement accessible par modélisation. Des indicateurs de l'état de fonctionnement du procédé, nommés résidus, sont établis par comparaison entre l'état théorique du système, estimé à l'aide du modèle, et son état mesuré. Un ensemble d'outils d'analyse des résidus permet d'accomplir les tâches du diagnostic. Toutefois, l'étape la plus contraignante dans la mise en oeuvre de ces méthodes est l'établissement du modèle et l'estimation de ses paramètres, qui est d'autant plus délicate pour les systèmes complexes.

Les **méthodes de diagnostic par reconnaissance de formes (RDF)**, par opposition aux méthodes internes, n'utilisent pas de modèle analytique du procédé. En effet, les méthodes par RDF s'appuient sur une représentation de l'état de fonctionnement d'un système à l'aide d'un ensemble de caractéristiques extraites des mesures sur le procédé. L'ensemble des caractéristiques et l'ensemble des modes de fonctionnement traduisent respectivement l'espace de représentation et l'espace de décision. L'opération de diagnostic consiste à associer toute observation sur le procédé à l'un des modes de fonctionnement.

Ainsi, le **choix d'une méthode de diagnostic** dépend essentiellement de la connaissance disponible du procédé étudié, sans toutefois oublier les considérations techniques et économiques. Cependant, parmi les méthodes présentées, les plus appropriées pour le

diagnostic des systèmes complexes sont celles basées sur l'approche par RDF puisqu'elles ont l'avantage de minimiser le besoin de connaissance à priori.

L'élaboration d'un système de diagnostic par RDF se réalise en trois phases [Zwin.95] ; La première, la *phase d'analyse* comprend l'analyse des données (pré traitement, ...), la définition d'un espace de représentation, la deuxième est la phase du choix d'une méthode de décision (discrimination entre classes), qui doit attribuer toute nouvelle donnée (ou mesure) à un mode de fonctionnement. La dernière est la *phase d'exploitation* où l'objectif est d'associer aux nouvelles observations les modes de fonctionnement.

La phase du choix d'une méthode de décision constitue une issue major pour l'élaboration d'un système de diagnostic par reconnaissance de forme, car on doit choisir une méthode qui offre les meilleures performances.

Notre travail consiste à étudier les performances d'une nouvelle méthode appelée : les supports à vaste marge ou en anglais : *support vector machines* (SVM). C'est une méthode de classification récemment développés par Vapnik [Vapn.98], et souvent traduite par les séparateurs à vaste marge. L'idée essentielle des SVM est d'utiliser des fonctions noyaux et de projeter ces 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. Ces algorithmes tentent de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes.

Ainsi, le mémoire est structuré en **5 chapitres**, suivis d'une conclusion générale.

Le premier chapitre rappelle brièvement, à travers une étude bibliographique, la maintenance industrielle, en présentant les avantages de la *maintenance prédictive (conditionnelle)*, les principales causes de défaillance affectant les machines tournantes, suivi par un état de l'art sur les méthodes d'analyse des signaux vibratoires.

Le deuxième chapitre aborde l'élaboration d'un système de diagnostic industriel par L'approche **Reconnaissance de Formes**, retenue dans le cadre de notre étude, Dans ce chapitre nous présentons également un panorama des méthodes de classifications utilisées dans le cadre du diagnostic industriel.

Le troisième chapitre présente une description des supports à vaste marge (SVM), en motionnant les principes, et les bases théoriques de cette méthode, ce chapitre précise les différents algorithmes d'apprentissage des SVM en détaillant beaucoup plus l'algorithme '*Sequential minimal optimisation*' (SMO), qui sera utilisé dans la partie expérimentale. Dans ce chapitre, nous abordons aussi le cas des SVM multi classes.

Le quatrième chapitre décrit quelques méthodes de sélection des paramètres. On détaillera, beaucoup plus, le principe des *algorithmes génétiques* qui seront utilisés, dans le cadre de notre travail, comme une méthode de sélection d'indicateurs dont l'objectif est de réduire la taille du vecteur forme.

Le cinquième chapitre, présente l'application expérimentale des SVM sur des signaux réels et simulés, en étudiant les performances de cette méthode pour différents cas (classification binaire, multi classes), avant et après sélection des indicateurs, et avec différents *noyaux*.

Nous terminerons par une conclusion générale sur cette étude, ainsi que quelques perspectives pour des études supplémentaires.

CHAPITRE I : MAINTENANCE ET ANALYSE VIBRATOIRE

I.1 Introduction

Dans ce chapitre, nous introduisons l'approche de maintenance, en discutant les différents types de maintenance, les causes de défaillances, ainsi que la localisation des défauts dans les machines tournantes. Dans la deuxième partie nous introduisons les diverses méthodes de prétraitement et d'analyse des signaux vibratoires et l'extraction des indicateurs. Parmi les méthodes proposées, quelques-unes seront considérées dans la partie expérimentale de ce mémoire. Il est clair que meilleur sera le prétraitement, plus simple et plus précise sera la classification des données présentées à l'entrée du classificateur.

Après avoir parcouru l'ensemble des méthodes d'analyse des données, nous justifierons pourquoi nous utilisons une méthode plutôt qu'une autre dans la partie expérimentale de ce mémoire.

I.2 Approche de maintenance

Le fonctionnement d'une installation, même dans des conditions normales entraîne un certain vieillissement de matériel, et parfois même des incidents ou accidents. Il est nécessaire pour en assurer une exploitation correcte de maintenir cette installation en bon état. La maintenance a un coût, et il est important de trouver le niveau optimum qui maintiendra au moindre coût l'installation à niveau sans mettre en danger la sécurité des personnes.

Avant de présenter les différents types de la maintenance, attardons-nous sur quelques définitions :

Le dictionnaire « *Le Robert* » [Robe.95], donne la définition suivante de la maintenance: "Action de maintenir, techniques et ensemble des opérations d'entretien destiné à accroître la fiabilité ou pallier des défaillances".

Nous pouvons nous contenter d'une description aussi peu précise, nous pouvons nous référer à la définition normative [Afno.01] : "Ensemble de toutes les actions techniques, administratives et de management durant le cycle de vie d'un bien, destinées à le maintenir ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise".

Le management de la maintenance concerne toutes les activités des instances de direction qui d'une part déterminent les objectifs, la stratégie et les responsabilités concernant la maintenance, et d'autre part les mettent en application par des moyens tels que la planification, la maîtrise et le contrôle de la maintenance, l'amélioration des méthodes dans l'entreprise (y compris dans les aspects économiques) [Math.05].

Quoique les travaux de recherche sur ce sujet soient légion, nous n'entrerons pas dans le détail de cette analyse et nous nous limiterons à la présentation de la partition classique entre maintenance corrective et maintenance préventive.

I.2.1 La maintenance corrective

La maintenance corrective est une forme délibérée d'entretien qui consiste, une fois la défaillance survenue, à en éliminer les effets et les causes. On distingue généralement [Afno.01] :

- **la maintenance palliative** : effectue une remise en état provisoire du matériel ayant subi un dysfonctionnement : c'est un dépannage. Cette maintenance est le plus souvent associée à des systèmes ne présentant pas d'impératif de sécurité.

- **la maintenance curative** : correspond à la remise en état de l'équipement et revêt un caractère définitif. L'équipement concerné par cette maintenance retrouve, après intervention, les caractéristiques qu'il avait avant l'apparition du problème.
- **la maintenance corrective d'amélioration** : représente une forme dérivée de la maintenance curative. Elle ne se limite pas à la remise en état du système mais implique en plus une modification de ses caractéristiques afin d'éviter la répétition du problème.

I.2.2 La maintenance préventive

La décision d'intervenir précède l'apparition du dysfonctionnement, ce caractère d'anticipation dépend de la présence de ressources appropriées. Ce mode de maintenance permet de diminuer le nombre de défaillances, il doit induire un gain économique substantiel né de la différence entre les coûts générés par l'intervention et la disponibilité qu'elle occasionne. On peut distinguer deux types essentiels de maintenance préventive [Zemo.03] :

- **La maintenance systématique** : elle est caractérisée par la connaissance des dates de visite, l'intervalle d'inspection étant déterminé arbitrairement ou en fonction des lois de comportement du système.
- **La maintenance conditionnelle** : ce type de maintenance comprend toutes les tâches de restauration de matériels ou de composants non défaillants, entreprise en application d'une évaluation d'état et de la comparaison avec un critère d'acceptation préétabli (défaillance potentielle).

Contrairement à la précédente, celle-ci n'est pas basée sur la statistique, mais sur l'induction scientifique. Par cet aspect, elle rappelle le principe fondamental des sciences expérimentales, l'interaction permanente entre l'observation des faits réels et l'élaboration de théories. Cette forme de maintenance possède aussi des objectifs bien plus ambitieux. Pour une machine, elle permet de déterminer d'une part, quel organe défaillant devra être remplacé et, d'autre part, la date à laquelle imposer l'intervention. Bien évidemment pour faire de telles prédictions, il est nécessaire d'accéder à une bonne compréhension du comportement interne de la machine, et par exemple à une bonne interprétation des phénomènes vibratoires.

En effet, il n'est pas possible de prédire l'avenir d'une machine sans connaître son état de santé au temps précis par rapport à un état de référence initial. Les avantages de ce type de maintenance sont :

- élimination d'arrêts imprévus (fiabilité et productivité accrues).
- élimination des dommages secondaires (un roulement défectueux non détecté à temps peut détruire une boîte de vitesses ou provoquer un incendie).
- élimination du gaspillage des pièces de rechange (pas de remplacement inutiles).
- réduction du stock de pièces de rechange (l'alarme est donnée suffisamment tôt pour que l'on puisse commander les pièces nécessaires).
- réduction du nombre des arrêts (en profitant d'un arrêt et production planifiée pour effectuer la réparation, en intervenant sur plusieurs machines lors d'un seul arrêt).
- amélioration de la sécurité des personnels, de la qualité du service.

Les avantages d'une maintenance conditionnelle bien conçue sont tellement évidents qu'elle devrait être implantée dans la plupart des entreprises.

I.2.3 Conclusion sur la maintenance

Il n'y a pas de maintenance prédictive sans un diagnostic minimum des défauts et de leur gravité. C'est pourquoi la première démarche d'une action de surveillance est de se

demander quels défauts sont susceptibles de se rencontrer sur la machine à surveiller. Dans ce qui suit nous discutons les causes de défaillances dans les machines tournantes, cette partie sera suivie par un état de l'art concernant les méthodes de traitement et d'analyse des signaux vibratoires engendrés par ces défauts.

I.3 Causes de défaillances et localisation des défauts dans les machines tournantes

Les statistiques concernant les causes de défaillances et la localisation des défauts dans les machines tournantes permettent de conclure que les organes les plus sensibles sont les engrenages et les roulements [Alatt.94], [More.95-1]. Les causes d'avaries sont multiples, et donnent lieu à des anomalies plus ou moins graves. Dans la section suivante, nous discutons les défauts d'engrenages et de roulements ; ces défauts apparaissant par fatigue, phénomène inévitable, auquel la maintenance préventive remédie par la détection avant destruction au cours du fonctionnement du système. Les défauts de fabrication ou de montage sont écartés.

I.3.1 Défauts d'engrenages

Les incidents caractéristiques des engrenages (Figure 1.1) sont liés aux dégradations de la denture (denture cassée, ou abîmée, usure uniforme ou non, *pitting* (piqûre), écaillage localisé ou réparti, mauvais centrage). On peut aussi observer de la corrosion sous frottement, qui se traduit par un enlèvement de métal lorsque l'engrenage est mal lubrifié, ou que les efforts sont importants. [Elba.99].

On peut distinguer deux grandes familles des défauts d'engrenages [More.95-2] :

I.3.1.1 Défauts uniformes

a) l'usure (abrasive et adhésive): l'usure est un phénomène local caractérisé par enlèvement de matière dû au glissement des deux surfaces de contact. Le développement de cette usure est lié à la charge mécanique à transmettre et à la vitesse de glissement, ainsi qu'à la présence d'éléments abrasifs dans le lubrifiant.

L'usure normale est inversement proportionnelle à la dureté superficielle de la denture et elle progresse lentement. Elle est négligeable pour les dentures de petit module et très durcies (60Hrc). Une usure anormale est produite lorsque le lubrifiant est pollué de particules abrasives ou lorsqu'il est corrosif. Elle conduit à un mauvais fonctionnement voir à une mise hors service.

b) Piquûres (Pitting): (Figure 1.1.a) 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 au 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 piquûres 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.

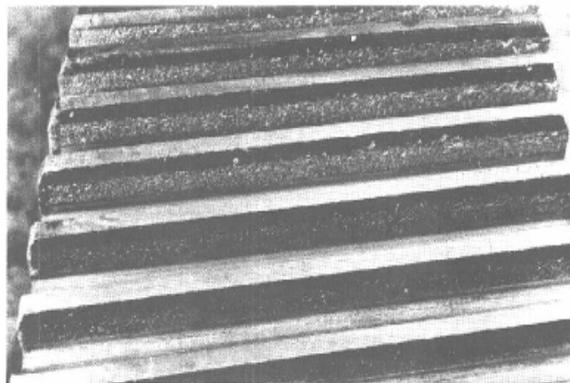
I.3.1.2 Défauts localisés

a) l'écaillage : (Figure 1.1.b) il se manifeste aussi sous forme de trous, mais ceux-ci sont beaucoup moins nombreux, plus profonds et plus étendus des piquûres. L'écaillage se trouve dans les engrenages cémentés, qui sont les plus répandus à l'heure actuelle car ils permettent de transmettre des couples importants avec de faibles dimensions. Ce type

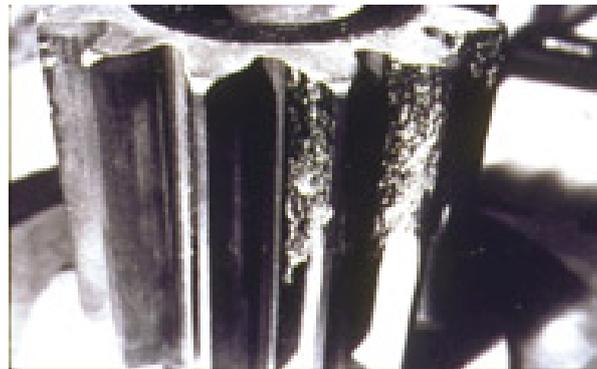
d'engrenage ne présente pratiquement pas de phénomène d'usure, l'écaillage qui est le principal défaut, évolue rapidement vers la rupture. La cause est connue : la pression superficielle est importante.

b) le grippage : il est la conséquence directe de la destruction brutale du film d'huile, sous l'effet de la température résultante d'un frottement sous charge. Le grippage est favorisé essentiellement par des vitesses élevées, de gros modules, un faible nombre de dents en contact. La probabilité de grippage est influencée par l'état physico-chimique du lubrifiant et par la condition de mise en service.

c) la fissuration : 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 dent du côté de la dent en traction.



a : Pitting sur les dents d'un engrenage



b: Défaut local d'écaillage

Figure 1.1 Défauts réels d'engrenage réparti et localisé.

Plus généralement, les détériorations superficielles d'engrenages sont le résultat de différents phénomènes pouvant intervenir au cours du fonctionnement du système.

Dans le tableau suivant nous avons récapitulé les différents défauts généralement rencontrés dans les engrenages et les modes de détection correspondants.

Défauts de denture	Signal temporel de vibration	Spectre associé
Défauts uniformes		
Usure	- Augmentation du niveau RMS - Modulation d'amplitude	- Augmentation du nombre et l'amplitude des raies d'engrènement et des harmoniques - Raies latérales
Pitting	- Idem, mais la détection semble meilleure par analyse d'huile	
Défauts localisés		
Fissure	- Impulsions - Modulation de phase	- Larges bandes - Raies latérales
Ecaillage	- Impulsions / chocs à l'état précoce	- Larges bandes
Fracture	- Chocs importants	- Larges bandes
Jeux de montage entre deux pignons (désalignement)		- Amplitude élevée à la 2 ^{ème} harmonique de la fréquence d'engrènement

Tableau 1.1 Synthèse des défauts d'engrenages [Beno.99].

I.3.2 Défauts liés aux arbres

Quel que soit le soin apporté à la construction des machines, il n'est pas possible de faire coïncider l'axe de rotation avec le centre de gravité de chaque tranche élémentaire du rotor, ce qui caractérise deux défauts majeurs dans les arbres, qui sont le balourd et le désalignement. Il en résulte que l'arbre en rotation est soumis à des efforts centrifuges qui le déforment. Ces efforts se traduisent par des vibrations liées à la fréquence de rotation f_r .

Les déséquilibres proviennent en général de défauts d'usinage, d'assemblage des rotors ou de montage. En fonctionnement, les rotors peuvent alors aussi se déformer sous l'effet d'échauffements dissymétriques [More.95-1].

I.3.2.1 Le balourd

Ce phénomène affecte toutes les machines qui possèdent une partie tournante, quelque soit leurs types, leurs dimensions, d'où le besoin apporté à leur construction. Il entraîne des vibrations qu'on trouve sur la quasi-totalité des machines surveillées. Les vibrations dues aux balourds ont des répercussions multiples sur les machines, allant d'une usure accélérée des paliers à la rupture catastrophique du rotor, en passant par les frottements ou encore des desserrements de liaison. Comme l'amplitude des forces de balourds croît proportionnellement avec le carré de la vitesse de rotation, on comprend aisément qu'à grande vitesse le plus petit déséquilibre puisse avoir des répercussions importantes [Alatt.94].

Dans le cas d'un réducteur ayant deux roues A et B l'effet de balourd qui a une composante radiale est de transmettre aux paliers fixes des forces vibratoires, synchrones à la rotation des deux roues A et B par l'intermédiaire des paliers. Cela se traduit sur le spectre de fréquence par deux composantes, l'une à la fréquence f_A de rotation de l'arbre A et l'autre à la fréquence f_B de l'arbre B. Quelques exemples de causes de déséquilibre (ou balourds) sont représentés sur la figure 1.2 [More.95-1].

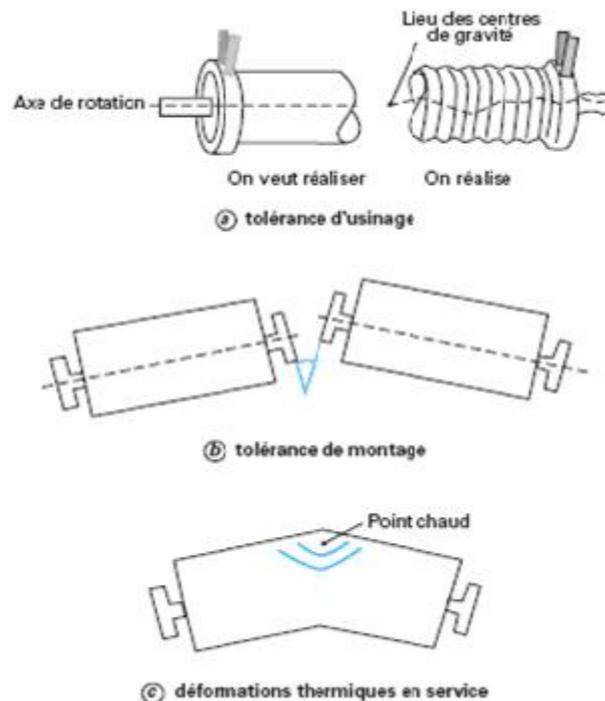


Figure 1.2 Quelques causes de déséquilibre (balourd) [Morel.95a].

I.3.2.2 Le désalignement

Le désalignement est un paramètre géométrique qui conditionne le contact des dents. Les grandes valeurs du désalignement peuvent entraîner la rupture du mécanisme ou des dentures. Dans le désalignement, il y a deux défauts : le défaut d'inclinaison et le défaut de déviation, le défaut d'inclinaison est un défaut d'alignement des arbres, il est caractérisé par un angle de rotation autour de l'axe définie perpendiculaire aux axes des deux roues et appartenant au plan défini par ces deux axes. Le défaut de déviation est caractérisé par un angle de rotation autour de l'axe défini par le défaut d'inclinaison. Ces deux défauts ont un effet différent sur le bruit d'engrènement ou sur le niveau vibratoire d'un couple d'engrènement [Alatt.94].

I.3.3 Défauts de roulements

Un roulement est constitué 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 sont des billes, des rouleaux ou des aiguilles, (on évoque ici le roulement à billes, le même développement peut être fait pour des rouleaux ou des aiguilles). La durée de vie du roulement est limitée par la fatigue superficielle des matériaux le constituant. Celle-ci est due aux contraintes créées par les passages répétitifs des billes sur les pistes.

Les principales avaries touchant les roulements peuvent être classés en deux catégories:

I.3.3.1. Défauts indépendants de la fatigue

Ces avaries peuvent être réduites et même éliminées par une bonne surveillance des causes de dégradation et un entretien approprié. Le tableau 1.2 résume ces défauts et les facteurs d'environnement s'y rapportant.

Mode d'avarie prématurée	Facteurs d'environnement
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 1.2 Défaut indépendant de fatigue [Feda.06]

I.3.3.2. Défauts dépendants de la fatigue

Le tableau 1.3 résume ces défauts, et les facteurs d'environnement s'y rapportant.

Mode d'avarie prématurée	Facteurs d'environnement
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

Tableau 1.3 Défaut dépendant de fatigue [Feda.06]

Ces avaries, dites aussi 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 [Alatt.94] [Beno.99].

Ce type d'avarie se manifeste par l'apparition de fissures qui croît lentement jusqu'à l'écaillage, C'est un phénomène aléatoire. Il existe deux 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.

I.3.3.3. Types de défauts sur les roulements

Nous distinguons quatre types de dégradations sur les roulements [Ypma.01], [Nand.99]:

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).

$$f_{d,bi} = \frac{N}{2} \left(1 + \frac{d_b}{d_m} \cos(\alpha) \right) (f_{rot} - f_{ext}) \quad (I.1)$$

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.

$$f_{d,be} = \frac{N}{2} \left(1 - \frac{d_b}{d_m} \cos(\alpha) \right) (f_{rot} - f_{ext}) \quad (I.2)$$

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.

$$f_{d,er} = \frac{1}{2} \frac{d_m}{d_b} \left(1 - \left(\frac{d_b}{d_m} \cos(\alpha) \right)^2 \right) (f_{rot} - f_{ext}) \quad (I.3)$$

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 $f_{d,ca}$.

$$f_{d,ca} = \frac{1}{2} \left(1 - \frac{d_b}{d_m} \cos(\alpha) \right) f_{rot} + \frac{1}{2} \left(1 - \frac{d_b}{d_m} \cos(\alpha) \right) f_{ext} \quad (I.4)$$

Théoriquement les défauts de roulements se manifestent par l'apparition de raies à des fréquences bien établies. Mais en pratique, l'analyse des défauts de roulements s'avère souvent complexes à cause de la variation de la vitesse et des efforts radiaux et axiaux engendrés par les pignons hélicoïdaux (variation de l'angle de contact et de la répartition des charges).

Aussi bien pour les engrenages que pour les roulements, les principaux défauts sont les écaillages de fatigue. Pour les engrenages, il peut survenir également une usure répartie sur toute la denture. Les anomalies sont localisées au niveau des contacts entre éléments mécaniques, et vont donc modifier la géométrie et les conditions de transmission de puissance. Pour un système de transmission à engrenages, la transmission est réalisée par obstacle (denture), et pour les liaisons par roulements, un mouvement relatif existe entre éléments mobiles et immobiles (arbre, bagues de roulements, éléments roulants, logements). Ces modes de transmission génèrent des vibrations naturelles. En présence de défauts, les vibrations sont modifiées, puisque les modifications de géométrie de contact provoquent des chocs plus importants que dans le cas d'un système sain.

I.4 Vibration : un bon indicateur d'état

Le fonctionnement des machines engendre des efforts qui seront souvent la cause des défaillances ultérieures (efforts tournants, turbulence, chocs, instabilité). Les efforts sont à leur tour causes de vibrations qui vont endommager les structures et les composants des machines. L'analyse de ces vibrations permet d'identifier les efforts dès qu'ils apparaissent, avant qu'ils n'aient causé de dommage irréversible. Elle permettra aussi après analyse d'en déduire l'origine et d'estimer les risques de défaillance. C'est sur ces concepts qu'est basée la maintenance prédictive. Pour la mettre en oeuvre, il faudra donc être capable de déterminer les causes de défaillance les plus fréquentes, d'évaluer leurs coûts, leur probabilité d'apparition, et de mettre en place une politique qui permet de détecter au plus tôt les symptômes [More.95-1]. En vue d'un diagnostic prédictif, il est nécessaire d'extraire des informations (indicateurs), pour la description d'un état de fonctionnement donné. A cet effet il faut avoir des connaissances suffisantes sur les méthodes d'analyse des signaux vibratoires.

I.5 Les méthodes d'analyse des signaux vibratoires

Comme les données sont obtenues à partir d'un appareil commercialisé, nous allons nous préoccuper essentiellement des étapes qui suivent. Le prétraitement des données est une étape importante car elle permet de sélectionner, dans l'espace de représentation, l'information nécessaire à l'application. Cette sélection passe par l'élimination du bruit dû aux conditions d'acquisition, par la normalisation des données, ainsi que par la suppression de la redondance. Le nouvel espace de représentation a une dimension plus petite mais demeure tout de même un espace de grande dimension. Les données subissant un prétraitement, dans ce travail de recherche, sont les signaux vibratoires. Il existe des outils dans l'environnement MATLAB permettant de traiter toutes ces données de manière efficace.

Dans la littérature, il existe quatre grandes catégories d'analyse des données vibratoires : analyse temporelle, analyse fréquentielle, analyse temps- fréquence et temps-échelle (ondelettes) [Andr.05].

I.5.1. Analyse temporelle

L'analyse dans le domaine temporel est basée directement sur le signal temporel lui-même. La distribution d'amplitude du signal peut s'exprimer sous forme d'une fonction de densité de probabilité. Pour cela, on découpe en plage de mesures de N intervalles I_j de longueur égale. On définit la probabilité $P_j(x)$ pour que l'amplitude x de la vibration soit comprise entre les bornes de l'intervalle I_j par le rapport du temps pendant lequel l'amplitude instantanée du signal a été comprise entre ces deux valeurs sur le temps total de mesure (Figure 1.3). [Elba.99] [Raad.03].

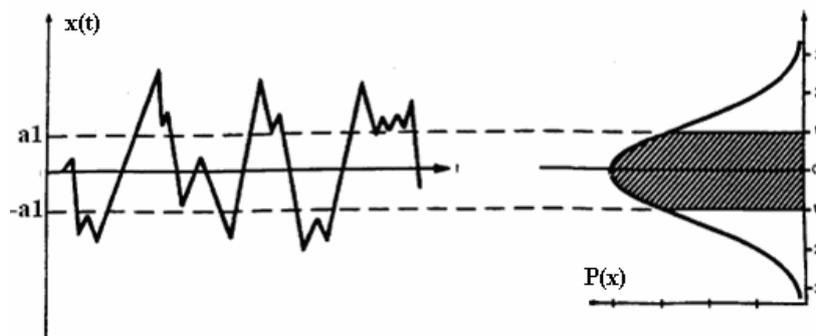


Figure 1.3 Établissement d'une courbe de densité de probabilité.
($a1$: valeur efficace du signal).

Le graphe de la figure 1.3 montre que l'allure générale du signal vibratoire peut être analysé à travers sa densité de probabilité.

$$p_j(x) = \sum_{i=1}^{i=n_j} \frac{\Delta t_{ij}}{T} \quad (I.5)$$

Pour obtenir des informations plus synthétique sur la forme de la courbe de distribution des amplitudes du signal sans la tracer, les méthodes traditionnelles calculent les caractéristiques statistiques suivantes [Andr.05] :

1-La moyenne : elle correspond au moment d'ordre 1, elle est donnée par :

$$M_1 = \bar{x} = \int_{-\infty}^{+\infty} x p(x) dx \quad (I.6)$$

2-La variance : elle correspond au moment d'ordre 2, et caractérise la dispersion des variables autour de la moyenne, elle est donnée par :

$$M_2 = \sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 p(x) dx \quad (I.7)$$

4-Le Skewness : c'est le moment d'ordre 3, il est donné par :

$$M_3 = S_k = \frac{1}{\sigma^3} \int_{-\infty}^{+\infty} (x - \bar{x})^3 p(x) dx \quad (I.8)$$

Le skewness caractérise le taux de dissymétrie de la distribution d'amplitudes du signal par rapport à la valeur moyenne.

5-Le kurtosis : le moment d'ordre 4, représente le taux d'aplatissement de la distribution d'amplitude, il donne une évaluation de l'importance du pic du sommet de la courbe. il est défini par :

$$M_4 = k = \frac{1}{\sigma^4} \int_{-\infty}^{+\infty} (x - \bar{x})^4 p(x) dx \quad (I.9)$$

6- La valeur efficace : c'est une valeur très caractéristique du signal, vu qu'elle a une relation directe avec l'énergie contenue dans celui-ci :

$$RMS = \sqrt{\frac{1}{T} \int_0^t x^2(t) dt} \quad (I.10)$$

7-Facteur de crête : c'est le rapport entre la valeur de crête et la valeur efficace (RMS).

$$F_{crête} = \frac{V_{crête}}{RMS} \quad (I.11)$$

Ces indicateurs sont appelés les *indicateurs temporels*.

Une autre approche très populaire dans le domaine temporel est le Moyennage temporel synchronisé (MTS), l'idée générale de MTS consiste à découper le signal vibratoire en segments de même longueur et d'effectuer une moyenne d'ensemble sur ces segments afin d'éliminer ou réduire le bruit, elle est donnée par :

$$\bar{s}(t) = \frac{1}{N} \sum_{n=0}^{N-1} s(t+nT), \quad 0 \leq t < T \quad (\text{I.12})$$

D'autres approches avancées appliquées dans l'analyse des signaux temporels sont les modèles paramétriques, l'idée générale de ces méthodes est de représenter le signal temporel par un modèle paramétrique et d'extraire des indicateurs basés sur les paramètres de ces modèles. Indiquons cependant que cette modélisation est une modélisation de représentation, les paramètres ne sont pas liés à la physique du phénomène (boîte noire). Parmi les modèles couramment utilisés on trouve :

a- Le modèle ARMA d'ordre p, q et noté ARMA (p, q) : il est donné par :

$$x_n = a_1 x_{n-1} + \dots + a_p x_{n-p} + \varepsilon_n - b_1 \varepsilon_{n-1} - \dots - b_q \varepsilon_{n-q}, \quad (\text{I.13})$$

Avec x_n : le signal.

a_i, b_i : les coefficients du modèle.

p, q : l'ordre du modèle.

b- Le modèle AR d'ordre p : c'est un cas particulier du modèle ARMA avec $q=0$, il est donné par :

$$x_n = - \sum_{k=1}^p a_k \cdot x_{n-k} + e_n \quad (\text{I.14})$$

Avec a_i : les coefficients du modèle autorégressif, p : l'ordre du modèle.

Le problème majeur lors du calcul du modèle autorégressif, est la sélection de l'ordre P . Pour résoudre ce problème, Aparna et al. [Apar.02] ont donné trois critères principaux pour le choix de l'ordre du modèle :

1. la variance ou l' RMS du signal erreur doit être minimale
2. la densité spectrale de puissance du signal obtenue par le modèle autorégressif (AR) doit être approximativement égale à celle obtenue par la transformée de Fourier discrète du signal réel (signal vibratoire, acoustique,....etc.).
3. le signal résiduel (différence entre le signal réel et celui obtenu par le modèle (AR) doit être constitué d'un bruit aléatoire, et donc, avoir un spectre de Fourier qui est un bruit blanc c'est-à-dire le spectre de fréquence doit être plat et ne contient pas de pics significatifs.

Une autre solution pour le choix de l'ordre du modèle autorégressif, a été donnée par Felkaoui et al [Felk.94], cette méthode est basée sur la décomposition en valeurs singulières (SVD) de la matrice de covariance, cette méthode se déroule en deux étapes essentielles :

a) Préparation de la matrice de covariance

Il s'agit d'identifier l'ordre d'un processus dynamique récurrent, sa sortie mesuré $y(i)$ est une fonction inconnue des $y(i-1), y(i-2), \dots$, précédents :

$$y(i) = -a_1 y(i-1), \dots, -a_p y(i-p) + e(i) \tag{I.15}$$

$y(i)$: sorties mesurées.

a_i : les paramètres réels du modèle.

$e(i)$: des échantillons d'un bruit blanc de moyenne nulle et de variance σ^2

Le signal prédit sur la base des p échantillons précédents mesurés à la sortie du processus s'écrit :

$$\hat{y}(i) = -\hat{a}_1 y(i-1), \dots, -\hat{a}_p y(i-p) \tag{I.16}$$

$\hat{y}(i)$: sortie prédite

\hat{a}_i : Paramètres estimés du modèle

Supposons maintenant qu'on a n mesures à la sortie, donc on peut écrire $n-p-1$ que l'équation (I.16) se réalise, la forme matricielle sera :

$$\begin{bmatrix} \hat{y}(p+1) \\ \hat{y}(p+2) \\ \cdot \\ \cdot \\ \hat{y}(n) \end{bmatrix} = \begin{bmatrix} -y(p) & -y(p-1) & \dots & -y(1) \\ -y(p+1) & -y(p) & \dots & -y(2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ -y(n-1) & -y(n-2) & \dots & -y(n-p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \cdot \\ \cdot \\ \hat{a}_p \end{bmatrix} \tag{I.17}$$

Et la forme condensée : $\hat{y} = H \hat{\theta}$

D'ici la matrice de covariance sera définie par :

$$Q(p) = \frac{1}{n-p} H^T H \tag{I.18}$$

b) le test

Dans la théorie du SVD toute matrice $m \times n$ réelle évaluée Q de rang p , où $p \leq \min(m, n)$ peut être décomposé comme $Q=URV^T$.

Où U est la matrice $m \times n$ avec $U^T U=I$, V est une matrice diagonale $n \times n$, et R est une matrice diagonale, les élément diagonaux α_i de R sont les valeurs singulières de Q arrangés dans un ordre croissant.

Toutes les valeurs singulières ne sont pas négatives et exactement p éléments entre eux sont strictement positifs

$$\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_p \geq \alpha_{p+1} = \dots = \alpha_n = 0 \tag{I.19}$$

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 utilise, jusqu'à ici, un modèle de perturbation additif donné par $Q_1=Q+E$, tandis que la matrice $Q(m \times n)$ est supposée avoir le rang $P \leq \min(m \times n)$, la matrice E est supposée avoir des valeurs petites mais elle est de rang plein. Alors la décomposition en valeurs singulières de Q_1 donnera une matrice :

$$R = \text{diag}(\beta_1, \dots, \beta_n) \tag{I.20}$$

Où $\beta_1 \geq \beta_2 \geq \dots \geq \beta_p \geq \beta_{p+1} \geq \dots \geq \beta_n = 0$ et $\beta_{p+1}, \dots, \beta_n$ seront habituellement petites, mais pas nécessairement nuls, ainsi β_p peut être également petite, le problème est de déterminer quelle valeur singulière est significative. En général nous voulons déterminer le rang correspond à l'ordre du modèle effectif en utilisant les valeurs singulières observées, ainsi le critère utilisé est le suivant : $\beta_p \gg \beta_{p+1}$

Où le rang est estimé égal à $(P_0=K)$ si

$$\frac{\beta_k}{\beta_{k+1}} = \max \frac{\beta_p}{\beta_{p+1}} \quad P = 1, \dots, P_{\max} \quad (I.21)$$

L'algorithme de recherche est le suivant :

1. former la matrice $Q(p)$ une seule fois pour $P=P_{\max}$.
2. faire la décomposition en valeurs singulières de la matrice de covariance Q .
3. et enfin utiliser le critère avec sa règle de sélection de l'ordre.

▪ Exemples d'utilisation des indicateurs

- Dans [sama.04-1] on trouve un couplage des réseaux de neurones artificiels avec les algorithmes génétiques pour le diagnostic d'engrenage d'une pompe menée par un moteur électrique, les signaux vibratoires issues de cette pompe sont divisés en 38 segments de 2048 points, trois ensembles d'indicateurs sont extraits de chaque segment :
 - 1^{er} ensemble : la moyenne, l'écart type, la variance, le skewness, le kurtosis, et les moments centrés.
 - 2^{eme} ensemble : Les mêmes indicateurs du premier ensemble sont extraits après intégral et dérivation du signal.
 - 3^{eme} ensemble : Les mêmes indicateurs sont extraits après filtrage du signal par deux types de filtres (passe-bas, passe-haut).
 Les performances du classificateur sont de l'ordre de 94.44 à 99.33% et de 100% après sélection des indicateurs par les algorithmes génétiques.
- Zhang et al., [Zhan.05] ont utilisé un ensemble d'indicateurs basé sur les moments et les cumulants pour la détection des défauts de roulement par les SVM et les réseaux de neurone artificiels, les algorithmes génétiques sont utilisés pour la sélection des indicateurs.
- Junsheng et al., [Juns.04] ont montré que l'utilisation des coefficients du modèle autorégressif (AR) comme indicateurs donne de bons résultats en classification des défauts de roulement. Dans ce travail, les signaux vibratoires ont été décomposés par la méthode EMD (Empirical Mode Decomposition).

I.5.2 Analyse fréquentielle

L'analyse dans ce domaine est basée sur la transformation du signal en domaine fréquentiel, l'avantage de l'analyse dans ce domaine par rapport au domaine temporel, est son aptitude d'identification et d'isolation facile de certains composants fréquentiels d'intérêt.

L'analyse conventionnelle la plus largement utilisée est l'analyse du spectre par la transformée de fourier elle est définie par :

$$X(f) = \int_{-\infty}^{+\infty} x(t).e^{-i2\pi ft} dt \tag{I.22}$$

L'outil le plus utilisé comme indicateur dans l'analyse spectrale est l'énergie [Andr05], elle est définie par :

$$E[X(f)X^*(f)] \tag{I.23}$$

Avec $X(f)$ est la transformée de fourrier du signal $x(t)$, 'E' est l'espérance mathématique, et (*) dénote le conjugué complexe de $X(f)$.

Il existe d'autres outils utiles pour l'analyse du spectre, qui sont l'analyse de l'enveloppe et l'analyse cepstrale.

I.5.2.1 L'analyse de l'enveloppe (aussi appelée démodulation de l'amplitude) :

Elle est basée sur La transformée d'Hilbert qui est un outil très utilisé pour le diagnostic des machines tournantes. [Sama.03],[Andr.05], la transformée d'Hilbert produit un signal temporel complexe $Z(t)$ définie par :

$$Z[X(t)] = x(t) + jH[x(t)] \tag{I.24}$$

L'enveloppe est l'amplitude de ce signal, elle représente une évaluation de la modulation présente dans le signal dû aux bandes latérales [Xian.06]. Elle révèle la modulation dans les signaux causés par un équipement défectueux. De plus, elle enlève le signal porteur et cela diminue l'influence de l'information sans apport, pour le but de découverte de défauts. La figure 1.4 représente le signal enveloppe d'une boite vitesse avec et sans défaut.

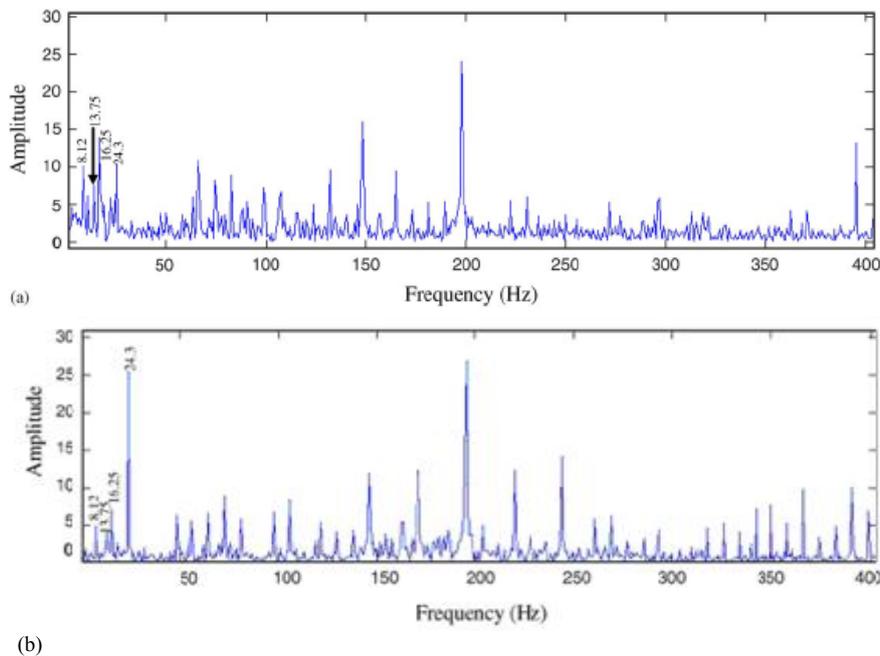


Figure1.4 Signal d'enveloppe d'une boite vitesse (a) sans défaut, (b) avec défaut

I.5.2.2 l'analyse cepstrale

C'est une méthode d'analyse qui a été développée pour le traitement de signal et elle a montré ces avantages par rapport à la transformée de Fourier rapide (FFT) dans certains cas [Cous.03], [capd.91]. Le cepstre a la capacité de détection des harmonique et les bandes latérales, La figure 1.4 représente le cepstre provenant d'une boîte de vitesses, le cepstre est défini comme étant la transformée de Fourier inverse du logarithme du spectre de puissance :

$$C(\tau) = F^{-1}[\log_{10} S(f)] \tag{I.25}$$

Avec F^{-1} est la transformé de Fourier inverse, $S(f)$ est le spectre de puissance

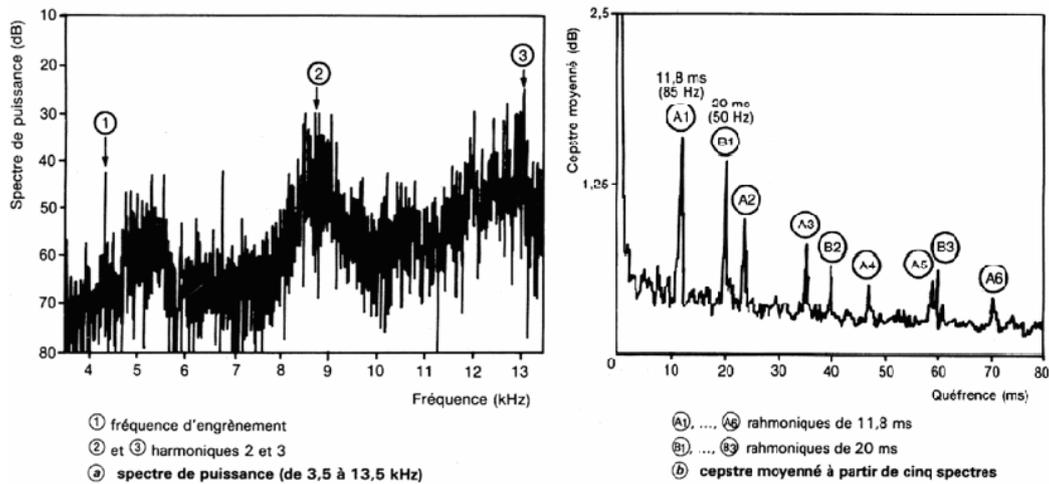


Figure 1.5 Exemple d'analyse cepstrale d'un signal vibratoire provenant d'une boîte de vitesses [Feda.06]

I.5.2.3 Exemples d'applications

- Dans [Roja.05] on trouve l'application des supports à vaste marge pour le diagnostic et la détection des défauts de l'élément roulant dans les roulements d'une pompe, ils ont appliqué la transformé rapide de fourrier (FFT) et il ont pris 33 points FFT comme indicateurs. La performance de classification a été comparée, en utilisant ces indicateurs fréquentiels avec des indicateurs statistiques extraits du signal temporel, qui sont basés sur les moments et les cumulants.
- Wong et al. [Wong.05] ont discuté la performance de classification des défauts d'engrangs en utilisant deux nouveaux ensembles d'indicateurs ; le premier est composé de quatre moments estimés de la densité spectrale de puissance (PSD) :

$$K_1 = \frac{1}{N'} \sum_{N'} P_{xx}(n)$$

$$K_2 = \frac{1}{N'-1} \sum_{N'} (P_{xx}(n) - K_1)^2$$

$$K_3 = \frac{1}{N' - K_2^{3/2}} \sum_{N'} (P_{xx}(n) - K_1)^3$$

$$K_4 = \frac{1}{N' - K_2^2} \sum_{N'} (P_{xx}(n) - K_1)^4$$

Le deuxième ensemble est formé de quatre moments estimés de la fréquence pondérée par la densité spectrale de puissance :

$$K_5 = \frac{1}{K_1} \sum_{N'} f(n) P_{xx}(n)$$

$$K_6 = \sqrt{\frac{\sum_{N'} [(f(n) - K_5)^2 P_{xx}(n)]}{K_1}}$$

$$K_7 = \frac{1}{K_1 K_6^3} \sum_{N'} [(f(n) - K_5)^3 P_{xx}(n)]$$

$$K_8 = \frac{1}{K_1 K_6^4} \sum_{N'} [(f(n) - K_5)^4 P_{xx}(n)]$$

I.5.3 Analyse temps -fréquence

Une limitation d'analyse du domaine fréquentiel et son incapacité à manier les signaux non stationnaires quand les défauts des machines se produisent, a donné lieu à l'analyse temps–fréquence, qui a été développé pour les signaux non stationnaires. L'analyse temps–fréquence traditionnelle utilise des distributions qui représentent l'énergie ou puissance des signaux en deux dimensions de temps et fréquence pour révéler les meilleurs modèles de diagnostic. La transformée de Fourier à court terme (STFT) ou spectrogramme [Andr.05] et la distribution Wigner–Ville (DWV), sont les distributions temps–fréquence les plus utilisées. La transformation de Wigner-Ville est définie par :

$$W_s(t, \nu) = \int_{-\infty}^{+\infty} S(t + \frac{\tau}{2}) S^*(t - \frac{\tau}{2}) . e^{-2\pi j \nu \tau} d\tau \tag{1.26}$$

Avec : s(t) signal complexe quelconque.

La figure 1.6 représente une distribution Wigner –Ville (DWV) pour deux niveaux de défauts d'engrenage (écaillage) et la distribution de l'énergie instantanée en fonction du temps

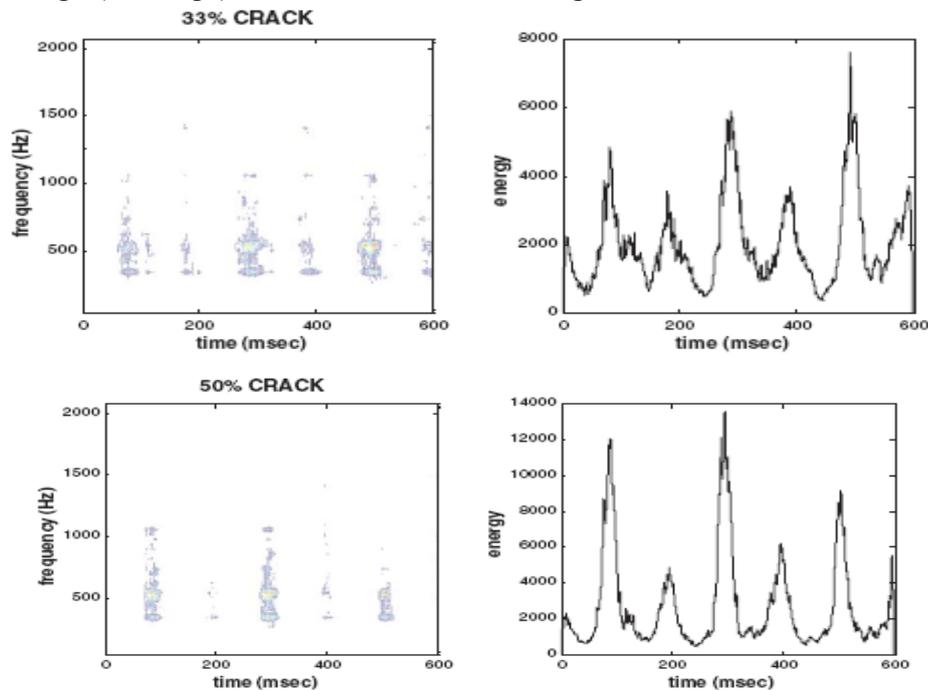


Figure 1.6 Une représentation temps fréquence (DWV) et l'énergie instantanée pour un pair d'engrenage a)30%, b) 50% écaillage de dent

Dans [Lout.06] on trouve l'utilisation de la distribution Wigner-Ville pour la détection des défauts d'écaillage de dents dans une boîte de transmission par engrenage, l'outil utilisé comme indicateur est l'énergie instantanée par unité de temps, elle est définie par :

$$E(t) = \int_{f_1}^{f_2} |W_{xx}(t, f)| df \quad (I.27)$$

La transformée de Wigner-Ville (plus généralement la distribution de Cohen) [Raad.03] possède de bonnes propriétés en fournissant une localisation précise de l'énergie dans le plan temps-fréquence

I.5.4 Analyse temps –échelle (ondelette) :

L'analyse en ondelettes consiste à décomposer le signal sur une base de fonctions à support borné. Les fonctions de base se déduisent d'une ondelette mère par dilatation et/ou contraction. On parle ainsi de décomposition temps-échelle [Raad.03]. La transformée en ondelettes d'un signal $s(t)$ est définie par :

$$C_{a, b} = \int_{-\infty}^{+\infty} S(t) \Psi_{a, b}(t) dt \quad (I.28)$$

On distingue deux types d'ondelettes[Raad.03]:

- ✓ l'ondelette continue de Morlet (qui correspond en fait à un filtrage du signal par un banc de filtres elle est définie par :

$$\Psi_{a, b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (I.29)$$

- ✓ et les ondelettes discrètes orthogonales qui garantissent l'unicité de la décomposition d'un signal.

Le choix de la meilleure 'ondelette' en fonction du problème considéré est important dans un souci d'interprétation physique de la décomposition.

Cette analyse est multi résolution, nous obtenons une bonne résolution fréquentielle en basses fréquences, et temporelle en hautes fréquences. Les coefficients de la décomposition, en nombre restreint, permettent de synthétiser le signal original avec une très bonne qualité.

Ce point est bien exploité dans les télécommunications et le traitement d'images, est d'un intérêt moindre en mécanique Ils peuvent être bien sûr utilisés pour la classification. On retrouve quelques applications en ce sens, par exemple pour le diagnostic de défauts dans les moteurs d'automobiles.

La bonne résolution en temps s'avère intéressante pour la détection de transitoires, de chocs, leur localisation temporelle étant très précise. Ainsi, plusieurs applications ont été faites pour détecter les défauts dans les mécanismes des machines alternatives et tournantes [Raad.03].

Pour les signaux d'engrenages, la transformée en ondelettes permet de mettre en évidence les non stationnarités et par conséquent, elle a été appliquée dans plusieurs études [Peng.04], [Sama.03], [Rafi.06], Par contre, pour l'analyse des machines tournantes, les représentations temps-échelles sont nécessairement difficilement interprétables à cause de la perte de la

notion de fréquences sauf en basses fréquences (BF), tellement utile pour l'interprétation mécanique.

1.5.5 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 l'identification des systèmes, et aussi pour la détection de défauts de roulement [Raad.03]. Cette hypothèse correspond à la nature physique des signaux issus des machines tournantes. 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. Plus généralement, 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, de nombreuses études montrent que les vibrations d'un système d'engrenages sont cyclostationnaires; notamment [Capd.92], [Bonn.05]. 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 [Boui.98]. Dans [Capd.95], il a été également montré que les liens qui apparaissent dans la corrélation spectrale, outil cyclostationnaire, permettent 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).

1.6 Conclusion

Dans ce chapitre nous avons présenté les différents types de maintenance, ainsi que la plupart des défauts responsables des pannes dans les systèmes mécaniques en général et des machines tournantes en particulier. Cette étude nous a permis de conclure que :

La majorité des pannes, sont causés en premier lieu par les défauts d'engrenage et de roulement.

En vue d'un diagnostic automatisé, il est nécessaire d'utiliser des indicateurs sensibles à l'apparition des défauts, et qui donnent la meilleure information sur le système surveillé. La majorité des travaux utilisent les descripteurs statistiques comme outil de diagnostic dans le domaine temporel.

L'analyse dans le domaine spectrale a présenté ces avantages en détection des défauts dans les machines tournantes, et son aptitude d'identification et d'isolation facile de certains composants fréquentiels d'intérêt.

L'analyse temps-fréquence, et temps-échelle, présentent un outil utile pour le diagnostic des machines tournantes, mais à cause de la complexité de ces méthodes, et l'extraction difficile des indicateurs, ces deux techniques seront écartées dans la partie expérimentale.

Les méthodes d'analyses des signaux vibratoires permettent d'extraire différents ensembles d'indicateurs qui seront utilisées comme entrées au classificateur, afin d'identifier un état de fonctionnement donné, ce genre de problème est le sujet de diagnostic par reconnaissance de formes qui sera abordé dans le deuxième chapitre.

CHAPITRE II : DIAGNOSTIC ET RECONNAISSANCE DE FORMES

II.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 présentons ensuite un panorama des méthodes de classification utilisées dans la RDF, en discutant brièvement les supports à vaste marge, cette méthode sera abordée avec plus de détail dans le prochain chapitre.

II.2 Principe de diagnostic par reconnaissance de formes

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 [Dubu.02]. 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.01] [Casi.03].

En reconnaissance de formes statistique 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 (Figure 2.1). \mathfrak{R}^M , autrement appelé espace de représentation.

Supposons 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 représentation 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. Cela revient à rechercher dans quelle zone de l'espace se situe la nouvelle observation (figure 2.2).

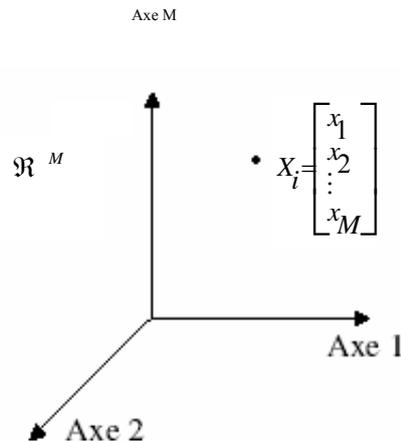


Figure 2.1 Caractérisation d'une observation par un vecteur forme représenté par un point dans \mathfrak{R}^M

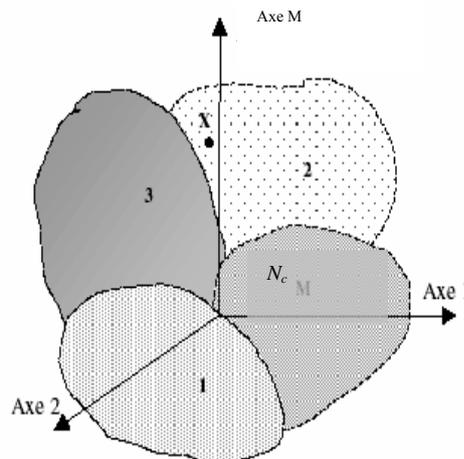


Figure 2.2 Objectif : associer une nouvelle observation X à l'une des classes

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 sur la machine. Cela est difficilement réalisable si la recherche de signature est basée sur un seul type d'analyse ou d'indicateur.

En générale la conception d'un système de diagnostic par RDF se déroule en trois phases [Beno.99], [Zwin.95] :

A. Phase d'analyse

Durant cette phase, toute l'expérience accumulée sur le processus doit être exploitée. De cette façon, on étudie les informations fournies par les capteurs et l'historique du processus.

De cette phase on doit sortir d'une définition du vecteur forme de paramètres, et la caractérisation des classes connues. De plus on dispose de l'ensemble d'apprentissage qui représente plusieurs modes de fonctionnement.

A l'issue de cette phase, l'espace de représentation de dimension M est choisi, et toute nouvelle observation sur le processus sera présentée par un point dans cette espace. On dispose également de la connaissance des N_c classes qui serviront de base au système de décision.

B. Mise en place du système de décision

Dans cette phase, on construit les frontières entre les classes. Il s'agit d'élaborer un système de décision qui décidera de l'appartenance ou non d'une nouvelle observation x à une des N_c classes. Notons que dans le système de décision choisi l'observation peut être soit attachée à une des classes connues, soit rejetée de toutes les classes.

Durant cette phase plusieurs méthodes de classification doivent être testées afin de ne garder que celle qui représente les meilleures performances.

Un panorama des méthodes de classification sera abordé dans la section (II.3) de ce chapitre.

C. Phase d'exploitation

Le vecteur forme et la règle de décision ayant été choisis, on réalise une implémentation du système de diagnostic afin de proposer une décision pour toute nouvelle observation effectuée.

Le système retenu ne doit pas être figé et doit pouvoir intégrer de nouvelles classes détectées et donc adapter sa règle de décision.

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. En général tout ou partie des modes de fonctionnement qualifiés d'anormaux n'ont jamais été rencontrés sur le système. Il est donc nécessaire de concevoir des systèmes de diagnostic adaptatif, dans les quel la connaissance incomplète peut être enrichie 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.

II.3 Panorama des méthodes de classification

La problématique de la reconnaissance des formes consiste à affecter les objets d'un ensemble A à des catégories ou classes prédéfinies. Ce type de question fait partie des problèmes de classification. Avant d'aborder une des méthodes de classification dans le cadre de notre travail, nous donnerons un panorama des méthodes de classification.

Les méthodes de classification font référence à l'existence de groupes ou classes de données et elles se divisent en deux groupes:

- Les méthodes de classification non supervisées (aussi 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.

- Les méthodes supervisées (aussi appelées «classificateurs») basées sur la notion d'apprentissage supervisé : méthodes utilisant un ensemble d'exemples où les classes d'appartenance sont connues au préalable. A partir de cet ensemble, des normes (ou règles) d'affectation seront définies.

Nous développerons ci dessous ces différentes méthodes en soulignant certains de leurs avantages et inconvénients.

II.3.1 Méthodes de classification non supervisées

Les problèmes de classification non supervisées ont été traités dans plusieurs références : [Duda.01], [bela.00], [Oukh.97], [Bess.05], [Zwin.95], [Corn.02], [Casi.03]. L'objectif de ces méthodes est de regrouper les individus en un nombre restreint de classes homogènes. Dans ce type de méthodes, les classes seront obtenues à l'aide des algorithmes formalisés et non par des méthodes subjectives. On distingue aussi les méthodes de classification non hiérarchiques et les méthodes de classification hiérarchiques.

II.3.1.1 Méthodes non hiérarchiques

Ce sont des méthodes qui produisent directement une partition en un nombre fixé de classes. Parmi ces méthodes, nous trouvons [Bela.00] :

II.3.1.1.1 Méthode de leader

Cette méthode considère chaque objet une seule fois. Lorsque le premier objet arrive, on lui attribue la première classe et il devient le leader de celle-ci. Ensuite, chaque fois qu'un nouvel objet se présente, on calcule sa distance par rapport aux leaders de chacune des classes existantes à cet instant, et on compare cette distance à un seuil. Si cette distance est inférieure au seuil fixé, on attribue au nouvel objet la classe du premier leader trouvé (pour lequel la distance calculée est inférieure au seuil), sinon une nouvelle classe est créée et le nouvel objet devient le leader de cette classe.

Cette méthode dépend de l'ordre de présentation des objets. Lorsque cet ordre n'est pas optimal, le nombre de classes augmente sensiblement. Par ailleurs, pour définir des nouveaux leaders, cette méthode utilise des distances, ce qui nous ramène au problème de la définition des métriques.

II.3.1.1.2 Méthode de k -means

Cette méthode est encore appelée *algorithme des centres mobiles*. Ce type d'algorithme, où la classe est représentée par son centre de gravité, a été étudié par plusieurs auteurs, à savoir l'algorithme k -means est l'un des algorithmes de clustering les plus connus. Il est basé sur la méthode des centroïdes (ou centres de gravité). Le principe de cette méthode est le suivant :

On se donne pour commencer, k centres arbitraires c_1, c_2, \dots, c_k où chaque c_i représente le centre d'une classe C_i . Chaque classe C_i est représentée par un ensemble d'individus plus proches de c_i que de tout autre centre. Après cette initialisation, on effectue une deuxième partition en regroupant les individus autour des m_j qui prennent alors la place des c_j (m_j est le centre de gravité de la classe C_j , calculé en utilisant les nouvelles classes obtenues). Le processus est ainsi réitéré jusqu'à atteindre un état de stabilité où aucune amélioration n'est plus possible.

Cette méthode est convergente et surtout avantageuse du point de vue calcul mais elle dépend essentiellement de la partition initiale. Il existe donc un risque d'obtenir une partition qui ne soit pas optimale mais seulement meilleure que la partition initiale. De plus, la définition de la classe se fait à partir de son centre, qui pourrait ne pas être un individu de l'ensemble à classer, d'où le risque d'obtenir des classes vides.

II.3.1.1.3 Méthode des nuées dynamiques

Cette méthode peut être considérée comme une généralisation de la méthode des centres mobiles. Le principe de la méthode est le suivant :

On tire au hasard k noyaux parmi une famille de noyaux (chaque noyau contient un sous-ensemble d'individus). Puis chaque point de l'ensemble d'apprentissage est affecté au noyau dont il est plus proche. On obtient ainsi une partition en k classes dont on calcule les noyaux. On recommence le processus avec les nouveaux noyaux et ainsi de suite jusqu'à ce que la qualité de la partition ne s'améliore plus. Cette méthode a l'avantage de traiter rapidement de grands ensembles d'individus. Elle fournit une solution dépendant de la configuration initiale et nécessite le choix du nombre de classes. En général le nombre de classes est fixé par l'utilisateur et l'initialisation est faite par un tirage au hasard. Pour comparer l'individu avec les noyaux, cette méthode utilise des distances, ce qui a l'inconvénient d'établir des métriques.

Les méthodes non hiérarchiques permettent de traiter rapidement de grands ensembles d'individus, mais elles supposent que le nombre des classes est fixé au départ. Si le nombre de classes n'est pas connu ou si ce nombre ne correspond pas à la configuration véritable de l'ensemble d'individus (d'où le risque d'obtenir des partitions de valeurs douteuses), il faut presque toujours tester diverses valeurs de k , ce qui augmente le temps de calcul. C'est pourquoi, lorsque le nombre des individus n'est pas trop élevé, on préfère utiliser les méthodes hiérarchiques.

II.3.1.2. Méthodes hiérarchiques

La classification hiérarchique consiste à effectuer une suite de regroupements en classes de moins en moins fines en agrégeant à chaque étape les objets ou les groupes d'objets les plus proches. Elle fournit ainsi un ensemble de partitions de l'ensemble d'objets. Cette approche utilise la notion de distance, qui permet de refléter l'homogénéité ou l'hétérogénéité des classes. Ainsi, on considère qu'un élément appartient à une classe s'il est plus proche de cette classe que de toutes les autres.

La principale difficulté de cette méthode est la définition du critère de regroupement de deux classes, c'est-à-dire la détermination d'une distance entre les classes. [Bela.00].

II.3.2 Méthodes de classification supervisées

Les méthodes d'affectation ou "classificateurs" sont caractérisées par la phase d'apprentissage qui consiste à établir des règles de classification à partir des connaissances disponibles a priori. Cette phase peut être réalisée à partir d'un apprentissage inductif ou déductif (par système expert). Le premier type d'apprentissage permet de passer de cas particuliers à des lois plus générales «si les hommes x , y , z , etc. sont mortels, alors on peut poser comme hypothèse d'induction que l'homme est mortel». Par contre le deuxième type permet de passer d'un cas général à un cas plus particulier «si l'hypothèse que tous les hommes sont mortels, est vrai, alors en conclusion Socrate, qui est un homme, est mortel». Les méthodes présentées dans ce chapitre utilisent soit l'apprentissage inductif soit l'apprentissage déductif mais pas les deux à la fois. Ces méthodes interviennent dans plusieurs domaines tels que la reconnaissance des formes, les statistiques, les réseaux connexionistes (réseaux de neurones artificiels), et l'intelligence artificielle... etc. [Belac.00][Duda.01]. Nous aborderons ici quelques-unes de ces méthodes dans chacun de ces domaines.

II.3.2.1 Méthodes d'apprentissage inductif

Les méthodes d'apprentissage inductif consistent à inférer des règles de décision à partir d'exemples des différentes classes. Ceci se fait dans le but d'une généralisation afin de prédire des

nouveaux cas, sur la base des paramètres les décrivant. Parmi les méthodes utilisant ce type d'apprentissage on trouve [Casi.03]:

II.3.2.1.1 Méthode des k plus proches voisins (k -ppv)

Le principe général de la méthode des k -ppv consiste à rechercher parmi l'ensemble d'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 [Corn.02].

Si $k = 1$, alors l'individu est affecté à la classe du plus proche voisin de l'ensemble T . Une variante de la règle de la majorité consiste à prévoir un seuil ' s ' au-dessus duquel une décision de rejet est prise. Ainsi, on peut rencontrer des cas où l'individu n'est affecté à aucune classe. En réalité, 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 x' les plus proches (figure 2.3).

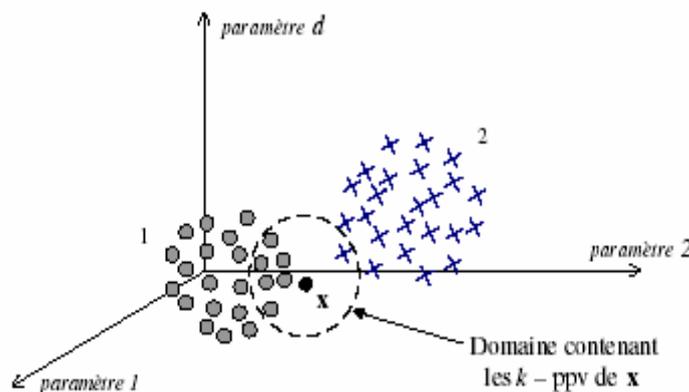


Figure 2.3 Illustration de l'effet de l'affectation par la règle des k -ppv [Casi03]

La méthode des k -ppv a l'avantage d'être très simple à mettre en oeuvre et d'utiliser directement l'ensemble d'apprentissage T . 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. Cependant, il est souvent, nécessaire de faire varier ce nombre k pour obtenir les meilleurs résultats possibles. 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 pour déterminer les voisins de l'individu à affecter, ce qui peut poser des problèmes si les dimensions à agréger ne sont pas homogènes.

II.3.2.1.2. Méthode bayésienne

Cette règle de décision consiste à affecter à chaque observation représentée par un vecteur paramètres Y à la classe la plus probable. L'approche bayésienne a pour but de minimiser la probabilité d'erreur de classification, c'est-à-dire elle utilise la valeur de la probabilité a posteriori qu'une observation x provienne de la classe w_i :

$$\Pr(w_i|x) = \frac{f(x|w_i).Pr(w_i)}{\sum_{j=1}^m f(x|w_j).Pr(w_j)} \quad (\text{II.1})$$

Où $f(x|w_i)$ est la densité de probabilité de \mathbf{x} dans la classe w_i .

La quantité au dénominateur est appelée densité de probabilité non conditionnelle à \mathbf{x} , ou loi de mélange.

La règle de Bayes définit l'affectation du vecteur \mathbf{x} à l'une des classes par [Casi.03] :

$$x \rightarrow w_i \text{ si } \Pr(w_i|x) = \max_{j=1..m} [\Pr(w_j|x)] \quad (\text{II.2})$$

Une caractéristique importante des données soumises à la méthode est la probabilité "Pr" avec laquelle les différentes classes apparaissent dans la population considérée. Elle est appelée *probabilité à priori*. En pratique, cette distribution est estimée à partir des fréquences observées dans les données, sauf si une connaissance à priori du domaine peut les fournir [Oukh.97].

II.3.2.1.3. Méthodes d'analyse discriminante

Les méthodes d'analyse discriminante ont été largement étudiées [Oukh.97], [Zwin.95], [Bela.00], [Corn.02], [Duda .01]. Une alternative pour résoudre le problème de décision est de fixer à priori le nombre et la forme des surface discriminantes permettant de séparer les classes . Ce genre de méthodes se base sur les travaux de Fisher (1936). Suivant les formes des classes, on peut trouver différents types de discrimination :

II.3.2.1.3.1 Discrimination linéaire

Elle consiste à séparer les classes par des frontières linéaires afin de regrouper les points à classer autour du centre de gravité de la classe (la moyenne de la classe), et à créer aussi des frontières linéaires entre les classes.

Dans le cas où on aurait n variables, la fonction de discrimination devient [Oukh.97] :

$$d(a) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1} \quad (\text{II.3})$$

Cette fonction dépend de paramètres w_1, \dots, w_n, w_{n+1} . La détermination de ces paramètres se fait par un algorithme d'apprentissage qui vise à satisfaire le critère associé au modèle. En fonction des données, le critère le plus utilisé pour ajuster ces paramètres est celui qui vise généralement à minimiser l'erreur de classification.

A. Discriminateur de Fisher

Il est basé sur la maximisation du critère de Fisher défini, pour chacun des paramètres du vecteur forme. Rappelons que ce critère tient compte à la fois de l'éloignement entre les classes et de leur compacité. Le but de l'approche est de rechercher le vecteur \mathbf{w} pour lequel le critère est maximal.

Prenons la formulation suivante :

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \quad (\text{II.4})$$

Le critère étant dérivable par rapport aux composantes de \mathbf{w} , sa valeur optimale sera obtenue en annulant son gradient. On obtient ainsi une solution directe pour \mathbf{w} :

$$w = S_w^{-1}(m_1 - m_2) \quad (II.5)$$

S_w et S_B sont respectivement les matrices de dispersion *intra - classe* et *inter - classes*, calculées pour les deux classes entre lesquelles la frontière séparatrice est recherchée :

$$S_w = N_1 S_{w_1} + N_2 S_{w_2} + \dots + N_c S_{w_c} = \sum_{i=1}^{N_c} N_i S_{w_i} \quad (II.6)$$

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

$$S_{w_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij} - m_i)(x_{ij} - m_i)^t \quad (II.7)$$

Avec

N_c : Nombre de classes

N_i : Nombre de vecteurs dans la classe ω_i

x_{ij} : $j^{\text{ième}}$ vecteur de la classe ω_i , m_i : Centre de gravité de la classe ω_i .

m : Centre de gravité de l'ensemble d'apprentissage : c'est le vecteur moyen total

$m = [m_1 \ m_2 \ \dots \ m_j \ \dots \ m_M]^t$ obtenu à partir de l'ensemble des échantillons $m_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$.

N : Nombre de vecteurs de l'ensemble d'apprentissage

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 :

$$S_B = \sum_{i=1}^{N_c} N_i (m_i - m)(m_i - m)^t \quad (II.8)$$

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

$$S = S_w + S_B \quad (II.9)$$

Etant donné que le centre de gravité m de l'ensemble est situé sur la frontière, alors la valeur du seuil w_0 peut être directement obtenue :

$$w_0 = -m^t w \quad (II.10)$$

Avec :

$$m = \frac{N_1 m_1 + N_2 m_2}{N} \quad (II.11)$$

B. Décision entre plusieurs classes

Une phase de décision en diagnostic par reconnaissance des formes est rarement limitée à une dichotomie. Au même titre que les approches statistiques, les méthodes d'analyse discriminante peuvent parfaitement être adaptées aux problèmes de décision entre plusieurs modes de fonctionnement. Le moyen le plus simple pour traiter le cas "multiclasses" est d'utiliser une approche par dichotomies successives :

$$\begin{cases} \text{si } g_{ij}(x) > 0 \text{ alors tester } g_{iu}(x) \text{ avec } (u = 1 \dots M \text{ et } u \neq j) \\ \text{si } g_{ij}(x) < 0 \text{ alors tester } g_{uj}(x) \text{ avec } (u = 1 \dots M \text{ et } u \neq i) \end{cases} \quad (\text{II.12})$$

Avec $g(x)$ est la fonction de décision ou règle de décision

Les frontières ayant été définies au préalable, il suffit d'appliquer séquentiellement la règle de décision entre deux classes w_i et w_j à toute nouvelle observation \mathbf{x} . La généralisation que nous avons utilisée est schématisée par l'arbre de décision représenté sur la figure 2.4. Chaque dichotomie testée (en accord avec (II.12)) offre deux solutions ($g > 0$ ou $g < 0$). Ainsi, $M - 1$ tests sont réalisés pour tout nouveau vecteur \mathbf{x} . Ce n'est que lors du test final entre la classe w_M et la classe ayant remporté la dichotomie précédente, que la classification de la nouvelle observation \mathbf{x} est réalisée [Casi.03].

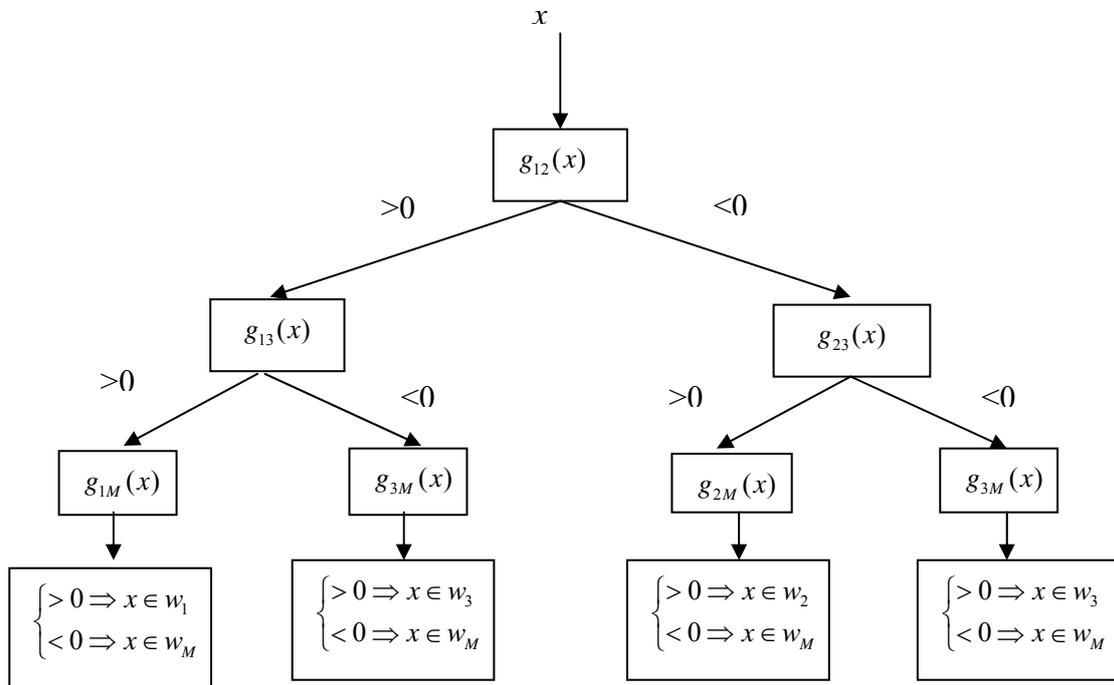


Figure 2.4 Généralisation de l'analyse discriminante au problème multiclasse [Bela.00]

Les rejets d'ambiguïté et de distance peuvent aisément être introduits dans les règles de décisions utilisant les frontières :

▪ **Rejet d'ambiguïté**

On comprend intuitivement qu'aucune décision ne puisse être prise concernant un point situé à la frontière entre deux classes. Le rejet peut être appliqué en définissant une marge de sécurité ϵ telle que :

$$X \rightarrow w_0 \text{ si } |g(x)| \leq \zeta \quad (\text{II.13})$$

▪ Rejet de distance

Comme pour la règle des $k - ppv$, il est possible de fixer un seuil pour la distance entre la nouvelle observation et sa classe d'affectation supposée. La décision entre classes est faite par le calcul direct des frontières séparatrices. Toutefois il est important de noter que ces approches de discrimination linéaires ne peuvent fournir des résultats acceptables que dans le cas où les modes seraient effectivement séparables. De plus deux problèmes peuvent se poser:

- 1) il n'est pas certain qu'il existe des frontières répondant aux équations.
- 2) si une solution existe et fournit de bons résultats avec l'ensemble d'apprentissage, cela ne signifie pas pour autant qu'elle soit parfaitement adaptée aux nouvelles observations. Les coefficients des équations mathématiques devront alors être actualisés [Casi.03].

II.3.2.1.3.2 Discrimination quadratique

Le principe de cette méthode est le même que celui développé précédemment excepté qu'au lieu de séparer les classes par des hyperplans, on les sépare par des surfaces qui ont généralement la forme ellipsoïde. La discrimination quadratique utilise plusieurs métriques (une par classe) pour mesurer la dispersion de chaque classe et la règle de décision est donnée comme suit [Casi.03] :

$$f(x) = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=1}^d w_{ij} x_i x_j \quad (\text{II.14})$$

II.3.2.1.4 Les supports à vaste marge (SVM)

Les supports à vaste marge (SVM) sont des algorithmes de classification récemment développés par Vapnik [Vapn.98], [Vapn.00]. L'idée essentielle 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 appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables. Ces algorithmes tentent de minimiser les erreurs d'apprentissage tout en maximisant la marge séparant les données des classes. 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 III)

II.3.2.1.5 L'approche des réseaux de neurones

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. 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. Plusieurs types de réseaux de neurones et plusieurs algorithmes d'apprentissage existent en littérature.

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 deux architectures les plus utilisées en surveillance industrielle, à savoir le Perceptron Multi Couches (PMC) et les Réseaux à Fonctions de base Radiales (RFR). La figure 2.5 montre l'architecture générale d'une application de surveillance par reconnaissance

des formes avec réseaux de neurones. L'expert humain joue un rôle très important dans ce type d'application.

Toute la phase d'apprentissage supervisé du réseau de neurones 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 [Zemo.03].

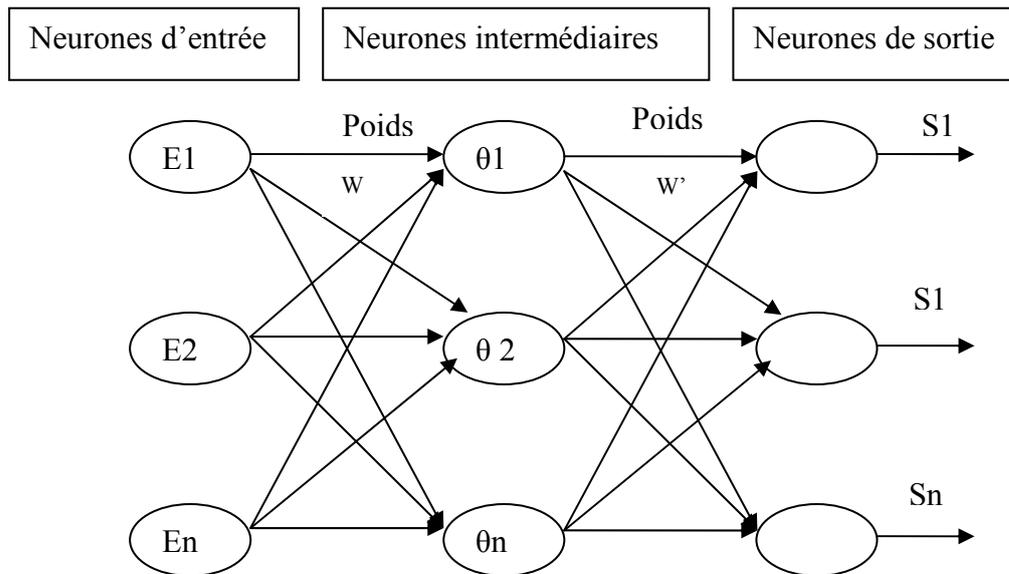


Figure 2.5 Structure d'un réseau de neurones artificiels [Zemo.03].

II.3.2.1.6 Affectation par arbres de décision

L'approche d'arbre de décision (AD) a été largement étudiée et appliquée dans le domaine de classification supervisée. Les arbres de décision sont une des techniques les plus populaires basées sur l'apprentissage supervisé. Cette méthode est très facile à mettre en oeuvre, de plus, on peut facilement interpréter les règles de décision issues de l'apprentissage. Nous considérons ici un exemple de classification de textes représentés par bags of word avec des poids binaires[Call.03].

Un AD est constituée, comme tout arbre, de noeuds et de feuilles. Chaque nœud contient un test et possède autant de descendants qu'il y a de valeurs possibles pour ce test, les tests consisteront simplement à vérifier la présence d'un terme dans le document analysé. On aura donc affaire à des arbres binaires. Aux extrémités des branches de l'arbre, on trouve des feuilles qui indiquent le résultat de la classification, c.à.d la classe que l'on assigne. Pour classer un nouveau document, il suffit de le soumettre à la racine de l'arbre et de le laisser parcourir les branches au fil des résultats des tests jusqu'à atteindre une feuille.

Dans le cadre des AD, l'entraînement consiste à créer l'arbre à partir de l'entraînement (training set). Il s'agit d'un processus récursif dans lequel les données d'entraînement seront utilisées pour déterminer l'attribut à tester lors de la création des noeuds.

Au départ, on dispose d'un ensemble de documents correspondants à l'ensemble d'entraînement complet. Sur la base de cet ensemble, on va déterminer un attribut (dans ce

cas, il s'agit de l'indice d'un terme) sur lequel le test de la racine de l'arbre va porter. Cet ensemble va ensuite être divisé en deux sous-ensembles contenant respectivement les données ayant et n'ayant pas satisfait au test. Le processus est alors répété sur ces deux sous-ensembles. Lorsque l'on ne peut plus trouver d'attribut permettant de subdiviser l'ensemble, une feuille portant le label majoritairement représenté par les exemples que l'on a reçus, est créée. L'idée générale est de créer des tests qui discriminent le plus possible les exemples d'apprentissage. En clair, cela veut dire que l'attribut sur lequel on fait le test doit séparer les exemples en deux ensembles de taille voisine.

II.3.2.2 Méthodes d'apprentissage déductif : Affectation par système expert

Les méthodes d'apprentissage déductif utilisent un raisonnement analytique qui est basé sur des inférences déductives dans le but est de transformer un ensemble de connaissance sous une forme désirée par l'utilisateur. Parmi les exemples utilisant ce type d'apprentissage on a les systèmes experts et la théorie des ensembles approximatifs (rough sets). Dans ce type d'apprentissage on présentera seulement les méthodes de classification utilisant les systèmes experts.

Un système expert a pour objectif de reproduire le comportement de l'expert lors de la résolution d'un problème, prenant appui sur une représentation des connaissances de ce dernier

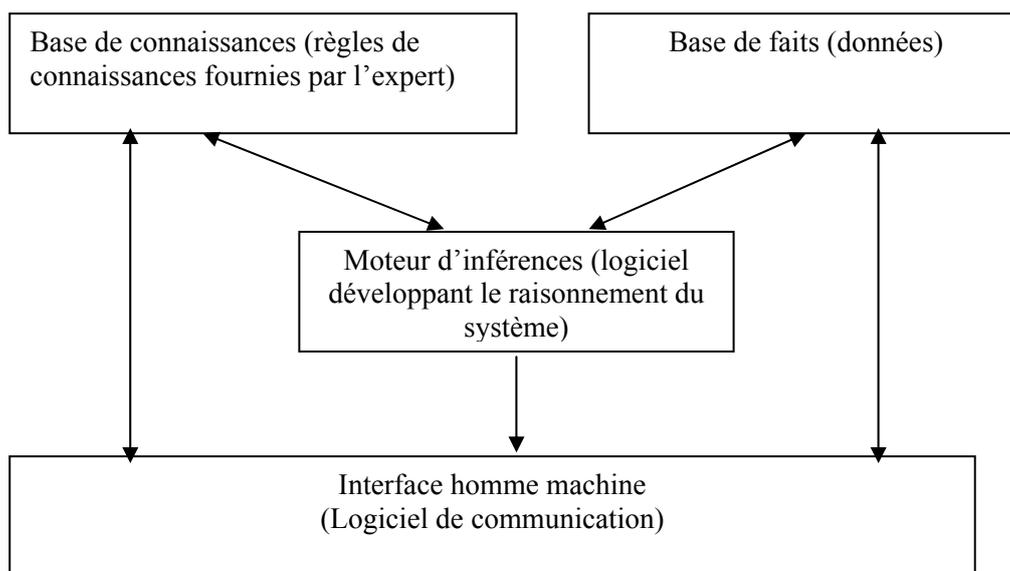


Figure 2.6 Architecture d'un système d'affectation par système expert

Les connaissances sont représentées par une base de règles et une base de faits. Les règles sont des assertions données sous formes d'implications. Elles peuvent être interprétées comme des conditions à réaliser pour déclencher une action donnée, et elles ont la forme suivante :

Si < conditions >, alors < actions/conclusion >.

La base des faits contient des assertions qui ne sont pas exprimées sous formes d'implications. Elle représente une connaissance relevant du cas particulier de l'individu à traiter, laquelle peut être fournie au système ou bien déduite par celui-ci. Pour affecter les individus aux différentes classes, le système cherche l'ensemble de règles applicables en effectuant un choix puis il applique la règle choisie et recommence le cycle. Le processus s'arrête lorsqu'il n'y a plus de règle applicable ou lorsque le but est atteint. Ce traitement est appelé moteur d'inférence. L'efficacité de ce raisonnement dépend de la pertinence du choix des règles.

L'affectation des individus se fait à l'aide d'un ensemble de règles comme dans les méthodes utilisant l'arbre de décision. Dans les systèmes experts, les classes et les règles d'affectation sont données par une expertise et non pas par un ensemble d'exemples (à l'inverse, les arbres de décision utilisent un ensemble d'exemples pour déterminer les classes et les règles d'affectation) [Bela.00].

II.4 Performance des méthodes de classification

La plus part des méthodes de classification mentionnées dans ce chapitre ont été largement appliquées dans plusieurs domaines. La question qu'on peut se poser est la suivante : comment évaluer les performances d'une méthode de classification ? En général, on divise l'ensemble de données disponible en deux sous-ensembles : l'un servira pour l'apprentissage et l'autre pour le test. L'ensemble d'apprentissage est utilisé pour déterminer les paramètres du modèle de classification, par exemple les paramètres w et b dans le cas supports à vaste marge. L'ensemble de test sert pour 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.

II.5 Conclusion

Dans ce chapitre nous avons montré le principe de diagnostic par reconnaissance de formes, en motionnant les phases principales pour la mise en ouvre de tel système de diagnostic, la première phase est liée à l'étude des informations fournies par les capteurs. La deuxième phase est la plus importante, là on doit choisir une méthode de décision qui offre une meilleure performance, Nous avons choisi dans le cadre de notre travail d'étudier les performances d'une nouvelle méthode d'apprentissage : les supports à vaste marge (SVM). La théorie, et les principes de cette méthode seront présentés dans le prochain chapitre.

CHAPITRE III : LES SUPPORTS À VASTE MARGE (SVM)

III.1 Introduction

Les « *Supports 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. Le principe est donc de trouver un classificateur, 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 [Vapn.98], [Vapn.00]. Elle s'est focalisée sur les propriétés de généralisation (ou prédiction) d'un modèle en contrôlant sa complexité. La dimension de Vapnik Chernovenkis (VC) est un indicateur du pouvoir séparateur d'une famille de fonctions associé à un modèle et qui en contrôle la qualité de prédiction. Le principe fondateur des SVM est justement d'intégrer à l'estimation le contrôle de la complexité c'est-à-dire le nombre de paramètres qui est associé dans ce cas au nombre de vecteurs de supports.

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 [Bess.05]:

- 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.
- le deuxième est le passage à la recherche de surfaces séparatrices non linéaires 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 machine à noyau ou *kernel machine*.

Sur le plan théorique, la fonction noyau définit un espace hilbertien, dit auto-reproduisant et isométrique par la transformation non linéaire de l'espace initial et dans lequel est résolu le problème linéaire. Cet outil devient largement utilisé dans de nombreux types d'application et s'avère un concurrent sérieux aux algorithmes les plus Performants. 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...). À noter que, sur le plan algorithmique, ces algorithmes sont plus pénalisés par le nombre d'observations, c'est-à-dire le nombre de vecteurs de supports potentiels, que par le nombre de variables. Néanmoins, des versions performantes des algorithmes permettent de prendre en compte des bases de données volumineuses dans des temps de calcul acceptables.

Le livre de référence sur ce sujet est celui de Vapnik [Vapn.98]. De nombreuses introductions et présentations des SVM sont accessibles sur des sites comme par exemple : www.kernel-machines.org.

III.2 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 :

$$(x_1, y_1), \dots, (x_i, y_i) \in \mathbb{R}^N \times \{\pm 1\} \quad (\text{III.1})$$

Où x_i représente la $i^{\text{ième}}$ observation de l'ensemble d'apprentissage et y_i son étiquette. Et soit l'ensemble de fonctions f_α défini :

$$\{f_\alpha : \alpha \in A\}, f_\alpha : \mathbb{R}^N \rightarrow \{\pm 1\} \quad (\text{III.2})$$

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 f_α qui approxime au mieux cette distribution consiste à minimiser le risque réel donné par :

$$R(\alpha) = \int \frac{1}{2} |f_\alpha(x) - y| dP(x, y) \quad (\text{III.3})$$

b) Risque empirique

Dans l'équation (III.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 :

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^l \frac{1}{2} |f_\alpha(x) - y| \quad (\text{III.4})$$

Si l'ensemble d'apprentissage est fini, la minimisation du risque empirique donné par l'équation III.4 ne garantit pas un minimum pour le risque réel [Ayat.04].

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 :

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \phi\left(\frac{h}{l}, \frac{\log(\eta)}{l}\right) \quad (\text{III.5})$$

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 :

$$\phi\left(\frac{h}{l}, \frac{\log(\eta)}{l}\right) = \sqrt{\frac{h(\log \frac{2l}{h}) - \log(\frac{\eta}{4})}{l}} \quad (\text{III.6})$$

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 (III.5) constitue un principe essentiel de la théorie des supports à vaste marge qui nécessite encore quelques remarques pertinentes.

Selon l'équation (III.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 :

$$S_1 \subset S_2 \subset \dots \subset S_n \subset \dots, \tag{III.7}$$

et dont les capacités respectives vérifient l'inégalité :

$$h_1 \leq h_2 \leq \dots \leq h_n \leq \dots, \tag{III.8}$$

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 $\{f_\alpha: \alpha \in A_n\}$ possibles, qui minimise le risque garanti pour désigner la limite supérieure de l'erreur de généralisation donnée dans l'équation (III.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

III.3 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.97] :

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,

$$L(x, \alpha) = f(x) + \sum_{i=1}^k \alpha_i f_i(x) \tag{III.9}$$

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 :

$$\min_x L(x, \alpha^*) = L(x_0, \alpha^*) = \max_{\alpha \geq 0} L(x_0, \alpha) \quad (\text{III.10})$$

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.

III.4 Principes des supports à vaste marge (SVM)

Les supports à vaste marge ou SVM en abrégé ont été le sujet de plusieurs travaux [Sama.04-2], [Yang.05-1], [Yang.05-2], [Jack.01], [Khar.02], les principes, et les bases théoriques de cette méthode ont été discutés dans plusieurs ouvrages, les plus connues sont [Vapn.98], [Vapn.00], [Burg.98].

III.4.1 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 :

$$(x_1, y_1), \dots, (x_i, y_i) \in \mathbb{R}^N \times \{\pm 1\} \quad (\text{III.11})$$

III.4.1.1 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 3.1 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 :

$$\begin{cases} w \cdot x_i + b \geq 1 & \text{si } y_i = 1 \\ w \cdot x_i + b \leq -1 & \text{si } y_i = -1 \end{cases} \quad (\text{III.12})$$

Ceci est équivalent à :

$$y_i(w \cdot x_i + b) \geq 1 \quad \text{Pour } i=1, \dots, m \quad (\text{III.13})$$

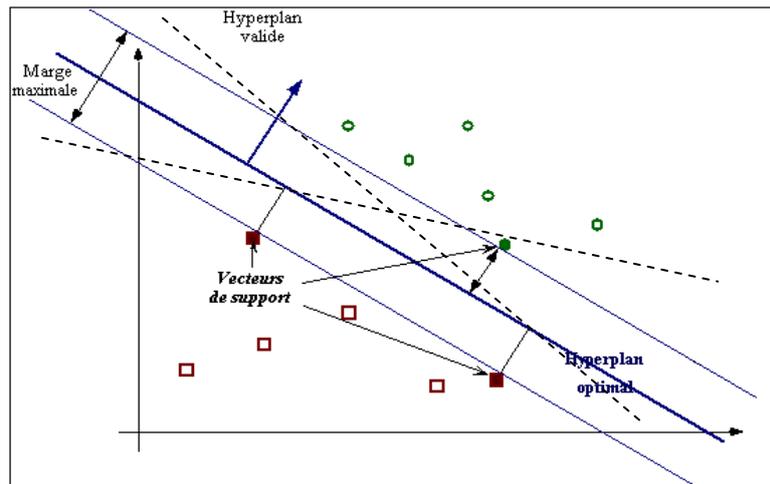


Figure 3.1 Exemple de séparation entre deux classes par les SVM [Corn.02]

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 d'un point à l'hyperplan est :

$$d(x) = \frac{|w \cdot x + b|}{\|w\|} \tag{III.14}$$

L'hyperplan optimal est celui pour lequel la distance aux points les plus proches (*marge*) est maximale. Cette distance vaut :

$$\begin{aligned} M &= \min_{x_i / y_i} \frac{w \cdot x + b}{\|w\|} - \max_{x_i / y_i} \frac{w \cdot x + b}{\|w\|} \\ &= \frac{1}{\|w\|} - \frac{-1}{\|w\|} \\ &= \frac{2}{\|w\|} \end{aligned} \tag{III.15}$$

Maximiser la marge revient donc à maximiser $\frac{2}{\|w\|}$ ce qui est équivalent à minimiser

$\frac{\|w\|^2}{2}$ sous la contrainte (III.13) Ceci est un problème de minimisation d'une fonction objective quadratique avec contraintes linéaires [Vapn.98] :

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \forall i \ y_i (w \cdot x_i + b) \geq 1 \end{cases} \tag{III.16}$$

L'introduction des multiplicateurs de Lagrange donne le lagrangien primaire qui s'écrit :

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [y_i (w^T x_i + b) - 1] \quad (III.17)$$

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 :

$$\begin{aligned} \frac{\partial L(w, b, \alpha)}{\partial w} &= 0 \\ \frac{\partial L(w, b, \alpha)}{\partial b} &= 0 \end{aligned} \quad (III.18)$$

Résoudre l'équation (III.18) donne :

$$\begin{aligned} w &= \sum_{i=1}^m \alpha_i y_i x_i \\ \sum_{i=1}^m \alpha_i y_i &= 0 \end{aligned} \quad (III.19)$$

m : nombre des points pour lesquels $\alpha_i > 0$.

En substituant l'expression de w dans l'équation (III.17) on peut écrire le problème dual sous la forme :

$$\begin{cases} \max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \forall i \alpha_i \geq 0 \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \quad (III.20)$$

Donc on peut écrire la fonction de décision des SVM linéaires sous la forme suivante :

$$f(x) = \text{sgn}(\sum_{i=1}^m \alpha_i^* y_i (x \cdot x_i) + b^*) \quad (III.21)$$

Où les coefficients α_i^* et b^* sont obtenus par résolution de l'équation (III.20).

III.4.1.2 Cas des données non linéairement séparables

Dans ce cas où les données sont non-linéairement séparables (figure3.2), 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 $i = 1, \dots, m$ appelées variables d'écart. Ces variables transforment l'inégalité (III.16) comme suit [Khar02] :

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum \zeta_i \\ \forall i \ y_i(w \cdot x_i + b) \geq 1 - \zeta_i \end{cases} \quad (III.22)$$

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 (III.20) par rapport à α_i sous les contraintes suivantes :

$$\begin{cases} \max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \forall i \ 0 \leq \alpha_i \leq C \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \quad (III.23)$$

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.

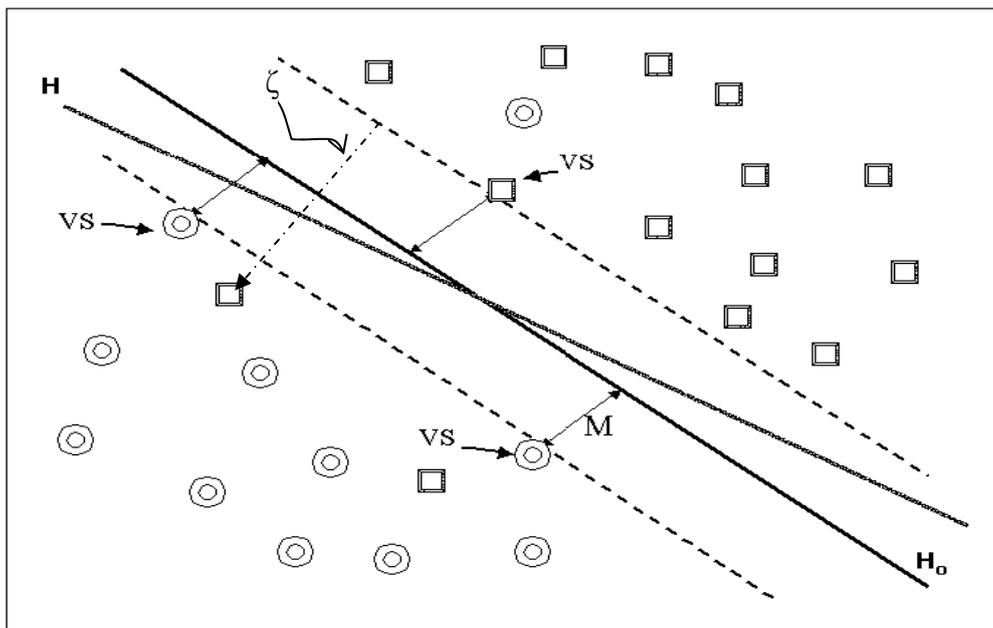
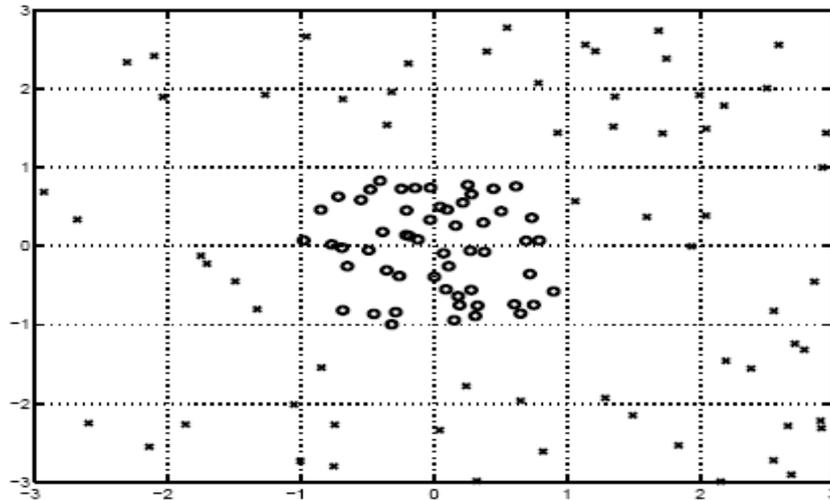


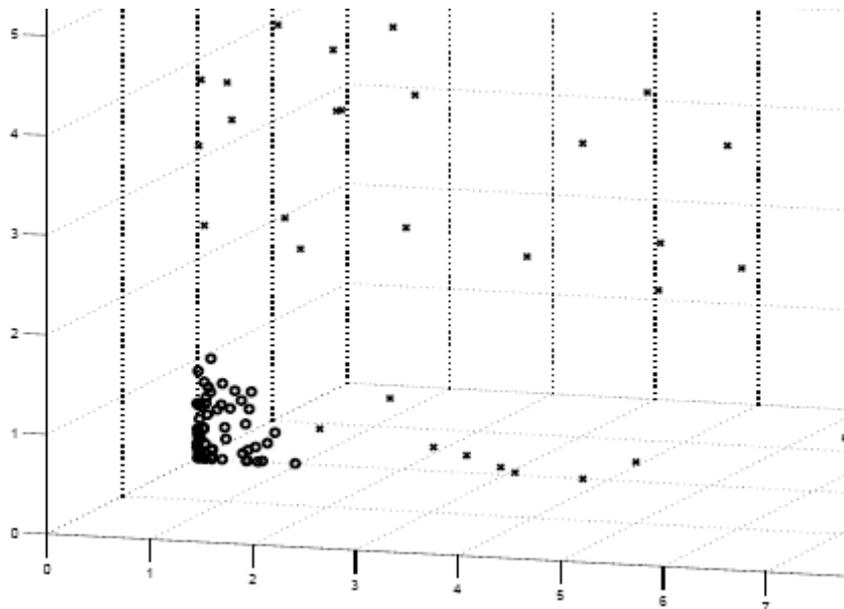
Figure 3.2 Hyperplans séparateurs dans le cas de données non-linéairement séparables (VS sont les Vecteurs de Supports).[Khar.02]

III.4.2 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. [Burg.98].



a) Espace de départ : exemples tirés aléatoirement dans \mathbb{R}^2 appartenant à deux classes non-linéairement séparables.



b : Espace augmenté : l'image des exemples de (a) dans \mathbb{R}^3 en utilisant la transformation Φ

Figure 3.3 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 [Khar.02].

Notons Φ une transformation non linéaire de l'espace d'entrée X en un espace de *redescription* $\Phi(x)$:

$$X = (x_1, \dots, x_d)^T \quad \text{à} \quad \Phi(X) = (\Phi(x_1), \dots, \Phi(x_d), \dots)^T \quad (\text{III.24})$$

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 :

$$\Phi(x_i) \cdot \Phi(x_j) = k(x_i, x_j) \tag{III.25}$$

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 [Khar02]:

▪ **Théorème de Mercer :**

Pour être sûr qu'une fonction symétrique $k(x, x')$ admet un développement de la forme suivante :

$$k(x, x') = \sum_{k=1}^{+\infty} \beta_k \Phi_k(x) \cdot \Phi_k(x') \tag{III.26}$$

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 :

$$\iint k(x, x') g(x) g(x') dx dx' \geq 0 \tag{III.27}$$

pour toute fonction $g \neq 0$ avec :

$$\int g^2(x) dx \geq 0 \tag{III.28}$$

Parmi les noyaux couramment utilisés par les chercheurs, on peut citer :

- Le noyau Polynomial : $k(x, x') = (x \cdot x' + 1)^q$ (III.29)
Où q est le degré du polynôme à déterminer par l'utilisateur.

- Le noyau RBF (*radial basis function*) : $k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$ (III.30)
Où σ est à déterminer par l'utilisateur.

- Le noyau Sigmoidal : $k(x, x') = \tanh(ax \cdot x' - b)$ (III.31)

Prenons l'exemple suivant [Ayat04]. Donc on a la transformation Φ tel que :

$$\begin{aligned} \Phi &: \mathfrak{R}^2 \rightarrow \mathfrak{R}^3 \\ X = (x_1, x_2) &\rightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2) \end{aligned}$$

$$\begin{aligned}
 k(u, v) &= \Phi(u) \cdot \Phi(v) \\
 &= (u_1^2, \sqrt{2}u_1u_2, u_2^2) \begin{pmatrix} v_1^2 \\ \sqrt{2}v_1v_2 \\ v_2^2 \end{pmatrix} \\
 &= (u_1^2v_1^2 + 2u_1v_1u_2v_2 + u_2^2v_2^2) \\
 &= (u_1v_1 + u_2v_2)^2 \\
 &= \left[(u_1, u_2) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right]^2 = (u \cdot v)^2
 \end{aligned}$$

Le Lagrangien se transcrit dans ce cas par :

$$L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{III.32}$$

l : taille de la base d'apprentissage.

Donc le problème d'optimisation (III.23) devient :

$$\begin{cases} \max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \forall i \quad 0 \leq \alpha_i \leq C \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \tag{III.33}$$

La résolution de ce problème d'optimisation conduit à la fonction de décision suivante :

$$f(x) = \text{sgn} \left(\sum_{i=1}^m \alpha_i^* y_i k(x_i, x) + b^* \right) \tag{III.34}$$

Où les coefficients α_i^* et b^* sont obtenus comme précédemment par résolution de l'équation (III.33).

III.4.2.1 Conditions de Karush-Kuhn-Tucker (KKT)

La résolution de l'optimisation quadratique de l'équation (III.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 :

$$\begin{aligned}
 \alpha_i = 0 &\Rightarrow y_i f(x_i) \geq 1 && \text{et } \zeta_i = 0 \\
 0 < \alpha_i < C &\Rightarrow y_i f(x_i) = 1 && \text{et } \zeta_i = 0 \\
 \alpha_i = C &\Rightarrow y_i f(x_i) \leq 1 && \text{et } \zeta_i \geq 0
 \end{aligned} \tag{III.35}$$

Les équations de III.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.04].

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.

III.4.2.2 Calcul du biais b

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 ξ_i correspondants sont nuls. L'égalité suivante est alors vérifiée :

:

$$y_i (b + \sum_{i=1}^l \alpha_i y_i k (x_i, x_j)) = 1 \quad (\text{III.36})$$

En considérant la moyenne calculée sur cet ensemble des vecteurs de supports, une valeur stable de b peut s'écrire :

$$b = \frac{1}{N_{sv}} (y_i - \sum_{i=1}^l \alpha_i y_i k (x_i, x_j)) \quad (\text{III.37})$$

Où N_{sv} représente le nombre de vecteurs de support considérés.

III.5 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 (III.33) dont la formulation matricielle s'écrit encore :

$$L_D = -\frac{1}{2} \alpha^T K \alpha + 1^T \alpha \quad (\text{III.38})$$

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.98] que nous utilisons pour l'entraînement du classificateur dans la partie expérimentale.

III.5.1 Méthode de chunking

La résolution de la fonction objective duale de l'équation (III.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 [Vapn.98] 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.

III.5.2 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 [Ayat04]. Osuna et al., [Osun.97] 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 [Joac.98].

III.5.3 Méthode de minimisation séquentielle : SMO

La méthode d'optimisation par minimisation séquentielle (de l'anglais «Sequential Minimal Optimization») proposée par Platt [Platt.98], peut être perçue comme le cas extrême des méthode de décomposition successive [Ayat.04]. 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 (III.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 :

$$y_1 \alpha'_1 + y_2 \alpha'_2 = y_1 \alpha_1 + y_2 \alpha_2 \quad (\text{III.39})$$

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 :

$$L_D(\alpha_1 + \alpha_2) = \alpha_1 + \alpha_2 - \frac{1}{2} G_{11} \alpha_1^2 - \frac{1}{2} G_{22} \alpha_2^2 - \frac{1}{2} G_{12} \alpha_1 \alpha_2 - \alpha_1 v_1 - \alpha_2 v_2 + c \quad (\text{III.40})$$

Où c est une constante qui ne dépend pas de α_1 et α_2 et où :

$$v_i = \sum_{j=3}^m \alpha_j G_{ij} \quad \text{pour } i = 1, 2 \quad (\text{III.41})$$

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 :

$$\begin{aligned} L_D(\alpha_2) = d - s \alpha_2 + \alpha_2 - \frac{1}{2} G_{11} (d - s \alpha_2)^2 - \frac{1}{2} G_{22} \alpha_2^2 \\ - G_{12} (d - s \alpha_2) \alpha_2 - (d - s \alpha_2) v_1 - \alpha_2 v_2 + c \end{aligned} \quad (\text{III.42})$$

Au maximum de $L_D(\alpha_2)$, nous devons satisfaire :

$$\begin{aligned} 0 = \frac{\partial L_D}{\partial \alpha_2} = 1 - s + s G_{11} (d - s \alpha_2) - G_{22} \alpha_2 \\ - d G_{12} + 2s G_{12} \alpha_2 + s v_1 - v_2 \end{aligned} \quad (\text{III.43})$$

Donc, pour se rendre au maximum de L_D , nous devons choisir pour α_2 la valeur α_2^u satisfaisant (u signifie «unconstrained») :

$$(G_{11} + G_{22} - 2s G_{12}) \alpha_2^u = d(s G_{11} - G_{12}) + s v_1 - v_2 + 1 - s \quad (\text{III.44})$$

En utilisant $d = \alpha_1 + s \alpha_2$, $s = y_1 y_2$, $G_{ij} = y_i y_j k_{ij}$ et :

$$\begin{aligned} K_i = \sum_{j=1}^m \alpha_j y_j k_{ji} \\ \eta = 2k_{12} - k_{11} - k_{22} \end{aligned} \quad (\text{III.45})$$

nous obtenons, après quelques lignes d'algèbre :

$$\alpha_2^u = \alpha_2 - y_2 \frac{(K_1 - y_1) - (K_2 - y_2)}{\eta} \quad (\text{III.46})$$

Or, lorsque nous aurons trouvé α_2' , la valeur de α_1' sera donnée par la contrainte :

$$\alpha_1' = \alpha_1 + s(\alpha_2 - \alpha_2') \quad (\text{III.47})$$

Ainsi on aura $0 \leq \alpha_1' \leq C$ SSI :

$$0 < \alpha_1 + s(\alpha_2 - \alpha_2') < C \quad (\text{III.48})$$

Cas : $S = 1$ c à d ($y_1 = y_2$) :

Dans ce cas on doit satisfaire :

$$0 < \alpha_1 + (\alpha_2 - \alpha_2') < C \quad (\text{III.49})$$

ce qui entraîne

$$(\alpha_1 + \alpha_2) - C \leq \alpha_2' \leq (\alpha_1 + \alpha_2) \quad (\text{III.50})$$

Or, puisqu'il faut également avoir $0 \leq \alpha_2' \leq C$ on a finalement que :

$$\max[0, (\alpha_1 + \alpha_2) - C] \leq \alpha_2' \leq \min[C, (\alpha_1 + \alpha_2)] \quad (\text{III.51})$$

Cas : $S = -1$ c à d ($y_1 \neq y_2$) :

de même nous obtenons :

$$\max[0, (\alpha_2 - \alpha_1)] \leq \alpha_2' \leq \min[C, C + (\alpha_2 - \alpha_1)] \quad (\text{III.52})$$

Dans tous les cas, nous devons avoir $U \leq \alpha_2' \leq V$ avec :

$$U = \begin{cases} \max[0, (\alpha_1 + \alpha_2) - C] & \text{si } y_1 = y_2 \\ \max[0, (\alpha_2 - \alpha_1)] & \text{si } y_1 \neq y_2 \end{cases} \quad (\text{III.53})$$

$$V = \begin{cases} \min[C, (\alpha_1 + \alpha_2)] & \text{si } y_1 = y_2 \\ \min[C, C + (\alpha_2 - \alpha_1)] & \text{si } y_1 \neq y_2 \end{cases} \quad (\text{III.54})$$

Considérant que α_2^u est donné par l'équation III.46, la nouvelle valeur α_2' est donnée par :

$$\alpha_2' = \begin{cases} V & \text{si } \alpha_2^u > V \\ \alpha_2^u & \text{si } U \leq \alpha_2^u \leq V \\ U & \text{si } \alpha_2^u < U \end{cases} \quad (\text{III.55})$$

Et, ayant obtenu α_2' , la nouvelle valeur α_1' est obtenu de l'équation (III.47). Typiquement l'algorithme débute avec $\alpha = 0$. Une paire de multiplicateurs est choisie et les nouvelles

valeurs sont obtenues selon les équations (III.55) et (III.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 (III.45) nous donne pour K_i :

$$K_i = \alpha_1 y_1 k_{1i} + \alpha_2 y_2 k_{2i} + \sum_{j=3}^m \alpha_j y_j k_{ji} \quad (\text{III.56})$$

et après la mise à jour de (α_1, α_2) , nous avons :

$$K'_i = \alpha'_1 y_1 k_{1i} + \alpha'_2 y_2 k_{2i} + \sum_{j=3}^m \alpha_j y_j k_{ji} \quad (\text{III.57})$$

Le nouveau K'_i est donc simplement reliée à l'ancien K_i par :

$$K'_i = K_i + (\alpha'_1 - \alpha_1) y_1 k_{1i} + (\alpha'_2 - \alpha_2) y_2 k_{2i} \quad (\text{III.58})$$

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 (III.37) on peut établir la fonction de décision III.34

III.6 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.

III.6.1 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 binaires, 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.06] :

$$f(x) = \arg \max \sum_{i \in \mathcal{I}^S} y_i \alpha_i^n k(x_i, x) + b^n \quad (\text{III.59})$$

Avec : $\sum_{i \in \mathcal{I}^S} y_i \alpha_i^n k(x_i, x) + b^n$ est la fonction de décision de l' $n^{\text{ième}}$ classificateur binaire.

La figure 3.4 illustre un exemple de classification de trois classes par l'approche un contre tous

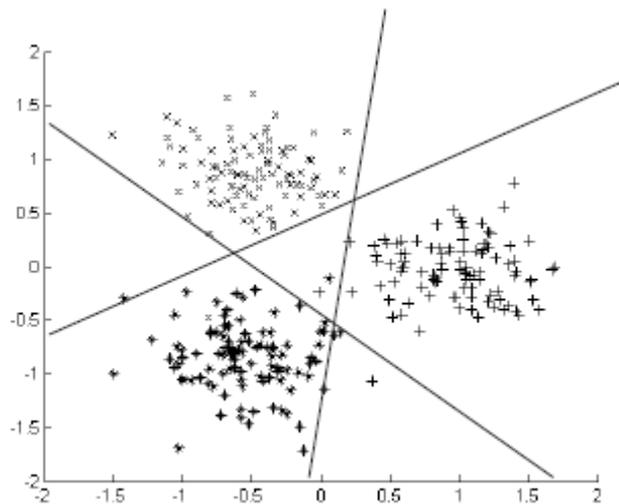


Figure 3.4 Problème à trois classes : frontières de décision linéaires dans la stratégie Un-contre-Tous

III.6.2 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.06].

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 conversée 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 :

$$u_{i,j} = F(f_{i,j}(x)) \quad (\text{III.60})$$

Où $f_{i,j}(x)$ est la fonction de décision du classificateur binaire impliquant la $i^{\text{ème}}$ classe et la $j^{\text{ème}}$ classe. L'estimation directe de la fonction F dans (III.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 (III.60) peut être écrite comme suit :

$$u_{i,j}(x) = \frac{1}{1 + \exp(Af_{i,j}(x) + B)} \quad (\text{III.61})$$

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 :

$$f_{\max}(x) = \arg \max_i \sum_{j:j \neq i} I(r_{i,j} > r_{j,i}) \quad (\text{III.62})$$

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 (III.63) :

$$P(y=1|x) = \frac{2 \arg \max_i \sum_{j:j \neq i} I(r_{i,j} > r_{j,i})}{K(K-1)} \quad (\text{III.63})$$

III.7 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 γ 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 multiclassées. 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.

CHAPITRE IV : SELECTION DES PARAMETRES DU VECTEUR FORME

IV.1 Introduction

Dans le premier chapitre nous avons abordé plusieurs techniques d'extraction de paramètres (indicateurs) des signaux vibratoires. Un ensemble donné de paramètres peut contenir des informations redondantes ou non pertinentes qu'il est intéressant de supprimer avant une éventuelle procédure de classification. Il est en effet important que, dans l'espace de représentation où s'effectuera la classification, les défauts appartenant à la même classe soient le plus groupés possible et qu'à l'inverse, les défauts issus de classes différentes soient situés dans des régions distinctes de l'espace.

La sélection d'un nombre réduit de paramètres du vecteur forme, pour représenter les signatures des défauts présente plusieurs intérêts. Elle permet tout d'abord de réduire le temps de calcul et la complexité des algorithmes de classification mis en jeu. Elle permet aussi de passer outre le phénomène de "malédiction de la dimension" qui se manifeste lorsqu'on trace les performances du classificateur en fonction de la dimension de l'espace d'entrée et du nombre d'observations de la base d'apprentissage. Oukhelou.L dans [Oukh.97], souligne que 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, faute de quoi les performances du classificateur sur des données non apprises se détériorent. Ce problème s'accroît lorsque la base de données est de taille réduite.

Dans ce chapitre nous allons présenter brièvement deux méthodes utilisées pour la sélection des paramètres, en détaillant beaucoup plus les algorithmes génétiques qui s'avèrent une méthode prometteuse d'optimisation à cause de sa large utilisation dans le domaine de la RDF, et dans quelques travaux de sélection des paramètres. Dans la partie expérimentale de ce mémoire, cette méthode sera utilisée pour la sélection du vecteur forme optimal.

IV.2 Sélection des paramètres par des méthodes séquentielles "constructive" et "destructive" (SFS & SBS)

Les méthodes de classement, que nous proposons de développer dans ce paragraphe permettent de remédier une sélection des paramètres les plus pertinents, en tenant compte du caractère multidimensionnel des paramètres qui sont ajoutés ou retirés successivement en se basant sur des critères de séparabilité entre classes. Une fois les paramètres ordonnés, il conviendra de n'en sélectionner que les d_r premiers. Dans la littérature, ces méthodes sont appelées *méthodes d'ordonnement*.

IV.2.1 Principe des algorithmes d'ordonnement

Une recherche exhaustive pour la sélection de d_r paramètres pertinents parmi les d originaux au sens d'un critère donné, impliquerait le test de toutes les combinaisons possibles soit :

$$\frac{d!}{(d-d_r)!d_r!} \text{ combinaisons.} \quad (\text{IV.1})$$

Calculer le critère pour toutes ces combinaisons devient vite prohibitif en temps de calcul et n'est possible que pour de très faibles valeurs de d . Les deux algorithmes d'ordonnement

conduisent à des solutions sous-optimales avec un coût en calcul considérablement réduit, sont : Procédure Séquentielle Constructive (SFS), et Procédure Séquentielle destructive (SBS)

IV.2.1.1 Procédure séquentielle constructive (SFS)

L'algorithme de recherche de type "Sequential Forward Selection" (SFS) fonctionne par ajouts successifs de paramètres. Partant d'un ensemble de d paramètres initiaux, on commence par sélectionner celui qui optimise le critère J de séparabilité entre classes. Ensuite, à chaque itération, on choisit parmi les caractéristiques restantes, celle qui en combinaison avec le sous ensemble sélectionné réalise une meilleure optimisation de J .

Soit $E = \{X_i | i = 1, \dots, d\}$ l'ensemble initial de paramètres. A la $k^{\text{ième}}$ étape, le sous ensemble E_k est tel que :

$$J(E_K) = \max \{J(E_{K-1} \cup X_i)\} \tag{IV.2}$$

$$X_i \in E_K - E_{K-1} \quad \text{Initialisation } E_0 = \phi \tag{IV.3}$$

On conçoit que les "dr" premiers paramètres choisis comme les plus pertinents par une telle procédure sont probablement différents de la solution optimale correspondant à une recherche exhaustive de dr paramètres parmi d .

IV.2.1.2 Procédure séquentielle destructive (SBS)

Cette technique dite "Sequential Backward Selection" (SBS) est duale de la précédente. Elle consiste à éliminer à chaque itération la caractéristique qui provoque la diminution la plus faible du critère de séparabilité entre classes et ce jusqu'à ce que $d-1$ paramètres soient éliminés. Le schéma de principe de cette méthode est représenté par la figure 4.1.

A partir de l'ensemble initial de paramètres $E = \{X_i | i = 1, \dots, d\}$, le sous-ensemble E_k obtenu à l'itération k est tel que :

$$J(E_K) = \max \{J(E_K - E_{K-1}) - X_i\} \tag{IV.3}$$

$$X_i \in E_K - E_{K-1} \quad \text{Initialisation : } E_0 = E \tag{IV.4}$$

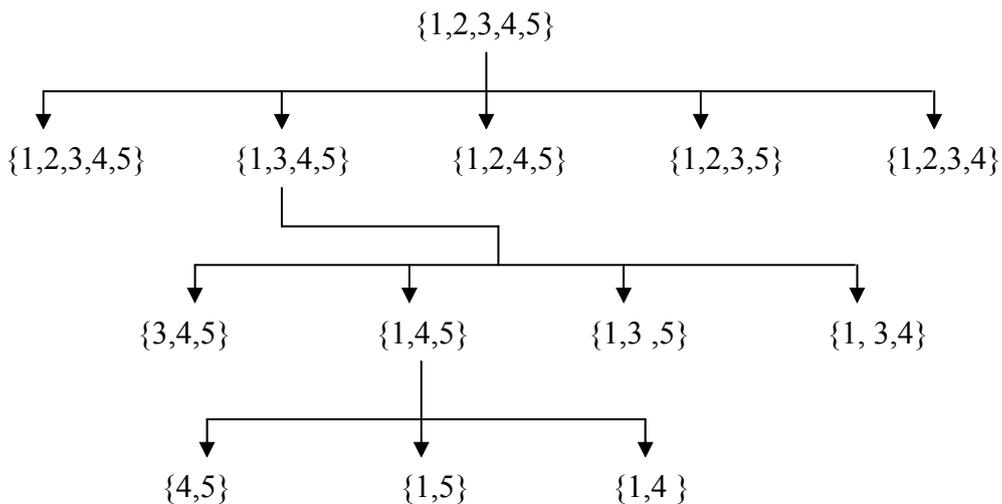


Figure 4.1 Schéma de principe de la sélection de paramètres par la méthode SBS, avec $d=5$ et $d'=2$.

On peut noter que contrairement à la procédure SFS, la procédure SBS doit être menée à son terme pour que l'ordonnancement soit exploitable. De plus, le nombre de calcul effectué sur des sous-espaces de grande taille est plus important. En conséquence, cette procédure possède un coût en calcul plus important que la procédure SFS.

Il existe des extensions de ces algorithmes qui consistent à ôter ou à rajouter non pas un paramètre à chaque itération mais q paramètres à la fois ce qui permet de diminuer les inconvénients liés à l'aspect séquentiel de ces procédures [Oukh.97].

IV.3 Les algorithmes génétiques

Les algorithmes génétiques (AGs) sont des algorithmes d'optimisation stochastique fondés sur les mécanismes de la sélection naturelle et de la génétique. Leur fonctionnement est extrêmement simple. On part avec une population de solutions potentielles (*chromosomes*) initiales arbitrairement choisies. On évalue leur performance (*fitness*) relative. Sur la base de ces performances on crée une nouvelle population de solutions potentielles en utilisant des opérateurs évolutionnaires simples : la sélection, le croisement et la mutation. On recommence ce cycle jusqu'à ce que l'on trouve une solution satisfaisante. Les algorithmes génétiques ont été discutés dans plusieurs ouvrages [Mich.99] [Davi.91].

IV.3.1 Principe

IV.3.1.1 Codification des données

Historiquement, le codage utilisé par les algorithmes génétiques était représenté sous forme de chaînes de bits contenant toute l'information nécessaire à la description d'un point dans l'espace d'état. Ce type de codage a pour intérêt de permettre la création des opérateurs de croisement et de mutation simples [Dura.04].

IV.3.1.1.1 Types de codage

Il existe trois principaux types de codage utilisables, et on peut passer de l'un à l'autre relativement facilement [Davi.91] :

- **Le codage binaire :**

C'est le plus utilisé. Chaque gène dispose du même alphabet binaire $\{0, 1\}$, Un gène est alors représenté par un entier long (32 bits), les chromosomes qui sont des suites de gènes sont représentés par des tableaux de gènes et les individus de notre espace de recherche sont représentés par des tableaux de chromosomes.

- **Le codage réel :**

Cela peut-être utile notamment dans le cas où l'on recherche le maximum d'une fonction réelle.

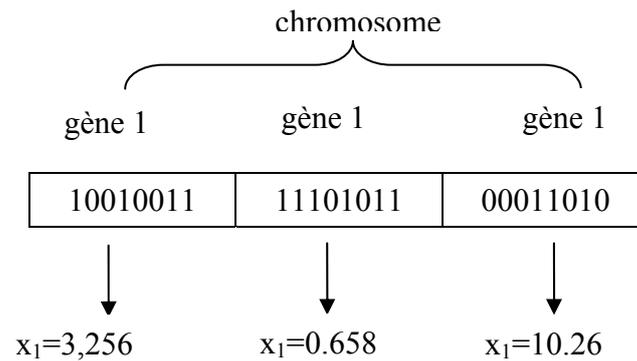


Figure 4.2 Illustration schématique du codage des variables réelles

- **Le codage de Gray :**

Dans le cas d'un codage binaire on utilise souvent la "distance de Hamming" comme mesure de la dissimilarité entre deux éléments de population, cette mesure compte les différences de bits de même rang de ces deux séquences. Et c'est là que le codage binaire commence à montrer ses limites. En effet, deux éléments voisins en terme de distance de Hamming ne codent pas nécessairement deux éléments proches dans l'espace de recherche. Cet inconvénient peut être évité en utilisant un "codage de Gray" : le codage de Gray est un codage qui a comme propriété qu'entre un élément n et un élément $n + 1$, donc voisin dans l'espace de recherche, un seul bit diffère.

IV.3.1.2 Génération aléatoire de la population initiale

Le choix de la population initiale d'individus conditionne fortement la rapidité de l'algorithme. Si la position de l'optimum dans l'espace d'état est totalement inconnue, il est naturel d'engendrer aléatoirement des individus en faisant des tirages uniformes dans chacun des domaines associés aux composantes de l'espace d'état, en veillant à ce que les individus produits respectent les contraintes. Si par contre, des informations a priori sur le problème soient disponibles, il paraît bien évidemment naturel d'engendrer les individus dans un sous domaine particulier afin d'accélérer la convergence. Dans l'hypothèse où la gestion des contraintes ne peut se faire directement, les contraintes sont généralement incluses dans le critère à optimiser sous forme de pénalités. [Dura.04]

Le problème major qui se pose est comment fixer la taille de cette population. Une population trop petite évolue probablement vers un optimum local peu intéressant. Une population trop grande met plus de temps pour converger vers des solutions envisageables. 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. Felkaoui.et al., [Felk.07] mentionnent que plusieurs chercheurs se sont penchés sur ce problème qui consiste à déterminer la taille optimale de la population pour atteindre la meilleure solution. Certains d'entre eux l'ont fixé entre 20 et 100 de manière empirique. D'autres essayent de l'ajuster en respectant le taux d'erreur de sélection ou de la faire varier d'une population à une autre. Par contre, d'autres chercheurs ont fixé le nombre d'individus à 50 et ont mentionné qu'il peut exister une relation entre la taille de la population et le nombre de gènes (bits). D'autres recommandent que la taille d'une population moyenne soit comprise entre 20 et 30.

IV.3.1.3 Fonction d'évaluation (fonction *fitness*)

Pour calculer le coût d'un point de l'espace de recherche, on utilise une *fonction d'évaluation*. L'évaluation d'un individu ne dépendant 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 : c'est le rôle de la fonction *fitness*. Cette méthode permet de s'assurer que les individus performants seront conservés, alors que les individus peu adaptés seront progressivement éliminés de la population. [Bena.02]. La fonction *fitness* que nous avons choisi d'utiliser dans la partie expérimentale est fondé sur la séparabilité entre classes et plus particulièrement sur les matrices de variance-covariance calculées directement sur l'ensemble des échantillons. Oukhelou.L dans [Oukh.97] souligne que d'autres critères basés sur les distances entre classes peuvent aussi être employés. Ils présentent toutefois l'inconvénient de ne pas refléter le recouvrement des classes et de ce fait, ils sont rarement utilisés en pratique. Par ailleurs, tous les critères qui s'intéressent à l'estimation de la probabilité d'erreur du système de classification ou à des mesures de distances entre lois de probabilité de différentes classes, supposent la connaissance -ou à défaut l'estimation- des densités de probabilité de chaque classe et restent plus difficiles à mettre en œuvre. Il convient de revenir aux définitions données dans le paragraphe (II.3.2.1.3.1) du deuxième chapitre concernant les matrices de variance-covariance interclasse S_B et intra classe S_W . Une solution importante peut être établie :

$$S = S_W + S_B \quad (\text{IV.5})$$

Où :

S : matrice de variance-covariance associée à la sous-matrice d'observation

S_W : matrice de variance-covariance intra classe

S_B : matrice de variance-covariance interclasse

Cette relation exprime le fait que pour un nuage de points donné (la matrice S étant par conséquent fixée), si la dispersion à l'intérieur des classes est faible, alors l'éloignement des classes est grand. L'obtention du sous-ensemble discriminant consiste alors à trouver les M' paramètres (combien et lesquels) qui permet de maximiser la mesure de dispersion interclasses 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 [Cazi.03].

Pour rendre compte de la répartition des classes dans l'espace sélectionné, les critères de sélection utilisent donc ces matrices S_W et S_B . Pour obtenir des classes bien compactes et donc séparées entre elles, on doit minimiser les variances à l'intérieur de chaque classe. Le premier critère à minimiser est défini comme suit :

$$J_1 = \text{tr} (S_W) \quad (\text{IV.6})$$

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 est alors défini, équivalent de J_1 mais après normalisation de la matrice intra classe :

$$J_2 = \text{tr} (S^{-1} S_W) \quad (\text{IV.7})$$

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 :

$$J_3 = \text{tr} (S_W^{-1} S_B) \quad (\text{IV.8})$$

Un autre critère est présenté dans [Casi.03], il consiste à maximiser :

$$J_4 = \text{tr} (S^{-1} S_B) \quad (\text{IV.9})$$

Il existe enfin un dernier critère qui permet d'examiner la dispersion entre les classes, en utilisant le déterminant de S_W et non sa trace :

$$J_5 = \frac{\det(S_W)}{\det(S)} \quad (\text{IV.10})$$

Les quatre premiers critères possèdent une propriété commune : ils varient tous de façon monotone avec la dimension du sous-espace d'observation, Casimir.R, dans [Casi.03] souligne qu'il n'existe pas d'indicateur permettant d'orienter le choix vers un critère plutôt qu'un autre. Si les solutions obtenues par les trois premiers critères sont différentes, elles devraient toutes être réexaminées a posteriori au vue des résultats de classification. On notera au passage que les deux derniers critères sont également invariants par transformation linéaire non singulière

IV.3.1.4 La sélection

La sélection est un opérateur clé sur lequel repose en partie la qualité d'un algorithme évolutionnaire. L'une des principales caractéristiques de tout opérateur de sélection est la pression sélective exercée sur la population. C'est un paramètre d'un maniement délicat, car opter pour une pression sélective trop forte revient à écraser dans l'œuf toute forme de diversité. Mais si l'on verse dans l'autre extrême, à savoir une pression sélective trop faible, l'algorithme tend à se rapprocher d'une simple recherche aléatoire. Il existe plusieurs méthodes de sélections parmi lesquelles [Bena.02] :

a. Sélection par roulette

C'est l'ancêtre de toutes les méthodes de sélection, et elle tire son nom de sa ressemblance avec la roulette ou la loterie. Elle consiste à affecter à chaque individu une probabilité d'être sélectionné, proportionnelle à sa fonction de mérite. Pour un problème de maximisation, on considère une population de taille N_p .

Soit : pour $i = 1, 2, 3, \dots, N_p$, f_i : la fonction d'adaptation de l'individu i .

$$F = \sum_{i=1}^N f_i \quad (\text{IV.11})$$

On pose F : la somme des valeurs des fonctions d'adaptation de l'ensemble des individus de la population. La probabilité P_i qu'a alors l'individu i d'être sélectionné est :

$$P_i = f_i/F \quad (\text{IV-12})$$

Il est bien entendu que le modèle le plus simple se fait avec remise, et qu'un individu peut être sélectionné un nombre de fois élevé. De cette manière, les meilleurs individus sont susceptibles de jouer un rôle déterminant dans les générations suivantes. Une certaine diversité est cependant maintenue, car même les individus les moins performants conservent une chance d'être choisis. En effet, $\forall i \text{ tq } 1 \leq i \leq N_p$, on a $P_i > 0$, ce qui revient à dire que même l'individu le moins adapté au problème considéré (ou en d'autres termes à l'environnement) a une probabilité non nulle d'être choisi. On dira de cette méthode qu'elle est *préservative*. Cela peut apparaître comme une propriété annexe, mais dès lors qu'on considère des environnements qui évoluent avec le temps, il est crucial de garder en réserve des individus apparemment "inadaptés" car leurs *atypismes* en fera de précieuses sources matérielles génétiques en cas de modification brusque du problème. C'est la une des forces des algorithmes évolutionnaires.

La biodiversité, ou du moins son analogue digital. Et même si les conditions du problème ne changent pas, une population uniformément homogène concentrée au niveau d'un optimum local ne peut en sortir que grâce à ces individus atypiques [Felk.07].

b. Sélection par Reste stochastique

La sélection stochastique avec reste s'inspire très fortement de la sélection par roulette, sauf qu'elle y ajoute un léger aspect déterministe. On commence par déterminer la probabilité $P_i = f_i/F$ qu'a chaque individu d'être sélectionné. Soit N_p la taille de la population. On pose alors $M_i = N \cdot P_i$, où M_i correspond au nombre moyen de copies qu'un individu peut espérer recevoir par une sélection de type roulette. On définit ensuite :

$n_i = E(M_i)$ où E désigne la partie entière ;

$r_i = M_i - E(M_i)$ les r_i sont les restes.

Le caractère déterministe de la sélection stochastique avec reste intervient alors, car on affecte automatiquement à chaque individu n_i copies dans la génération suivante comme [Mich.99] :

$$\sum_{i=1}^N n_i \leq N \quad (\text{IV-13})$$

$$\delta = N - \sum_{i=1}^N n_i \quad (\text{IV-14})$$

Il reste en générale un nombre de copies à affecter pour compléter la population de la génération suivante. Pour attribuer ces copies restantes, le processus utilisé est une roulette classique, qui s'appuie non pas sur les valeurs P_i mais sur les restes r_i .

c. Sélection par rang

Toute une famille de méthodes utilise une sélection qui n'est pas directement proportionnelle à la valeur de la fonction de mérite, mais dépend plutôt du rang de l'individu dans la population. Le classement par rang, ou *ranking*, a principalement pour but d'éviter que les individus avec des solutions de mérite très élevées ne perturbent le processus stochastique en obtenant un nombre de copies trop important.

d. Sélection par tournoi

Une sélection par tournoi consiste, chaque fois que l'on veut choisir un individu, à sélectionner un sous-ensemble de la population et à ne conserver que le meilleur élément du sous-ensemble. Si T est la taille du tournoi, on sélectionne aléatoirement T individus dans la

population et on les fait entrer en compétition : seul le vainqueur (donc celui qui a la meilleure fonction de mérite) est retenu. Le processus est réitéré N fois, où N est la taille de la population. Cette méthode est caractérisée par une pression sélective en général plus forte que les méthodes proportionnelles (pour qu'un individu peu performant puisse être sélectionné, il faut que ses adversaires soient encore moins bons que lui). Elle donne des résultats particulièrement bons dans les cas où le caractère un peu trop aléatoire de la roulette introduit un biais.

e. Modèle élitiste

Le modèle élitiste consiste à recopier automatiquement le meilleur individu de chaque génération dans la suivante. En effet, le caractère aléatoire de la sélection fait que rien ne garantit que la meilleure solution soit conservée. L'élitisme permet donc de s'assurer que la meilleure solution de chaque génération ne peut que s'améliorer avec le temps. Le modèle élitiste n'est pas une méthode de sélection en soi, mais plutôt une variante qu'on ajoute aux méthodes déjà citées. [Mich.99]

f. Sélection par troncature

Il existe également des algorithmes évolutionnaires qui considèrent que la sélection ne doit pas être laissée au hasard. L'une des plus classique étant la sélection par troncature (*truncation selection*) où on choisit de manière déterministe les $T\%$ meilleurs individus d'une génération pour générer la suivante. Cela revient en quelque sorte à généraliser le modèle élitiste à tout le processus de sélection [Mich.99].

IV.3.1.5 Le croisement

Le croisement est le premier opérateur qui intervient au niveau du processus de reproduction. Il est vital, en ce sens, que c'est grâce à lui qu'effectuent les échanges d'information au sein de la population. Chaque fois qu'un individu est retenu par le processus de sélection, on détermine aléatoirement s'il doit participer à un croisement (avec une probabilité P_c) ou être transmis directement à l'opérateur de mutation...

Soit V un individu choisi par l'opérateur de sélection. Soit r un nombre aléatoire, $0 \leq r \leq 1$. On a: Si $r \leq P_c$ alors le chromosome est sélectionné pour le croisement. Le croisement est la manière la plus répandue de créer de nouveaux individus : on choisit deux individus (parents) et on les combine pour en générer deux autres (enfants). Cette combinaison utilise la notion de points de coupures qui correspond aux locus au niveau desquels s'effectuera l'échange de matériel génétique [Davi.91].

a. Croisement simple à un point

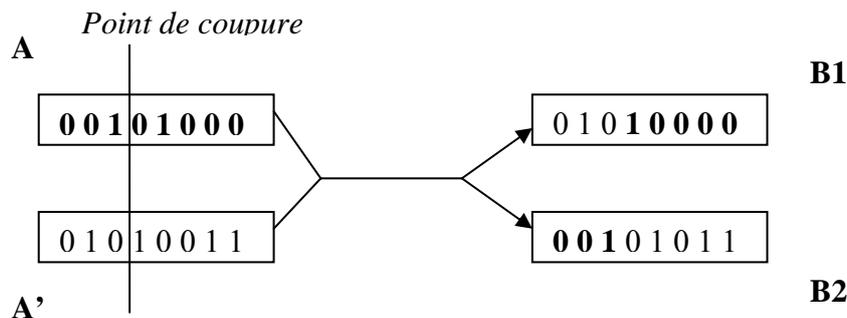


Figure 4.3 Opérateur de croisement, un seul point de coupure

Le croisement à un point est l'opérateur de croisement le plus simple et le plus utilisé. Il consiste à sélectionner un point de coupure, puis à subdiviser le génotype de chacun des parents en deux parties de part et d'autre de ce point. Les fragments obtenus sont alors échangés pour créer les génotypes des enfants. Ainsi, la figure 4.3 présente le cas où le point de coupure tombe après le troisième locus ; les parents A et A' produisent alors les enfants B1 et B2. [Davi.91].

b. Croisement multipoints

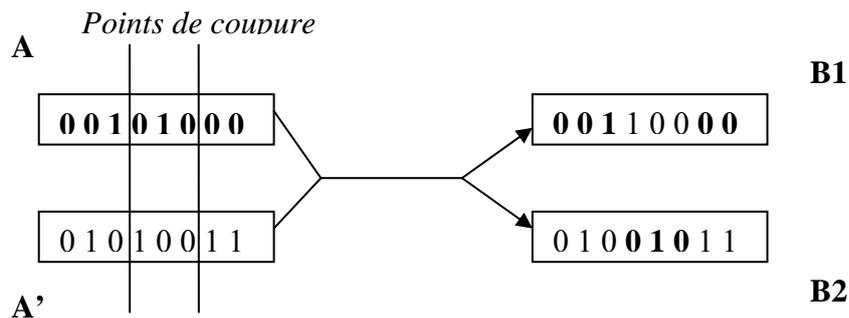


Figure 4.4 Opérateur de croisement à deux points de coupures

Le croisement multipoints est une généralisation du croisement à un point en ce sens qu'au lieu de choisir un seul point de coupure, on en sélectionne k . l'intérêt d'une telle approche apparaît rapidement, car il devient possible en un seul croisement d'échanger des *building blocks* qui sont situés au milieu du chromosome sans avoir à procéder par étapes. Bien sur la contrepartie est que c'est une technique qui est plus destructrice par rapport à ces mêmes *buildings blocks*, puisqu'ils peuvent se retrouver scindés en plusieurs fragments sans intérêt. La figure 4.4 représente un croisement à deux points, dans le cas où les points de coupure tomberaient après les troisièmes et sixième locus : les parents A et A' produisent alors les enfants B1 et B2.

C. Croisement uniforme

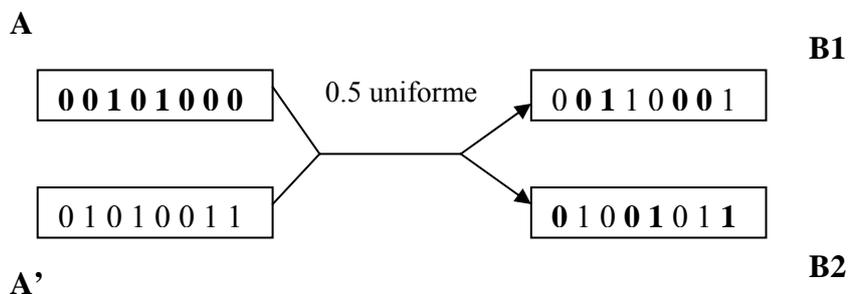


Figure 4.5 Exemple de croisement 0.5-uniforme

Soient A et A' les parents, et $B1$ et $B2$ les enfants. Pour chaque locus de $B1$, on détermine aléatoirement (avec une probabilité P_c) si l'allèle est fourni par A ou par A' . Le second enfant $B2$ se voit affecter l'allèle du parent qui n'a pas été retenu. Si $B(k)$ dénote la valeur de l'allèle au niveau du locus k pour le chromosome B , n la longueur du chromosome et r un nombre aléatoire, avec $0 \leq r \leq 1$, on a :

$$\forall K, 1 \leq K \leq n :$$

$$\text{Si } r \leq P_c \text{ Alors } B_1(K) = A(K) \text{ et } B_2(K) = A'(K)$$

$$\text{Si } r > P_c \text{ Alors } B_1(K) = A'(K) \text{ et } B_2(K) = A(K)$$

La figure 4.5 montre un exemple de croisement 0.5-uniforme (chaque parent à une chance égale de transmettre son allèle à chacun des enfants pour tout les locus). Il est généralement admis que les valeurs de P_c comprises entre 0.8 et 0.95 donnent les résultats les plus satisfaisants dans la plus plupart des cas. Certains problèmes nécessitent cependant la mise en oeuvre de représentations particulières plus complexes, de type matriciel par exemple. Dans ce cas, les opérateurs de croisement sont également plus complexes mais obéissent néanmoins à un principe commun, l'échange d'information entre individus. Il est à noter que si les croisements présentés ici mettent tous en jeux deux parents, ce n'est pas toujours le cas. Il existe en effet des opérateurs de croisement qui font intervenir trois parents.

IV.3.1.6 la mutation

La mutation est une modification aléatoire de la valeur d'un gène qui se produit avec une probabilité fixée. Elle est primordiale puisqu'elle donne la garantie que tout point de l'espace de recherche peut être atteint lors du processus d'exploitation. Si P_m est la probabilité de mutation,

$0 \leq r \leq 1$ un nombre tiré au hasard et que l'on ait $r < P_m$ pour le quatrième locus par exemple, alors [Davi.91]:



Figure 4.6 Opérateur de mutation

Comme pour le taux de croisement, la recherche d'un taux de mutation optimal a suscité de nombreux travaux. On retiendra juste que les valeurs classiques de P_m oscillent entre 0.001 et 0.01.

IV.3.1.7 Critère d'arrêt

Le critère de convergence est la condition qui fait arrêter le déroulement des générations d'un AG. On arrête la recherche lorsqu'on arrive à la solution optimale de la fonction d'évaluation ou une valeur plus proche de cette dernière. Pour effectuer ces conditions, on a plusieurs méthodes. On cite les deux critères les plus utilisés [Felk.07] :

- la différence entre les valeurs maximums de deux générations successives. Si la valeur absolue de cette différence est inférieure ou égale à ε (ε est une valeur faible dépendante de la nature du problème) le programme s'arrête.
- le programme s'arrête lorsque la différence entre le maximum et la moyenne du fitness est inférieure ou égale à une valeur défini par :

$$\max(f_{eval}) - V_{moy}(f_{eval}) \leq \mu * \max(f_{eval}) \quad (IV.15)$$

IV.4 Conclusion

Ce chapitre a été organisé selon trois sections principales ; la première décrit la méthode séquentielle "constructive" (SFS), la deuxième présente la méthode "destructive" (SBS), et dans la dernière, nous avons discuté, avec détails, le principe des algorithmes génétiques. Ces derniers seront utilisés dans la partie expérimentale, pour la sélection du vecteur forme optimal.

La réduction de l'espace de représentation permettra de diminuer le temps de classification pour toute nouvelle observation et d'éviter la dégradation des performances de la règle de décision. Les paramètres sélectionnés vont jouer un rôle important dans la phase d'apprentissage. En effet, ceux-ci devront caractériser au mieux les différents modes de fonctionnement. En d'autres termes, les paramètres jugés comme étant les plus pertinents, ou les plus informatifs, seront ceux qui permettront de distinguer les classes dans l'espace de représentation. L'optimisation du système de reconnaissance de formes consiste à maximiser la performance (meilleur taux de reconnaissance calculé) et à minimiser la complexité du classificateur (diminuer le nombre des entrées).

CHAPITRE V : PERFORMANCES DES SVM

V.1 Introduction

Dans le présent chapitre, nous donnons une description des systèmes mécaniques étudiés, ainsi que leurs modes de fonctionnement, nous présentons également les signaux vibratoires sur les quels nous allons travailler. Cette étape sera suivie par la préparation des observations qui seront, par la suite, utilisées comme entrée aux SVM, afin d'examiner les performances de cette méthode de classification.

Avant d'étudier les performances des SVM sur les signaux réels, nous précétons par des applications sur des signaux aléatoires, et des signaux d'engrenage simulés.

Les premiers signaux réels que nous proposons d'étudier proviennent d'un réducteur de CETIM « *Centre d'Etudes Techniques des Industries Mécaniques* », ces signaux ont été fournis à plusieurs laboratoires de traitement du signal et d'images à travers le monde.

Les deuxièmes signaux ont été fournis par la NAVY américaine, ces signaux sont acquis d'une instrumentation de boîte de vitesse d'un hélicoptère.

Durant cette partie expérimentale, nous avons adapté à notre problème une boîte à outil MATLAB appelée : *Statistical Pattern Recognition Toolbox for Matlab*¹ [Vojt.04]. L'annexe 1 montre quelques exemples de programmes de classification. L'algorithme principal utilisé pour l'entraînement des SVM est celui de Platt : *Sequential Minimal Optimisation* (SMO), pour plus de détail voir [Platt.98].

V.2 Applications sur des signaux aléatoires

Avant d'étudier les performances des SVM, nous précétons par une étude préliminaire sur des signaux simulés, le but de cette étude est de tester la validité de la méthode d'estimation des paramètres du modèle AR.

Les premiers signaux utilisés pour former les fichiers de données sont les signaux donnés par les équations suivantes :

$$y_1 = 2.76.7y(i-1) - 3.8106y(i-2) + 2.6535y(i-3) - 0.9238.y(i-4) \quad (V.1)$$

$$y_2 = 1.8y(i-1) - 0.9y(i-2) \quad (V.2)$$

Afin d'avoir un nombre suffisant pour l'apprentissage des SVM pendant la tâche de classification, plusieurs individus (signaux) appartenants à deux classes différentes, ont été créés en ajoutant, d'une manière aléatoire, un bruit gaussien au signaux $y_1(i)$ et $y_2(i)$. La base de données obtenue est composée de 100 individus dans chaque classe. Le niveau de bruit ajouté aux signaux est déterminé par un rapport signal sur bruit (SNR), qui est défini par [Felk94] :

$$SNR = 10 \log \left(\frac{\sum_{i=1}^n y^2(i)}{\sum_{i=1}^n e^2(i)} \right) \quad (V.3)$$

<http://cmp.felk.cvut.cz/cmp/software/stprtool/index.html>

Avant de tester les performances de notre méthode de classification (SVM), nous précédon's d'abord par une étude pratique sur l'influence du niveau de bruit présent dans le signal, puis le choix de l'ordre du modèle AR, sur l'estimation des paramètres de ce modèle. La figure 5.1 représente le signaux temporels $y_1(i)$ et $y_2(i)$ pour différents niveaux de bruit

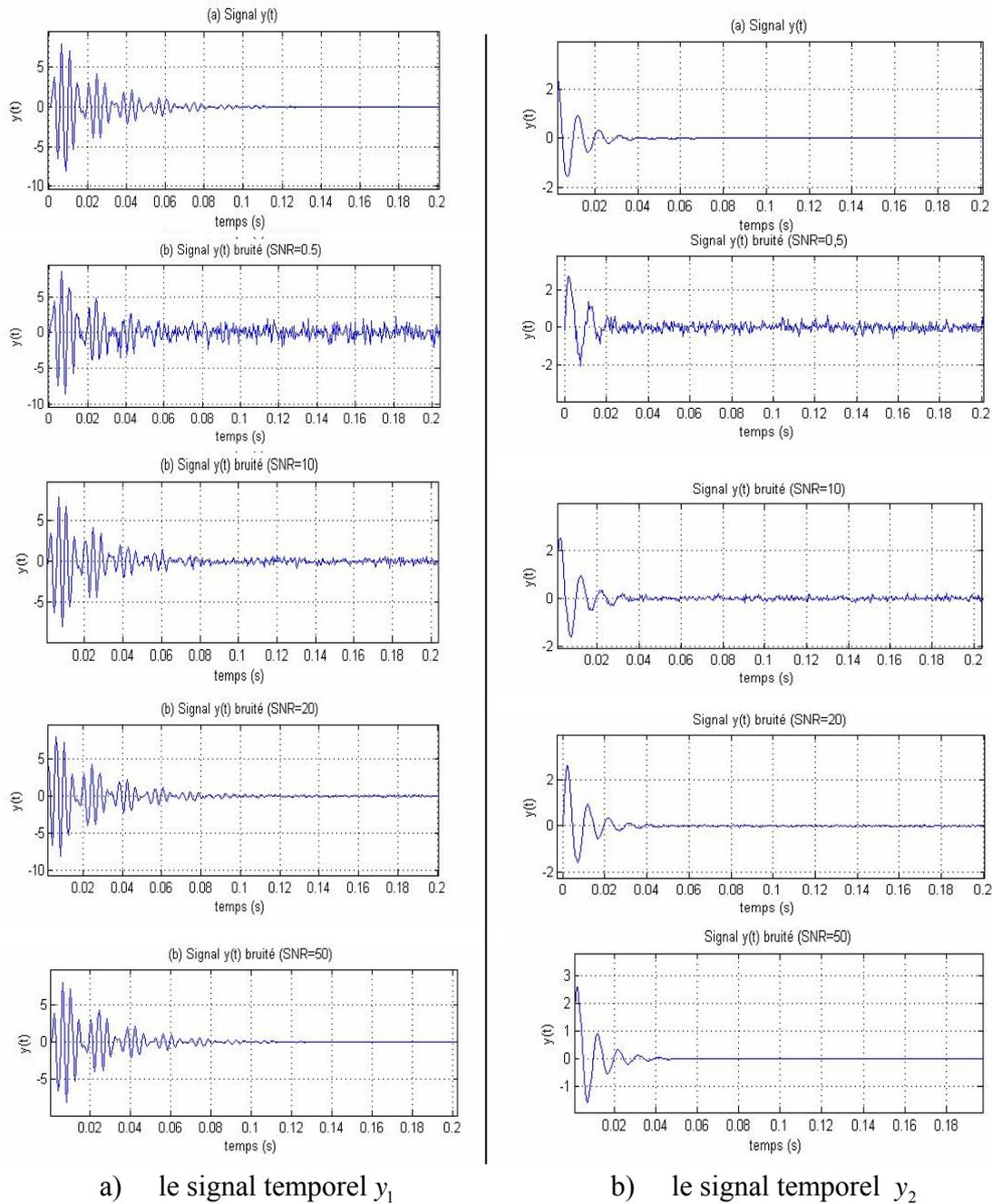


Figure 5.1 les signaux temporels y_1 et y_2 pour différents niveaux de bruit

V.2.1 L'erreur d'estimation des paramètres du modèle en fonction du rapport signal sur bruit (SNR)

Les coefficients a_i sont calculé à l'aide de la fonction « IVAR » de l'environnement MATLAB.

La figure 5.2 représente l'erreur d'estimation des coefficients a_i du signal y_1 (équation V.1) en fonction de SNR. E_1, E_2, E_3, E_4 sont la différence entre les valeurs réels et estimés des coefficients a_1, a_2, a_3, a_4 respectivement. D'après cette figure l'erreur d'estimation des ces coefficients par la fonction « IVAR » varie en fonction de SNR pour les quatre coefficients. On peut constater que cette erreur est nulle pour un rapport SNR > 25db pour les quatre coefficients. Dans cet intervalle, la fonction « IVAR » peut estimer les coefficients réels du model autorégressif.

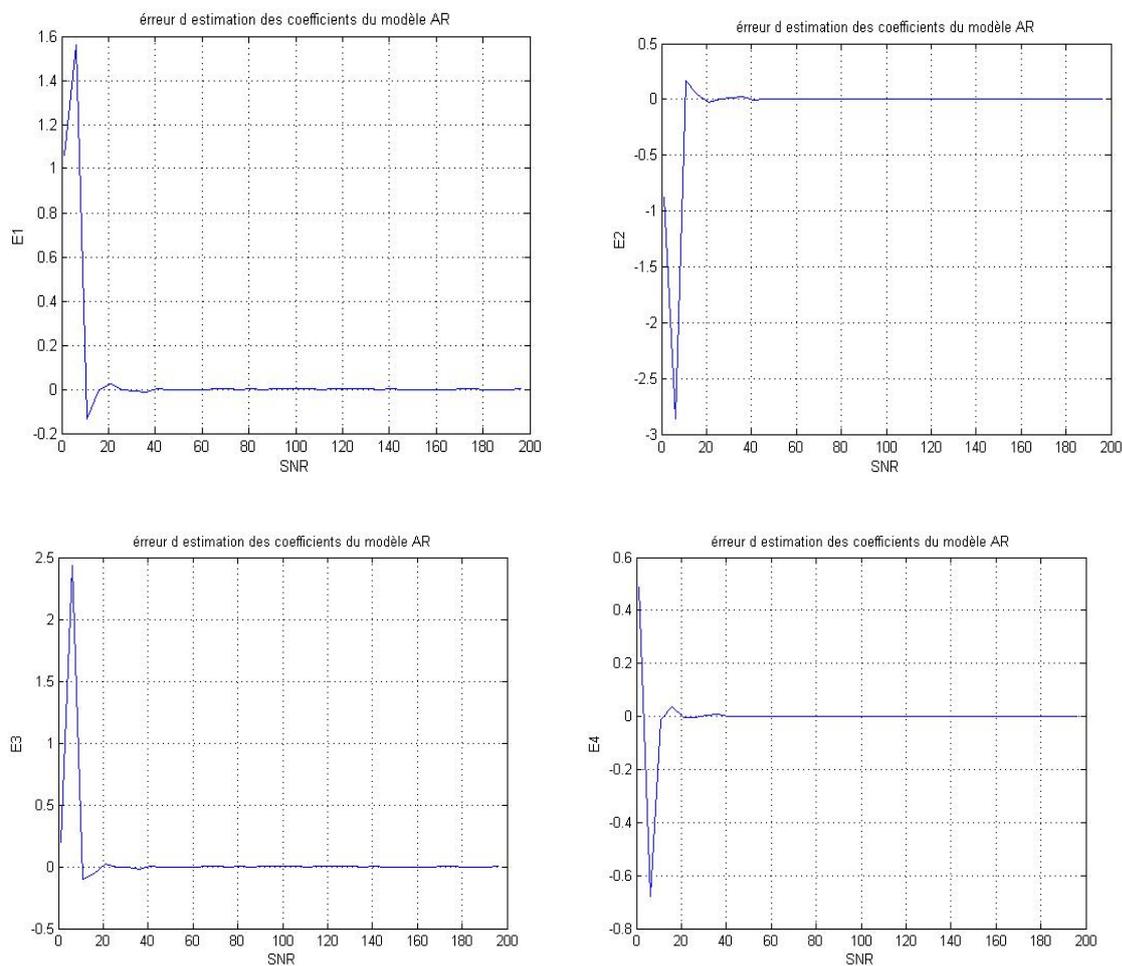


Figure 5.2. Evolution de l'erreur d'estimation des coefficients du modèle AR en fonction du rapport signal/bruit (SNR)

Donc, on peut conclure qu'avec un signal affecté par un niveau élevé de bruit, il est difficile d'établir le modèle autorégressif de ce signal, car dans ce cas, la fonction « IVAR » n'estime pas les valeurs réelles des coefficients. D'où la nécessité d'effectuer des prétraitements tel que le filtrage pour réduire le bruit, afin de donner lieu à une bonne estimation des paramètres, et par conséquent, une possibilité d'utilisation de ces paramètres comme entrée au classificateur.

V.2.2 Effet du choix de l'ordre sur l'estimation des coefficients du modèle AR

Dans cette étape, nous avons essayé de mettre en évidence l'effet du choix de l'ordre P du modèle AR sur l'estimation des coefficients a_i . Nous avons pris le cas du signal y_1 , donné par l'équation (V.1). Plusieurs modèles, d'ordres différents, ont été établis. Chaque fois on calcule l'erreur entre les valeurs réelles (équation (V.1), et les valeurs estimées par la fonction IVAR pour différents ordres.

La figure 5.3 présente l'erreur de l'estimation des coefficients a_1, a_2, a_3 , et a_4 pour différents ordres P (20,15,10,8,6,4)

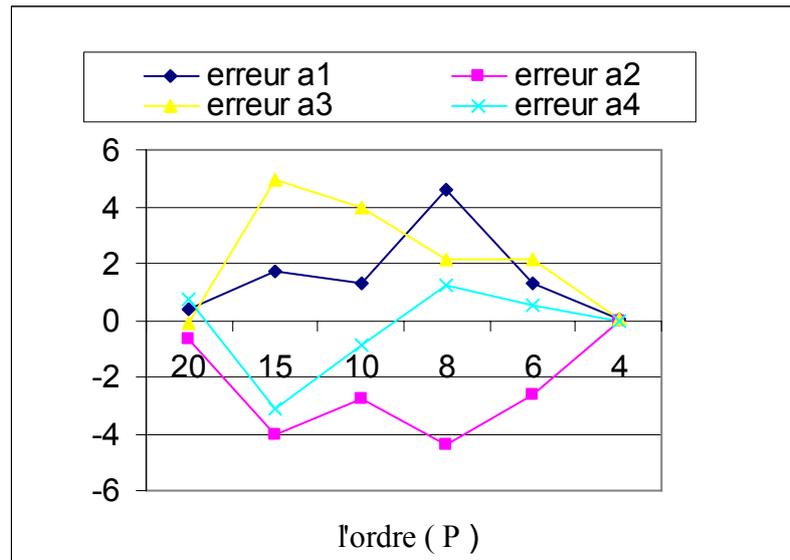


Figure 5.3 Erreurs d'estimation des coefficients du modèle AR en fonction de l'ordre du modèle

D'après cette figure, le choix de l'ordre P joue un rôle très important dans l'estimation des coefficients du modèle ; cette figure montre qu'arrivant à une valeur de ' P ' proche de la valeur réelle du modèle ($P=4$), l'erreur d'estimation diminue considérablement, et tend vers zéro pour cette valeur. Dans ce cas la fonction IVAR estime bien les coefficients a_i .

En conclusion on peut dire que, pour estimer les paramètres réels d'un modèle autorégressif, la première étape qu'on doit appliquer est le filtrage du signal afin de réduire niveau de bruit. La deuxième étape est le choix de l'ordre du modèle par des méthodes adéquates.

V.3 Conduite des différents étapes d'apprentissage et de test des SVM

Voici donc un résumé de la démarche à suivre pour entraîner et tester les SVM :

- 1- La première étape concerne le traitement de signal pour la préparation des données que l'on doit classer. Ces données seront divisées en deux ensembles ; le premier pour former la matrice d'apprentissage BA, le deuxième pour composer la matrice de test BT, ces deux ensembles sont stockés dans deux tableaux X de type observations/variables de la forme suivante :

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

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

On a alors N observation, représentées en lignes et M indicateur, 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 : EA pour l'apprentissage, et ET pour le test. Le vecteur EA est de taille Na , égale au nombre d'observations de la matrice d'apprentissage, le vecteur ET est de taille Nt , 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 pour chaque observation les valeurs d'étiquettes suivantes : 1 pour la première classe, et 2 pour 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 valeur (-1 ou 1). Pour résoudre ce problème nous utilisons la conversion suivante :

$$e' = e * 2 - 3$$

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, mais pour la méthode « un contre un », si le $i^{\text{ème}}$ classificateur binaire utilise par exemple les données de la $3^{\text{ème}}$ et la $4^{\text{ème}}$ classe, les étiquettes que utilise ce classificateur, sont changés temporellement pour avoir des valeurs 1 ou 2. Pour la méthode « un contre tous », la première classe prend l'étiquette 1, et l'étiquette 2 pour le reste des classes.

2- 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'.

5- l'algorithme SMO utilise la matrice d'apprentissage EA et le vecteur d'étiquettes ET pour résoudre le problème d'optimisation rappelé dans le troisième chapitre par l'équation (III.33). la résolution de 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 d'apprentissage pour prédire les sorties y_i de chaque observation x_i . Si le signe de la sortie est positif, y prend la valeur 1, dans le cas contraire, elle 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).

9- On peut donc calculer la performance des SVM par :

$$per = \frac{ny_b}{ny_i} \times 100$$

ny_b : nombre de sorties bien prédites.

ny_i : nombre total de sorties désirées (étiquettes).

L'organigramme donné par la figure 5.4 résume ces différentes étapes de classification sans sélection des indicateurs.

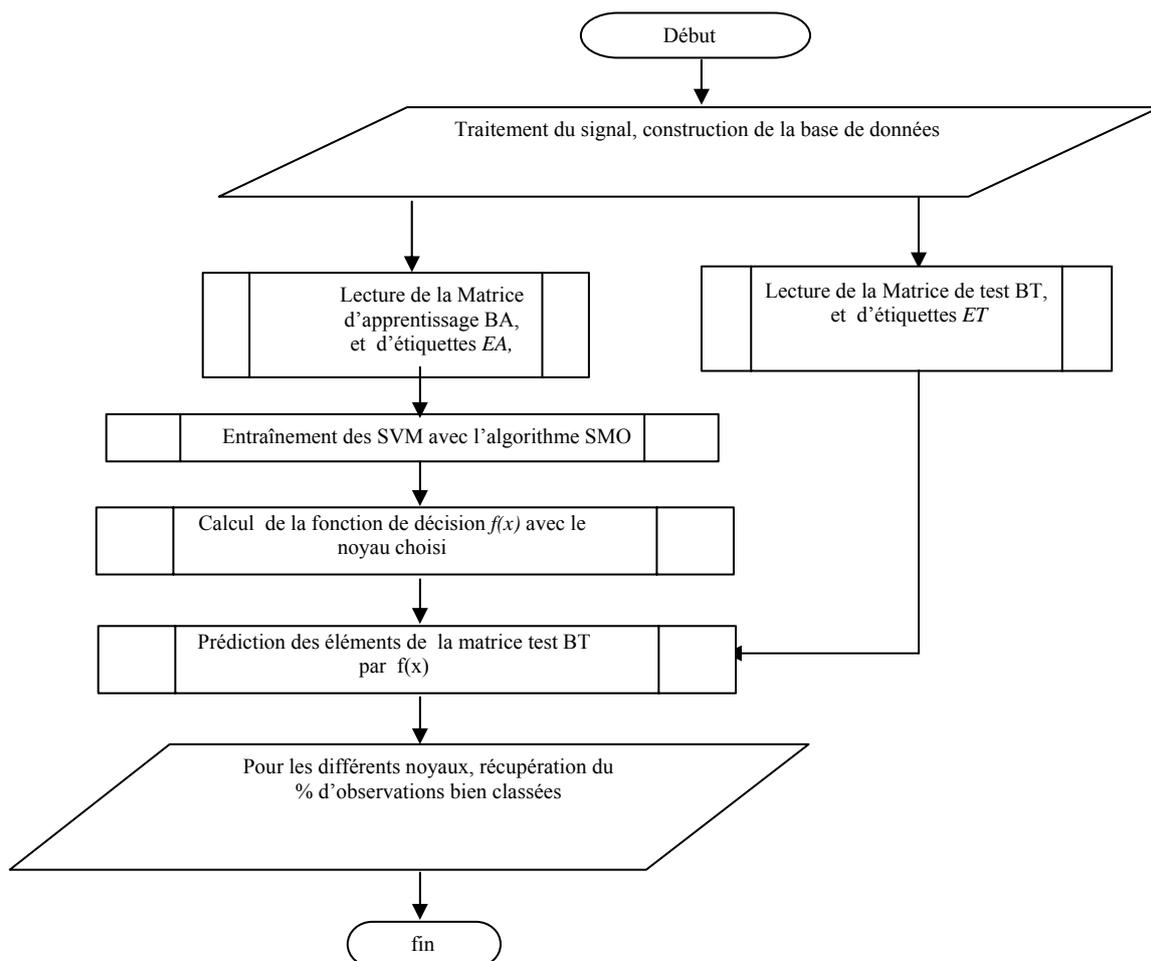


Figure 5.4. Organigramme de classification par les SVM sans sélection d'indicateurs. Utilisation du vecteur initial.

V.3.1 Classification des signaux aléatoires

Les signaux créés précédemment par les équations (V.1) et (V.2) sont traités par la fonction « IVAR » de l'environnement MATLAB, pour estimer les coefficients a_i du modèle AR. Plusieurs bases de données de $(200 \times P)$ sont obtenues, avec $P=20,15,10,4,2$, et pour différentes valeurs de $SNR=0.5,5,10,20,50,100,200$. Chaque base est divisée en deux sous ensembles ; le premier est de 150 individus, utilisé pour entraîner les SVM, le deuxième formé des 50 restante est utilisé pour le test. Les différents étapes de l'organigramme de la figure 5.4 ont été suivie pour avoir les performances de classification des SVM pour différentes valeurs de SNR et pour différents ordres du modèle autorégressif AR. Le noyau que nous avons utilisé est un noyau RBF (fonction à base radiale) de largeur $\sigma=5$, Avec une constante de régularisation $C=100$.

Les résultats obtenus sont rapportés sur le tableau 5.1

SNR	Ordre				
	20	15	10	4	2
200	72%	84%	82%	98%	100%
100	68%	58%	80%	98%	100%
50	68%	72%	78%	98%	100%
20	72%	80%	84%	98%	100%
10	68%	76%	82%	98%	100%
5	64%	80%	84%	88%	100%
0.5	72%	70%	84%	92%	100%

Tableau 5.1 Performance des SVM en fonction de niveau de bruit et l'ordre du modèle AR

Les performances des SVM varient en fonction de SNR qui indique le niveau de bruit présent dans les signaux, et de l'ordre P qu'on choisi. Nous pouvons constater que nous avons eu une performance de 92 à 98% avec $P=4$ et une performance de 100% avec $P=2$. Cela peut être expliqué par la bonne estimation des coefficients du modèle AR pour des valeurs de p proches de la valeur réelle du modèle, on peut constater aussi que les résultats obtenues en choisissant le meilleur ordre, le bruit n'a pas un effet considérable, sauf pour un niveau élevé (basses valeurs de $SNR < 10\text{db}$).

V.3.2 Classification des signaux d'engrenages simulés

Dans cette partie, des signaux d'engrenages simulés sont utilisé pour tester les performances des SVM en utilisant les coefficients du modèle AR comme indicateurs. Les signaux temporels qui forment la première classe (*sans défaut*) sont créés à partir d'un modèle mathématique présenté dans [Xian.06], ils sont donnés par l'équation (V.4) dans un intervalle de temps d'échantillonnage $t \in [0,0.5]$.

$$y_1 = \cos(2 \cdot \pi \cdot 1560 \cdot t) + \cos(2 \cdot \pi \cdot 1600 \cdot t) \quad (\text{V.4})$$

En présence d'un défaut de type écaillage, ce signal sera modulé par 40 et 80Hz, cette dernière n'apparaît que dans l'intervalle de temps [0.1, 0.3]. Donc les signaux temporels qui forment la classe (*avec défaut*) sont donnés par l'équation suivante :

$$y_2 = (1 + \sin(2\pi \cdot 40 \cdot t)) \cdot \cos(2\pi \cdot 1560 \cdot t + \sin(2\pi \cdot 40 \cdot t)) + 1.5 \cdot \text{rand}(t) + \begin{cases} (1 + \sin(2\pi \cdot 80 \cdot t)) \cdot \cos(2\pi \cdot 1600 \cdot t + \sin(2\pi \cdot 80 \cdot t)) & t = [0, 0.099] \\ \cos(2\pi \cdot 1600 \cdot t) & t = [0.1, 0.3] \\ (1 + \sin(2\pi \cdot 80 \cdot t)) \cdot \cos(2\pi \cdot 1600 \cdot t + \sin(2\pi \cdot 80 \cdot t)) & t = [0.301, 0.5] \end{cases} \quad (\text{V.5})$$

La figure 5.5 représente les signaux temporels donnés par les équations V.4, et V.5.

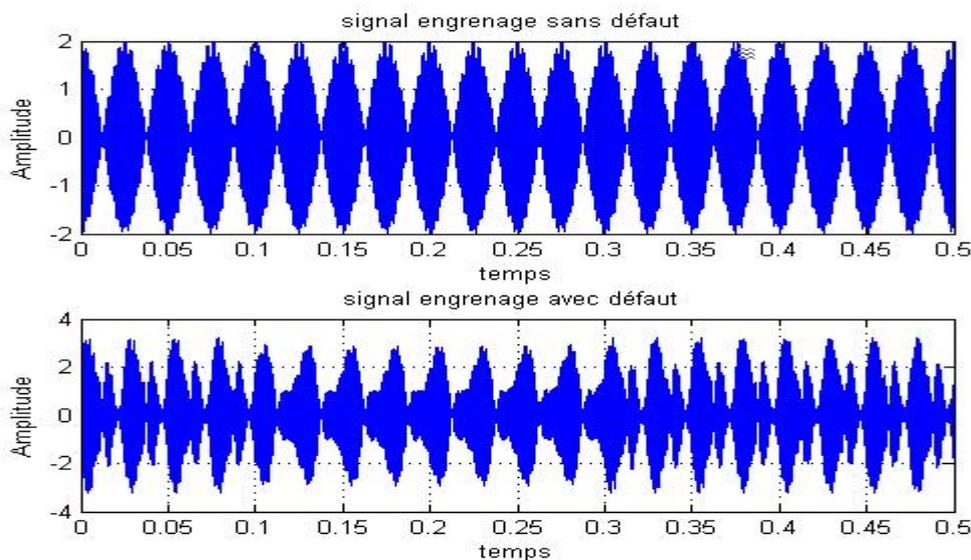


Figure 5.5 signaux d'engrenage simulés

Pour avoir suffisamment d'individus dans les deux classes, une base de données des signaux a été obtenue en ajoutant, à chaque fois, un bruit aléatoire aux signaux y_1 et y_2 . Cette base est formée de 200 signaux représentant deux classes différentes ; la première est obtenue en utilisant l'équation (V.4), la deuxième par l'équation (V.5). (100 signal dans chaque classe)

V.3.2.1 Extraction des indicateurs

Dans cette étape, l'ordre du modèle AR est inconnu, alors les coefficients a_i sont calculés pour différents valeurs de P (2,4,6,8,10,12,14,16,18,20), et pour une valeur de SNR égale à 20. Nous obtenons alors, 10 matrices de $(200 \times P)$, où P est l'ordre choisi du modèle. Pour chaque base de données, l'algorithme d'apprentissage des SVM est entraîné avec 150 individus, or les 50 restantes sont utilisées pour le test.

V.3.2.2 Performances des SVM en fonction du SNR pour différents ordres

Les SVM sont entraînés et testés avec les différentes bases de données, Le noyau que nous avons choisi d'utiliser, est un noyau RBF. La figure 5.6 représente la variation des performances des SVM avec les différentes bases de données (différents ordres), et pour différentes largeurs du noyau RBF

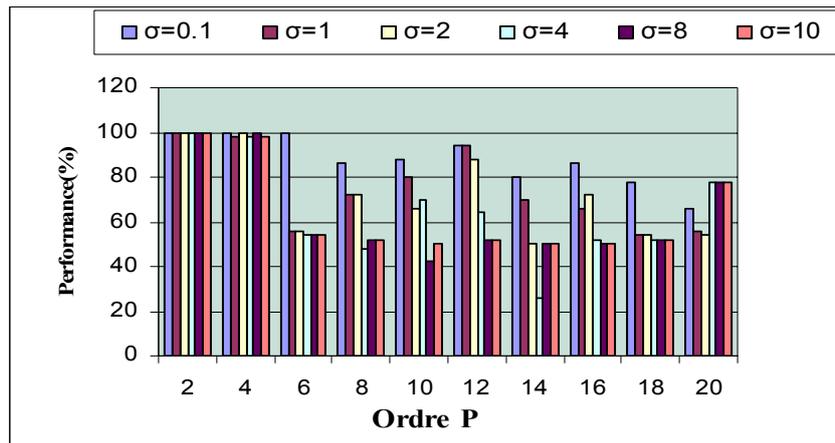


Figure 5.6 Performances des SVM pour un SNR=20

Les résultats obtenus montrent que les performances des SVM varient en fonction de l'ordre P du modèle choisi. Nous avons obtenu une meilleure performance de classification (pourcentage des bien prédits) de 98% à 100% avec des valeurs de P inférieures à 6 pour les différentes valeurs de σ . Ceci nous permet de dire que les coefficients du modèle autorégressif, calculés pour un ordre P inférieur à 6 sont des bons indicateurs d'état.

V.4 Conclusion

Cette première étude nous a permis de conclure que la bonne classification passe toujours à travers une bonne préparation des données d'apprentissage. Les résultats obtenus avec les coefficients du modèle AR sont encourageants, mais cela reste toujours conditionné par un bon choix du modèle représentant le signal.

Le choix de l'ordre P est un facteur très important pour une meilleure estimation des paramètres a_i du modèle AR. Donc une bonne estimation de ces coefficients offre des informations précises sur le signal traité, et par conséquent elle facilite leur utilisation comme entrée au classificateur.

V.5. Application sur des signaux réels

Dans cette section nous allons étudier les performances des SVM avec une application sur des signaux réels, dans la première partie, comme pour les signaux simulés, tout l'ensemble d'indicateurs calculés participe à l'entraînement et au test des SVM. Dans la deuxième partie, seulement, un sous ensemble des indicateurs est sélectionné pour entraîner et tester les SVM. Cette sélection est faite sur la base du pouvoir séparateur de chaque indicateur. A cet effet nous proposons une méthode de sélection basée sur un couplage des algorithmes génétiques et l'analyse discriminante linéaire (ADL).

V.5.1. Banc d'essai CETIM

Les vibrations d'engrenages sur lesquelles nous avons fait les tests proviennent d'un réducteur du CETIM (figure 5.7). (*Centre d'Etudes Techniques des Industries Mécaniques, 52 av. Felix Louat, 60300 Senlis*), fonctionnant 24h/24h. Le dimensionnement des roues d'engrenages cimentés trempés, ainsi que les conditions de fonctionnement (vitesse, couple) sont fixés de façon à obtenir un écaillage sur toute la largeur d'une dent. Tous les jours après l'acquisition des signaux vibratoires, le banc est arrêté, pour expertiser l'état des dentures des roues. Les essais du réducteur s'étalent sur 12 jours avec l'enregistrement d'un signal vibratoire chaque jour.

Le système est composé d'un moteur, d'un réducteur de bouclage de rapport 40/42 et du réducteur testé, de rapport 20/21. Le réducteur testé a pour vitesse de rotation 1000 tr/min, soit une fréquence de rotation 16,67 Hz. Or, les deux roues composant ce réducteur ont un nombre de dents quasiment identique. Leurs fréquences de rotation sont donc très proches l'une de l'autre. Nous avons:

La fréquence d'engrènement des deux roues $f_{eng} = 333.48 \text{ Hz}$

La fréquence d'échantillonnage $f_e = 20 \text{ kHz} = 20000 \text{ Hz}$

La période d'échantillonnage : $\Delta t = 1/f_e = 5.10^{-5} \text{ s}$

Au cours de l'expérimentation, l'engrenage de test passe de l'état de bon fonctionnement à celui d'un engrenage détérioré.

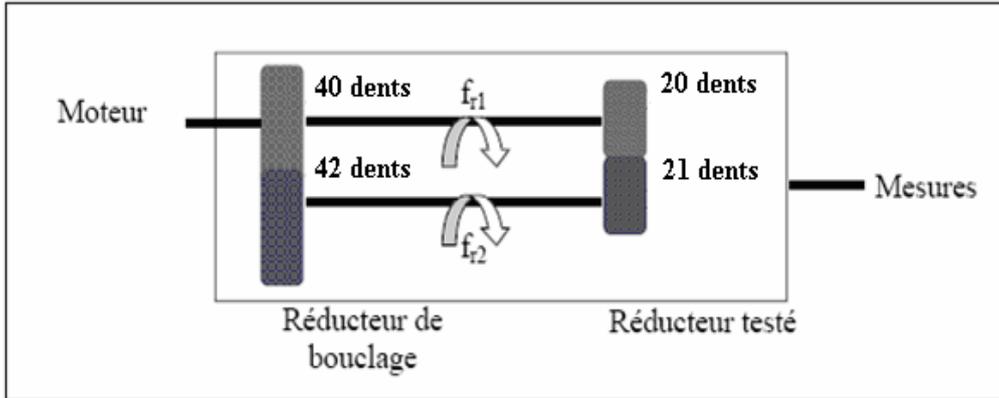


Figure 5.7 Schéma de dispositif d'enregistrement

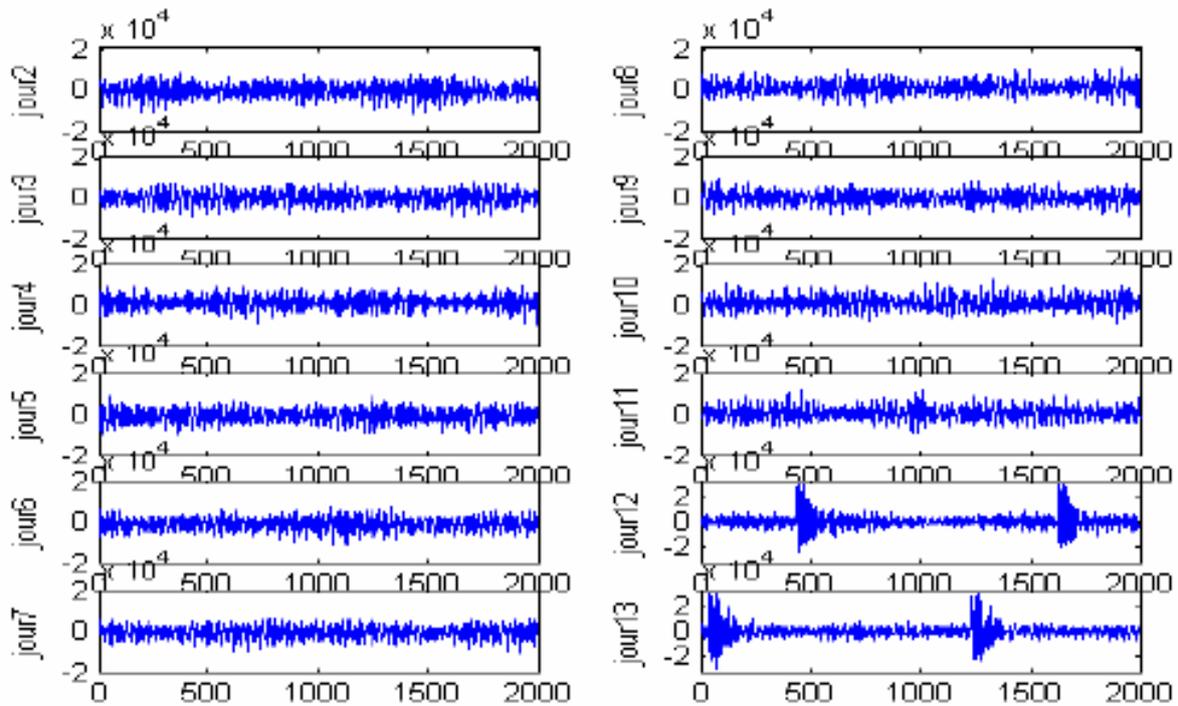


Figure 5.8 Signaux temporels du train d'engrenage

La seule remarque possible sur les signaux temporels de la figure 5.8 consiste à souligner l'apparition d'un motif, lié au défaut avec l'apparence d'un choc, de manière périodique. La mesure de cette période nous donne environ $0.0625s$ soit une fréquence de $16 Hz$. Cette remarque nous permet uniquement de penser qu'un défaut apparaît sur l'une des roues mais sans pouvoir toutefois déterminer laquelle des deux est atteinte. A cet effet nous considérons que les signaux de 2^{ième} au 10^{ième} jour, forment la première classe (sans défaut), et que le 12^{ième} et le 13^{ième} jour appartiennent à la deuxième classe (avec défaut).

V.5.1.1 Préparation des observations

V.5.1.1.1 Division des signaux

Pour avoir suffisamment d'individus dans les deux modes de fonctionnement (sans et avec défaut) et puisque nous disposons d'enregistrements d'une longueur de $N_s=60160$ points, chaque signal est décomposé en K_t tranches de N_M points avec, éventuellement, un recouvrement des tranches. Tous les prétraitements et traitements du signal peuvent être effectués. Nous procédons alors au choix de K_t et N_M . 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 $T_2=0.063s$. Si nous prenons un coefficient de 5. La longueur nécessaire de la tranche sera : $Tt = 0.063 \times 5 = 0.315s$ soit un nombre d'échantillons nécessaire par tranche de :

$$Ne = \frac{Tt}{\Delta t} = 1260 * 5 = 6300 \text{ pts} . \quad \text{Où } 1260 \text{ est le nombre de points par tour.}$$

Si nous prenons pour N_M une valeur supérieure à Ne , $N_M = 2^{13} = 8192$ échantillons. Alors

le nombre de tranche dans le cas d'un recouvrement 50 %, soit $s_1 = \frac{N_M}{2}$, est de :

$$K_t = \frac{N_s - s_1}{N_M - s_1} \cong 13 \text{ tranches} \tag{V.6}$$

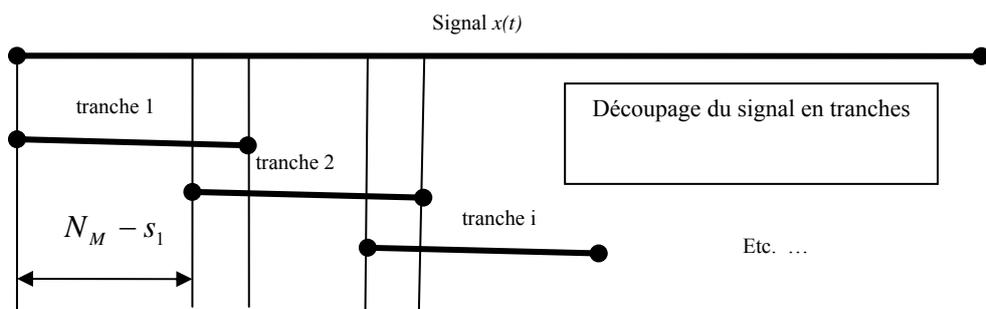


Figure .59. Découpage d'un signal en tranches ou sections

V.5.1.1.2 Filtrage

Pour éviter l'effet de masque qui entache fortement la sensibilité des indicateurs "larges bandes" 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 machine et des défauts déjà existants, 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.

V.5.1.1.3 Décimation

Pour Chaque jour, nous disposons de 13 tranches de signaux, échantillonnées à 20000 Hz. Par le filtrage ci-dessus, nous avons limité notre étude à une fréquence maximale de 1083 Hz. Les tranches des signaux sont, donc, suréchantillonnées. 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 le théorème de Shannon ($f_e \geq 2 \cdot f_{\max}$) est donc bien vérifié. Le vecteur décimé est 4 fois plus court.

V.5.1.1.4 Extraction des indicateurs

Après les prétraitements précédents, et une analyse préliminaire, nous avons choisi de calculer les indicateurs temporels et fréquentiels 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 ($1.f_e = 333.33$ Hz, $2.f_e = 666.66$ Hz, et $3.f_e = 999.99$ Hz). Les bandes de calcul sont alors les suivantes: [250-417 Hz], [583-750 Hz] et [916-1083 Hz].

a) Les indicateurs temporels

Dans chaque tranche nous avons choisi de calculer les indicateurs statistiques 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 les trois bandes précitées pour chaque tranche. La matrice des données obtenue est de (156 individus \ 27 indicateurs).

b) Les coefficients du modèle AR

L'idée générale est de représenter le signal temporel par un modèle paramétrique et d'extraire des indicateurs basés sur les paramètres de ce modèle. Ici nous avons choisi d'utiliser le modèle autorégressif (AR) d'ordre P , et d'extraire les coefficients a_i du modèle établi pour chaque tranche.

Lors du calcul des coefficients a_i du modèle autorégressif pour les signaux simulés, nous avons montré que le choix de l'ordre P a un effet considérable sur l'estimation de ces coefficients, pour résoudre ce problème nous avons utilisé un algorithme basé sur la méthode proposée dans [Felak94], décrite dans le premier chapitre. Le résultat obtenu par cet algorithme nous donne pour un modèle AR représentant une tranche l'ordre 19, donc nous avons estimé 19 coefficients. Ce qui nous donne pour tous les jours, une matrice de (156 individus \ 19 indicateurs).

c) Les indicateurs fréquentiels

La figure.5.10 montre la densité spectrale de puissance (DSP) pour six jours différents. On constate bien une augmentation d'énergie pour le 12^{ème} 13^{ème} jour comparée au autres cas. De ce fait, les énergies dans les bandes et autour des fréquences d'engrènements seront d'une grande utilité.

Pour extraire les indicateurs fréquentiels, nous avons choisi de calculer les énergies autour des fréquences caractéristiques dans les bandes [250-417 Hz], [583-750 Hz], [916-1083 Hz], [1250-1417] et dans la bande totale [250-1417]. La matrice obtenue est, donc, composée de (156 individus \ 5 indicateurs).

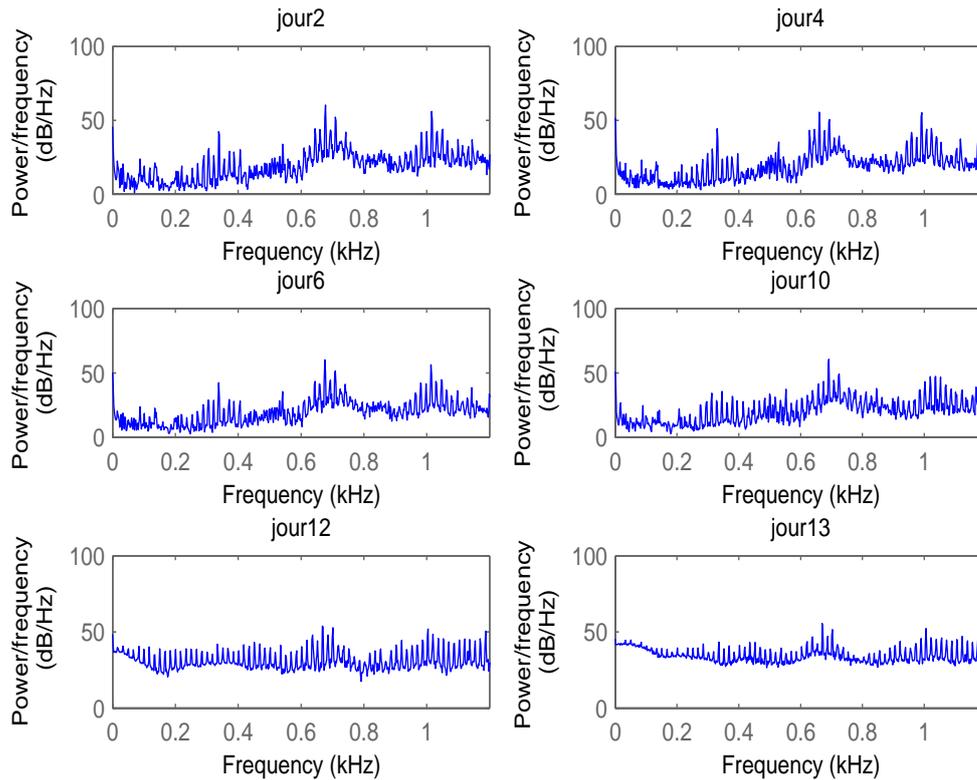


Figure 5.10 DSP pour différents signaux

V.5.1.1.5 Normalisation

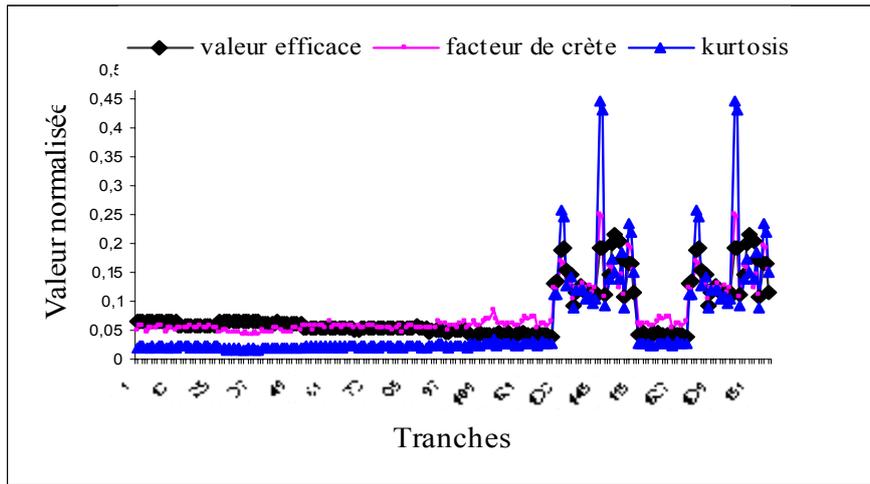
Dans la littérature (par exemple voir [lind.00]) la normalisation des données a un effet significatif pour l'accélération, et le succès de l'entraînement des SVM. La normalisation a pour but d'éviter qu'un paramètre ne prenne trop d'importance par rapport aux autres simplement parce qu'il varie dans de plus grandes proportions, on procède généralement à une normalisation des données en divisant les composantes de chaque paramètre par son écart-type et en les centrant. Donc avant d'entraîner les SVM, chaque ligne des matrices d'apprentissage BA, et de test BT, sont normalisées en accord avec la formule suivante :

$$x_i = \frac{x_i - m_x}{\sigma_x} \tag{V.7}$$

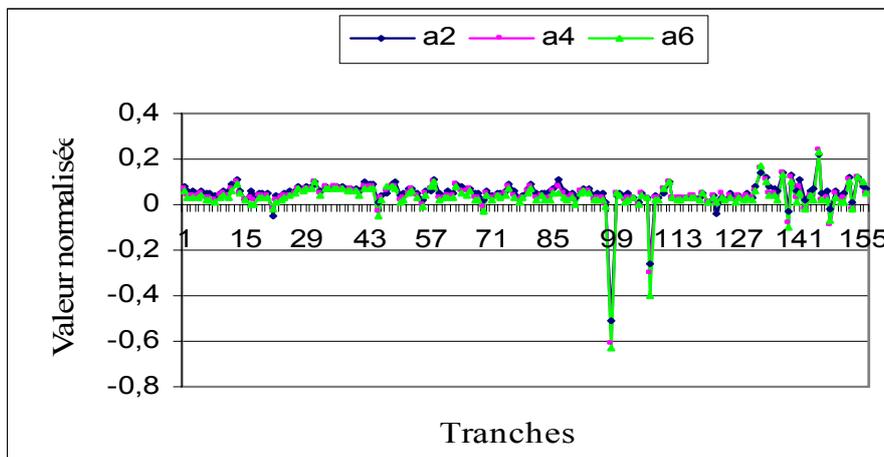
Où m_x : la moyenne

σ_x : La variance

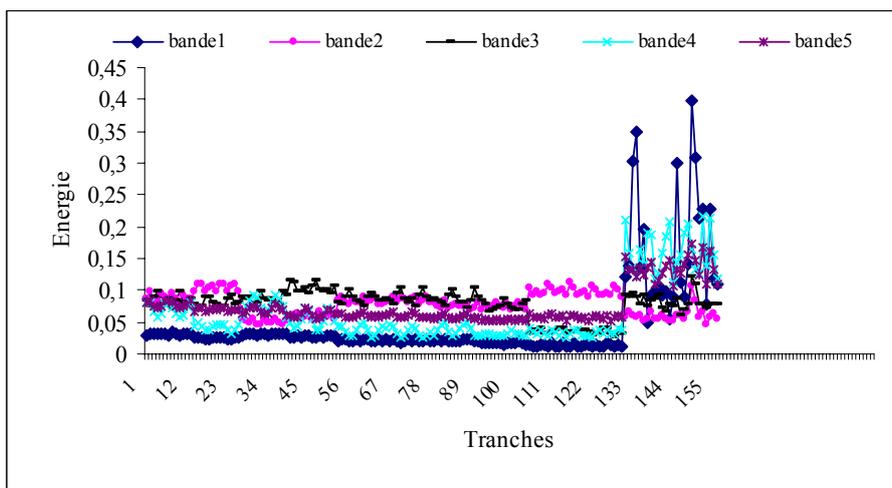
Donc Les trois ensembles de données obtenus, forment une matrice totale (BDT) de (156individu \ 51 indicateurs). La figure 5.11 représente l'évolution de quelques indicateurs de chaque ensemble.



a) les indicateurs temporels calculés dans la bande [250-417]



b) les coefficients du modèle AR



c) les indicateurs fréquentiels

Figure 5.11 Evolution de quelques indicateurs pour toutes les tranches (12 jours)

Cette figure montre l'évolution des indicateurs pour les douze jours considérés (156 tranches). On peut constater qu'arrivant à l'observation 131, une variation brusque apparaît pour la majorité des indicateurs, car à partir de cette observation on rentre dans l'intervalle des observations appartenantes à la deuxième classe (de 131 jusqu'à 156), cette variation, caractérise ainsi une dégradation de l'état de l'engrenage et l'apparition d'un mode de fonctionnement avec défaut

V.5.1.2 Performances des SVM (sans sélection)

L'étape d'extraction des indicateurs nous a permis d'avoir une matrice de données formée par les trois ensembles précédents, et qui est utilisée comme entrée au classificateur (SVM). Chaque ensemble est composé de 156 individus, dont 84 sont utilisés pour l'entraînement, et les 72 restants pour le test.

Un des grands problèmes rencontrés lors de la classification par les SVM est le choix des valeurs des paramètres du noyau permettant une bonne classification, un autre problème est lié au choix de la valeur de la constante de régularisation C , car le choix de ces deux paramètres dépend de la nature des problèmes traités. Pour résoudre ces deux problèmes nous avons utilisé la méthode de validation croisée (en anglais « cross validation »). Cette technique consiste à utiliser deux ensembles indépendants, en partitionnant les données d'apprentissage en deux ensembles distincts ; le premier pour l'apprentissage, le deuxième pour la validation, pour avoir les meilleurs paramètres nous utilisons deux vecteurs de paramètres ; le premier est composé de N paramètres de noyau, le deuxième est composé de M valeurs de la constante de régularisation C . Les SVM sont entraînés chaque fois avec deux paramètres différents. Nous obtenons alors $N*M$ modèles (fonctions de décisions). Le meilleur modèle obtenu (celui qui donne un faible taux d'erreur), qui est établi avec les meilleurs valeurs des paramètres du noyau et de la constante C . Nous utilisons ce modèle pour prédire les sorties des observations de la matrice de test BT. La figure 5.12 illustre la méthode de validation croisée dans le cas de la recherche des meilleurs paramètres du noyau et de la constante de régularisation qui minimisent le taux d'erreurs

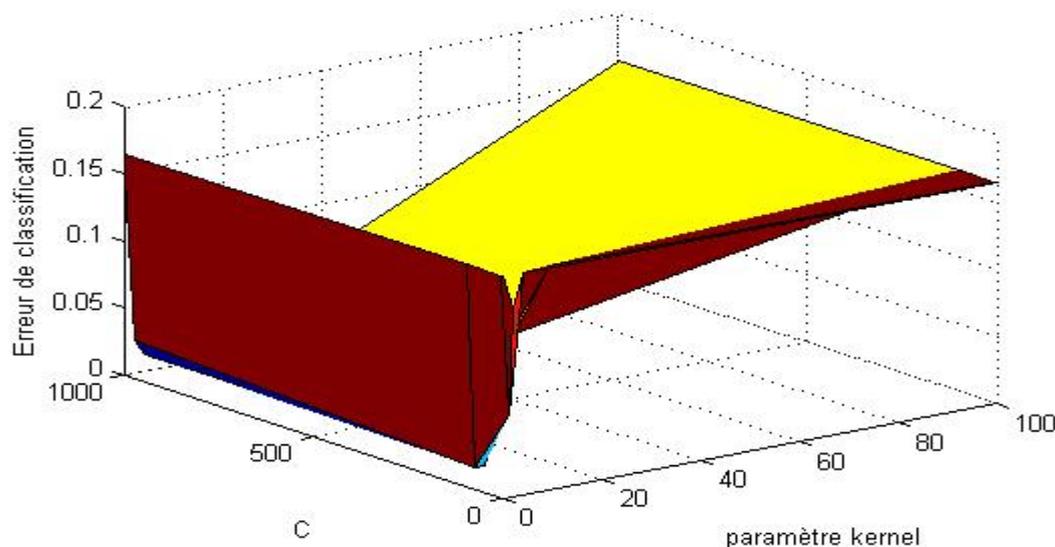


Figure 5.12 illustration de la validation croisée pour les données de CETIM

Cette figure montre l'évolution du taux d'erreur de classification des données de CETIM en fonction de paramètre du noyau, et de la constante de régularisation C . le noyau utilisé est une fonction à base radiale (RBF), la valeur optimale de la largeur σ est égale à 2, avec une

valeur optimale de C égale à 100. Donc le meilleur modèle, que nous avons utilisé pour prédire les sorties des observations de la base de test BT, est celui établi en utilisant les valeurs optimales de ces deux paramètres.

Le tableau 5.2 représente les performances des SVM pour les trois ensembles d'indicateurs et pour différents noyaux (*kernel*) : polynomial, RBF, et sigmoïdal. Notons ici que les paramètres optimaux de chaque noyau ont été choisis par la méthode de validation croisée

Ensembles d'indicateurs	Nbre VS	Noyau	Performance en test (%)
1 ^{er} ensemble (27)	11	Polynomial	95.83
	5	RBF	100
	12	Sigmoïdal	100
2 ^{ième} ensemble (19)	15	Polynomial	86.11
	11	RBF	97.22
	26	Sigmoïdal	83.33
3 ^{ième} ensemble (5)	4	Polynomial	100
	6	RBF	100
	9	Sigmoïdal	100
Totale (51)	23	Polynomial	98.61
	10	RBF	100
	30	Sigmoïdal	97.22

Tableau 5.2 Performances des SVM .

Avec:

1^{er} ensemble : les indicateurs temporels

2^{ième} ensemble : les coefficients du modèle AR.

3^{ième} ensemble : les indicateurs fréquentiels.

Nbre VS : nombre des vecteurs de supports.

Les résultats montrent que le meilleur pourcentage des bien prédits (performance en test) est obtenu avec les indicateurs fréquentiels (100%) pour les noyaux : RBF, polynomial, et sigmoïdal, tandis que l'utilisation de la somme des trois ensembles d'indicateurs n'a pas amélioré les performances du classificateur, cela peut être expliqué par l'effet de redondance d'indicateurs. Pour éliminer cet effet, il est nécessaire d'utiliser des méthodes de sélection des indicateurs les plus pertinents.

Les résultats montrent que le noyau RBF donne la meilleure performance pour les trois types d'indicateurs, cela est favorisé par un bon choix du largeur σ par la méthode de validation croisée, car ce paramètre détermine le rayon de l'hypersphère qui enferme les données dans l'espace multidimensionnel ; si la valeur de σ est trop petite, l'espace enfermé est aussi trop petit, donc on risque de perdre une partie significative des données, aussi une grande valeur peut causer un chevauchement entre les classes.

V.5.2 Boite de vitesse d'un hélicoptère « NAVY Westland data »

Le second ensemble de signaux étudiés provient de l'instrumentation d'une boite de vitesse d'un hélicoptère de la NAVY américaine (Westland data). L'annexe 2 présente une photo du système étudié ainsi que sa position dans le moteur. L'objectif est d'étudier les performances des SVM dans la classification des données appartenantes à plusieurs classes

(les SVM multi classes). Nous disposons des signaux caractérisant le système pour sept défauts différents et d'un ensemble provenant du système sain. La figure 5.13 présente un schéma simplifié du système.

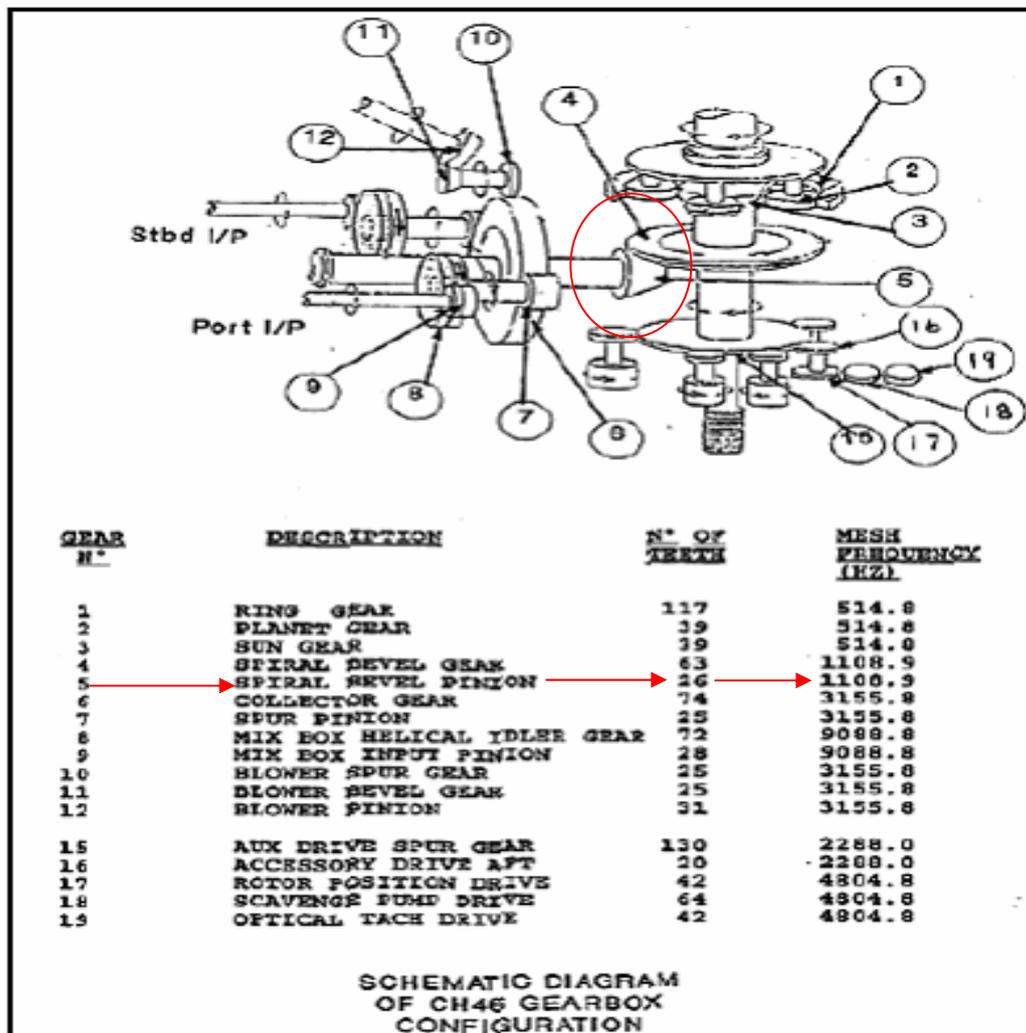


Figure 5.13 Schéma simplifié de la boîte de vitesse de l'hélicoptère CH46. Numérotation des éléments et caractéristiques

Le tableau 5.3 récapitule les sept défauts disponibles ainsi que leur nature (écaillage, fissure...) et les pièces concernées.

Numéro de défaut	Pièces concernées par le défaut	Description du défaut
2	2	Ecaillage de l'Epicyclic Planet Gear
3	5	Pitting de l'Input Pinion
4	5	Ecaillage de l'engrenage conique
5	9	Ecaillage de l'Helical Input Pinion
6	8	Fissure de l'Helical Idler Gear
7	6	Fissure du Collector Gear
8	5 et 6	Fissure du Quill Shaft

Tableau 5.3 Description et position des défauts.

Pour chaque défaut, huit accéléromètres enregistrent, en différents endroits de la boîte de vitesse, les vibrations engendrées par le système. Les mesures sont effectuées à vitesse constante, pour neuf couples différents. Le tableau ci dessous présente les données disponibles en fonction du couple et du défaut considéré. Le défaut numéro 9 correspond en fait au système sain.

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

Tableau 5.4 Ensemble des données disponibles en fonction du couple et du défaut.

Chaque signal comporte 412464 points, échantillonnés à 103116Hz. Il est à noter que pour chaque défaut, nous disposons généralement de trois signaux ; le premier provenant du système sain, le second caractérisant le défaut partiel, et le dernier dû au défaut établi. Dans cette étude nous précédon, comme dans la section V.5.1.1.1, par la divisions des signaux afin d'avoir un nombre suffisant d'observations pour l'apprentissage et le test des SVM.

Dans la suite de ce chapitre, nous nous limiterons à l'étude d'un seul défaut particulier : l'écaillage de l'engrenage conique (défaut #4). L'annexe 3 présente des photos de l'engrenage avant les tests ainsi que des vues des différents niveaux de défaut mis à notre disposition.

La figure 5.14 présente un schéma simplifié de l'élément atteint par le défaut que nous souhaitons étudier ainsi que la position de l'élément dans le système. L'accéléromètre le plus proche ainsi que la notation des fréquences utilisées par la suite y sont également présentés.

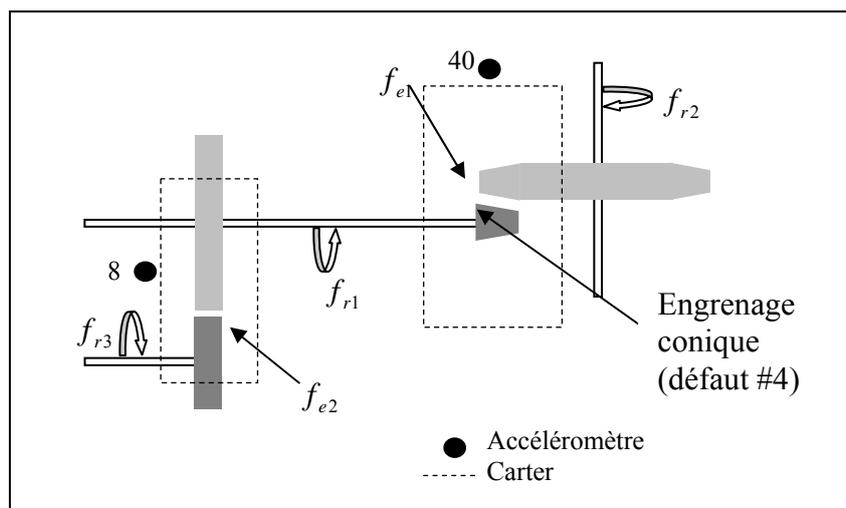


Figure 5.14 Schéma simplifié du système

Fréquence d'échantillonnage : $fe = 103116Hz$
 Nombre de points total est : $N_T = 412464$ points
 Les fréquences d'engrènement : $fe_1 = 1108.9 Hz$. $fe_2 = 3155.8 Hz$
 Nombre de dents du pignon conique 5 : $n_{d1} = 26$ dents

$$fr_1 = \frac{fe_1}{n_{d1}} = \frac{1108.9}{26} = 42.65Hz$$

Le nombre de points par période N_p est :

$$N_p = \frac{fe}{fr_1} = \frac{103116}{42.65} = 2417 pts \tag{V.8}$$

La période T est : $T = N_p * \Delta T = N_p * \frac{1}{fe} = 0.023s$

La figure 5.15 représente les signaux temporels recueillis par le capteur 4 pour un couple de 45%.

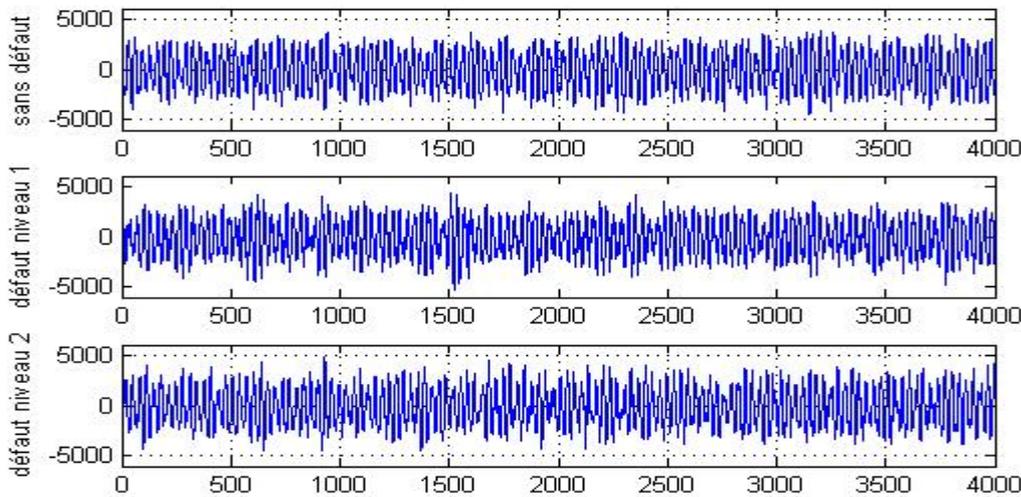


Figure 5.15 Signaux temporels du capteur 4, couple 45%, pour différents états du pignon conique

V.5.2.1 Préparation des observations

V.5.2.1.1 Prétraitement des signaux

Dans cette partie les étapes de prétraitement restent les mêmes que celles des signaux de CETIM. Nous procédons alors 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 $T = 0.023s$. Si nous prenons un coefficient de 20, la longueur nécessaire de la tranche sera : $Ne = 2417 \times 20 = 48340$ point.

Si nous prenons pour N_M une valeur supérieure à Ne , $N_M = 2^{16} = 65536$ échantillons. Alors

le nombre de tranches dans le cas d'un recouvrement 50 %, soit $s_1 = \frac{N_M}{2}$, est de :

$$K_t = \frac{N_s - s_1}{N_M - s_1} \cong 11 \text{ tranches} \tag{V.9}$$

Pour le défaut étudié, nous disposons de 9 signaux représentant l'état sain, 6 signaux pour le défaut niveau1, et 9 signaux pour le défaut niveau 2. Après la division de chaque signal en 11 tranches nous obtenons 264 tranches échantillonnées à 103166 Hz. le signal de chaque tranche a donc été centré puis filtré par un filtre passe-bande, là nous avons limité notre étude à une fréquence maximale de 5000 Hz. 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.

V.5.2.1.2 Extraction des indicateurs

Dans cette étape nous avons choisi d'extraire les trois ensembles des 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) autour des 3 principales harmoniques ($1.f_e = 1108.9$ Hz, $2.f_e = 2217.8$ Hz, et $3.f_e = 3266.7$ 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 (264individus \ 27indicateurs)

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 reste la même que pour les signaux de CETIM, nous avons obtenu pour une bande fréquentielle de 426 Hz l'ordre du modèle représentant est égale 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 (264individus \ 6indicateurs)

3^{ème} ensemble : dans cette étape, nous avons choisi de calculer les énergies autours des fréquences caractéristiques dans les trois bandes précités, dans la bande [4222,4648], et la bande totale [895-1417]. La matrice obtenue est, donc, composée de (264 individus \ 5indicateurs).

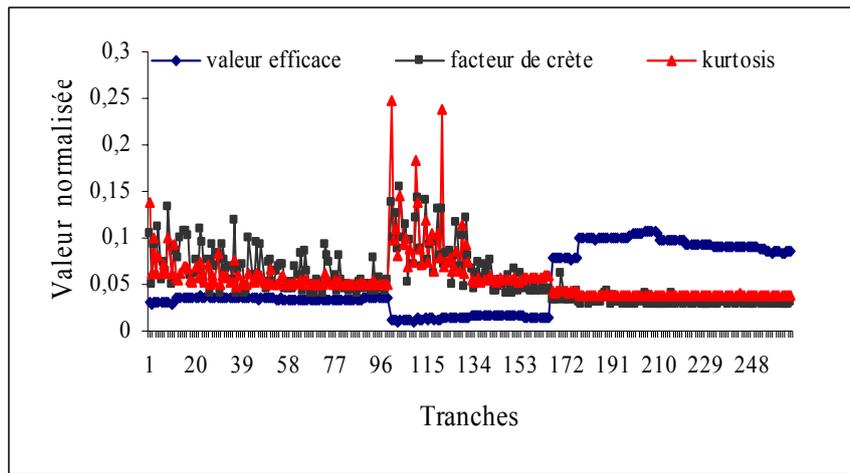
Ces trois ensembles forment une matrice totale de (264individus \ 38 indicateurs) ces individus appartiennent au trois classes suivantes :

1^{ière} classe (sans défaut) : composée de 99 individus.

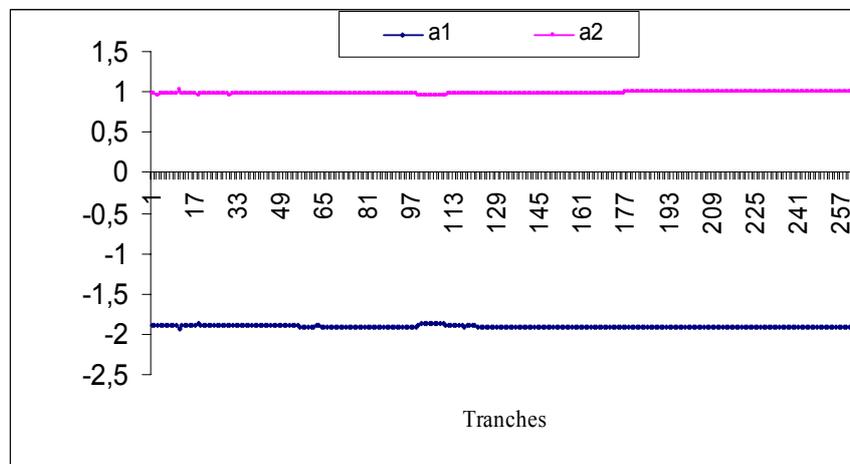
2^{ième} classe (défaut niveau1) : composée de 66 individus.

3^{ième} classe (défaut niveau2) : composée de 99 individus

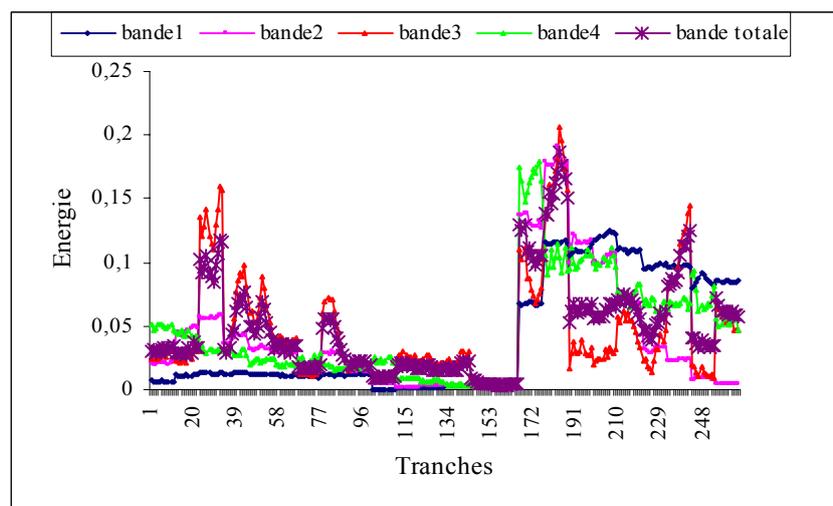
Toutes les données ont été normalisées en accord avec (V.7). La matrice totale obtenue est divisée en deux sous ensembles : Le premier formé de 144 individus (6 tranches de chaque signal), utilisé pour l'entraînement des SVM, le deuxième composé de 120 (les 5 tranches restantes de chaque signal), utilisé pour le test. La figure 5.16 représente l'évolution de quelques indicateurs de chaque ensemble d'indicateurs.



a) les indicateurs temporels calculés dans la bande [895-1321]



b) les coefficients du modèle AR calculés dans la bande [895-1321]



c) les indicateurs fréquentiels

Figure 5.16 Evolution de quelques indicateurs pour toutes les tranches

Cette figure représente l'évolution de quelques indicateurs de chaque ensemble d'indicateurs calculé. On peut constater une variation des indicateurs temporels, et fréquentiels dans les trois intervalles suivants :

[1,99] : 99 tranches de la première classe (sans défaut)

[99,165] : 66 tranches de la deuxième classe (défaut niveau 1)

[166,256] : 99 tranches de la troisième classe (défaut niveau 2)

cette variation, caractérise une dégradation de l'état de l'engrenage pour le deux niveaux de l'écaillage. On peut également remarquer que la variation des coefficients du modèle AR, est très petite cela peut être expliquée par le calcul des coefficient dans des bandes très étroites (426Hz), contrairement au signaux de CETIM, où nous avons calculé ces coefficients sur le long de chaque tranche.

V.5.2.3 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*.

V.5.2.3.1 Un contre un

Cette méthode est basée sur des combinaison de $q_j(x)$ classificateurs binaires tel que :

$$q_j(x) = \begin{cases} 1 & \text{pour } f(x) \geq 0 \\ 2 & \text{pour } f(x) < 0 \end{cases} \quad (\text{V.10})$$

Dans le cas étudié, nous avons trois classes, les classificateurs 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 classificateur binaire ne peut affecter que les observation appartenant au 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 C classes. Soit $q_j : x \subset \mathcal{R}^n \rightarrow \{y_j^1, y_j^2\}$ $j = 1, \dots, g$ un classificateur de g classificateurs binaires, qui affecte l'observation x à la classe $y_j^1 \in \mathcal{Y}$ ou $y_j^2 \in \mathcal{Y}$, et soit $v(x)$ un vecteur de $[c \times 1]$ votes de classes \mathcal{Y} lors de la classification de l'observation x .

Le vecteur $v(x) = [v_1(x), \dots, v_c(x)]^T$ est calculé comme suit :

$$v_y(x) = 0 \quad \forall y \in \mathcal{Y}$$

Pour $j=1$ jusqu'à g faire

Si $y = q_j(x)$ alors

$$\quad \quad \quad v_y(x) = v_y(x) + 1$$

Fin si

Fin pour

Cette méthode affecte l'observation x à la classe $y \in \mathcal{Y}$ qui a la majorité des votes :

$$y = \arg \max_{y=1\dots g} v_y(x) \quad (\text{V.11})$$

Le tableau 5.5 représente les performances de chaque classificateur binaire ainsi que la performance totale obtenue en utilisant la règle de vote majoritaire, pour les différents noyaux : linéaire polynomial, RBF, et sigmoïdal.

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

Tableau 5.5 Performances des SVM avec la méthode « un contre un » pour les différents noyaux (sans sélection)

Le tableau 5.5 compare les résultats obtenus avec les différents noyaux. Les symboles (perf 1-2) (perf 1-3) (perf 2-3) désignent respectivement les performances des classificateurs binaires établis pour les paires de classes (1et2) (2et3), et (1et3), 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 la meilleure performance (91.11% à 100%) est obtenue avec le classificateur binaire établi entre la première et la troisième classe, cela est expliqué par le niveau développé du défaut et qui est traduit sur la figure 5.16 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 5.17 représente l'hyperplan optimal séparant les trois classes

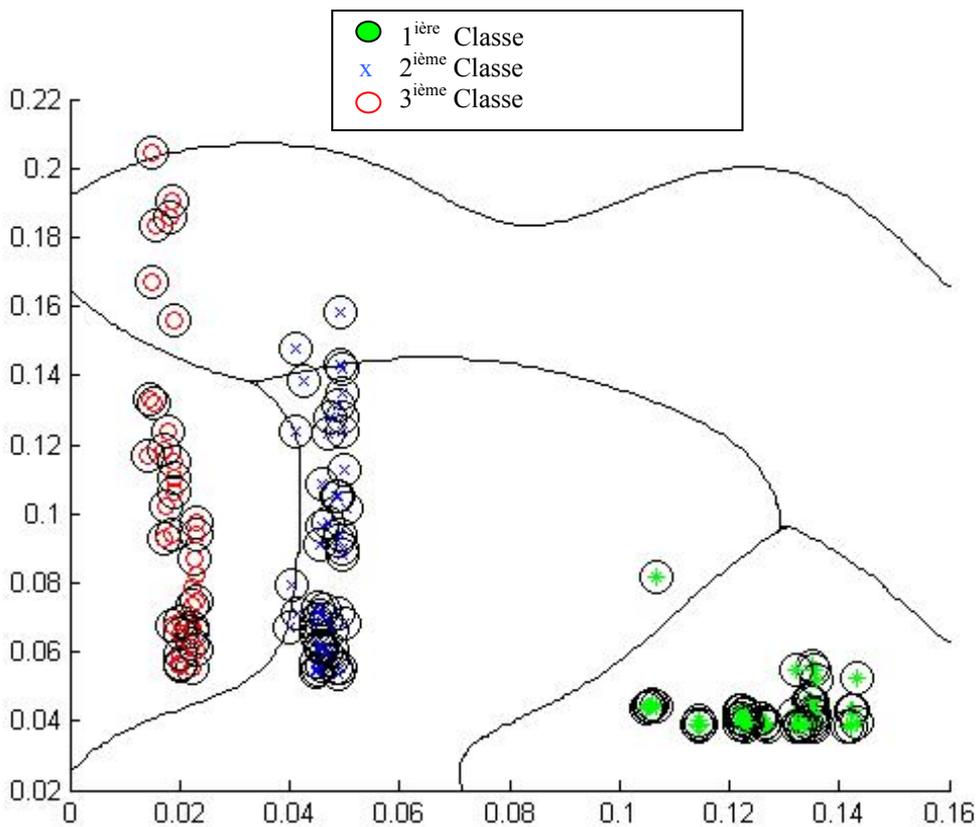


Figure 5. 17 Séparation des trois classes (avec un noyau RBF)

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’écailage de l’engrenage.

V.5.2.3.2 Un contre tous

La deuxième méthode que nous avons utilisé est « un contre tous ». L’entrée du classificateur « un contre tous » est composée de la matrice d’apprentissage BA et de test BT, les vecteurs d’étiquettes : EA pour l’apprentissage, ET pour le test . Ces deux vecteurs sont composés des étiquettes $y_i \in y = \{1, 2, \dots, Nc\}$ avec Nc : le nombre des classes.

Le problème de classification multi classes est résolu par la méthode un contre tous, en transformant le problème en séries de Nc classificateurs binaires. Chaque classificateur binaire est entraîné et tester en suivant les mêmes étapes données par l’organigramme de a figure 5.4. chaque classificateur 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 :

$$y'_i = \begin{cases} 1 & \text{pour } y = y_i \\ 2 & \text{pour } y \neq y_i \end{cases} \quad (V.12)$$

Pour les données que nous avons traité, trois classificateurs binaires ont été entraînés et testés (1 contre tous, 2 contre tous, 3 contre tous). Les résultats obtenus sont représentés sur le tableau 5.6

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

Tableau 5.6 Performances des SVM avec la méthode « un contre tous » pour les différents noyaux (sans sélection)

Les symboles (perf 1-tous) (perf 2-tous) (perf 3-tous) désignent respectivement les performances des classificateurs 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. 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.

V.6 Sélection des paramètres du vecteur forme (indicateurs)

Dans la littérature 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.

Avec un grand nombre d'indicateurs, qui sont considérés dans notre problème (plus de 50 indicateurs pour les signaux de CETIM), le calcul de tous ces indicateurs peut être considéré comme un gaspillage de temps; évidemment, il est plus que préférable d'être capable de calculer tous ces indicateurs. Le problème est de déterminer quels indicateurs fournissent l'information, et les quels sont sensibles à l'apparition d'un défaut dans le système étudié.

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.

Les algorithmes ont été montrés pour être extrêmement robuste, et offrent une bonne exactitude sans exiger toute connaissance antérieure.

V.6.1 Sélection par AG

L'AG utilisé dans cette expérience est un AG simple. Les expériences sont programmées dans l'environnement MTLAB. L'AG utilise une dimension de la population de taille N_p , et commence avec une génération aléatoire des chromosomes. Chaque chromosome est de taille N_b égale au nombre d'indicateurs utilisés. C'est un compromis entre une recherche complète, généralement à travers une population grande, et accomplir le besoin désirée dans un temps du calcul raisonnable. Un modèle élitiste de la population est utilisé, c'est à dire le meilleur individu dans la population antérieure est gardé dans la nouvelle population, et ce qui empêche la décroissance de la performance du AG avec la croissance du nombre des générations.

V.6.1.1 Architecture proposée du système de sélection par AG

La figure 5.18 donne une architecture générale de notre système de sélection des primitifs (indicateurs), se système est basé sur les algorithmes génétiques comme méthode d'optimisation, et le critère de la trace tiré de l'analyse discriminante linéaire (LDA) comme fonction d'évaluation. Nous commençons avec une population aléatoire, Les individus de cette population vont ensuite être évalués sur la base d'une fonction d'adaptation (fitness), être sélectionnés, se reproduire et subir des mutations. C'est un processus itératif qui prend généralement fin lorsque on arrive à un critère d'arrêt.

La conduite des différentes étapes de l'algorithme de sélection sont décrites ci-dessous ; de l'acquisition du signal, en passant par l'évaluation, la reproduction et enfin la sélection de la solution optimale.

V.6.1.2 Conduite des différentes étapes du système de sélection

V.6.1.2.1 Préparation des donnés

Cette étape a été abordée dans la section V.5.1.1 pour les signaux de CETIM, et la section V.5.2.1 pour les signaux de WESTLAND, elle concerne la préparation des observations, en commençant par la division des ces signaux en N_t tranches, passant par le filtrage et la décimation. La dernière phase est celle de l'extraction des indicateurs. A la fin de cette étape nous obtenons une base de données totale BDT de forme (Individus\ indicateurs) et des vecteurs d'étiquettes des classes EA et ET. La base BDT est divisée en deux sous ensembles ou base distinctes. Ces données servent comme entrée au classificateur SVM; la première BA utilisé pour l'apprentissage (calcul de la fonction de décision), la deuxième BT utilisé pour tester les performances des SVM.

L'algorithme de sélection proposée tend à chercher parmi tous les indicateurs calculés, quels sont les indicateurs les plus pertinents et qui offrent une meilleure discrimination entre les classes et par conséquent une meilleure performance de classification.

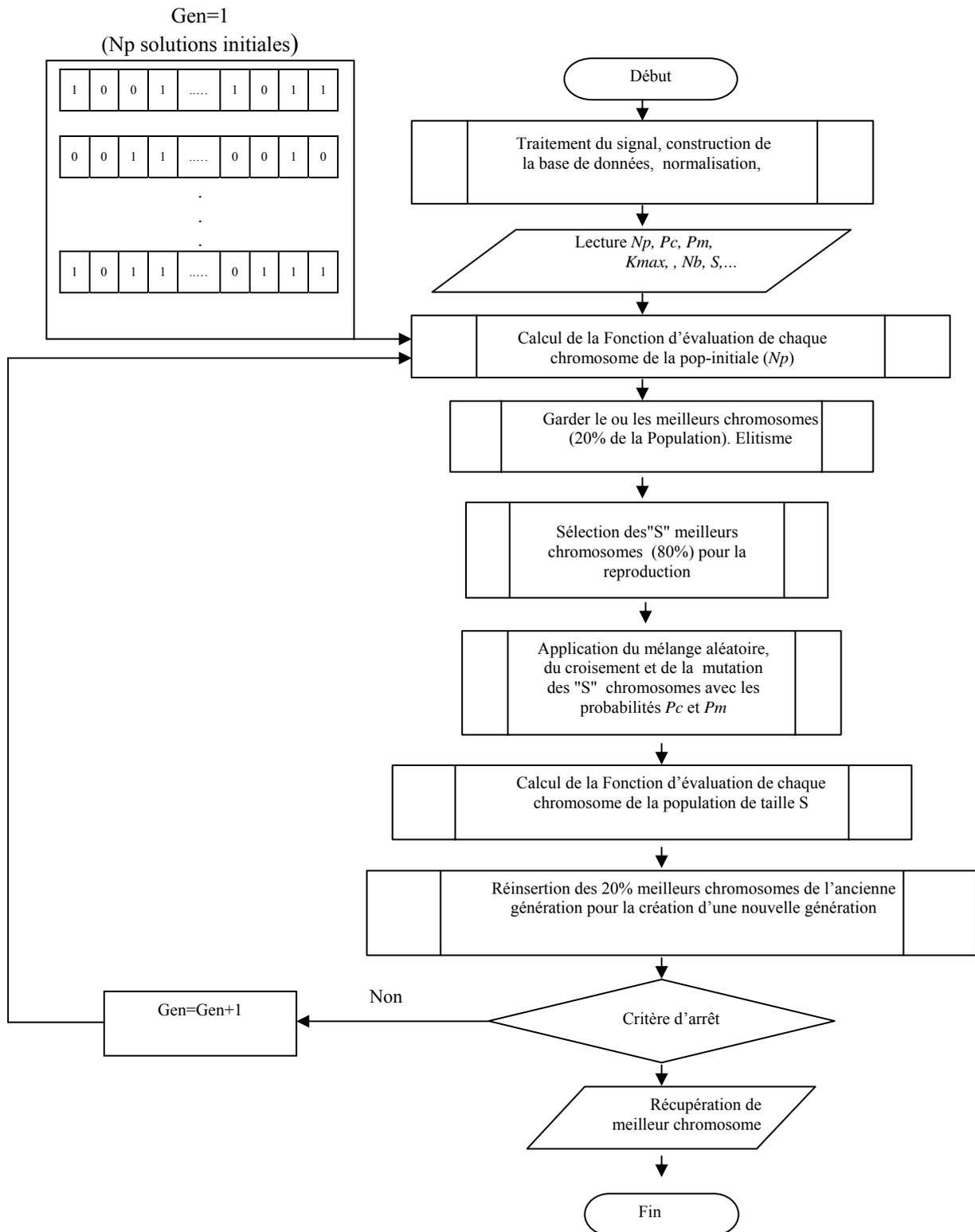


Figure 5.18 Organigramme de sélection du vecteur forme optimal par Algorithmes Génétiques simples

V.61.2.2 Codification des chromosomes de contrôle des indicateurs

Pour résoudre notre problème de sélection des indicateurs. Nous avons choisi d'utiliser le codage binaire, à cause de la simplicité et l'utilisation large de ce type de codage, où le nombre de bits dans chaque chromosome (individu) correspond au nombre d'indicateurs possibles qui peuvent être inclus dans l'ensemble de départ. Ainsi, une chaîne binaire est composée par M bits. Si le $j^{\text{ième}}$ ($i=1, \dots, M$) bit a une valeur unitaire, alors l'indicateur correspondant doit être choisi et participe à la classification sinon si le $j^{\text{ième}}$ bit a une valeur 0, alors l'indicateur correspondant doit être exclu. La figure 5.19 représente un exemple d'un chromosome codant un individu candidat de la population. Il spécifie les indicateurs à prendre en considération comme entrée du système de classification et participe à la décision, ils sont le 1er, 4ème, 5ème, ...

1	0	0	1	1	1	0	0	0	1	1
---	---	---	---	---	-------	---	---	---	---	---	---

Figure 5.19 Un exemple de chaîne bits (chromosome)

Ce type de codage a pour intérêt de permettre de créer des opérateurs de croisement et de mutation simples qui seront appliqués aux cours des différentes générations du processus itératif.

V.61.2.3 Génération de la population initiale

La population initiale est constituée d'un ensemble d'individus (chromosomes) générés aléatoirement. Cependant rien n'empêche d'utiliser des résultats et des solutions existantes pour former la population initiale. En d'autres termes, un ensemble de solutions connues préalablement peut être injecté dans la population initiale.

Le nombre d'individus d'une population ou la taille de la population constitue un paramètre important pour l'AG qu'il faudra déterminer. La représentation de la population P est :

$$P = [P_1 P_2, \dots P_i \dots P_{Np}]$$

Où P_i représente le $i^{\text{ième}}$ chromosome dans la population et Np représente le nombre de chromosomes dans la population. Tous les chromosomes dans notre cas ont été générés aléatoirement.

V.6.1.2.4 Calcul de la fonction d'évaluation (fitness)

Nous utiliserons dans ce travail une approche génétique qui optimise le choix des paramètres par la minimisation d'une fonction coût. Cette dernière est choisie d'après le critère de la Trace tiré de l'Analyse Discriminante Linéaire. La figure 5.20 représente l'organigramme de calcul de la fonction d'évaluation « trace » par LDA

Le critère de la trace est une fonction objective tiré de l'ADL. 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 paramètres de départ, un sous-ensemble de M' paramètres 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 (classification supervisée).

L'obtention du sous-ensemble discriminant consiste alors à trouver les M' paramètres (combien et lesquels) qui permet de maximiser la mesure de dispersion interclasses 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.

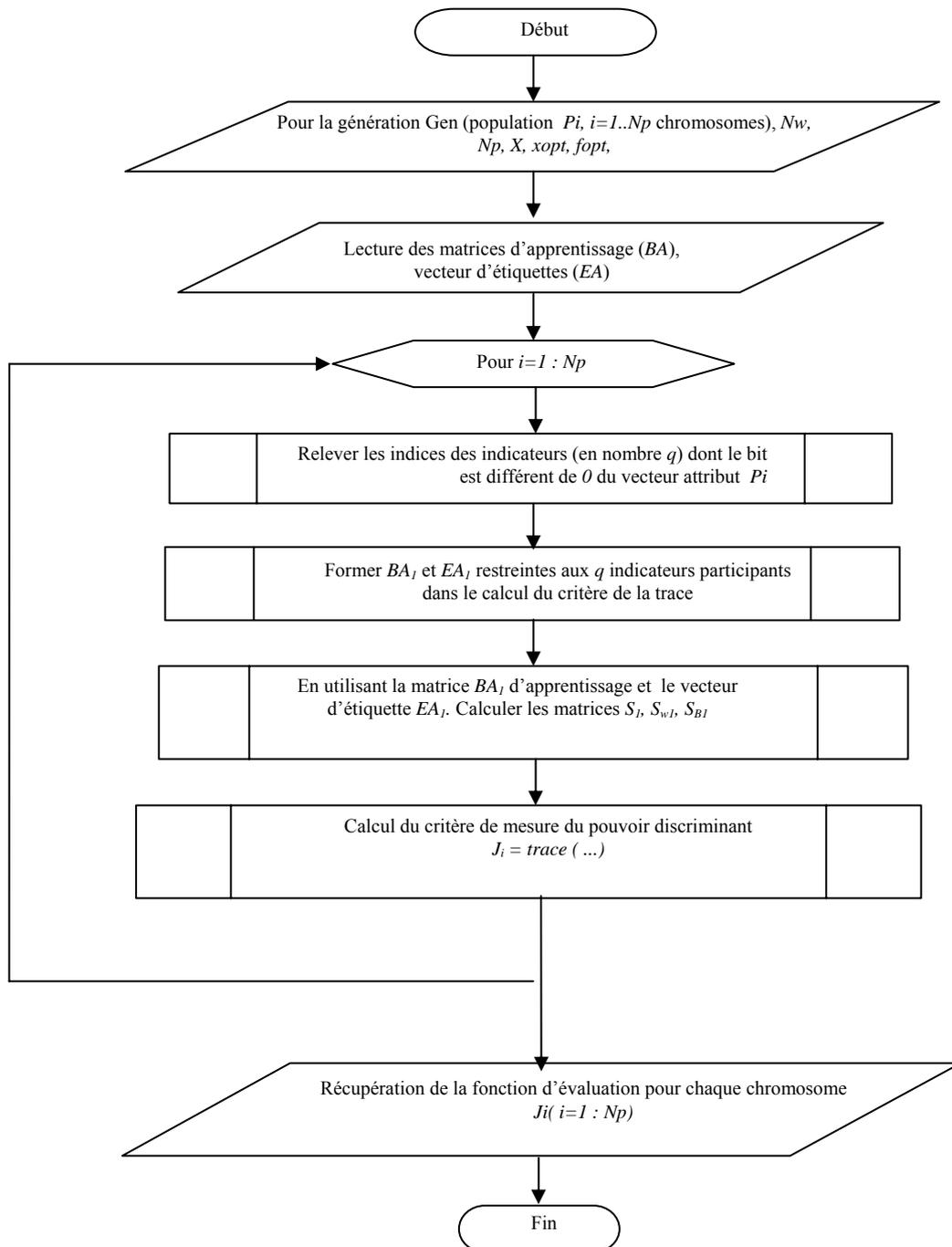


Figure 5. 20 Organigramme de Calcul de la fonction d'évaluation « trace » par ADL

V.6.1.2.5 Nombre de génération

C'est un chiffre que nous devons fixer. Il est préférable qu'il soit assez grand afin de mieux visualiser la convergence de la solution. Certains auteurs 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 utilisé un nombre de 500 générations.

V.6.1.2.6 Le critère d'arrêt

Les étapes de l'AG sont répétées durant t cycles, l'arrêt de l'algorithme est fixé d'après un *critère d'arrêt*. On peut avoir plusieurs critères d'arrêt :

- Le nombre de génération fixé initialement a été atteint.
- La valeur de la fonction d'adaptation a atteint une valeur fixée *a priori*.
- L'absence d'évolution de la valeur de la fonction d'adaptation des individus d'une population à une autre.
- Les chromosomes ont atteint un certain degré d'homogénéité.

Dans le cadre de notre travail, nous avons utilisé le premier critère avec un nombre de 500 générations

V.6.1.2.7 Sélection des individus pour la reproduction

Durant la phase de reproduction de l'AG, des individus de la population sont sélectionnés d'après la méthode de sélection choisie et sont recombinaisonnés produisant des enfants de la prochaine génération. Cette phase utilise des mécanismes de reproduction qui sont : le *croisement* (crossover) et la *mutation*.

Pour générer de nouveaux descendants (enfants), dans notre cas, des parents sont sélectionnés, en fonction de leur fonction d'évaluation, puis des mécanismes de *croisement* (crossover) et la *mutation*, sont appliquées. Pour sélectionner les meilleurs chromosomes dans une population, certaines techniques proposées par différents auteurs sont discutées dans la section IV.3.1.4 du quatrième chapitre. 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, peut-être, 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 seront pas reproduits dans la prochaine génération. C'est pour cela que le modèle (stratégie) élitiste a été également adopté.

V.6.1.2.8 Le croisement

Après avoir sélectionné les individus les mieux adaptés pour le processus de reproduction, ils subissent l'opération de croisement qui nécessite deux parents afin d'effectuer l'échange des gènes entre eux. Cet échange permet de former deux descendants possédant des caractéristiques issues des deux parents. Chaque individu se voit attribuer une même probabilité P_c de participer à un croisement. Les probabilités de croisement les plus couramment utilisées appartiennent à l'intervalle $[0.6, 0.95]$.

V.6.1.2.9 La mutation

Cet opérateur est appliqué sur chaque chromosome issu de l'opération de croisement ou appartenant à une population. L'action de l'opérateur de mutation consiste à changer ou à permuter des valeurs des gènes du chromosome C avec une probabilité P_m . Cette probabilité de mutation est assez faible en pratique. On peut observer à un moment du cycle de l'AG que des individus générés n'évoluent plus. Pour les faire évoluer, l'opération de mutation peut modifier les valeurs des gènes pour constituer des individus non similaires.

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. En outre, pour les fixer, ces probabilités de mutation et de croisement dépendent de la nature de la fonction objective du problème à résoudre. Felkaoui et al., dans [Felk.07] ont cité quelques exemples tels que :

- Pour une population de taille grande $N_p = 100$, $P_m = 0.001$ et $P_c = 0.6$
- Pour une population de taille petite $N_p = 30$, $P_m = 0.01$ et $P_c = 0.9$.

Dans notre cas nous utilisons 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 : $P_m = 1/M$, où M est la taille du chromosome en gènes.

V.7 Performances des SVM avec les indicateurs sélectionnés

La figure 5.21 représente l'organigramme de classification par les SVM utilisant le vecteur forme optimal

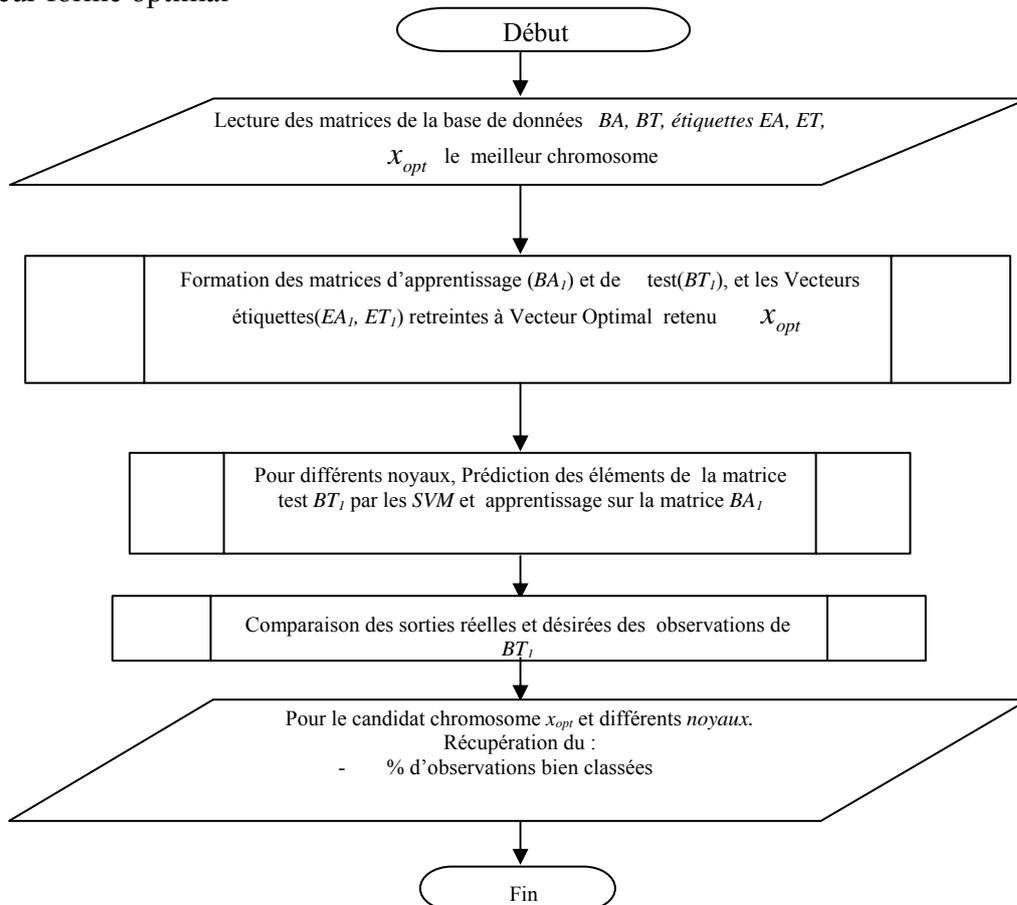


Figure 5.21 Organigramme de classification et de validation par les SVM utilisant le vecteur forme optimal

Pour évaluer les performances des SVM après sélection des indicateurs, la matrice de données présentée à l'entrée du classificateur est composée de $M' < M$ indicateurs. Où M est le nombre initial des indicateurs, et M' le nombre des indicateurs sélectionnés par le couplage AG_ADL

V.7.1 Résultat et discussions

Pour notre étude nous avons choisi pour l'algorithme de sélection proposé, les paramètres des AG suivants :

- Opérateur de sélection d'observations : sélection par troncature avec une stratégie généralisée du modèle élitisme, taille de la population : $N_p = 30$, taux de mortalité : 20%
- Nombre de génération maximale : $k_{\max} = 500$.
- Fonction d'évaluation : le critère de la trace (S_w).
- Taille du chromosome : $N_b = 51$ pour les signaux de CETIM, et 38 pour WESTLAND.
- Opérateur de croisement : 1 point, probabilité $P_c = 95\%$.
- Opérateur de mutation : probabilité $P_m = \frac{1}{N_b}$ ($\frac{1}{51}$ pour CETIM et $\frac{1}{38}$ pour Westland).
- Critère d'arrêt (atteindre k_{\max}).

V.7.1.1 Le banc d'essai de CETIM (avec sélection)

Le tableau ci-dessous représente les performances des SVM pour différents noyaux avec le vecteur forme optimal (Xo_{opt}) retenu par l'algorithme de sélection.

Noyau (Kernel)	Le vecteur optimal	Nombre des 'SV'	Performance en test (%)	mal classés	C
linéaire	2 7 11 12 13 14 15 22 25 29 30 36 37 51	12	100	-	100
polynomial	1 5 7 10 11 13 16 20 22 29 32 36	4	100	-	100
RBF	1 4 10 11 15 16 23 24 30 33 34	13	100	-	10
sigmoïdal	3 5 7 11 12 13 14 15 16 20 24	14	98.61	-	1000

Tableau 5.7 Performances des SVM pour les différents noyaux avec les indicateurs sélectionnés

Les performances des SVM sur la base de test, en utilisant le vecteur forme optimal, sont représentées par le tableau 5.7. En comparaisons avec le tableau 5.2 Les résultats obtenus montrent, une amélioration de la performance de la classification, pour les différents noyaux. Cette performance atteint 100%. (Toutes les observations sont correctement affectées aux classes correspondantes) avec un nombre des indicateurs réduit pour les noyaux linéaire, polynomial, et RBF et une performance de 98.61 avec le noyau sigmoïdal . La séparation des deux classes est, donc, réalisée avec un taux d'erreur très faible. On peut aussi remarquer que pour les différents noyaux, les 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. Les résultats expérimentaux obtenus montrent de bonnes performances de la méthode de sélection des paramètres. En effet, les vecteurs attribut optimaux proposés par l'algorithme sont suffisamment représentatifs. Pour voir les numéros des indicateurs sélectionnés dans la base de données (voir Annexe4). La figure 5.22 donne une représentation visuelle de l'hyperplan séparant les données de CETIM, après sélection des indicateurs, en utilisant un noyau sigmoïdal

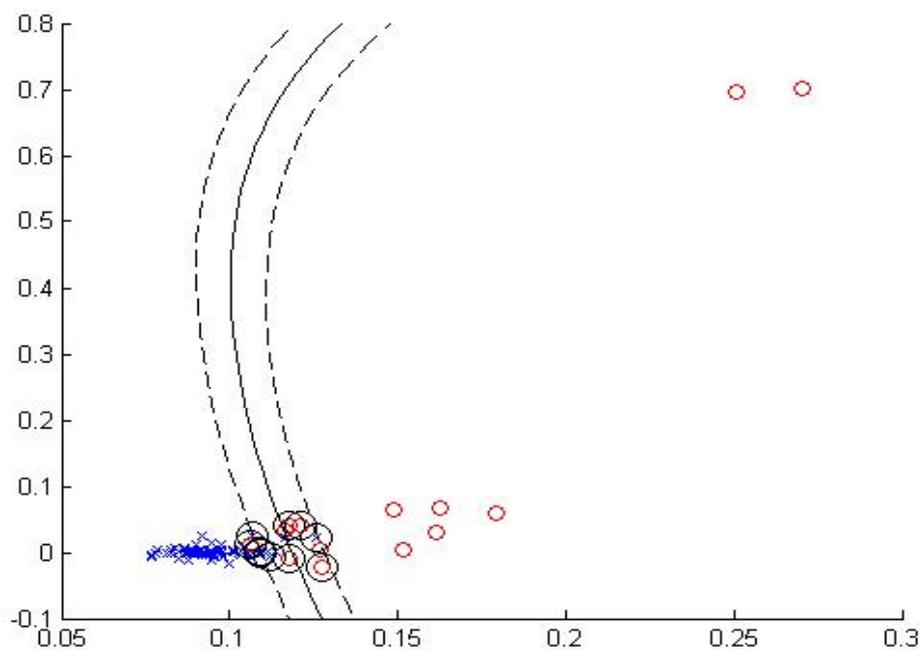


Figure 5.22 Séparation des deux classes avec un noyau sigmoïdal ($a=1$, $b=4$)

Cette figure représente l'hyperplan optimale qui sépare les deux classes (en bleu : la classe sans défaut ou de fonctionnement sain, et en rouge la classe avec défaut). On peut constater que le vecteur forme optimal retenu par l'algorithme de sélection est suffisamment représentatif, et les classes peuvent être séparés

V.7.1.2 NAVY Westland data (avec sélection)

V.7.1.2.1 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 5.8 représente les performances des SVM binaires établis pour chaque pair de classes et la performance totale en utilisant la règle de vote majoritaire

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

Tableau 5.8 Performances des SVM avec la méthode « un contre un » pour les différents noyaux, avec les indicateurs sélectionnés

En comparaison avec les performances obtenues en utilisant le vecteur forme initiale (Tableau 5.5), les résultats du tableau 5.8 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 été 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 les noyaux RBF, et polynomial.

Les résultats obtenus montrent aussi, pour le noyau sigmoïdal, qu'une diminution du nombre d'indicateurs de (de 38 à 9) n'entraîne qu'une perte « admissible » en performance de 5 % (de 95,83 % à 90,83). On peut constater également que la majorité des indicateurs sélectionnés sont les indicateurs temporels, qui ont donc, un pouvoir discriminant important. Pour voir leurs numéros dans la base de donnée (voir annexe5)

V.7.1.2.2 Un contre tous

Le tableau 5.9 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.

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

Tableau 5.9 Performances des SVM avec la méthode « un contre tous » pour les différents noyaux, avec les indicateurs sélectionnés

Avec les différents noyaux utilisés, on peut constater que la performance des classificateurs binaires est améliorée en comparaison avec le tableau 5.6. 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. Ça est toujours la cause de la génération aléatoire de la population aléatoire. On peut constater également que la majorité des indicateurs sélectionnés, sont les indicateurs temporels 6 7 8 9 15 17 23 24 et 25. (Voir annexe 5).

V.7.1.3 Les vecteurs de supports

L'examen du nombre des vecteurs de supports utilisés dans le 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, pendant que peu de vecteurs, indique que l'hyperplan est relativement facile à identifier, et par conséquent, la performance sera relativement bonne. Prenons comme exemple pour les données de Westland, les résultats des tableaux 5.5 et 5.7, on peut constater que le nombre des vecteurs de supports varie en fonction des noyaux choisis. Avec le noyau sigmoïdal, la fonction de décision est établi avec 22 vecteurs de supports, et 26 en utilisant un noyau RBF, ce qui explique les performances obtenues avec ces deux noyaux en comparaison avec le noyaux linéaire et polynomial. On peut constater également que le noyau polynomial nécessite un nombre élevé de vecteurs de support pour établir la fonction de décision. En comparant les résultats du tableaux 5.5 (sans sélection) et le tableau 5.8 (après sélection du vecteur forme) On peut constater qu'après sélection des

indicateurs, le nombre des vecteurs de 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 des vecteurs de supports pour déterminer l'hyperplan optimal est faible.

V.9 Conclusion

Dans ce chapitre 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.

Dans la première étape nous avons testé les performances des SVM sur des signaux aléatoires, et des signaux d'engrenage simulés, en utilisant les coefficients du modèle AR comme indicateurs. Dans cette première étude les résultats obtenus sont encourageantes. Ensuite notre étude a été menée sur des signaux d'engrenage réels ; les premiers signaux étudiés sont ceux de CETIM, dans ce cas nous avons étudié les performances des SVM binaires (séparation entre deux classes), les performances des SVM multiclassés ont été également étudiés en utilisant les signaux des de 'Westland'. Nous avons obtenu de bons résultats avec les deux approches utilisés (un contre un, et un contre tous).

La deuxième partie de ce chapitre a été consacrée à l'étude des performances des SVM, après sélection des indicateurs du vecteur forme optimal, pour les deux ensembles des signaux réels. Les résultats obtenues montrent une amélioration des performances dans la plus part des cas et une diminution admissible dans certain cas mais avec un nombre d'indicateurs très réduit.

CONCLUSION GENERALE ET PERSPECTIVES

Pour aborder l'étude, nous avons établi un état de l'art sur les causes de défaillances dans les machines tournantes, et les méthodes d'analyse des signaux vibratoires permettant d'identifier les défauts. Ce travail préliminaire a mis l'approche d'analyse des signaux mesurés sur les bancs d'essais en cours de fonctionnement. Les analyses sont réalisées, dans le domaine temporel et fréquentiel.

Dans ce travail nous avons limité notre étude sur les défauts d'engrenages qui sont la cause principale de défaillance dans la plus part des machines tournantes. Les informations issues de ces analyses permettent souvent une identification efficace des défaillances. Toutefois ces approches trouvent leurs limites en présence d'importants bruits de mesure qui noient l'information pertinente.

Dans la deuxième partie nous avons examiné les performances des SVM. Là, nous avons également évalué l'utilité de cette technique dans le diagnostic des défauts d'engrenages sur la base signaux vibratoire simulés, et des données réels disponibles.

Les résultats obtenus montrent que les SVM ont de bonnes performances en classification des signaux simulé et réels pour les deux cas étudiés (classification binaire, multi classes) et avec différents noyaux.

L'utilisation de la méthode de validation croisée nous a permis de bien choisir les paramètres des noyaux et par conséquent, nous avons obtenus de bonnes performances des SVM.

Dans La dernière partie nous avons mis au point un système de sélection des indicateurs. Ce système a été réalisé par un couplage des algorithmes génétiques (AG) et l'analyse discriminante linéaire (ADL). Ce système de sélection nous a permis d'avoir un ensemble d'indicateurs suffisamment représentatif. Dans ce cas nous avons montré que la sélection est capable de réduire la taille du vecteur forme, et d'augmenter les performances des SVM dans la plus part des cas.

Les travaux réalisés dans le cadre de ce mémoire pourraient donner lieu à des études supplémentaires sur plusieurs points :

- Calculer d'autres paramètres avec des méthodes de traitement de signal plus performantes, pour améliorer la pertinence du vecteur forme.
- Améliorer les méthodes utilisées en exploitant les rejets de distance pour rendre le système de décision totalement adaptatif, c'est à dire autoriser un nouvel apprentissage du système de décision dès que de nouveaux événements ont modifié la vie du processus.
- Utiliser d'autres méthodes pour les SVM multiclsses telle que DAGSVM (Directed Acyclic Graph Support Vector Machines)

Annexe1

La Boite à outils « Statistical Pattern Recognition Toolbox for Matlab (STPRTool) » :

Example1 :Classification binaire (deux classes) :

```
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
cerror(ypred,tst.y)                 % calcul de l'erreur de
                                     classification
```

Example 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
```

Example 3 :Multiclass (un contre tous)

```
data = load('pentagon');
options = struct('ker','rbf','arg',1,'C',10,'verb',1);
model = oaasvm(data,options);        %fonction de décision 'un contre tous'
figure;
ppatterns(data); ppatterns( model.sv.X);
pboundary( model );
```

Annexe2

Moteur de l'hélicoptère et position de ces différents composants

La photo ci-dessous présente le moteur de l'hélicoptère instrumenté. Elle nous permet de repérer ses différents composants et notamment le système étudié : la boîte de vitesse CH46



Westland Research Paper RP907

**Final Report on CH-46
Aft Transmission Seeded
Fault Testing**

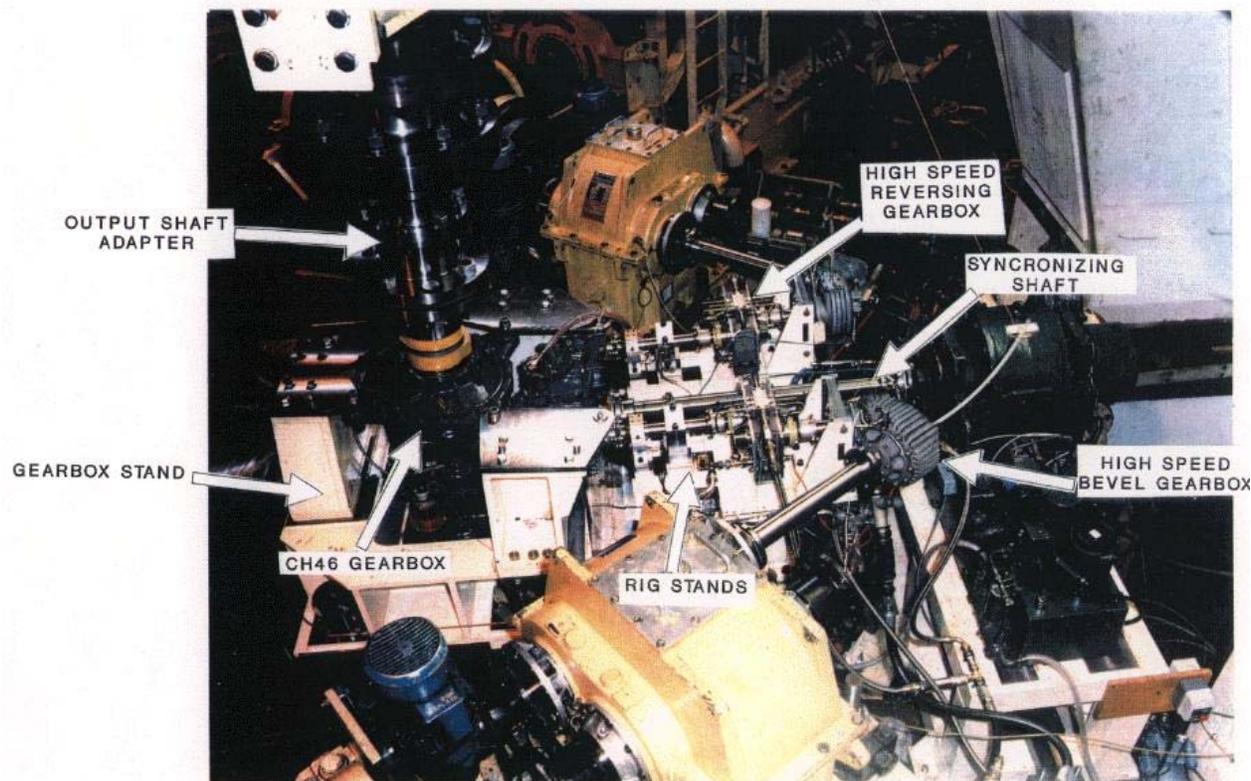


Figure 1. *Moteur de l'hélicoptère et position de ces différents composants*

Annexe3

L'engrenage conique (défaut #4)

Cette annexe présente trois photos de l'engrenage conique (défaut #4) et un schéma de son emplacement dans le système. Elles nous permettent de voir la pièce saine ainsi que les deux niveaux d'écaillage mis à notre disposition. La première figure nous permettra notamment de repérer l'endroit où l'écaillage sera provoqué



Figure1. *Vue d'ensemble de l'engrenage conique sain.*



Figure 2. *Généralisation de l'écaillage.*



Figure 3. *Écaillage généralisé.*

Annexe 4 : Classement des indicateurs dans la base de données CETIM

Bandes de calcul	Les indicateurs temporels																										
	Bande [250-417Hz]									Bande [583-750 Hz]									Bande[916-1083 Hz].								
indicateur	VE	FC	SK	K	M5	M6	M7	M8	M9	VE	FC	SK	K	M5	M6	M7	M8	M9	VE	FC	SK	K	M5	M6	M7	M8	M9
Numéro dans la base de données	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

indicateur	Coefficients du modèle AR																		
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19
Numéro dans la base de données	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46

Avec :

VE : valeur efficace

FC : facteur de crête

SK : skewness

K : Kurtosis

Mi : moment d'ordre i

Ai : les coefficients du modèle AR

Ei : l'enregie dans la i^{ème} bande

Bandes de calcul	Les indicateurs fréquentiels				
	Bande [250-417 Hz]	Bande [583-750 Hz]	Bande [916-1083Hz],	Bande [1250--1417Hz]	Bande [250-1417Hz]
indicateur	E1	E2	E3	E4	E5
Numéro dans la base de données	47	48	49	50	51

Annexe 5 : Classement des indicateurs dans la base de données Westland

Bandes de calcul	Les indicateurs temporels																										
	Bande [895-1321 Hz]									Bande [2004-2430 Hz]									Bande [3113-3539 Hz]								
indicateur	VE	FC	SK	K	M5	M6	M7	M8	M9	VE	FC	SK	K	M5	M6	M7	M8	M9	VE	FC	SK	K	M5	M6	M7	M8	M9
Numéro dans la base de données	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

Bandes de calcul	Coefficients du modèle AR					
	Bande [895-1321 Hz]		Bande [2004-2430 Hz]		Bande [3113-3539 Hz]	
Indicateur	a1	a2	a1	a2	a1	a2
Numéro dans la base de données	28	29	30	31	32	33

Avec :

- VE : valeur efficace
- FC : facteur de crête
- SK : Skewness
- K : Kurtosis

- Mi : moment d'ordre i
- ai : les coefficients du modèle AR
- Ei : l'enregie dans la i^{ème} bande

Bandes de calcul	Les indicateurs fréquentiels				
	Bande [895-1321 Hz]	Bande [2004-2430 Hz]	Bande [3113-3539 Hz]	Bande [4222,4648Hz]	Bande [895-4648Hz]
indicateur	E1	E2	E3	E4	E5
Numéro dans la base de données	34	35	36	37	38

REFERENCES BIBLIOGRAPHIQUES

[Afno.01] X 60-319 norme européenne ICS : Terminologie de la maintenance. AFNOR , Juin 2001.

[Alatt.94] Alattas Mohamed « *Maintenance des machines tournantes : signature de défauts d'engrenages droit et hélicoïdaux* », Thèse de doctorat , INSA de Lyon, 1994.

[Andr.05] Andrew K. Jardine .S, Daming Lin, & Dragan Banjevic « *A review on machinery diagnostics and prognostics implementing condition-based maintenance*», Mechanical Systems and Signal Processing, 22 September 2005.

[Apar.02] Aparna Dode , Mallikarjun Rao «*Pattern Recognition of Acoustic Emission Signals from PZT ceramics* », Journal of Nondestructive Testing Vol. 7 N 09 .2002

[Ayat.04] Ayat Nedjemeddine «*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.

[Bela.00] Belacel Nabil « *Méthodes de Classification Multicritère Méthodologie et Applications à l'Aide au diagnostic Médical*», Thèse de doctorat, Institut de statistique et de recherche opérationnelle, 2000.

[Bena.02] Benahmed .N « *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.

[Beno.99] Benouaghran Ramzi, « *Contribution au diagnostic des machines tournantes complexes. Application à un laminoir* », Thèse de doctorat, UTC de Compiègne, Avril 1999.

[Bess.05] Bessel. P « *Data mining II. Modélisation Statistique & Apprentissage* » publications du laboratoire de statistique et probabilités, Université Paul Sabatier, Toulouse, Mars 2005.

[Bonn.05] Bonnardot Frédéric « *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.

[Boui.00] Bouillaut Laurent « *Approches cyclostationnaire et non-linéaire pour l'analyse vibratoire de machines tournantes : Aspect théoriques et applications au diagnostic* », Thèse de doctorat,UTC de Compiègne, Novembre 2000.

[Burg.98] Burges C.J. «*A tutorial on support vector machines for pattern recognition*», Data Mining and Knowledge Discovery 2 (2) (1998) 121–167.

[Call.03] Callut.J « *Implémentation efficace des Support Vector Machines pour la classification* », Mémoire présentée en vue de l'obtention du grade de Maître en Informatique université libre de Bruxelles.2003.

Bibliographie

[Capd.91] Capdessus. C, Sidahmed M «*Analyse des vibrations d'un engrenage : cepstre, corrélation, spectre*», Journal de Traitement de signal, Volume 8, N° 5, 1991, pp.365-372.

[Capd.92] Capdessus. C «*Aide au diagnostic des machines tournantes par traitement du signal* », Thèse de doctorat , Institut National Polytechnique de Grenoble, 1992.

[Capd.95] 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.

[Casi.03] Casimir. R «*Diagnostic des défauts des machines asynchrones par reconnaissance des formes*», Thèse de doctorat , Ecole Centrale de Lyon, 2003.

[Corn.02] Cornuéjol.A, Miclet.L «*Apprentissage artificiel ,concept et algorithmes*», Edition OYROLLES,2002.

[Cous.03] Cousinard O, Marconnet. P «*Détection de l'endommagement d'un engrenage par l'emploi de l'analyse cepstrale et de la détection d'enveloppe : Application et validation industrielles pour le diagnostic d'un réducteur fonctionnant à faible vitesse* », 16ème Congrès Français de Mécanique, septembre 2003.

[Davi.91] Davis L.Ed «*Hundbook of genetic algorithms*», New York : Van Nostrand Reinhold,1991.

[Dubu.02] Dubuisson.B «*diagnostic, intelligence artificielle, et reconnaissance des formes* » Paris, Hermès science publications, 2001.

[Duda.01] Duda R. O, Hart P. E , & Stork D. G. «*Pattern Classification*», (2nd edition). New York: Wiley-Interscience, 2001.

[Dura.04] Durand. N «*Algorithmes génétiques et autres outils d'optimisation appliqués à la gestion de trafic aérien*», Thèse de doctorat université de Bruxelles. 2004.

[Elba.99] El badaoui Mohamed, «*Contribution au diagnostic vibratoire des réducteurs complexes à engrenages par l'analyse cepstrale* », Thèse de l'Université Jean Monnet, 9 Juillet 1999.

[Feda.06] Fedala .S «*Le diagnostic vibratoire automatisé : comparaison des méthodes d'extraction et de sélection du vecteur forme*», Thèse de Magister IOMP, Université Ferhat Abbas Sétif Algérie, Novembre 2006.

[Felk.94] Felkaoui .A, Fortas B, Apostoliouk. A «*Sur la sélection de l'ordre dans l'analyse spectrale moderne des processus linéaires*», ICSS'94 (1994).

[Felk.07] Felkaoui.A, Bekka.R.E.H, Antoni. J, Ménad .S «*Etude des performances de la sélection d'indicateurs par couplage algorithmes génétiques – analyse discriminante linéaire*» Soumis au journal européen des systèmes automatisés, janvier 2007.

[Khar.02] Kharoubi Jamel « *Etude de Techniques de Classement - Machines à Vecteurs Supports - pour la Vérification Automatique du Locuteur* » Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 2002.

[Jack.01] Jack.L.B, Nandi A.K « *Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms* », Mechanical systems and signal processing (2002)16(2-3),373-390.

[Joac.98] Joachims. T « Text categorization with support vector machines : learning with many relevant features ». 10th European Conference on Machine Learning ECML-98, pages 137-142, 1998.

[Juns.04] Junsheng. C, Dejie Yu, & Yang Yu « A fault diagnosis approach for roller bearings based on EMD method and AR model ». Mechanical Systems and Signal Processing 20 (2006) 350-362

[Lind.00] Lindsay Burnett Jack « Applications of artificial intelligence in machine condition monitoring » Thèse de doctorat, Université de Liverpool, 2000.

[Lout.06] Loutridis S.J. « *Instantaneous energy density as a feature for gear fault detection* », Mechanical Systems and Signal Processing 20 (2006) 1239-1253.

[Lure03] Lurette.C « *Développement d'une technique neuronale auto-adaptative pour la classification dynamique de données évolutives. application a la supervision d'une presse hydraulique.* » thèse de doctorat , Université des Sciences et Technologies de LILLE.2003

[Math.05] Mathieu Glade « *Modélisation des coûts de cycle de vie : prévision des coûts de maintenance et de la fiabilité Application à l'aéronautique* », thèse de doctorat , école centrale de Lyon 2005.

[Mich.99] Michalewicz. Z « *Genetic Algorithms + Data Structures = Evolution Programs* ». Springer Verlag, collection *Artificial Intelligence*, 3rd Edition, New York. 1999.

[More.95-1] Morel Jacques « *Surveillance vibratoire et maintenance prédictive* », Techniques de l'Ingénieur R 6100, pp2.20.1995.

[More.95-2] Morel Jacques « *Vibration des machines et diagnostic de leur état mécanique* », Edition Eyrolles, Paris 1995.

[Nand.99] Nandi Subhasis, Toliyat Hamid « *Condition monitoring and diagnosis of electrical machines- a review* », IEEE 1999.

[Oukh.97] Okhelou Latifa « *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, 4 Juillet 1997, pp 211.

[Peng.04] Peng Z.K., Chu .F.L « *Application of the wavelet transform in machine condition monitoring and fault diagnostics* » Mechanical Systems and Signal Processing 18 (2004) 199-221

[Platt.98] Platt.J.C «*Sequential Minimal Optimization:A Fast Algorithm for Training Support Vector Machines*» Technical Report MSR-TR-98-14 April 21, 1998.

[Platt.99] Platt.J.C «*Probabilistic outputs for support vector machines and comparisons to Regularized Likelihood Method*» Advances in Large Margin classifiers, MITPress, 1999.

[Raad.03] Amani Raad « *Contributions aux statistiques cycliques d'ordre supérieur : applications au diagnostic des défauts d'engrenage* », Thèse de doctorat, Université Technologie de Compiègne, 2003.

[Rafi.06] Rafiee J. Arvani. F, Harifi A, Sadeghi M.H «*Intelligent condition monitoring of a gearbox using artificialneural network*» Mechanical Systems and Signal Processing, 4 Aout 2006

[Robe.95] Dictionnaire de la langue française, nouveau petit robert. Mars 1995.

[Roja.05] A. Rojas, Nandi A. K «*Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines*», Mechanical Systems and Signal Processing, 10 May 2005.

[Sama.03] Samanta B.and Al-Balushi K.R «*Artificial neural network based fault diagnostics of rolling element bearings using time-domain features*», Mechanical Systems and Signal Processing (2003) 17(2), 317–328.

[Sama.04-1] Samanta B «*Artificial neural networks and genetic algorithms for gear fault detection*», Mechanical Systems and Signal Processing 18 (2004) 1273–1282.

[Sama.04-2] Samanta.B «*gear fault detection using artificial neural networks and support vector machines with genetic algorithms*», Mechanical systems and signal processing,18(2004),625-644.

[Vapn.98] Vapnik .V « *Statistical Learning Theory*»: wiley Interscience publication, NewYork, 1998.

[Vapn.00] Vapnik .V « *TheNature of Statistical Learning Theory*», 2nd edition. springer, 2000.

[Vrab.03] Vrabie Valeriu, « *Statistiques d'ordre supérieur : applications en géophysique et électrotechnique* »,Thèse de doctorat Université POLITEHNICA de Bucarest 2003.

[Vojt.04] Vojtěch F and V Hlaváček «*Statistical Pattern Recognition Toolbox for Matlab*», CZECH TECHNICAL UNIVERSITY; juin 2004.

[Wong.05]. Wong M.L.D, Jack,and L.B, Nandi A.K. « *Modified self-organising map for automated novelty detection applied to vibration signal monitoring* », Mechanical Systems and Signal Processing 24 January 2005.

[Xian.06] Xianfeng Fan, Ming J. Zuo «*Gearbox fault detection using Hilbert and wavelet packet transform*», Mechanical Systems and Signal Processing 20 (2006) 966–982.

[Yann.97] 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.

[Yang.05-1] Yang.B.S,Hiang.w.w «*Condition classification of small reciprocating compressor for refrigerators using artificial neural network and support vector machines*»,Mechanical systems and signal processing,19 (2005),371-390.

[Yang.05-2] Yang.B.S,Hiang.w.w « *Cavitation detection of butterfly valve using support vector machines*»,Journal of Sound and Vibration 287 (2005) 25–43.

[Ypma.01] Ypma Alexander « *Learning methods for machine vibration analysis and health monitoring* », Thèse de l'université technique de Delft -pays bas, 12 Novembre 2001, pp221.

[Zemo.03] Zemouri Mohamed Ryad «*Contribution à la surveillance des systèmes de production à l'aide des réseaux de neurones dynamiques :Application à la e-maintenance* » Thèse de doctorat ,Université de Franche-Comté.2003.

[Zhan.05] L. Zhang, L. B. Jack, A. K. Nandi «*Fault detection using genetic programming*», Mechanical Systems and Signal Processing 19 (2005) 271–289.

[Zhou.06] ZHOU. R «*Feature extraction of musical content for automatic music transcription*», These de doctorat , Chinese Academy of Science, Beijing, Chine.2006.

[Zwin.95] Zwingelstein G., « *Diagnostic des défaillances. Théorie et pratique pour les systèmes industriels*», Traité des Nouvelles Technologies - Série Diagnostic et Maintenance. Editions Hermès, Paris, 1995.

الملخص

الحوامل ذات الهامش الواسع هي طريقة جديدة للتلقين، اقترحها "فابنيك" في 1995. الفكرة الأساسية تنص على نقل المعطيات من فضاء الدخول، و التي تنتمي إلى قسمين مختلفين ولا يمكن فصلها خطيا، إلى فضاء الخصائص بحيث يصبح من الممكن فصلها خطيا. في هذا الفضاء، تستعمل تقنية بناء المستوي الفاصل الأمثل لحساب دالة التقرير التي تفصل القسمين. في هذه الأطروحة درسنا الحوامل ذات الهامش الواسع كطريقة تصنيف من أجل التشخيص الآلي للألات الدوارة. الهدف الأخر من هذه الدراسة يختص بمرحلة التحليل والتي يتم فيها استخراج المؤشرات و اختيار أحسنها من حيث القدرة على الفصل بين القسمين، و في هذا الصدد استعملنا مزوجة بين الخوارزميات الجينية و التحليل الفاصل الخطي.

كلمات المفاتيح

الاهتزازات الميكانيكية، الألات الدوارة، التشخيص الآلي، التصنيف، معالجة الإشارة، تحليل المعلومات

Résumé

Les supports à vaste marge (SVM) est une nouvelle technique d'apprentissage statistique proposée par V. Vapnik en 1995. L'idée essentielle des SVM consiste à projeter les données de l'espace d'entrée (appartenant à deux classes différentes) non linéairement séparables dans un espace de plus grande dimension appelé espace de caractéristiques de façon à ce que les données deviennent linéairement séparables. Dans cet espace, la technique de construction de l'hyperplan optimal est utilisée pour calculer la fonction de décision séparant les deux classes.

Dans ce travail de thèse, nous avons étudié les SVM comme technique de classification pour le diagnostic vibratoire des machines tournantes. Un autre intérêt de cette étude sera accordé à la phase d'analyse qui représente l'extraction et la sélection du vecteur forme. Dans ce concept, un couplage des algorithmes génétiques (AG) et l'analyse discriminante linéaire (ADL) est utilisé pour la sélection de paramètres par ordre de pertinence au sens de discrimination entre les classes.

Mots clés : SVM, vibrations mécaniques, machines tournantes, diagnostic automatisé, classification, traitement de signal, traitement de données

Abstract

Support Vector Machines (SVM) is a new technique of statistical learning, proposed by V. Vapnik in 1995. The main idea of SVM consists in projecting data of the entrance space (belonging to two different classes) non-linearly separable in a space of bigger dimension called space of feature so that data become linearly separable. In this space, the technique of construction of the optimal hyperplan is used to calculate the function of decision separating the two classes.

In this work of thesis, we studied the SVMs as technical of classification for the vibratory rotating machine diagnosis. Another interest of this study will be granted to the phase of analysis that represents the extraction and the selection of the vector form. In this concept, a coupling of genetic algorithm (AG) and the linear discriminant analysis (LDA) will be used for the selection of parameters by order of relevance to the sense of discrimination between classes.

Key words: SVM, mechanical vibrations, rotating machines, automated diagnosis, classification, signal processing, treatment of data