

III.1 Introduction

Avec l'apparition des ordinateurs l'homme a découvert un moyen d'effectuer diverses tâches avec deux capacités non négligeables que lui ne possède pas : la rapidité et la précision.

Cependant l'exécution d'une tâche pour l'ordinateur nécessite sa programmation préalable par l'homme. Cette caractéristique fait apparaître les ordinateurs comme des machines exécutant des ordres « aveuglement » et l'homme n'a pas désespéré de voir un jour construire une machine à son image, c'est à dire intelligente, capable d'apprendre, de raisonner, de réfléchir sans son intervention.

Ce sont des recherches basées sur le fonctionnement du cerveau qui ont constituées le point de départ de cette gigantesque recherche.

Des travaux de neurobiologistes ont, en effet, révélé que le cerveau est constitué d'un nombre extrêmement élevé d'unités de traitement élémentaire de l'information (les neurones biologiques) fortement interconnectées. Ces recherches sont l'origine de l'inspiration des réseaux de neurones artificiels[20].

La commande, en temps réel, des onduleurs multi-niveaux par la stratégie d'élimination d'harmoniques, exige d'une part, la détermination de tous les angles d'amorçage des interrupteurs, et d'autre part, le stockage de toutes les valeurs obtenues. En effet, cette opération nécessite une très grande capacité mémoire, du fait que le calcul est effectué dans certain cas avec un pas de 10^{-3} concernant le taux de modulation 'm', aussi la réalisation de cette mémoire devient coûteuse[4].

Pour éviter l'usage des mémoires électroniques, une autre solution est préconisée consiste à utiliser la technique des réseaux de neurones artificiels. Ces derniers sont largement appliqués dans les domaines de la classification, mémorisation, filtrage et approximation[4].

Dans ce chapitre, nous allons surtout exploiter leur propriété d'approximation universelle. L'objectif est de reproduire toutes les caractéristiques d'allumage non linéaires, déjà calculées dans le cas étudiés pour les onduleurs à sept niveaux. Pour cela, sont mis en œuvre des réseaux multicouches entraînés par l'algorithme d'apprentissages pour assurant l'approximation de la caractéristique optimisée de l'onduleur à sept niveaux.

III-2 Définition

Les réseaux de neurones formels sont des structures (le plus souvent simulées par des algorithmes exécutés sur des ordinateurs, parfois sur des machines ou même des circuits spécialisés) qui prennent leur inspiration dans le fonctionnement élémentaire des systèmes nerveux, il sont utilisés essentiellement pour résoudre des problèmes de classification, de reconnaissance de formes, d'association, d'extraction de caractéristique et d'identification [20].

III-3 Historique

Le concept des réseaux de neurones n'est pas nouveau, l'idée était de concevoir un système ayant comme tâche de modéliser la bio-physiologie du cerveau. Cette modélisation tente d'expliquer comment le cerveau opère et fonctionne [21].

Le but de la recherche sur les réseaux de neurones n'est pas de créer des machines qui traitent l'information plus rapidement que les calculateurs traditionnels, mais c'est de créer des machines qui se montrent supérieures dans les domaines où le cerveau humain dépasse ces calculateurs [22].

En 1943, **Culloch et Pitts** adoptèrent les affirmations de **James** (1890) et formalisèrent une description du neurone qui est l'élément fondamental de tous les réseaux.

En 1949, **Hebb** introduit la notion de " plasticité synaptique ", c'est-à-dire, le mécanisme de modification progressive des couplages entre neurones responsables de changements permanents de leurs propriétés collectives, ce qu'on appelle " **apprentissage**".

En 1969, **Minsky et Papert** démontrèrent, dans leur ouvrage intitulé " **Perceptrons** ", un certain nombre de théorèmes sur les limitations d'un réseau mono-couche et conclurent que ces limitations se généralisent pour les réseaux multicouches. Ceci a poussé de nombreux chercheurs à abandonner cette voie pour se diriger vers l'intelligence artificielle qui semblait un domaine plus prometteur.

Dans le début des années 80, les travaux, de **Rosenblatt** sur le perceptron et ceux de **Widrow et Hoff** sur les algorithmes adaptatifs sont s'avèrent les plus importants [22].

III.4 Le Modèle Neurophysiologique

L'élément de base du système nerveux central est le neurone. Le cerveau se compose d'environ mille milliards de neurones, avec 1000 à 10000 synapses (connexions) par neurone. Le neurone est une cellule composée d'un corps cellulaire et d'un noyau (Figure

(III.1)). Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique.

C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone). L'information est traitée alors par le corps cellulaire. Les signaux transmis par les neurones cheminent des fibres nerveuses appelées Axones. La transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angströms entre l'axone du neurone afférent et les dendrites du neurone efférent. La jonction entre deux neurones est appelée synapse[20].

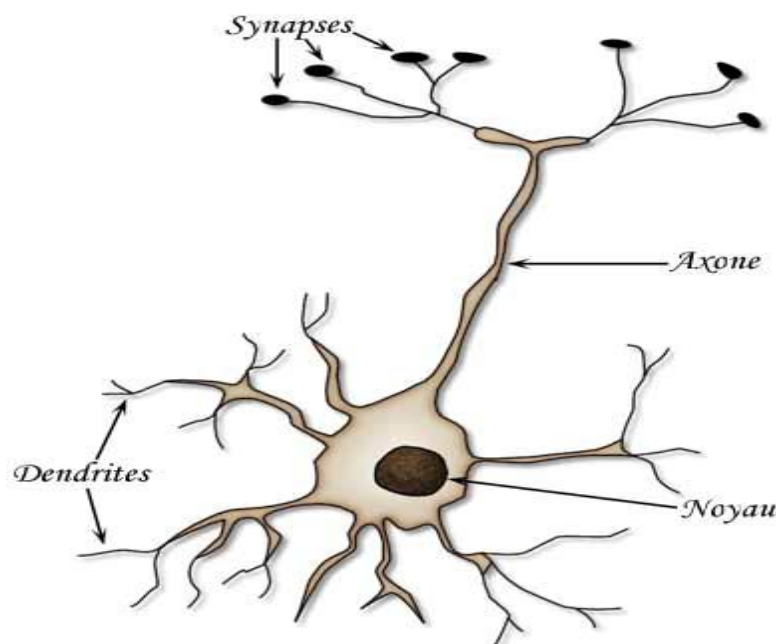


Fig (III.1) Les éléments constituant le neurone biologique

Chaque neurone reçoit des signaux excitateurs ou inhibiteurs par ses dendrites. Ces signaux pondérés sont combinés dans le corps cellulaire. Le potentiel résultant est comparé au seuil interne, s'il est supérieur à celui-ci, le neurone provoque l'émission dans l'axone d'un train d'impulsions vers les synapses, dans le cas contraire il reste inactif.

III.5 Le Modèle Mathématique (Neurone Artificiel)

Un neurone artificiel est une modélisation simplifiée du neurone biologique, c'est un processeur élémentaire qui reçoit un nombre d'entrées en provenance des autres neurones ou de l'extérieur. A chacune de ces entrées est associée un poids W abréviation de *weight* (poids en anglais) représentatif de la force (ou bien de la pondération) de la connexion.

La figure (III.2) montre le modèle du neurone artificiel.

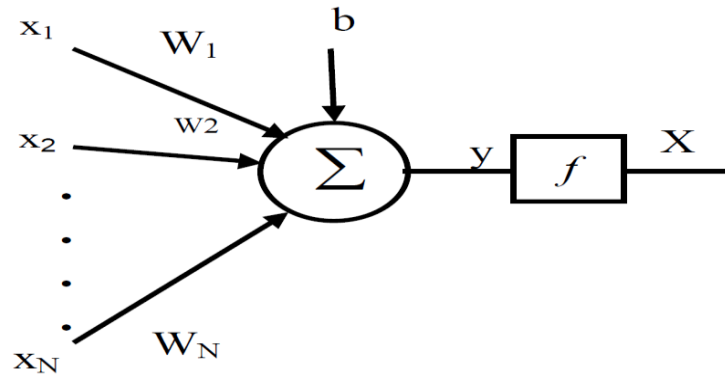


Fig (III.2) Modèle du neurone artificiel

La sortie du neurone est une somme pondérée de ses entrées plus un biais, multipliée par une fonction d'activation qui peut être linéaire ou non linéaire (identité, sigmoïde, tangente hyperbolique, etc.) [22]. Une fonction d'activation linéaire aboutira à une structure linéaire, qui défait le but d'une mise en œuvre des réseaux de neurones.

$$y = \sum_{i=0}^n W_j x_j + b \quad / \quad X = f(y) \quad (\text{III. 1})$$

x_j est la valeur de la j^{eme} entrée du neurone, W_j est le poids synaptique correspondant au neurone j , b désigne le biais et $f(.)$ la fonction d'activation.

III.6 Fonctions D'activation ou De Seuillage

Elle établit une correspondance entre les valeurs présentées en entrée du neurone et la valeur de sortie, elle est présentée généralement par une non linéarité[20].

<i>Catégorie</i>	<i>Type</i>	<i>Equation</i>	<i>Allure</i>
Seuil	Heaviside	$h(x)=1 \text{ si } x>0$ $hx)=0 \text{ si } x \leq 0$	

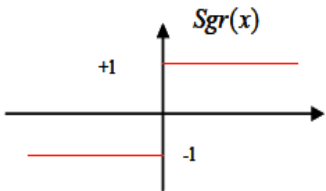
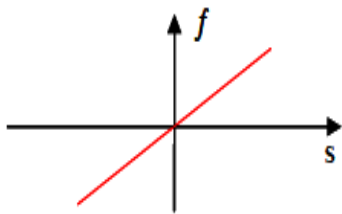
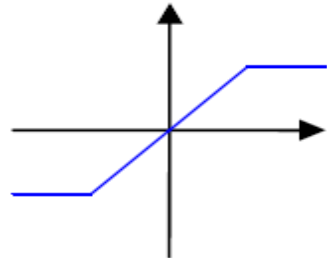

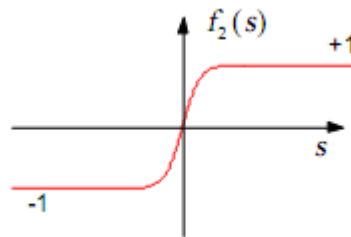
	Signe	$sgr(x)=1 \text{ si } x>0$ $sgr(x)=-1 \text{ si } x\leq 0$	
Linéaire	Identité	$f(x)=x$	
	Saturation	$f(k,x)=1 \text{ si } x<-1/k$ $f(k,x)=-1 \text{ si } x>=1/k$ $f(k,x)=x \text{ sinon}$	
Sigmoïde	Positive (logistique)	$f_1(s) = \frac{1}{1 + e^{-s}}$	
	Symétrique (type tanh)	$f_2(s) = \frac{1 - e^{-s}}{1 + e^{-s}}$	

Tableau (III.1) les fonctions d'activation les plus utilisées

La fonction la plus utilisée est de type sigmoïde [23]. Elle est définie par :

$$f(y) = \frac{1 - e^{-\delta \cdot y}}{1 + e^{-\delta \cdot y}} \quad (\text{III. 2})$$

Où δ dénote le paramètre de la sigmoïde qui définit le degré de la non linéarité, $f(y) \in [-1, 1]; \forall y \in \mathbb{R}$.

La représentation de la fonction sigmoïde pour différentes valeurs de δ est donnée par la figure (III.3).

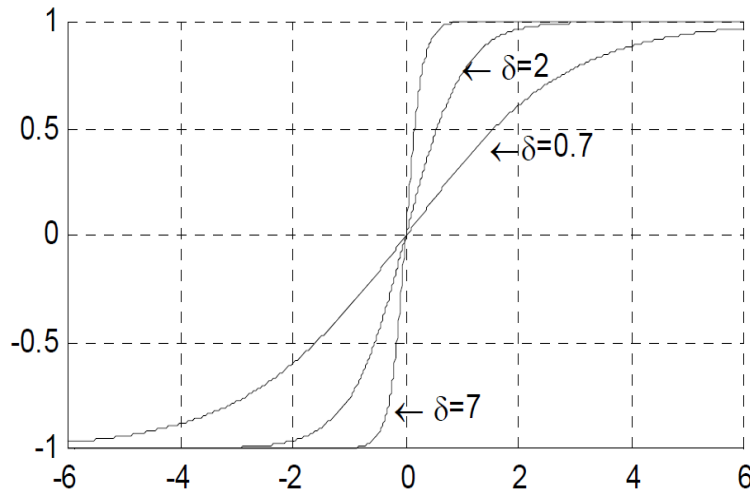


Fig (III.3) La fonction d'activation sigmoïde

III.7 Mise En Œuvre Des Réseaux Neuronaux

La procédure de développement d'un réseau neuronal peut être décrite en quatre étapes principales : [20]

Etape 1 : fixer le nombre de couches cachées

Mis à part les couches d'entrée et de sortie, l'analyste doit décider le nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation ; avec une couche cachée, il est capable, avec un nombre suffisant de neurones, d'approximer toute fonction continue. Une seconde couche cachée prend en compte les discontinuités éventuelles.

Etape 2 : déterminer le nombre de neurones par couches cachées

Chaque neurone supplémentaire permet de prendre en compte des profils spécifiques des neurones d'entrée. Un nombre plus important permet donc de mieux coller aux données

présentées mais diminue la capacité de généralisation du réseau. Il n'existe pas de règle générale mais des règles empiriques. La taille de la couche cachée doit être :

- ✓ Soit égale à celle de la couche d'entrée.
- ✓ Soit égale à 75% de celle-ci.
- ✓ Soit égale à la racine carrée du produit de nombre de neurones dans la couche d'entrée et de sortie.

Etape 3 : choisir la fonction d'activation

Nous considérerons la fonction sigmoïde pour le passage de la couche d'entrée à la couche cachée. Le passage de cette dernière à la couche de sortie sera soit linéaire, soit sigmoïde (logistique) selon nos types de variables.

Etape 4 : choisir l'apprentissage

L'apprentissage par rétro-propagation nécessite la détermination du paramètre d'ajustement des poids synaptiques à chaque itération. La détermination du critère d'arrêt est aussi cruciale dans la mesure où la convergence peut passer par des minimums locaux.

III.8 Propriété Fondamentale Des Réseaux De Neurones

La propriété d'approximation universelle, justifie dans une large mesure, l'utilisation des réseaux de neurones dans les domaines des sciences de l'ingénieur. La propriété d'approximation universelle des réseaux de neurones peut s'énoncer comme suit [4] :

Pour toute fonction déterministe suffisamment régulière, il existe au moins un réseau de neurone non bouclé, possédant une couche de neurones cachés et un neurone de sortie linéaire, qui réalise une approximation de cette fonction et ces dérivées successive, au sens de moindres carrées, avec une précision arbitraire.

Ce théorème alors ne s'applique que si les fonctions d'activités des neurones cachés sont continués et bornées, telles que les fonctions sigmoïdes et les fonctions à base radiale.

Nous signalons, que cette propriété n'est pas spécifique aux réseaux de neurones. Les polynômes, les séries de Fourier possèdent aussi cette particularité. La différence avec les réseaux de neurones réside dans leur parcimonie. En effet, pour obtenir une approximation d'une précision donnée, les réseaux de neurones utilisent moins de paramètres que les approximateurs usuels. En particulier, le nombre de paramètres varie essentiellement de manière linéaire en fonction du nombre de variable de la fonction que l'on cherche à

approcher, alors qu'il varie beaucoup plus rapidement avec la dimension de l'espace des entrées dans le cas des approximateurs usuels.

En pratique, les RNA sont donc avantageux, par rapport à d'autres approximateurs conventionnels, surtout pour des systèmes à plus d'une variable. Les réseaux de neurones offrent donc les avantages suivants [4] :

- l'estimation des fonctions est d'autant plus rapide que le nombre de paramètres à calculer est petit ;
- avec un minimum d'exemple à l'entrée du réseau, il est possible d'approximer une fonction délicate, et difficile à traiter analytiquement. Le fait d'utiliser moins de coefficients que les méthodes classiques de régression, permet donc une économie en nombre d'exemples d'entrées, ce qui peut être particulièrement important lorsque d'acquisition des exemples est coûteuse ou lente.

Il en résulte de tout ça, que :

- les RNA sont des outils statiques de régression qui permettent l'approximation, au sens des moindres carrés, de toute fonction non linéaire suffisamment régulière ;
- cette approximation est caractérisée par une économie, en nombre de coefficients, donc en nombre d'exemples.

III.9 Les Architectures Neuronales

L'intérêt des neurones réside dans les propriétés qui résultent de leur association en réseaux, c'est-à-dire de la composition des fonctions non linéaires réalisées par chacun des neurones. La plupart des réseaux de neurones ont une topologie définie sous forme de couches. Il existe quelques exceptions lorsque le réseau n'est pas explicitement défini sur plusieurs couches (comme par exemple certaines mémoires associatives), mais elles peuvent alors être considérées comme n'ayant qu'une seule couche. L'architecture du réseau peut alors être décrite par le nombre de couches, le nombre de neurones dans chaque couche et par les connexions entre les neurones. On distingue alors deux types de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés. [20]

III.9.1 Les réseaux de neurones non bouclés

Dans les réseaux de neurones non bouclés, dit aussi statiques, non récurrents ou réseaux à couches, l'information circule des entrées vers les sorties dans un seul sens, c'est-à-dire la sortie d'un neurone ne peut pas être injectée ni directement à son entrée ni indirectement à

travers d'autres neurones, donc la sortie courante n'a aucune influence sur les sorties futures (réseaux acycliques). Il existe des réseaux monocouches tels que le Perceptron et l'Adaline et des réseaux à plusieurs couches dits les MLP. [20]

III.9.1.1 Le Perceptron

Le Perceptron est le premier modèle de réseau les neurones a été introduit par McCulloch et Pitts en 1949. Les Perceptrons sont des réseaux monocouches qui contiennent en entrée des neurones entièrement connectés à d'autres neurones organisés en sortie par une couche de connexions modifiable de poids. Les neurones de ce modèle sont binaires. Ils reçoivent des signaux excitateurs et inhibiteurs provenant des neurones en amont afin d'effectuer une sommation. Si cette somme est supérieure à un seuil, le neurone est dans un état actif et émet un signal vers les autres neurones. Si ce n'est pas le cas, il est dans un état inactif et n'émet aucun signal.[20]

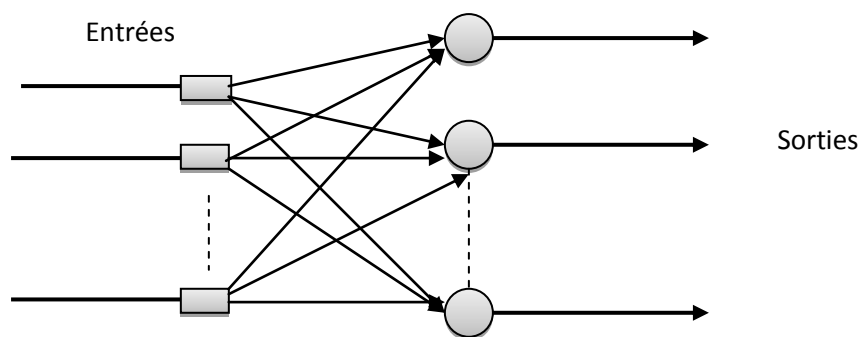


Fig (III.4) Le perceptron monocouche

III.9.1.2 L'ADALINE (Adaptive Linear Element)

Ce modèle est composé d'un seul neurone de type McCulloch dont la fonction d'activation est linéaire. La principale différence entre les deux modèles (perceptron et Adaline) se situe au niveau de la règle d'apprentissage car l'erreur dans le premier est calculée sur la sortie non linéaire alors que dans le deuxième, elle est évaluée sur la sortie linéaire.[17]

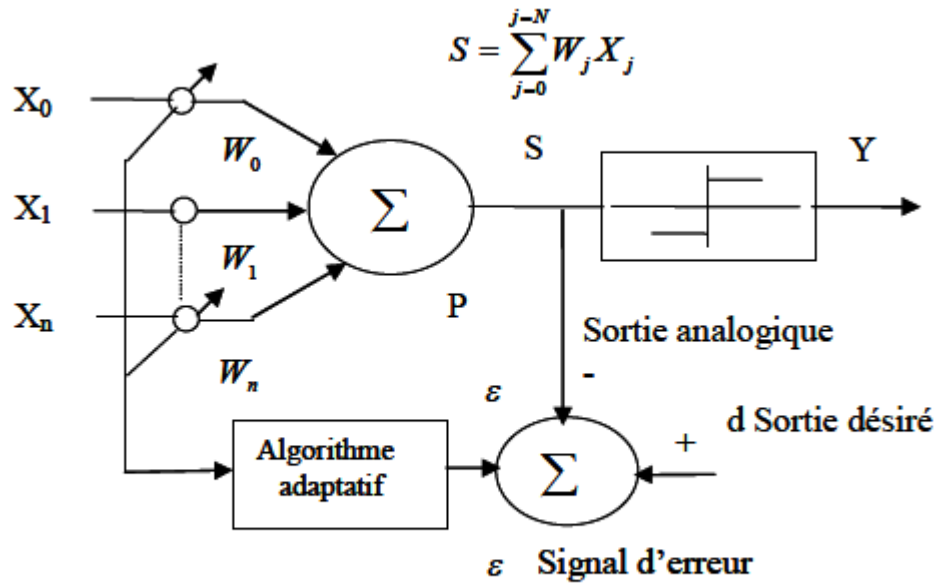


Fig (III.5) : Représentation de l'Adaline

III.9.1.3 Le Perceptron Multi-Couches (MLP)

Dans les réseaux MLP « Multi Layer Perceptron », les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avales. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et à celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de l'information (de l'activation) au sein d'un réseau. On appelle couche d'entrée l'ensemble des neurones d'entrée, couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelées couches cachées. L'architecture du réseau multicouche MLP est représentée par la figure (III.6).

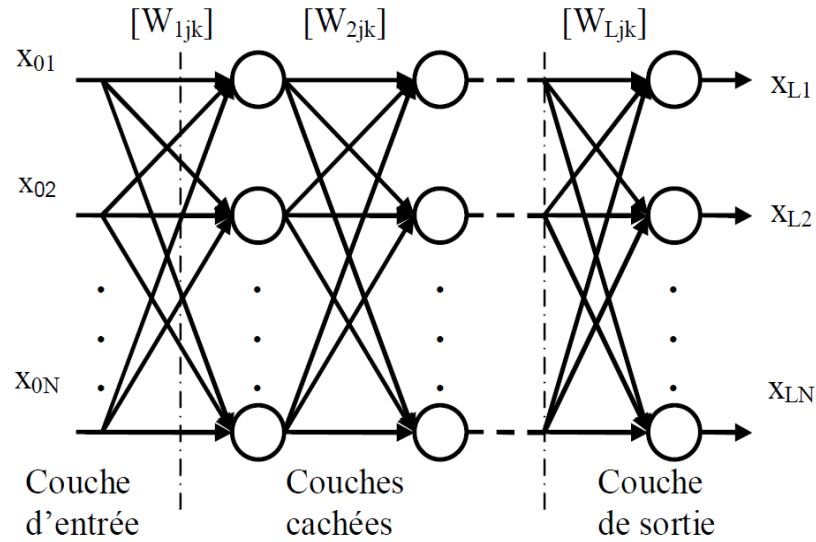


Fig (III.6) Architecture du perceptron multicouche MLP

Le réseau MLP contient N entrées, $(L-1)$ couches cachées de N_i neurones chacune ($1 < i < L-1$) et une couche de sortie contenant N_L neurones. L'indice de la couche est i , x_{ik} est la sortie du neurone k de la couche i , W_{ijk} est le poids reliant la sortie $x_{(i-1)j}$ au neurone k de la couche i , N_i est le nombre de neurones dans la couche i . Avec ces notations, la sortie x_{ik} du neurone (i, k) sera donnée par:

$$y_{ik} = \sum_{j=1}^{N(i-1)} W_{ijk} x_{(i-1)j} + b_{ik} / X_{ik} = f(y_{ik}) \quad (\text{III. 3})$$

Le filtrage non linéaire du MLP dépend énormément du nombre de couches cachées, le réseau peut avoir plusieurs de ces couches[23]. Cependant, trop de couches cachées compliquera l'apprentissage et augmentera le coût de calcul de la structure et peu de couches cachées sera inadéquat pour créer la non linéarité suffisante.

Il a été démontré qu'un réseau MLP (Multi Layer perceptron) avec seulement deux couches cachées dont la fonction d'activation est de type sigmoïde et une sortie avec une fonction d'activation linéaire peut approximer n'importe quelle fonction continue pourvu qu'un nombre suffisant de neurones soit fourni. Cette propriété est dite propriété d'approximation universelle [23].

III.9.1.4 Les Réseaux Rbf (Fonctions Radiales De Base)

Le réseau RBF (noté RBF pour Radial Basis Function) est un réseau de neurone supervisé, ou il s'agit d'une spécialisation d'un MLP. Il est constitué uniquement de trois couches :

La couche d'entrée: Elle retransmet les inputs sans distorsion.

La couche RBF : Couche cachée qui contient des neurones RBF.

La couche de sortie : Simple couche qui contient une fonction linéaire.

Chaque neurone RBF contient une gaussienne qui est centrée sur un point de l'espace d'entrée. La fonction gaussienne permet aux neurones de ne répondre qu'à une petite région de l'espace d'entrée, région sur laquelle la gaussienne est centrée.[20]

La sortie du réseau RBF est simplement une combinaison linéaire des sorties des neurones RBF multipliées par le poids de leur connexion. Il y'a quatre paramètres principaux dans un réseau RBF :

- ✓ Le nombre de neurones RBF (dans l'unique couche cachée).
- ✓ La position des centres des gaussiennes de chacun des neurones.
- ✓ La largeur de ces gaussiennes.
- ✓ Les poids des connexions entre les neurones RBF et les neurones de sortie.

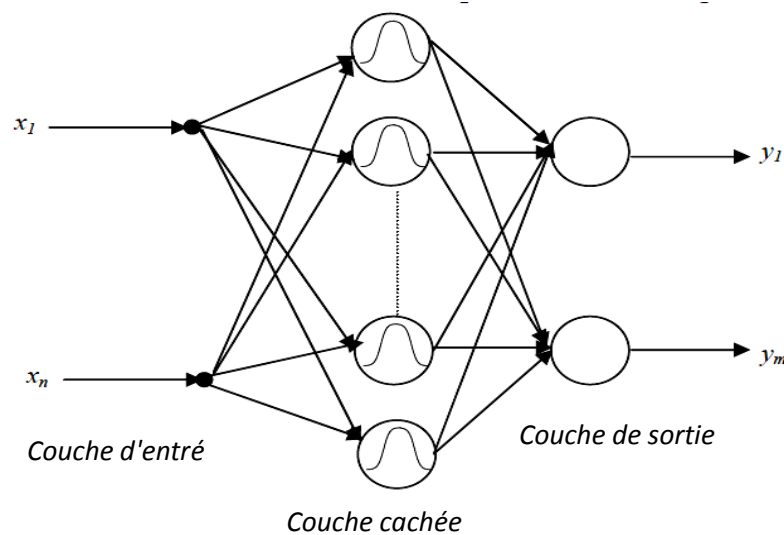


Fig (III.7) Réseau de neurones de type RBF

La sortie d'un neurone de la couche cachée est donnée par :

$$y_i = f(\|x - c_i\|) = \exp\left(-\frac{1}{2} \frac{\|x - c_i\|^2}{\sigma_i^2}\right) \quad (\text{III.4})$$

Et la sortie est définie par :

$$y_i = \sum_j^n w_{ij} \|x - c_i\| \quad ; \quad i = 1, \dots, n \quad (\text{III.5})$$

Les réseaux RBF sont très utilisés pour la classification et l'approximation des fonctions. Toutefois le problème de choix du nombre des neurones cachés se pose néanmoins toujours.

III.9.2 Les réseaux de neurones bouclés

Les réseaux récurrents peuvent contenir des chemins bouclés, passant plusieurs fois par un même neurone. Ils présentent donc une dynamique complexe due aux multiples rétroactions internes. Ce type de réseau a donc des capacités théoriques supérieures à celles des réseaux non récurrents. Il existe plusieurs types de réseaux récurrents tels que le modèle de HOPFIELD. Un exemple de réseaux bouclés est donné par La Figure (III.8). [20]

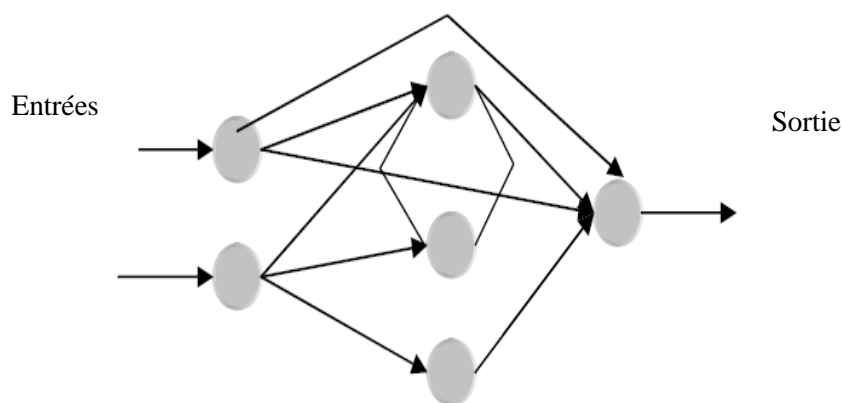
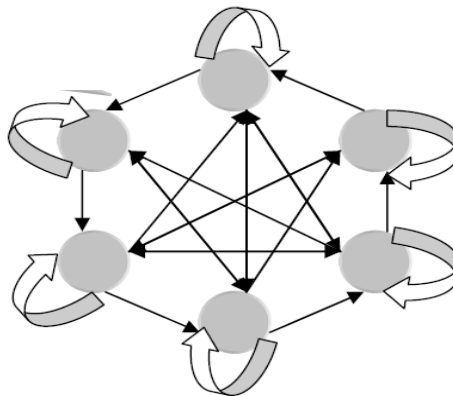


Fig (III.8) Exemple d'un réseau bouclé



Fig(III.9) Réseau entièrement bouclé

La Figure (III.9) montre un Réseau à connexion complète : c'est la structure d'interconnexion la plus générale. Chaque neurone est connecté à tous les neurones du réseau (et à lui-même).

III.10 Apprentissage Des Réseaux De Neurones

L'apprentissage est la propriété la plus intéressante des réseaux neuronaux. Durant cette phase, le comportement du réseau est modifié à travers les poids en fonction d'un jeu de données présentées en entrée, jusqu'à l'obtention du comportement désiré.

L'apprentissage peut être supervisé ou non supervisé selon la présence ou l'absence de la réponse désirée $d(n)$.

III.10.1 Apprentissage supervisé

L'apprentissage est dit supervisé lorsque le réseau est forcé à converger vers un état final précis, ce qui nécessite la connaissance *à priori* de la réponse désirée $d(n)$. La méthode la plus utilisée est la rétro propagation du gradient [23]. Elle consiste à présenter des exemples au réseau, calculer sa sortie, ajuster les poids de façon à réduire l'écart entre cette sortie et la réponse désirée pour satisfaire un certain critère de performance, figure (III.10).

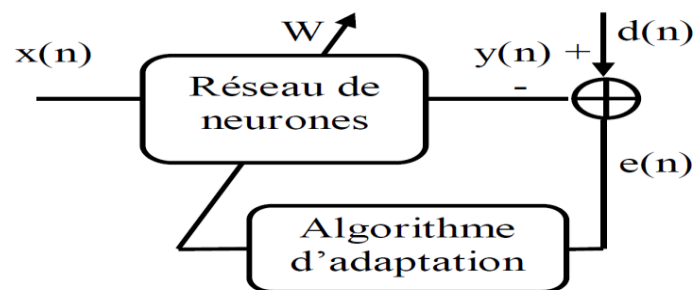


Fig (III.10) Mode d'apprentissage des réseaux de neurones supervisé

III.10.2 Apprentissage non supervisé

Dans l'apprentissage non supervisé, seules les valeurs d'entrée sont disponibles et le réseau est laissé libre de converger vers n'importe quel état final. La connaissance *à priori* de la sortie désirée n'est pas nécessaire, la procédure d'apprentissage est basée uniquement sur les valeurs d'entrées. Le réseau s'auto-organise de façon à optimiser une certaine fonction de coût, figure (III.11).

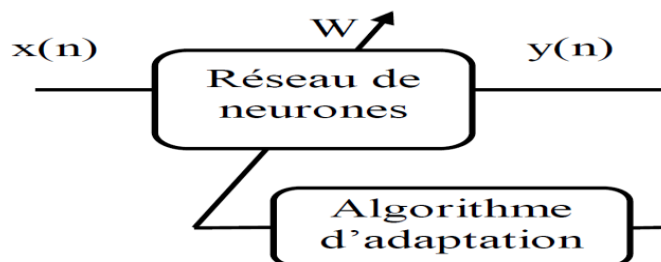


Fig (III.11) Mode d'apprentissage des réseaux de neurones non supervisé

III.10.3 L'algorithme de rétro propagation

La rétro propagation du gradient BP (Back Propagation) est une généralisation de la règle de « Widrow-Hoff » pour un réseau multicouche .Inventée par Rumelhart, Hinton et Williams en 1986 [23]. C'est un algorithme supervisé dont l'objectif est d'ajuster les poids du réseau MLP de façon à minimiser une fonction de coût différentiable, telle que l'erreur quadratique entre la sortie du réseau et la sortie désirée :

$$E(n) = \|d(n) - x_L(n)\|^2 \quad (\text{III. 6})$$

Où

$x_L(n)$: est la sortie du réseau au temps n ,

$d(n)$: la sortie désirée au temps n .

L'algorithme de BP assure la descente du gradient sur le critère d'erreur pour atteindre le minimum. Le gradient de $E(n)$ est calculé pour tous les poids de la manière suivante :

$$\frac{\partial E(n)}{\partial W_{ijk}} = \frac{\partial E(n)}{\partial y_{ik}} \cdot \frac{\partial y_{ik}}{\partial W_{ijk}} = \frac{\partial E(n)}{\partial y_{ik}} \cdot x_{(i-1)j} \quad (\text{III. 7})$$

Dans le cas de la couche de sortie ($i = L$) on évalue le terme d'erreur de sortie δ_{Lk} comme suit :

$$\delta_{Lk} = \frac{\partial E(n)}{\partial y_{Lk}} = 2f'(y_{Lk})(d_k - x_{Lk}) \quad (\text{III. 8})$$

Où

$f'(\)$: est la dérivée de la fonction d'activation : $f'(x) = \frac{df(x)}{dx}$.

Pour les couches cachées, le terme d'erreur δ_{ik} du neurone (i, k) est donné par :

$$\delta_{ik} = f'(y_{ik}) \sum_{j=1}^{N(i+1)} \delta_{(i+1)j} W_{(i+1)kj} \quad (\text{III. 9})$$

La modification des poids et des biais est obtenue selon les équations suivantes :

$$W_{ijk}(n+1) = W_{ijk}(n) + \mu \delta_{ik} x_{(i-1)j} + \alpha (W_{ijk}(n) - W_{ijk}(n-1)) \quad (\text{III. 10})$$

$$b_{ik}(n+1) = b_{ik}(n) + \mu \delta_{ik}$$

Où

μ : est le pas d'apprentissage,

α : le terme d'inertie (momentum).

Le choix du pas d'apprentissage influe beaucoup sur la rapidité de convergence, un pas trop petit ralenti l'apprentissage, un pas trop important provoque un risque d'instabilité. Il est même possible que l'algorithme rencontre un minimum local.

La figure (III.12) représente une forme 3D d'une fonction de coût E ayant des minimum locaux et un minimum global, où sa projection sur le plan de base formé par deux variables (poids W_1 et W_2) donne un contour de plusieurs niveaux de la fonction de coût. La figure (III.13) représente la Minimisation de la fonction de coût E par la méthode du gradient.

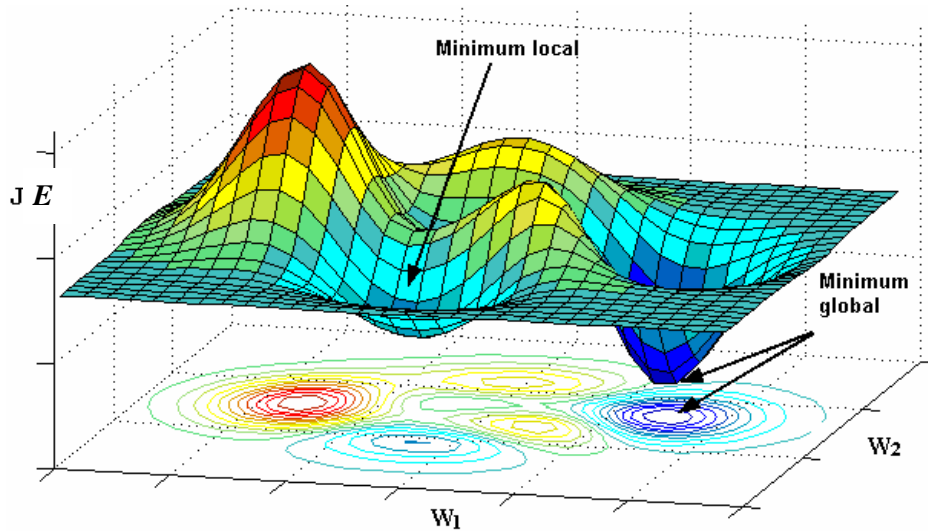


Fig (III.12) Représentation de la fonction de coût E d'un neurone à deux entrées pondérées W_1 et W_2 .

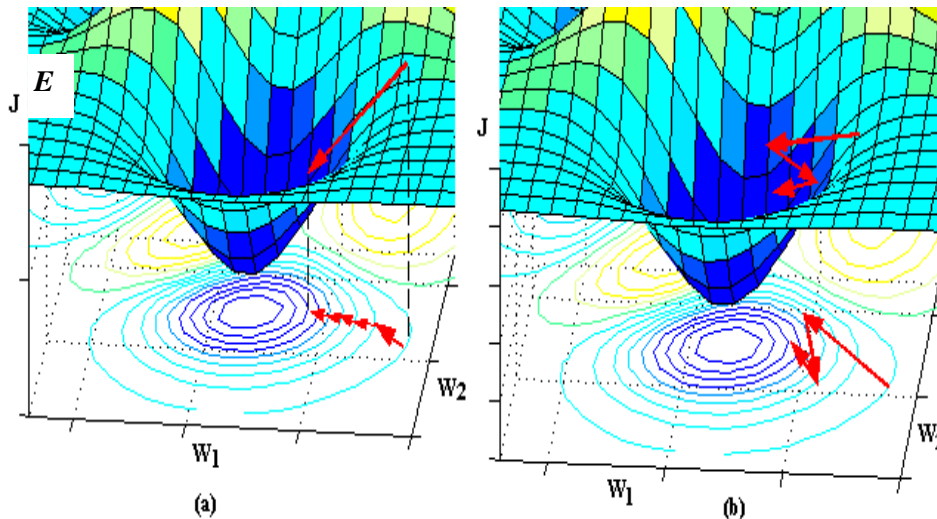


Figure (III.13) Minimisation de la fonction de coût E par la méthode du gradient.

(a) Pas du gradient est petit, convergence lente, le minimum global peut être atteint.

(b) Pas du gradient est grand, convergence rapide, le minimum global est rarement atteint

Le terme d'inertie ($0 < \alpha < 1.0$) permet de sortir du minimum local et de limiter les oscillations lors de l'apprentissage ; en prenant en compte les évolutions des étapes précédentes et donc de converger plus rapidement [23].

La phase d'apprentissage est souvent arrêtée lorsque l'erreur calculée sur l'ensemble de la base d'apprentissage est inférieure à un seuil déterminé par l'utilisateur. Il est possible d'arrêter l'apprentissage en fixant une limite au nombre d'itérations.

III.10.4 Apprentissage du réseau RBF

L'apprentissage des réseaux RBF est composé d'une règle d'apprentissage non supervisée pour l'adaptation des centres et d'une règle d'apprentissage supervisé pour l'adaptation des poids.

III.10.4.1 Apprentissage des centres

L'apprentissage des centres se fait selon les étapes suivantes :

- 1- Présenter un signal x_n au réseau
- 2- Calculer la distance entre le vecteur d'entrée et les centres

$$D_i(n) = \|x(n) - c_i(n)\|. \quad (\text{III.11})$$

- 3- Déterminer le centre le plus proche du signal d'entrée I de façon que :

$$D_I(n) = \min D_I(n) \quad (\text{III. 12})$$

- 4- Mettre à jour les centres c_I selon l'équation suivante :

$$c_I(n+1) = c_I(n) + \mu(x(n) - c_I(n)) \quad (\text{III. 13})$$

- 5- Aller à l'étape 1 [19].

III.10.4.2 Adaptation des poids

L'adaptation des poids est réalisée par la règle de rétro propagation du gradient, elle est utilisée sur la seule couche de sortie :

- 1- Présenter une paire des entrées et de sortie désirée (x_n, d_n) .
- 2- Calculer l'erreur entre la sortie du réseau et la sortie désirée : $e_j(n) = d_j(n) - y_i$
- 3- Mise à jour des poids de la couche de sortie selon l'équation :

$$w_{kj}(n+1) = w_{kj}(n) + \alpha e_j(n) \phi_k(\|x - c_k\|) \quad (\text{III.14})$$

α est une constante petite positive

4- Aller à l'étape 1 [23].

III.11 Généralisation d'un réseau MLP (Validation croisée)

L'un des nombreux intérêts des réseaux MLP pour l'approximation de fonctions est leur capacité à généraliser, c'est à dire leur capacité à donner une bonne réponse à une donnée qui n'appartient pas à la base d'apprentissage qui a servi à l'éduquer.

Cette capacité à traiter des données non connues de la base d'apprentissage, la faculté de généralisation, permet de faire de l'interpolation ou de l'extrapolation.

Pour estimer la qualité d'un réseau de neurones on fait généralement appel à la racine carrée de l'erreur quadratique moyenne.

$$MSE_i = \sqrt{\frac{1}{p} \sum_{i=1}^p (y_i - \hat{y}_i)^2} \quad (\text{III. 15})$$

Avec :

p : Le nombre d'échantillons dans la base d'apprentissage.

y_i : La composante i de la sortie désirée.

\hat{y}_i : La composante estimée (calculée) du neurone i de la couche de sortie du réseau MLP.

Pour analyser le comportement du Réseau de Neurones, on observe donc indépendamment ces deux quantités y_i et \hat{y}_i respectivement. L'indicateur d'un bon réseau sera le taux de généralisation : c'est-à-dire avoir le même MSE mais sur une base d'exemples autre que celle utilisée durant l'apprentissage (la base de test ou de généralisation).

C'est à travers la valeur du MSE que l'on peut connaître la capacité de généralisation d'un réseau de neurones face à des données hors de la base d'apprentissage. Malheureusement, ce taux de généralisation n'est pas accessible directement, on cherche donc à l'estimer. Pour cela, la base de données est partagée en deux :

- la base d'apprentissage sur laquelle on configure le réseau pour réaliser au mieux la tâche qui lui est attribuée.
- La base de généralisation sur laquelle on teste l'aptitude du réseau à bien traiter des données hors de la base d'apprentissage.

Nous avons évoqué plusieurs fois le problème du sur apprentissage, qui est provoqué par la capacité d'un réseau de neurones, possédant un nombre d'unités de mémorisation plus que

nécessaire (on parle aussi de ‘ sur paramètre ’), à apprendre parfaitement les exemples de la base d’apprentissage.

Typiquement, l’évolution de l’erreur quadratique sur la base d’apprentissage en fonction du nombre de cycles d’apprentissage, se comporte comme la courbe A de la figure (III.14). Sur cette même figure est également représentée l’évolution de l’erreur quadratique en généralisation calculée à partir d’une base de données différente de la base d’apprentissage (courbe B).

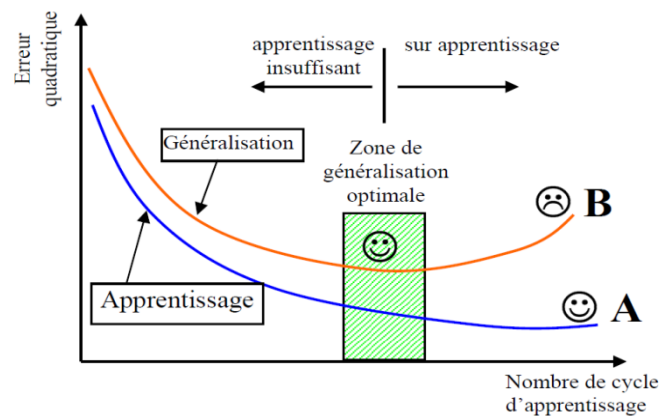


Fig (III.14) Evolution de l’erreur quadratique de l’apprentissage et de généralisation.

Afin d’arrêter l’apprentissage juste avant que ne se produise ce phénomène de sur apprentissage, plusieurs méthodes ont été proposées.

La plus simple consiste à disposer de trois bases de données distinctes : une base d’apprentissage, une base de test et une base dite de « validation croisée » [24].

Cette dernière base est utilisée pendant l’apprentissage afin d’examiner le comportement du réseau pour des données qui lui sont inconnues. Ainsi, l’apprentissage est arrêté lorsque l’erreur sur cette courbe B (base de validation croisée) atteint un minimum.

Notons que cette technique nécessite d’avoir suffisamment de données pour constituer trois bases à la fois représentatives et distinctes. Dans le cas où on ne disposerait pas de suffisamment de données, une technique a été développée pour n’utiliser que la base d’apprentissage ; elle consiste à effectuer deux apprentissages successifs.

Pour le premier apprentissage, la base d’apprentissage est divisée en deux de manière à constituer une plus petite base d’apprentissage et une base de validation croisée.

Le nombre de cycles d'apprentissage nécessaire pour parvenir à la zone de généralisation optimale est conservé. Ce nombre est ensuite utilisé pour stopper l'apprentissage avec la base d'apprentissage globale.

III.12 Les champs d'application des RNA

Les propriétés d'apprentissage et d'approximation des réseaux de neurones ont permis une importante application de ces derniers dans les différents domaines pratiques, notamment, les domaines suivants : [20]

- ✓ Identification des processus
- ✓ Contrôle des processus
- ✓ Classification des signaux
- ✓ Mémoires associatives
- ✓ ...etc.

III.13 Conclusion

Ce chapitre a fait l'objet de quelques définitions et généralités sur les réseaux de neurones artificiels. Nous avons rappelé les concepts les plus importants pour comprendre l'intérêt des réseaux de neurones comme outil de classification.

Les réseaux de neurones présentés dans ce chapitre sont des systèmes paramétriques non linéaires qui relient des variables d'entrée à des variables de sortie. Les paramètres sont déterminés statistiquement lors d'une phase d'apprentissage à partir d'une base de données appelée base d'apprentissage.

L'objectif assigné est l'exploitation des réseaux de neurones pour l'approximation des caractéristiques de commutation des onduleurs multi-niveaux. Nous avons déterminé les réseaux capables de reproduire les angles de commutation assurant l'élimination des harmoniques non désirés pour l'onduleur à 7 niveaux.

Le quatrième chapitre comprend tous les résultats de simulation.