POLYPUBLIE

Approche multiéchelle pour l'approximation de la fonction discriminante d'un réseau de quantrons

Jean-François Connolly

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2005
Type	Mémoire ou thèse / Dissertation or Thesis Mémoire de maîtrise/Master's thesis
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/7602/
Conditions d'utilisation Terms of use	© Tous droits réservés / All rights reserved
Référence Citation	Connolly, J.-F. (2005). Approche multiéchelle pour l'approximation de la fonction discriminante d'un réseau de quantrons [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7602/

Détails de la thèse
Thesis details

Directeurs de recherche Advisors	Richard Labib
Programme Program	Non spécifié

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

APPROCHE MULTIÉCHELLE POUR L'APPROXIMATION DE LA FONCTION DISCRIMINANTE D'UN RÉSEAU DE QUANTRONS

JEAN-FRANÇOIS CONNOLLY

DÉPARTEMENT

DE MATHÉMATIQUES

ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(MATHÉMATIQUES APPLIQUÉES)

AOÛT 2005

© Jean-François Connolly, 2005.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-16769-4
Our file Notre référence
ISBN: 978-0-494-16769-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

APPROCHE MULTIÉCHELLE POUR L'APPROXIMATION DE LA FONCTION DISCRIMINANTE D'UN RÉSEAU DE QUANTRONS

présenté par : CONNOLLY, Jean-François

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

M. ADJENGUE Luc-Désiré, Ph.D., président

M. LABIB Richard, Ph.D., membre et directeur de recherche

M. GAGNON Michel, Ph.D., membre

À mes parents.

REMERCIEMENTS

Je voudrais tout d'abord remercier mon directeur de recherche, Richard Labib, pour m'avoir fait découvrir les passionnants domaines que sont les réseaux de neurones et les ondelettes. Son aide et son soutien financier ont été indispensables à la réalisation de cette maîtrise. J'aimerais également souligner ma reconnaissance envers mes parents, Marie et Richard, pour leur support et encouragements tout au long de mes études. Et finalement, j'aimerais remercier ma conjointe, Laurence, pour son aide à la correction de ce document et surtout pour son éternelle patience.

RÉSUMÉ

Un problème courant dans le domaine de l'intelligence artificielle est d'obtenir le maximum d'une fonction dite d'approximation. Une nouvelle architecture de réseau, développée dernièrement, génère une fonction d'onde complexe qu'il faut optimiser de manière à en trouver le maximum global. Pour résoudre ce problème, nous proposons une approche multiéchelle d'approximation utilisant une parabole tronquée comme fonction de base.

Basée sur les méthodes multirésolutions développées dans le cadre des ondelettes, cette approche multiéchelle est motivée par deux facteurs : la présence du quantron comme modèle du neurone et la méthode de Laplace.

Nous allons voir l'approximation à une seule échelle ainsi que l'approximation multiéchelle proprement dite. Pour ces deux approches, la convergence sera démontrée et les résultats seront analysés.

ABSTRACT

Finding an approximation function's maximum is a common problem in artificial intelligence. A new type of neural network generates a wave function of high complexity that needs to be optimized in order to find its global maximum. To resolve this problem, we present a multiscale approximation scheme that uses compactly supported parabolas.

Based on multiresolution analysis developed in the wavelet's theory, this multiscale scheme is motivated by two elements : the quantron as the neuron's model and the Laplace's method.

We will see an approximation method on only one scale at a time followed by the multiscale approach. For both of these methods, the convergence will be proved and the results analysed.

TABLE DES MATIÈRES

DÉDICACE iv

REMERCIEMENTS v

RÉSUMÉ vi

ABSTRACT vii

TABLE DES MATIÈRES viii

LISTE DES TABLEAUX x

LISTE DES FIGURES xi

INTRODUCTION 1

CHAPITRE 1 – MOTIVATIONS 6

1.1 Le quantron 6

1.2 Méthode de Laplace 12

CHAPITRE 2 – CADRE THÉORIQUE 14

2.1 Espaces d'approximation 14

2.1.1 Espaces de Banach et Hilbert 14

2.1.2 Bases orthogonales des espaces de Hilbert 16

2.1.3 Bases de Riesz 17

2.2 Analyse multirésolution 18

CHAPITRE 3 – DE L'APPROXIMATION À L'OPTIMISATION 21

3.1 Approximation à une échelle 22

3.1.1 Base d'approximation 22

3.1.2 Stabilité de la base 35

3.1.3 Reconstruction et convergence 41

3.2 Approximation multiéchelle 55

CHAPITRE 4 – APPLICATIONS ET DISCUSSION 63

4.1 Approximation à une échelle 64

4.2 Approximation multiéchelle 70

4.2.1 Conditions d'application 71

4.2.2 Exemples 73

CONCLUSION 83

BIBLIOGRAPHIE 85

LISTE DES TABLEAUX

Tableau 4.1 : Erreur quadratique absolue d'approximation d'une fonction $\rho(t)$ pour des valeurs de j variant de 0 à 7. 66

Tableau 4.2 : Coefficients non nuls d'une approximation pour $j = 4$. 66

Tableau 4.3 : Sommet des paraboles d'une approximation pour $j = 4$. 66

Tableau 4.4 : Erreur quadratique absolue d'approximation d'une fonction P(t) pour des valeurs de j variant de 0 à 7 69

Tableau 4.5 : Erreur quadratique absolue d'approximation d'une fonction $\rho(t)$ pour des valeurs de j variant de 2 à 4. 75

Tableau 4.6 : Erreur quadratique absolue d'approximation d'une fonction P(t) positive pour des valeurs de j variant de 2 à 4 77

Tableau 4.7 : Erreur quadratique absolue d'approximation d'une fonction P(t) pour des valeurs de j variant de 2 à 5 81

LISTE DES FIGURES

Figure 1.1 : Synapse d'un neurone 7

Figure 1.2 : Représentation du niveau de variation de polarisation dans la membrane d'une synapse pour un seul potentiel d'action. 8

Figure 1.3 : Représentation d'un quantron à quatre entrées. 10

Figure 1.4 : Représentation du niveau de variation de polarisation dans la membrane pour 10 potentiels d'actions. 12

Figure 3.1 : Projection de P(t) sur le sous-espace V_3. 22

Figure 3.3 : Effet des paramètres p et h sur $\theta(t)$ 26

Figure 3.4 : Propriété de translation et de dilatation de la fonction de base 28

Figure 3.5 : Exemple de fonction $\theta_{j,n}(t)$ pour différentes valeurs de j et n. 29

Figure 3.6 : Calcul de l'erreur du sommet d'une parabole centrée à $t = 0$ au sommet d'une parabole centrée à $t = \tau$ 32

Figure 3.7 : Erreur d'approximation en fonction de τ 33

Figure 3.8 : Reconstruction de $\theta_{0,0}(t)$ avec $\theta_{1,0}(t)$ et $\theta_{1,1}(t)$ 34

Figure 3.9 : Indépendance linéaire. 36

Figure 3.10 : Transformée de Fourier de $\theta(t)$. 38

Figure 3.11 : La fonction $\beta(\omega)$. 41

Figure 3.12 : Projection non-orthogonale de P(t) sur V_5 42

Figure 3.13 : Approximation discrète de P(t) pour $j = 5$. 43

Figure 3.14 : Intervalle défini pour la fonction d'erreur. 44

Figure 3.15 : Organisation des sous-espaces V_j pour une approximation multiéchelle. 56

Figure 3.16 : Reconstruction de $g_j(t)$. 57

Figure 3.17 : Approximation $g_0(t)$ et fonction résiduelle $r_0(t)$. 57

Figure 3.18 : Approximation $g_1(t)$ et fonction résiduelle $r_1(t)$. 58

Figure 3.19 : Approximation $g_2(t)$ et fonction résiduelle $r_2(t)$. 58

Figure 3.20 : Approximation $g_3(t)$ et fonction résiduelle $r_3(t)$. 59

Figure 4.1 : Approximation d'une fonction $\rho(t)$ avec des paraboles et avec des fonctions de Haar 65

Figure 4.2 : Approximation d'une fonction P(t) positive avec des paraboles et des fonctions de Haar 68

Figure 4.3 : Approximation d'une fonction P(t) partiellement positive avec des paraboles et des fonctions de Haar. 70

Figure 4.4 : Approximation $g_3(t)$ d'une fonction $\rho(t)$ et fonction résiduelle $r_3(t)$. 72

Figure 4.5 : Approximation $g_2(t)$ d'une fonction $\rho(t)$ pour des paraboles et des fonctions de Haar 73

Figure 4.6 : Approximation $g_3(t)$ d'une fonction $\rho(t)$ pour des paraboles et des fonctions de Haar 74

Figure 4.7 : Approximation $g_4(t)$ d'une fonction $\rho(t)$ pour des paraboles et des fonctions de Haar 74

Figure 4.8 : Approximation $g_2(t)$ d'une fonction P(t) pour des paraboles et des fonctions de Haar 76

Figure 4.9 : Approximation $g_3(t)$ d'une fonction P(t) pour des paraboles et des fonctions de Haar 76

Figure 4.10 : Approximation $g_4(t)$ d'une fonction P(t) pour des paraboles et des fonctions de Haar.....77

Figure 4.11 : Approximation $g_2(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.....79

Figure 4.12 : Approximation $g_3(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.....80

Figure 4.13 : Approximation $g_4(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.....80

Figure 4.14 : Approximation $g_5(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.....81

INTRODUCTION

La plus grande improvisation de l'esprit humain est celle qui donne de l'existence à ce qui n'en a pas.

Napoléon Bonaparte

« Parce que vous saurez distinguer ce qui est possible de ce qui ne l'est pas, vous n'entreprendrez rien qui ne puisse être mené à bonne fin. »[1] En disant ces mots, Sun Tsu discutait de la capacité des bons généraux de la Chine antique à reconnaître diverses situations et à poser des actions en fonction de leurs expériences passées; il faisait allusion à leur capacité d'apprendre et de s'adapter à des situations nouvelles. L'origine de cette aptitude d'apprentissage est le cerveau. D'un point de vue purement scientifique, le cerveau peut être considéré comme un système complexe, non-linéaire, et surtout extrêmement puissant. En fait, la force du cerveau humain est de savoir organiser ses composants de base, les neurones, en réseaux efficaces permettant un traitement parallèle de l'information, et ce contrairement aux ordinateurs conventionnels qui fonctionnent de manière séquentielle.

Pour entrevoir l'amplitude de la complexité du cerveau, on peut comparer les éléments de base de ce dernier avec l'élément de base d'un microprocesseur : les

[1] Sun Tsu, *L'Art de la Guerre*, traduction de P. Amyot, Paris, 1772.

portes logiques fabriquées en silice. La première chose que l'on peut constater est que le neurone en tant que tel n'est pas aussi puissant qu'on le croit. Typiquement, le neurone est environ un million de fois plus lent que la porte logique; la période de temps pour que celui-ci effectue une action est de l'ordre de la milliseconde (10^{-3} s) alors que du côté des portes logiques, cette même période de temps est plutôt de l'ordre de la nanoseconde (10^{-9} s). Pour compenser la fréquence d'opération relativement lente du neurone, le cerveau en possède une quantité phénoménale. Shepherd et Koch (1990) estimèrent le nombre de neurones dans le cortex humain à environ 10 milliards et le nombre de connections à plus de 60 trillions (10^{12}). Ce réseau extrêmement dense fait en sorte que le cerveau possède une structure très efficace. Plus spécifiquement, l'efficacité énergétique du cerveau est d'environ 10^{-16} joules par opération par seconde comparativement 10^{-6} joules pour les meilleurs ordinateurs de l'époque (Faggin, 1991). En guise d'exemple concret de la puissance du cerveau dans son ensemble, celui-ci accomplit chaque jour, de manière routinière, des tâches de reconnaissance de formes (par exemple, la reconnaissance de visages dans un milieu à priori inconnu) en une période d'environ 100 à 200 millisecondes. Une tâche de moindre complexité pourrait prendre des jours à effectuer avec un ordinateur conventionnel (Haykin, 1996).

Évidemment, le grand intérêt de comprendre le fonctionnement du cerveau est de reproduire artificiellement son énorme potentiel. Les premières avancées dans ce domaine ont été faites par McCulloch et Pitts (1943), qui décrivent le neurone comme étant une unité de calcul logique faisant partie d'un réseau. Ils démontrent

qu'avec un nombre suffisant de ces unités, un réseau pourrait, en principe, résoudre n'importe quelle fonction.

Quelques temps après, Hebb (1949) introduit le concept de règle d'apprentissage pour les modifications au niveau des connexions entre chaque neurone : les synapses. Il soutient que les connexions dans le cerveau sont constamment modifiées au fur et à mesure qu'une personne apprend à effectuer différentes tâches.

Neuf ans plus tard, Rosenblatt (1958) propose avec son perceptron une approche d'apprentissage supervisée avec preuve de convergence. Le perceptron est alors un simple neurone avec des poids synaptiques qui pondèrent la valeur de chaque entrée auquel on ajoute un biais. La somme des entrées et du biais est traitée par une fonction d'activation qui introduit la non-linéarité du neurone et limite l'amplitude de la sortie. Encore une fois, l'apprentissage se fait en modifiant les poids synaptiques du neurone.

Dans la même année, Widrow et Hoff (1960) introduisent l'algorithme d'apprentissage du moindre carré, qui permet l'apprentissage d'un réseau possédant une couche de neurones. Cet algorithme n'est en fait que la minimisation de l'erreur quadratique de la sortie en fonction de la valeur des poids synaptiques. C'est donc à cette époque que l'on voit apparaître un des problèmes auxquels nous sommes confrontés dans cette recherche, soit l'optimisation globale d'une fonction dans le cadre d'un réseau de neurones. Contrairement à Widrow et Hoff, notre but ne sera pas la minimisation, mais la maximisation d'une fonction.

Malgré ces avancées dans les réseaux de neurones, Minsky et Papert (1969) mettent en évidence les limites des réseaux faits à partir de perceptrons. Afin de repousser les limites flagrantes de ce dernier et d'augmenter le potentiel des réseaux de neurones artificiels, les considérations biologiques furent mises de côté au profit d'algorithmes mathématiques développés en fonction de problèmes spécifiques.

Labib (1999) revient à des considérations biologiques et présente un nouveau modèle de neurone : le quantron. Dans ce modèle de neurone, la polarisation de la membrane à l'entrée du neurone doit atteindre un certain seuil pour que celui-ci puisse transmettre l'information (i.e. émettre des impulsions). Dans le cadre de l'algorithme d'apprentissage, présentement en développement, il est nécessaire de savoir si un neurone conduit ou non. En d'autres mots : si le maximum global de la fonction modélisant la polarisation de la membrane du neurone atteint le seuil de transmission.

Encore une fois, nous sommes confrontés à un problème d'optimisation globale de fonctions. Par contre, comme les fonctions correspondantes aux impulsions du quantron sont complexes, nous avons besoin d'en faire une approximation. Toujours dans le cadre de l'algorithme d'apprentissage, l'approximation doit être faite avec une fonction dérivable, donc analytique et il faut que cette fonction puisse être optimisée avec une méthode exigeant qu'elle soit formée de paraboles positives. En optimisant la fonction approximée, nous serons ainsi en mesure de déterminer si chacun des éléments d'un réseau de quantrons transmettent de l'information et nous pourrons simplifier le réseau.

Cette recherche a comme objectif de développer une méthode d'approximation d'une fonction quelconque, qui est en fait une somme de polarisations d'une membrane d'un neurone modélisé par un quantron, de manière à ce que cette approximation soit formée uniquement de paraboles positives. La méthode proposée se fonde sur une approche multiéchelle. Inspirée d'une analyse multirésolution, cette approche consistera à décomposer une fonction à l'aide de produits scalaires entre cette fonction et un noyau (une fonction de base) constitué d'une parabole positive à support compact. La décomposition de la fonction se fera sur tout le domaine à l'aide de translations de ce noyau et le changement d'échelle se fera par la dilatation du noyau.

En premier lieu, nous allons expliquer les deux éléments qui justifient l'utilisation de notre approche, soit le quantron et la méthode de Laplace. Ensuite nous donnerons les définitions nécessaires à notre approche concernant les divers éléments des espaces d'approximation ainsi que la définition d'une analyse multirésolution. Ensuite l'approche proposée sera expliquée, sa convergence sera démontrée et nous verrons en quoi nous n'avons pas une approche multirésolution, mais bien une approche multiéchelle. Finalement, des résultats seront présentés et accompagnés d'une discussion.

CHAPITRE 1 – MOTIVATIONS

Dans cette section nous allons détailler les deux éléments qui nous motivent à employer la méthode d'approximation proposée. Ces deux éléments sont la nature du quantron ainsi que la méthode de Laplace.

1.1 Le quantron

Le quantron est un modèle de neurone basé sur des signaux ondulants induits par l'accumulation de molécules biologiques : les neurotransmetteurs. Ceux-ci seront responsables de la transmission de l'information au niveau de la connexion entre chaque neurone, c'est-à-dire la synapse.

D'un point de vue biologique, la synapse est un espace vide entre le neurone pré-synaptique et le neurone post-synaptique. Lorsqu'un potentiel d'action (signal électrique représentant de l'information) est émis par le neurone pré-synaptique, des neurotransmetteurs sont libérés dans la synapse. La combinaison neurotransmetteur-récepteur modifie alors le niveau d'excitation du neurone post-synaptique de façon à augmenter ou diminuer sa tendance à générer des potentiels d'action (voir figure 1.1).

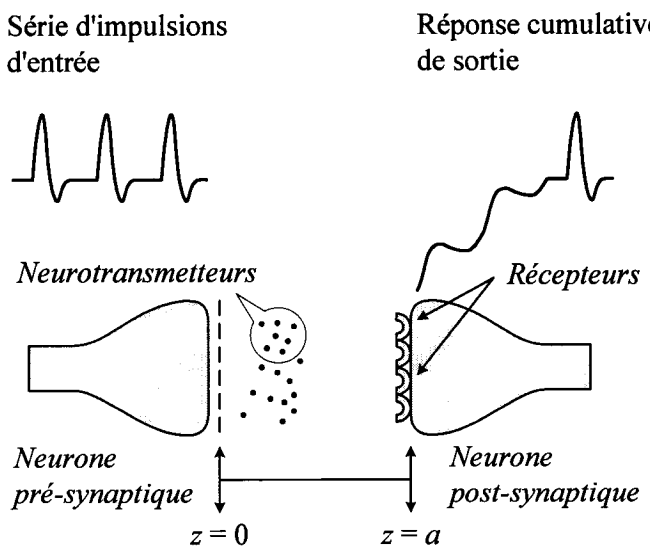


Figure 1.1 : Synapse d'un neurone.[2]

Il a été montré que la tension $V(t)$ induite dans la membrane post-synaptique peut être exprimée par :

$$V(t) = 2w\left(1 - \Phi\left(\frac{\ln a}{\sqrt{t}}\right)\right), \qquad (w \in \mathbb{R},\ a > 1,\ t > 0) \tag{1.1}$$

où w est une constante de proportionnalité qui pondère la valeur des entrées et a un paramètre représentant la largeur de la fente synaptique. La fonction $\Phi(\cdot)$ est la fonction de répartition de la variable aléatoire gaussienne centrée réduite. Après un temps S, l'association neurotransmetteur-récepteur se termine et la membrane se décharge à la même vitesse qu'elle s'est chargée. Le niveau de polarisation global dans le neurone post-synaptique est donné par :

[2] Figure tirée de : LABIB R. (1999), « New single neuron structure for solving nonlinear problems », *Proc. Of the 1999 international Joint Conference on Neural Networks*, manuscript #0476, 4 pages.

$$\rho(t)=V(t)(u(t)-u(t-S))+ \\ (V(S)-V(t-S))(u(t-S)-u(t-2S)) \tag{1.2}$$

où $u(\cdot)$ est la fonction échelon (également appelée fonction Heaviside). La figure 1.1 montre $\rho(t)$ pour différentes largeurs de fente synaptique a, de temps d'interruption S et de poids synaptique w. Dans le cas d'un poids synaptique w négatif, la polarisation de la membrane post-synaptique est négative, ce qui réduit les chances de transmission d'information.

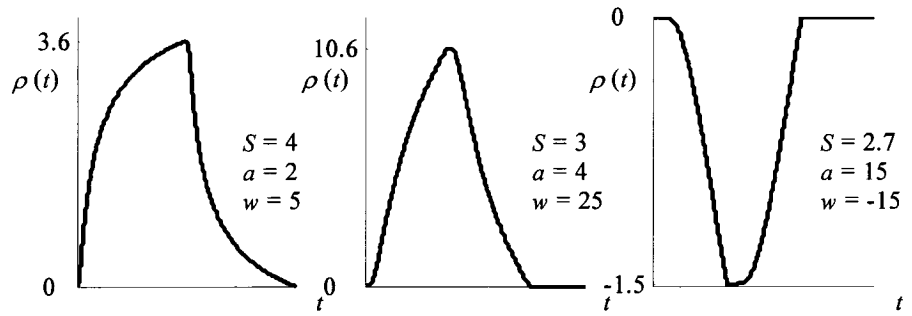


Figure 1.2 : Représentation du niveau de variation de polarisation dans la membrane d'une synapse pour un seul potentiel d'action.

La propagation de la part du neurone post-synaptique se fait seulement si la sommation temporelle de la fonction $\rho(t)$ est assez grande pour atteindre le seuil Γ nécessaire au déclenchement du potentiel d'action post-synaptique (Labib, 1999).

Dans un neurone biologique, les potentiels d'action surviennent pleinement ou pas du tout, c'est pourquoi ce n'est pas l'amplitude de ces derniers, mais bien leur

fréquence d'arrivée qui détermine s'il y a une réponse de la part du neurone post-synaptique. En considérant la période x (> 0) qui s'écoule entre deux potentiels d'action pré-synaptiques et le délai θ précédant l'arrivée du premier potentiel d'action, la réponse totale $R(t)$ du neurone post-synaptique devient :

$$R(t) = \sum_{k=0}^{\infty} \rho(t - \theta - k\,x). \tag{1.3}$$

La fréquence pré-synaptique des signaux d'entrée est alors donnée par x^{-1}. À ce moment-ci, tel que mentionné précédemment, la sortie Ω du quantron stimulé par une entrée x dépend de l'atteinte ou non du seuil Γ. Cette sortie est alors :

$$\Omega = \begin{cases} 0 & \text{si } R(t) < \Gamma \quad \forall t \in [0,\infty) \\ \alpha & \text{sinon} \end{cases} \tag{1.4}$$

où la période s'écoulant entre deux potentiels d'action post-synaptique du quantron α est définie par

$$\alpha = \inf\{\, t > 0 : R(t) = \Gamma \,\}. \tag{1.5}$$

Dans le cadre d'un réseau, une fois que la dimension de la fente synaptique a et le seuil Γ sont spécifiés, la tâche à accomplir consiste à trouver les paramètres w (poids), S (temps d'interruption) et θ (délai de premier passage) de manière à résoudre un

problème spécifique. Un seul neurone est alors la cible de plusieurs potentiels d'action et la polarisation de la membrane peut être exprimée par :

$$\mathrm{P}(t) = \sum_{k=1}^{n} R_n(t) \tag{1.6}$$

où P est un ρ majuscule. La figure 1.2 illustre un quantron avec quatre entrées.

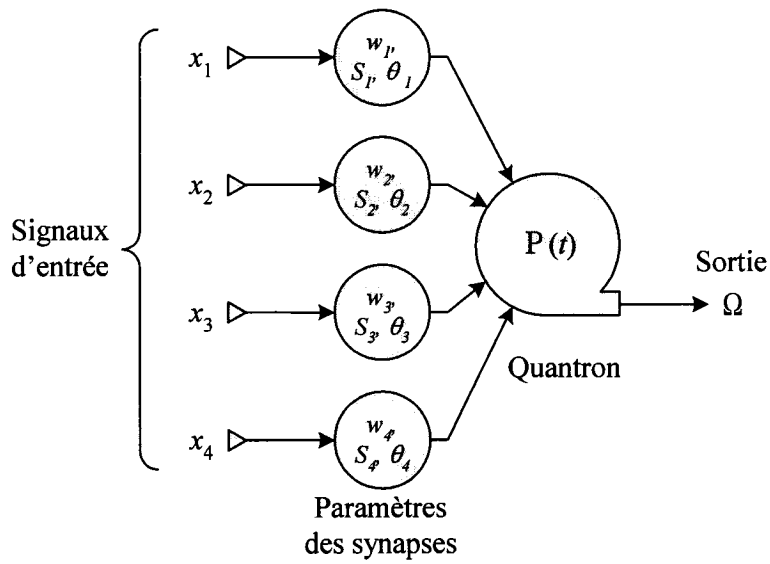


Figure 1.3 : Représentation d'un quantron à quatre entrées.

L'algorithme d'apprentissage du quantron cible deux actions : la simplification du réseau et l'ajustement de la synapse dont les paramètres sont S, θ et w. Étant donné que dans un réseau, le quantron transmet de l'information si et seulement si la fonction P(t) atteint le seuil Γ, il est nécessaire d'optimiser cette fonction pour savoir s'il y aura une possibilité de réponse de la part du quantron. De cette manière, le

réseau pourra être simplifié en éliminant les neurones superflus (qui ne conduiront jamais). De plus, l'ajustement des différents paramètres du quantron nécessite que l'approximation soit dérivable, donc analytique.

La complexité des fonctions entourant le quantron nécessite une approximation, qui ne sera en fait qu'un compromis nous permettant de travailler avec une fonction plus simple. L'élément justifiant l'utilisation d'une approche multiéchelle est la nature corpusculaire des fonctions entourant le quantron. En effet, chaque entrée affecte la polarisation de la membrane post-synaptique telle que décrite par $R(t)$, qui est un ensemble de fonctions à support compact (définies sur un intervalle où les fonctions $\rho(t-\theta-k\,x)\neq 0$). La sommation des quatre entrées du quantron de la figure précédente forme la fonction P(t) et le support de cette fonction est également compact (voir figure 1.3). Cette dernière est un phénomène transitoire et du même coup, non périodique. Elle n'est donc pas idéale pour l'utilisation d'une approximation de Fourier. Il est alors normal qu'une méthode d'approximation multiéchelle, ayant la même nature corpusculaire offre de meilleures performances.

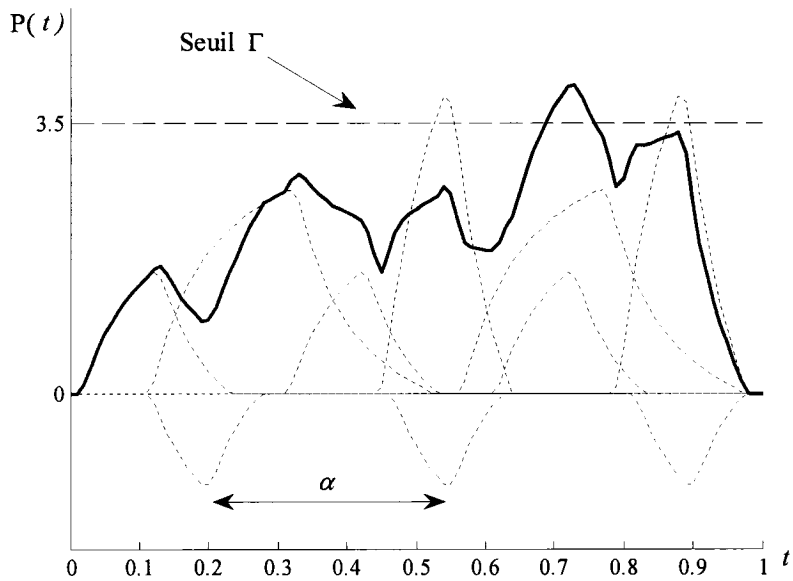


Figure 1.4 : Représentation du niveau de variation de polarisation dans la membrane pour 10 potentiels d'actions.

1.2 Méthode de Laplace

En utilisant une méthode développé par Pierre-Simon de Laplace (1774) appelée méthode de Laplace, il est possible, sous certaines conditions (domaine d'application, image positive, etc.), d'estimer le maximum global d'une fonction en étudiant l'expression $\int_{-\infty}^{\infty} e^{p \cdot f(t)}\, dt$ où p est une constante qui tend vers l'infini.

Comme les seules fonctions exponentielles intégrables analytiquement sur tout le domaine $\mathbb{R}$ sont des fonctions où $f(t)$ est un polynôme d'ordre deux ou moins, cette méthode n'est pas d'une grande utilité. Par contre, le fait d'utiliser une approche

multiéchelle et d'approximer $f(t)$ à l'aide d'un ensemble de paraboles tronquées nous permet d'exprimer cette dernière comme une somme :

$$f(t) = \sum_{k=1}^{n} \theta_k(t) \tag{1.7}$$

où les fonctions $\theta_k(t)$ sont des polynôme d'ordre deux. Nous pourrons alors utiliser la méthode de Laplace.

CHAPITRE 2 – CADRE THÉORIQUE

Notre objectif étant l'approximation de fonctions, il faut définir les différents espaces d'approximation utilisés, leurs caractéristiques et leurs bases. Ensuite, nous donnerons la définition d'une analyse multirésolution.

2.1 Espaces d'approximation

2.1.1 Espaces de Banach et Hilbert

Pour définir la distance entre la fonction à approximer et l'approximation, nous travaillons dans un espace de Banach, c'est à dire une espace vectoriel complet qui admet une norme vérifiant les propriétés suivantes :

$$1)\ \forall f \in \mathrm{H},\ \|f\| \geq 0 \quad \text{et} \quad \|f\| = 0 \ \Leftrightarrow\ f = 0, \tag{2.1}$$

$$2)\ \forall \lambda \in \mathrm{C}\ \|\lambda f\| = |\lambda| \|f\|, \tag{2.2}$$

$$3)\ \forall f, g \in \mathrm{H},\ \|f + g\| \leq \|f\| + \|g\|\ . \tag{2.3}$$

Comme nous voulons travailler dans un espace muni d'un produit scalaire afin de pouvoir faire des projections, définir des angles et l'orthogonalité, nous travaillerons dans un espace de Banach muni d'un produit scalaire, soit un espace de Hilbert noté **H**. Les propriétés du produit scalaire sont :

$$1)\ \forall \lambda_1, \lambda_2 \in \mathrm{C},\ \langle \lambda_1 f_1 + \lambda_2 f_2,\ g \rangle = \lambda_1 \langle f_1,\ g \rangle + \lambda_2 \langle f_2,\ g \rangle, \tag{2.4}$$

$$2)\ \langle f,\ g \rangle = \langle g,\ f \rangle^*, \tag{2.5}$$

$$3)\ \langle f,\ f \rangle \geq 0 \text{ et } \langle f,\ f \rangle = 0 \quad \Leftrightarrow \quad f = 0. \tag{2.6}$$

Dans cette recherche, nous travaillons dans l'espace $L^2(\mathbb{R})$ des fonctions f continues et d'énergie finie sur $\mathbb{R}$. Dans cet espace, qui est également un espace de Hilbert, le produit scalaire de deux fonctions continues $f(t)$ et $g(t)$ est défini comme étant

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)\, g^*(t)\, dt\ . \tag{2.7}$$

Comme nous ne travaillerons qu'avec des fonctions réelles, la définition utilisée sera

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)\, g(t)\, dt\ . \tag{2.8}$$

À ce moment-ci, $\| f \| = \langle f, f \rangle^{1/2}$ est une norme qui est définie par

$$\| f \| = \left(\int_{-\infty}^{\infty} \left(f(t) \right)^2 dt \right)^{1/2} < \infty. \tag{2.9}$$

Comme cet espace est un espace d'énergie finie la norme existe et est toujours inférieure à l'infini. Toujours dans l'espace $L^2(\mathbb{R})$, l'erreur quadratique entre deux fonctions f_1 et f_2 est calculée selon

$$\varepsilon = \| f_1 - f_2 \|^2 = \int_{-\infty}^{\infty} (f_1(t) - f_2(t))^2 \, dt \, . \tag{2.10}$$

Si $\varepsilon = 0$, alors f_1 et f_2 sont égales dans $L^2(\mathbb{R})$, ce qui veut dire que $f_1(t) = f_2(t)$ sur presque tout $t \in \mathbb{R}$.

2.1.2 Bases orthogonales des espaces de Hilbert

Il est important de définir les bases orthogonales, car idéalement c'est avec celles-ci que l'on voudrait travailler. Dans un espace de Hilbert **H**, une famille $\{e_n\}_{n\in\mathrm{N}}$ est orthogonale si

$$\langle e_n, e_p \rangle = 0 \tag{2.11}$$

pour $n \neq p$. Cette famille $\{e_n\}_{n\in\mathrm{N}}$ formera une base orthogonale de **H** s'il existe une suite $\lambda[n]$ telle que pour tout $f \in \mathbf{H}$

$$\lim_{N\to\infty} \left\| f - \sum_{n=0}^{N} \lambda[n] e_n \right\| = 0 \tag{2.12}$$

où

$$\lambda[n]=\frac{\langle f,e_n\rangle}{\|e_n\|^2}. \tag{2.13}$$

De plus, si $\|e_n\|=1$ pour tout $n\in\mathbb{N}$, la base sera orthonormée. Un bon exemple de base de Hilbert orthonormée est une série de Fourier.

2.1.3 Bases de Riesz

S'il n'est pas possible d'avoir une base orthogonale, l'idéal est de travailler avec une base de Riesz. Une base de Riesz est une famille de vecteurs $\{e_n\}_{n\in\mathbb{N}}$ linéairement indépendants qui engendre un espace de Hilbert de dimension finie dans lequel les conditions d'orthogonalité ont été relâchées. Comme l'orthogonalité n'est plus présente, une condition d'équivalence d'énergie doit être imposée pour assurer la stabilité de la base, ce qui est primordial lors de l'approximation de fonctions en utilisant une approche multiéchelle. Avec

$$f=\sum_{n=0}^{\infty}\langle f,e_n\rangle e_n, \tag{2.14}$$

cette condition exige qu'il existe $A > 0$ et $B > 0$ tels que, pour tout $f\in\mathbf{H}$, nous ayons

$$A\| f \|^2 \leq \sum_n |\langle f, e_n \rangle|^2 \leq B\| f \|^2 . \qquad (2.15)$$

Dans le cas des bases orthogonales, A = B = 1.

2.2 Analyse multirésolution

L'analyse multirésolution, qui est la base de notre étude, est une forme particulière de représentation multiéchelle de fonctions qui est appliquée principalement dans le cadre des ondelettes et qui permet de ne traiter que les détails intéressants d'une résolution donnée. La définition mathématique des analyses multirésolutions a été introduite par Mallat (1989) et Meyer (1992). Pour éviter toute confusion lors du changement de résolution dyadique (du facteur de 2), nous faisons référence à une résolution 2^j qui est l'équivalent d'une échelle 2^{-j}.

L'ensemble des sous-espaces fermés $\{ V_j \}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$ forme alors une analyse multirésolution si ces sous-espaces possèdent les propriétés suivantes :

1) Invariance à la translation : V_j est invariant pour toutes translations proportionnelles à une résolution 2^j ;

$$\forall (j,n) \in \mathbb{Z} \ , \ f(t) \in V_j \quad \Leftrightarrow \quad f(t - 2^{-j} n) \in V_j . \qquad (2.16)$$

2) Dilatation par 2 : Le sous espace V_j est l'image de V_{j-1} par une dilatation d'un facteur 2 ;

$$\forall j \in \mathbb{Z} \ , \ f(t) \in V_j \quad \Leftrightarrow \quad f\left(\frac{t}{2}\right) \in V_{j-1}. \tag{2.17}$$

3) Inclusion : Une approximation à une résolution 2^j contient toutes les informations nécessaires au calcul d'une approximation à une échelle plus grossière 2^{j-1} ;

$$\forall j \in \mathbb{Z} \ , \ V_{j-1} \subset V_j . \tag{2.18}$$

4) Intersection nulle : Lorsque la résolution du sous-espace V_j devient infiniment petite, nous perdons complètement notre signal ;

$$\lim_{j \to -\infty} V_j = \bigcap_{j=-\infty}^{\infty} V_j = \{ 0 \}. \tag{2.19}$$

5) Densité dans $L^2(\mathbb{R})$: Dans le même sens que 5), si la résolution de V_j devient infiniment grande, nous reconstituerons exactement notre signal ;

$$\lim_{j \to \infty} V_j = \text{ Adhérence}\left(\bigcup_{j=-\infty}^{\infty} V_j \right) = L^2(\mathbb{R}). \tag{2.20}$$

6) Base de Riesz : Il existe une fonction $\theta(t)$ tel que $\{ \theta(t-n) \}_{n \in \mathbb{Z}}$ forme une base de Riesz de V_0.

La fonction $\theta(t)$ peut s'interpréter comme une cellule unitaire et l'existence d'une base de Reisz assure la stabilité numérique de notre approximation. Pour avoir une analyse multirésolution proprement dite, la base du sous-espace V_0 doit être orthogonale et sera construite à partir d'une base de Riesz orthogonalisée que l'on appelle fonction d'échelle et qui sera notée $\phi(\cdot)$. La base de Riesz est orthogonalisée pour donner la fonction d'échelle par la relation suivante :

$$\hat{\phi}(\omega) = \frac{\hat{\theta}(\omega)}{\sqrt{\beta(\omega)}} \tag{2.21}$$

où,

$$\beta(\omega) = \sum_{k=-\infty}^{\infty} \left|\hat{\theta}(\omega - 2\pi k)\right|^2 . \tag{2.22}$$

En généralisant V_0 à tous les sous-espaces, la famille $\{\phi_{j,n}\}_{j,n\in\mathbb{Z}}$ est une base orthonormée de V_j où

$$\phi_{j,n}(t) = \frac{1}{\sqrt{2^{-j}}}\phi\left(\frac{t - 2^{-j}n}{2^{-j}}\right). \tag{2.23}$$

À ce moment-ci, l'approximation à une résolution 2^j est la projection orthogonale de f sur V_j et elle s'obtient par la décomposition sur la base orthonormée de la fonction d'échelle :

$$P_{V_j} f = \sum_{n=-\infty}^{\infty} \langle f, \phi_{j,n} \rangle \phi_{j,n} . \tag{2.24}$$

CHAPITRE 3 – DE L'APPROXIMATION À L'OPTIMISATION

En l'an 48 avant Jésus-Christ, Pompée établit son camp sur une colline de Pharsale et attend César pour la bataille qui allait mettre un terme à la guerre civile faisant rage entre les tenants de ces deux généraux romains. « Pompée prêta serment de ne revenir que vainqueur et tout le monde célébrait d'avance la victoire, car il leur semblait impossible qu'un général si habile donnât des assurances qui fussent sans fondement. »[3] Mais Pompée subit néanmoins la défaite aux mains de César.

L'erreur que Pompée fit avant de cette bataille fut de ne pas estimer convenablement la valeur de l'armée ennemie et l'astuce de César. Bien qu'ici il ne soit pas question de guerre, notre but est de faire ce que Pompée n'a pas fait : examiner une situation (décrite dans notre cas par des équations mathématiques) et utiliser une méthodologie spécifique pour paramétrer les éléments concernés (les quantrons) dans le but de fournir une approximation qui soit la plus juste possible de la fonction P(t). La méthode d'approximation employée sera une méthode multiéchelle, basée sur les analyses multirésolutions. L'approche multiéchelle nous permet de répondre au critère de la problématique : fournir une approximation analytique formée de paraboles positives. La démarche sera expliquée pour l'approximation à une échelle avant de poursuivre avec l'extension multiéchelle.

[3] César, *Guerre civile*, III, 88

3.1 Approximation à une échelle

3.1.1 Base d'approximation

La fonction de base, notée $\theta(t)$, nous permet de décomposer une fonction quelconque, notée $f(t)$, en effectuant un produit scalaire entre les deux fonctions concernées selon l'équation 2.8. Un bon exemple est le cas des approximations de Fourier, où la projection de $f(t)$ se fait avec des sinusoïdes de différentes fréquences. Dans le cas de la polarisation de la membrane d'un quantron post-synaptique, la projection se fait avec une série de paraboles tronquées qui engendre les sous-espaces d'approximation V_j.

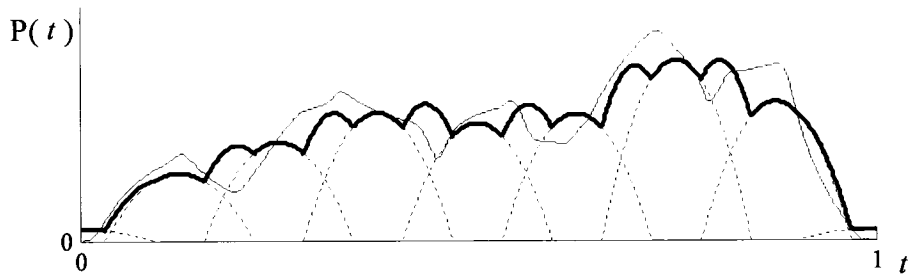


Figure 3.1 : Projection de P(t) sur le sous-espace V_3.

L'idée derrière la construction d'une parabole comme fonction de base $\theta(t)$ est de reproduire les conditions similaires pour l'application du théorème de la valeur moyenne. D'après Swokowski (1993), ce théorème, développé dans le cadre de l'intégrale telle que définie par Riemann, stipule que si $f(t)$ est continue sur un

intervalle fermé [*a*, *b*], alors il existe une point *z* de l'intervalle ouvert]*a*, *b*[pour lequel

$$f(z) = \frac{1}{b-a}\int_a^b f(t)\,dt\,. \tag{3.1}$$

Le lemme 3.1 nous permet d'utiliser ce théorème pour définir l'intégrale de Riemann en termes d'approximation multirésolution. La preuve de ce lemme utilise la première fonction développée dans le cadre des analyses multirésolutions par Haar (1910). La fonction de base de Haar ($\phi(t)$ sur la figure 3.2) est définie par $u(t)-u(t-1)$, où $u(\cdot)$ est la fonction échelon. Avec la structure des analyses multirésolutions, les fonctions de Haar qui engendrent les sous-espaces V_j sont définies par :

$$\phi_{j,n}(t) = \begin{cases} \sqrt{2^j} & \text{si } 0 \le 2^j t - n \le 1 \\ 0 & \text{sinon.} \end{cases} \tag{3.2}$$

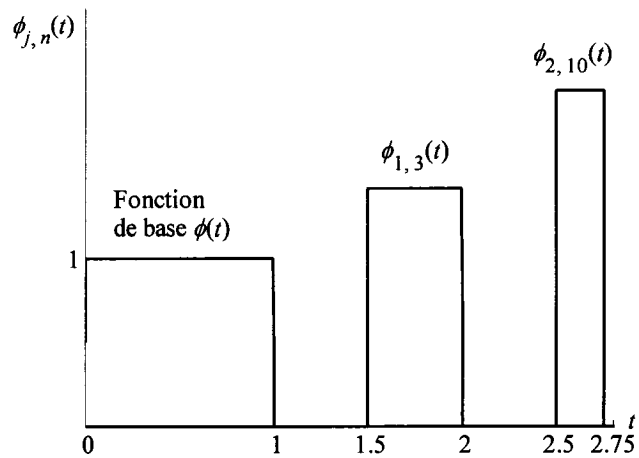


Figure 3.2 : Fonctions d'échelles de Haar pour différentes résolutions.

Lemme 3.1 : En utilisant comme fonction d'échelle $\phi_{j,n}(t)$ la fonction de Haar telle que définie par l'équation 3.2, le résultat de l'approximation à une résolution j fixe est une somme de Riemann.

Preuve : Définissons un rectangle $\eta(t)$ d'une somme de Riemann comme étant le résultat de l'approximation avec une fonction de Haar :

$$\eta(t) = \left\langle f(t), \phi_{j,n}(t) \right\rangle \phi_{j,n}(t), \tag{3.3}$$

où j et n sont des constantes. Le résultat du produit scalaire est :

$$\begin{aligned} \left\langle f(t), \phi_{j,n}(t) \right\rangle &= \int_{-\infty}^{\infty} f(t)\phi_{j,n}(t)\,dt \\ &= \sqrt{2^j} \int_{2^{-j}n}^{2^{-j}(n+1)} f(t)\,dt. \end{aligned} \tag{3.4}$$

Une fois que le coefficient obtenu à partir du produit scalaire est multiplié par la fonction $\phi_{j,n}(t)$, nous avons notre rectangle :

$$\eta(t) = \begin{cases} 2^j \displaystyle\int_{2^{-j}n}^{2^{-j}(n+1)} f(t)\,dt & \text{si} \quad 0 \le 2^j t - n \le 1 \\ 0 & \text{sinon.} \end{cases} \tag{3.5}$$

Or, ce résultat est exactement le même que le résultat du théorème de la valeur moyenne qui nous permet de trouver une valeur $f(z)$ située dans l'intervalle $[2^{-j}n, 2^{-j}(n+1)]$ telle que :

$$f(z) = \frac{1}{(2^{-j}(n+1) - 2^{-j}n)} \int_{2^{-j}n}^{2^{-j}(n+1)} f(t)\,dt$$

$$= 2^{j} \int_{2^{-j}n}^{2^{-j}(n+1)} f(t)\,dt \qquad (3.6)$$

La hauteur d'un rectangle $\eta(t)$ d'une somme de Riemann est la valeur moyenne de la fonction sur l'intervalle $[2^{-j}n, 2^{-j}(n+1)]$ de la fonction d'échelle de Haar, et ce, pour tout le domaine de la fonction $f(t)$. La reconstruction de $f(t)$ à partir d'une analyse multirésolution est alors une somme de Riemann.

■

Les éléments intéressants qu'impliquent le lemme 3.1 sont les propriétés que possède la fonction de Haar $\phi(t)$ et avec lesquelles nous allons construire notre fonction de base $\theta(t)$. Ces propriétés sont les suivantes:

1) Le maximum de la fonction a une valeur de 1 : $\max\{\phi(t)\} = 1$; (3.7)

2) La normalité : $\int_{-\infty}^{\infty} \phi(t)\,dt = 1$. (3.8)

Dans le cas le plus général, notre parabole est décrite par l'équation :

$$\theta(t) = \begin{cases} at^2 + bt + c & \text{si } p_1 \le t \le p_2 \\ 0 & \text{sinon} \end{cases} \tag{3.9}$$

où a, b et c sont des constantes et où p_1 et p_2 sont les zéros de la parabole. Pour réduire le nombre de paramètres, nous centrons notre parabole sur l'origine, ce qui donne $p_2 = -p_1 = p$ et l'équation 3.9 devient

$$\theta(t) = \begin{cases} h\left(p^2 - t^2\right) & \text{si } -p \le t \le p \\ 0 & \text{sinon} \end{cases} \tag{3.10}$$

où p et h sont les paramètres qui ajustent les dimensions de notre parabole en largeur et en hauteur respectivement (voir figure 3.3).

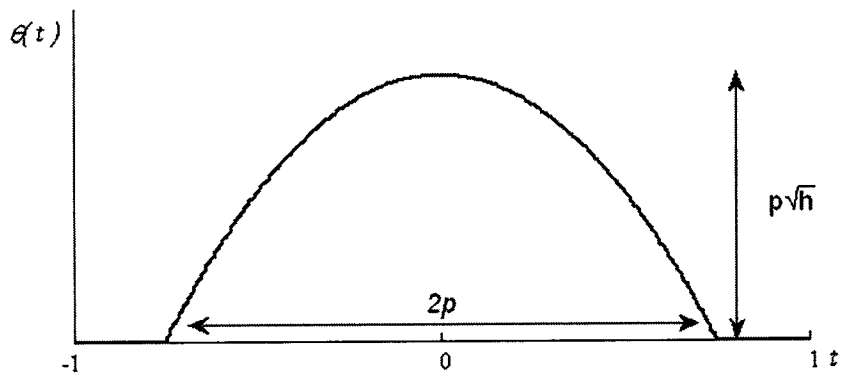


Figure 3.3 : Effet des paramètres p et h sur $\theta(t)$.

En appliquant la première propriété (éq. 3.7), nous avons : $\max\{\theta(t)\} = \theta(0) = 1$ et

$$\theta(0) = h(p^2 - 0^2) = 1 \qquad \Leftrightarrow \qquad p = \frac{1}{\sqrt{h}}. \tag{3.11}$$

Avec la deuxième propriété (éq. 3.8), nous obtenons :

$$\begin{aligned} \int_{-\infty}^{\infty} \phi(t)\,dt &= \int_{-p}^{p} \theta(t)\,dt \\ &= \int_{-p}^{p} h(p^2 - t^2)\,dt \\ &= \frac{4hp^3}{3} \end{aligned} \tag{3.12}$$

et

$$\frac{4hp^3}{3} = 1. \tag{3.13}$$

En solutionnant 3.10 et 3.12 on obtient $p = 3/4$ et $h = 16/9$ pour finalement avoir la fonction de base suivante :

$$\theta(t) = \begin{cases} 1 - \dfrac{16t^2}{9} & \text{si } -\frac{3}{4} \le t \le \frac{3}{4} \\ 0 & \text{sinon.} \end{cases} \tag{3.14}$$

L'application de la propriété d'invariance sur toutes les translations de longueur proportionnelle à la résolution (éq. 2.16) donne une fonction paramétrée par la résolution 2^j et une translation n :

$$\theta_{j,n}(t)=\begin{cases} 1-\dfrac{16\left(t-\dfrac{n}{2^j}\right)^2}{9} & \text{si } -\tfrac{3}{4}\le t-\dfrac{n}{2^j}\le\tfrac{3}{4} \\ 0 & \text{sinon.} \end{cases} \tag{3.15}$$

L'application de la propriété de dilatation (éq. 2.17) permettra le changement de résolution en agrandissant les détails d'un facteur 2 lorsqu'on augmente la résolution. La fonction est alors paramétrée par le changement de résolution 2^j et donne :

$$\theta_j(t)=\begin{cases} 1-\dfrac{16\left(2^j t\right)^2}{9} & \text{si } -\tfrac{3}{4}\le 2^j t\le\tfrac{3}{4} \\ 0 & \text{sinon.} \end{cases} \tag{3.16}$$

La figure 3.4 illustre bien ces deux propriétés.

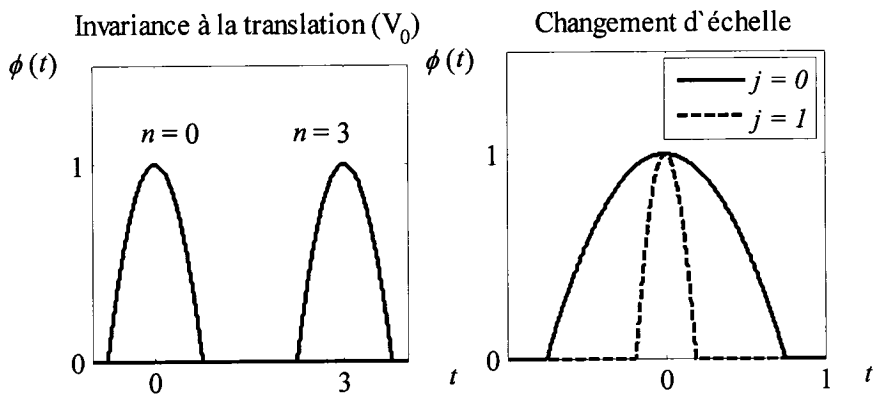


Figure 3.4 : Propriété de translation et de dilatation de la fonction de base.

Une fois que ces deux éléments sont incorporés à la fonction de base et que le terme $\sqrt{2^j}$ est ajouté pour la conservation d'énergie entre les différentes résolutions, nous

avons une famille de fonctions pouvant engendrer les sous-espaces V_j tel qu'illustré à la figure 3.5 :

$$\theta_{j,n}(t)=\begin{cases}\sqrt{2^j}\left(1-\dfrac{16\left(2^j t-n\right)^2}{9}\right) & \text{si } -\tfrac{3}{4}\le 2^j t-n\le\tfrac{3}{4}\\ 0 & \text{sinon.}\end{cases} \tag{3.17}$$

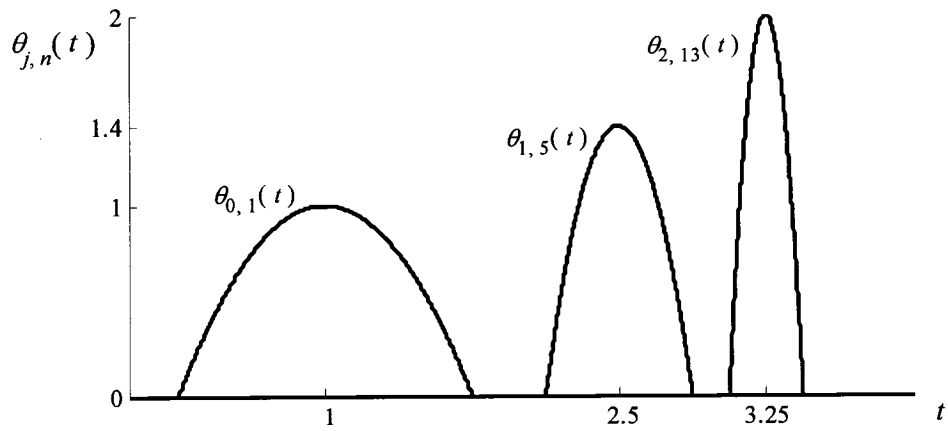


Figure 3.5 : Exemple de fonction $\theta_{j,n}(t)$ pour différentes valeurs de j et n.

Bien que nous ayons une fonction base qui possède les propriétés permettant à la fonction de Haar de reproduire un élément d'une somme de Riemann et bien que nous nous soyons munis d'une structure respectant l'esprit des analyses multirésolutions pour reproduire la sommation en question, nous devons nous assurer que le sommet de la parabole indique véritablement un point sur la courbe. Le théorème suivant permet d'établir ce résultat.

Théorème 3.1 : Le sommet d'une parabole telle que décrite par $\theta_{j,n}(t)$ qui approxime la fonction $f(t)$ et qui est centrée sur un point c tend vers $f(c)$.

Preuve : Lorsque la résolution j tend vers l'infini, la largeur du support de notre parabole (où $\theta_{j,n}(t) \neq 0$) tend vers zéro. Sur cet intervalle très petit, la fonction à approximer se ramène à une droite ayant comme équation $f(t) = mx + b$. En un point c, la valeur moyenne de cette droite sera :

$$\begin{aligned} f(c) &= \frac{1}{(c+i-c+i)} \int_{c-i}^{c+i} (mt+b)\,dt \\ &= \frac{1}{2i}\left(\frac{m\left((c+i)^2-(c-i)^2\right)}{2} + 2bi \right) \\ &= mc + b, \end{aligned} \tag{3.18}$$

soit la valeur de $f(t)$ au centre de l'intervalle, à $t = c$.

Selon la propriété 1 des analyses multirésolutions (éq. 2.16), un déplacement c est exprimé en fonction de n par la relation $c = 2^{-j}n$. Alors, l'approximation de $f(t)$ à l'aide de la fonction de base $\theta_{j,n}(t)$ donne une parabole d'une hauteur $\sqrt{2^j}$, déplacée de $n = 2^j c$ et multipliée par le produit scalaire entre $f(t)$ et $\theta_{j,n}(t)$. La valeur du sommet de cette parabole centrée sur un point c est :

$$\begin{aligned}
\theta_{j,n}(c) &= \sqrt{2^j}\left\langle f(t), \theta_{j,n}(t)\right\rangle \\
&= \sqrt{2^j}\int_{\left(2^{-j}c-2^{-j}\frac{3}{4}\right)}^{\left(2^{-j}c+2^{-j}\frac{3}{4}\right)} (mt+b)\sqrt{2^j}\left(1-\frac{16}{9}\left(2^j t-2^j c\right)^2\right)dt \\
&= \frac{2^j t}{54}\Big[\ 27mt-24mt^3 4^j+64mt^2 2^j c-48mt^2c^2 \\
&\qquad\qquad +54b-32bt^2 4^j+96bt\,2^j c\ \Big|_{\left(2^{-j}c-2^{-j}\frac{3}{4}\right)}^{\left(2^{-j}c+2^{-j}\frac{3}{4}\right)} \\
&= mc+b. \qquad (3.19)
\end{aligned}$$

soit la valeur de la fonction au point c.

■

En effectuant des translations de $2^{-j}n$, où n est un entier, nous avons un recouvrement de deux paraboles, donc une base qui n'est pas orthogonale. Un autre aspect important de notre démarche est de trouver un pas qui minimise l'erreur quadratique d'approximation due au recouvrement. Pour ce faire, nous approximons la fonction $f(t) = 1$. Tel qu'illustré à la figure 3.6, le calcul de l'erreur se fait à une résolution $j = 0$, du sommet d'une parabole centrée à l'origine au sommet d'une autre parabole centrée à une distance τ de l'origine.

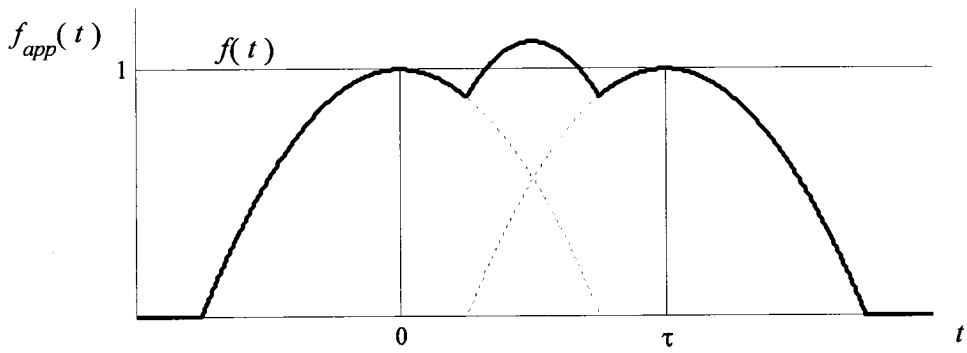


Figure 3.6 : Calcul de l'erreur du sommet d'une parabole centrée à $t = 0$ au sommet d'une parabole centrée à $t = \tau$.

Avec une erreur quadratique telle que définie par l'équation 2.10, la convergence de l'approximation $f_{app}(t)$ vers f dans **H** signifie que :

$$f_{app} = f \quad \Leftrightarrow \quad \varepsilon = 0 \, . \tag{3.20}$$

Dans ce cas spécifique, la fonction $f(t)$ varie selon τ et nous utilisons donc une erreur quadratique absolue qui sera une fonction de τ. Cette erreur absolue est décrite par :

$$\varepsilon(\tau) = \frac{\| f(t) - f_{app}(t) \|^2}{\| f(t) \|^2} = \frac{\int_0^\tau \left(f(t) - f_{app}(t) \right)^2 dt}{\int_0^\tau \left(f(t) \right)^2 dt}$$

$$= \frac{\int_0^\tau \left(f(t) - f_{app}(t) \right)^2 dt}{\tau} \tag{3.21}$$

où

$$f_{app}(t) = \theta_{0,0}(t) + \theta_{0,0}(t-\tau). \tag{3.22}$$

Comme il y a recouvrement pour des valeurs de $\tau = [0, 1.5]$, la fonction d'erreur montrée à la figure 3.7, devient :

$$\varepsilon(\tau) = \begin{cases} \dfrac{\int_0^{\tau} \left(f(t) - f_{app}(t)\right)^2 dt}{\tau} & \text{si} \quad 0 \le \tau \le 1.5 \\ 0 & \text{sinon} \end{cases} \tag{3.23}$$

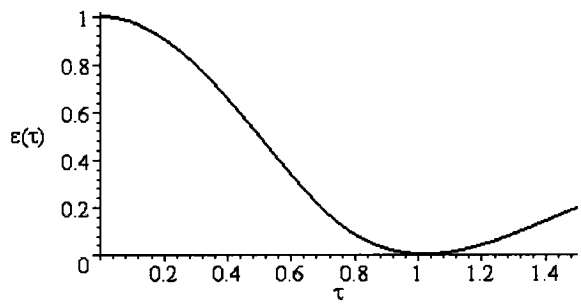


Figure 3.7 : Erreur d'approximation en fonction de τ .

Une fois optimisée, cette fonction indique un pas optimal de $\tau = 1.019$. Comme cette valeur est très proche de $\tau = 1$, que la variation de $\varepsilon(\tau)$ entre ces deux valeurs de τ n'est pas grande et par souci de simplicité, on conserve une incrémentation unitaire de n pour la translation des paraboles.

Tout est maintenant en place pour le prochain lemme qui va séparer l'approche multiéchelle formée de paraboles tronquées d'une analyse multirésolution. En effet, le lemme suivant indique que les sous-espaces engendrés par cette base ne respectent

pas la propriété d'inclusion (éq. 2.18). Par conséquent, l'approximation multiéchelle faite avec la base $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ n'est pas une analyse multirésolution.

La propriété d'inclusion est généralement vérifiée si on peut construire $\theta_{j,n}(t)$ à partir d'une combinaison linéaire de $\theta_{j+1,n}(t)$. Dans le cas de Haar, comme le démontre la figure 3.8, la simplicité de la fonction fait en sorte que la solution est triviale.

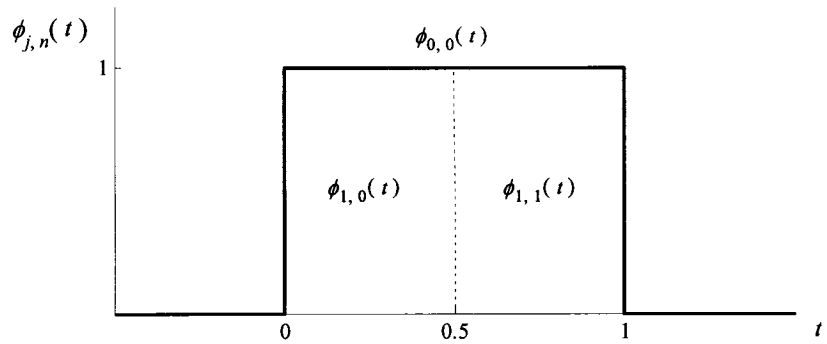


Figure 3.8 : Reconstruction de $\theta_{0,0}(t)$ avec $\theta_{1,0}(t)$ et $\theta_{1,1}(t)$.

Lemme 3.2 : Les sous-espaces V_j engendrés par la famille de fonctions $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ ne satisfont pas la propriété d'inclusion : $\forall j \in Z$, $V_j \subset V_{j+1}$.

Preuve : Pour qu'il y ait inclusion, on doit pouvoir construire la fonction $\theta_{j,0}(t)$ avec un ensemble de fonctions $\theta_{j+1,n}(t)$. Par contre, comme la fonction de base est une parabole délimitée par l'intervalle [$2^{-j}(n$ - ¾), $2^{-j}(n$ + ¾)], elle possède une discontinuité à chaque extrémité de cet intervalle. La fonction $\theta_{j,0}(t)$, qui est une

fonction continue sur l'intervalle [-¾, ¾], ne peut donc pas être exactement reconstruite.

■

3.1.2 Stabilité de la base

La non-orthogonalité de la base des sous-espaces V_j et le fait que nous n'ayons pas une analyse multirésolution n'empêchent pas que nous ayons besoin d'une approximation stable. Une condition nécessaire mais non suffisante de la stabilité d'une base est l'indépendance linéaire des vecteurs formant la base en question. Le lemme suivant nous donne cette propriété importante pour une incrémentation unitaire de n.

Lemme 3.3 : L'ensemble de fonctions qui engendrent un sous-espace V_j forment un système linéairement indépendant.

Preuve : Le pas de translation pour une résolution j de $2^{-j}n$, où $n \in \mathbb{Z}$, fait en sorte que seulement deux paraboles à la fois vont se superposer. Si on considère deux paraboles $\theta_{j,\,n+k}(t)$ et $\theta_{j,\,n+k+2}(t)$, comme l'indique la figure 3.9, le seul moyen d'aller chercher l'information comprise dans l'intervalle $[2^{-j}(n + k + ¾), 2^{-j}(n + k + 2 - ¾)]$ est d'inclure une troisième parabole : $\theta_{j,\,n+k+1}(t)$.

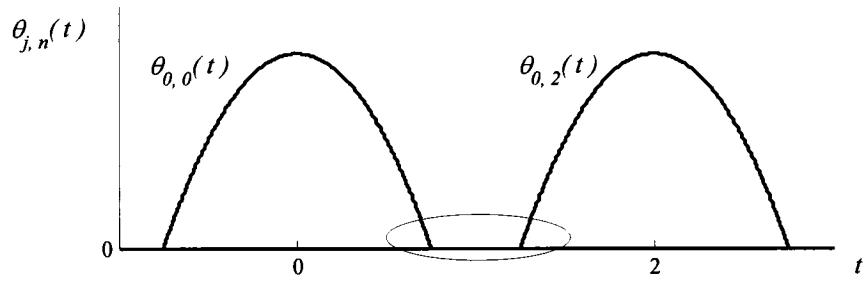


Figure 3.9 : Indépendance linéaire.

Ainsi, l'équation

$$\sum_{n=-\infty}^{\infty} c_{j,n}\,\theta_{j,n}(t) = 0 \quad \Leftrightarrow \quad c_{j,n} = 0 \quad \forall n \in \mathrm{N} \tag{3.24}$$

définissant l'indépendance linéaire est satisfaite.

■

L'indépendance linéaire n'est qu'un prérequis à l'élément qui va assurer la stabilité des approximations avec les fonctions $\theta_{j,n}(t)$: la présence d'une base de Riesz. À partir de la définition d'une base de Riesz (éq. 2.15), Daubechies (1992) a démontré qu'une condition nécessaire et suffisante pour qu'une famille de fonctions $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ soit une telle base d'un sous espace de $L^2(\mathbb{R})$ est si et seulement si il existe $A > 0$ et $B > 0$ tels que :

$$\forall \omega \in [-\pi, \pi], \quad \frac{1}{B} \le \beta(\omega) \le \frac{1}{A} \tag{3.25}$$

où $\beta(\omega)$ est définie par l'équation 2.22.

Le théorème suivant démontre la présence d'une base de Riesz qui engendre le sous-espace V_0, et qu'avec 2.16 et 2.17, la stabilité de la décomposition de $f(t)$ sur V_0 s'étend à tous les sous-espaces V_j.

Théorème 3.2 : La famille de fonctions $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ est une base de Riesz qui engendre le sous-espace V_0.

Preuve : Pour établir que $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ est une base de Riesz, on doit satisfaire deux conditions : premièrement, nos fonctions doivent former un ensemble de vecteurs linéairement indépendants, et deuxièmement, il faut prouver qu'il existe bel et bien des bornes A et B qui satisfont l'équation 3.25. Le lemme 3.3 nous permet de respecter la première condition. Pour respecter la deuxième, il faut commencer par définir la transformée de Fourier de $\theta(t)$, illustrée à la figure 3.10, telle que:

$$\begin{aligned}\hat{\theta}(\omega) &= \int_{-\infty}^{\infty} \theta(t) e^{i\omega t}\, dt \\ &= \int_{-\frac{3}{4}}^{\frac{3}{4}} \left(1-\frac{16}{9}t^2\right) e^{i\omega t}\, dt \\ &= -\frac{8}{9}\frac{4i\left(e^{3/4\, i\omega} - e^{-3/4\, i\omega}\right) + 3\omega\left(e^{3/4\, i\omega} + e^{-3/4\, i\omega}\right)}{\omega^3}\end{aligned} \tag{3.26}$$

En simplifiant les nombres complexes :

$$\hat{\theta}(\omega) = \frac{64}{9}\frac{\sin\left(\frac{3\omega}{4}\right)}{\omega^3} - \frac{16}{3}\frac{\cos\left(\frac{3\omega}{4}\right)}{\omega^2} \tag{3.27}$$

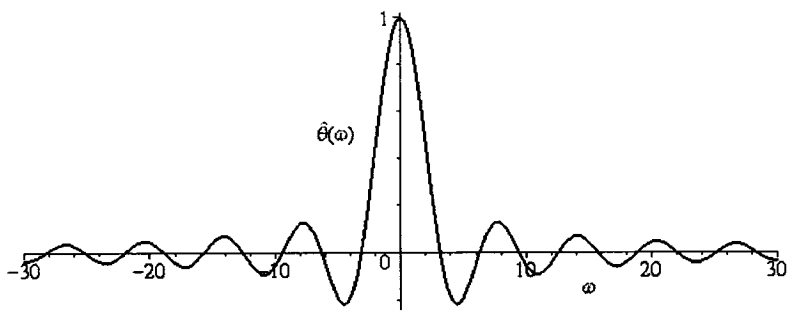


Figure 3.10 : Transformée de Fourier de $\theta(t)$.

À partir de la transformée de Fourier de notre fonction de base, le terme à l'intérieur de la somme formant $\beta(\omega)$ (éq. 2.22), nommé $b(\omega, k)$ est calculé de la manière suivante :

$$\begin{aligned} b(\omega,k) &= \left|\hat{\theta}(\omega - 2\pi k)\right|^2 \\ &= \hat{\theta}(\omega - 2\pi k)^2 \\ &= \left(\frac{64}{9}\frac{\sin\left(\frac{3(\omega-2\pi k)}{4}\right)}{\omega^3} - \frac{16}{3}\frac{\cos\left(\frac{3(\omega-2\pi k)}{4}\right)}{\omega^2}\right)^2 . \end{aligned} \tag{3.28}$$

En effectuant la multiplication et les substitutions trigonométriques appropriées dans l'équation 3.28 :

$$b(\omega,k)=\frac{4096}{81}\frac{\sin\left(\frac{3\omega}{4}\right)^2\cos\left(\frac{3\pi k}{2}\right)^2}{(\omega+2\pi k)^6}-\frac{2048}{27}\frac{\sin\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\pi k}{2}\right)^2}{(\omega+2\pi k)^5}$$

$$+\frac{256}{9}\frac{\cos\left(\frac{3\omega}{4}\right)^2\cos\left(\frac{3\pi k}{2}\right)^2}{(\omega+2\pi k)^4}+\frac{4096}{81}\frac{\cos\left(\frac{3\omega}{4}\right)^2\sin\left(\frac{3\pi k}{2}\right)^2}{(\omega+2\pi k)^6}$$

$$+\frac{2048}{27}\frac{\sin\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\omega}{4}\right)\sin\left(\frac{3\pi k}{2}\right)^2}{(\omega+2\pi k)^5}+\frac{256}{9}\frac{\sin\left(\frac{3\omega}{4}\right)^2\sin\left(\frac{3\pi k}{2}\right)^2}{(\omega+2\pi k)^4}$$

$$+\frac{2048}{27}\frac{\sin\left(\frac{3\omega}{4}\right)^2\sin\left(\frac{3\pi k}{2}\right)\cos\left(\frac{3\pi k}{2}\right)}{(\omega+2\pi k)^5}-\frac{2048}{27}\frac{\cos\left(\frac{3\omega}{4}\right)^2\sin\left(\frac{3\pi k}{2}\right)\cos\left(\frac{3\pi k}{2}\right)}{(\omega+2\pi k)^5}$$

$$+\frac{8192}{81}\frac{\sin\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\omega}{4}\right)\sin\left(\frac{3\pi k}{2}\right)\cos\left(\frac{3\pi k}{2}\right)}{(\omega+2\pi k)^6}$$

$$-\frac{512}{9}\frac{\sin\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\omega}{4}\right)\sin\left(\frac{3\pi k}{2}\right)\cos\left(\frac{3\pi k}{2}\right)}{(\omega+2\pi k)^4}. \qquad (3.29)$$

Dans l'expression 3.29, les valeurs paires de k impliquent que $\sin(3\pi\ k/2) = 0$ et $\cos(3\pi\ k/2)^2 = 1$; les valeurs impaires de k impliquent de leur côté que $\cos(3\pi\ k/2) = 0$ et $\sin(3\pi\ k/2)^2 = 1$. À ce moment-ci, le terme $b(\omega, k)$ se sépare en deux termes, soit $b_{pair}(\omega,k)$ et $b_{impair}(\omega,k)$ tels que:

$$b_{pair}(\omega,k)=\frac{4096}{81}\frac{\sin\left(\frac{3\omega}{4}\right)^2}{(\omega+2\pi k)^6}-\frac{2048}{27}\frac{\sin\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\omega}{4}\right)}{(\omega+2\pi k)^5}+\frac{256}{9}\frac{\cos\left(\frac{3\omega}{4}\right)^2}{(\omega+2\pi k)^4}, \qquad (3.30)$$

$$b_{impair}(\omega,k)=\frac{4096}{81}\frac{\cos\left(\frac{3\omega}{4}\right)^2}{(\omega+2\pi k)^6}+\frac{2048}{27}\frac{\sin\left(\frac{3\omega}{4}\right)\cos\left(\frac{3\omega}{4}\right)}{(\omega+2\pi k)^5}+\frac{256}{9}\frac{\sin\left(\frac{3\omega}{4}\right)^2}{(\omega+2\pi k)^4}. \quad (3.31)$$

Dès lors, la somme

$$\beta(\omega)=\sum_{k=-\infty}^{\infty} b(\omega,k) \quad (3.32)$$

peut être effectuée en divisant celle-ci en deux sommations distinctes, soit une pour les valeurs paires de k et une autre pour les valeurs impaires de k:

$$\beta(\omega)=\sum_{k=-\infty}^{\infty} b_{paire}(\omega,2k)+\sum_{k=-\infty}^{\infty} b_{impaire}(\omega,2k+1) \quad (3.33)$$

où :

$$\sum_{k=-\infty}^{\infty} b_{paire}(\omega,2k)=\frac{2}{1215}+\frac{992}{1215}\cos\left(\frac{\omega}{4}\right)^4+\frac{224}{1215}\cos\left(\frac{\omega}{4}\right)^2 \quad (3.34)$$

et

$$\sum_{k=-\infty}^{\infty} b_{paire}(\omega,2k)=\frac{406}{405}+\frac{992}{1215}\cos\left(\frac{\omega}{4}\right)^4-\frac{736}{405}\cos\left(\frac{\omega}{4}\right)^2. \quad (3.35)$$

Finalement, nous avons :

$$\beta(\omega)=\frac{244}{243}+\frac{1984}{1215}\cos\left(\frac{\omega}{4}\right)^4-\frac{1984}{1215}\cos\left(\frac{\omega}{4}\right)^2 \quad (3.36)$$

qui, en simplifiant, donne :

$$\beta(\omega) = \frac{4}{243}\left(61 - \frac{124}{5}\sin\left(\frac{\omega}{4}\right)^2\right). \tag{3.37}$$

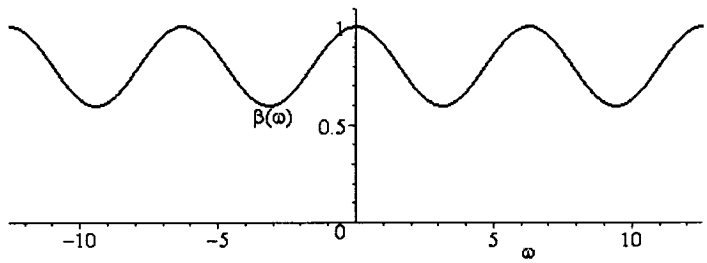


Figure 3.11 : La fonction $\beta(\omega)$.

La périodicité de $\beta(\omega)$ fait en sorte qu'il existe des valeurs A et B qui satisfont l'équation 3.25, démontrant ainsi l'existence d'une base de Riesz. Ces valeurs sont A = 244/243 et B = 724/1215.

■

3.1.3 Reconstruction et convergence

Bien que la famille de fonctions $\{\theta_{j,n}(t)\}_{j,n\in\mathrm{N}}$ forme une base de Riesz, l'absence d'analyse multirésolution fait en sorte que l'approximation d'une fonction $f(t)$ est la projection non-orthogonale sur le sous-espace V_j illustrée à la figure 3.12 et définie par :

$$f_j(t) = \sum_{n=-\infty}^{\infty} \left\langle f(t), \theta_{j,n}(t) \right\rangle \theta_{j,n}(t) \tag{3.38}$$

où

$$\left\langle f(t), \theta_{j,n}(t) \right\rangle = \int_{-\infty}^{-\infty} f(t)\theta_{j,n}(t)\,dt\,. \tag{3.39}$$

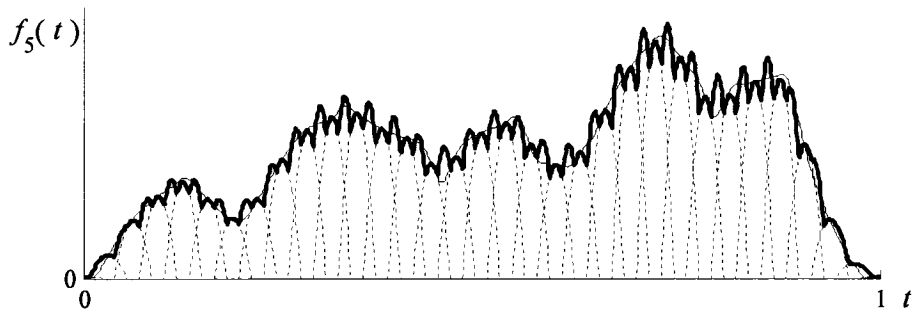


Figure 3.12 : Projection non-orthogonale de P(t) sur V_5.

Une fois cette décomposition faite, nous obtenons pour une résolution j fixe un vecteur de coefficients dont la $n^{\text{ième}}$ valeur est, encore une fois, le résultat du produit scalaire entre $f(t)$ et $\theta_{j,n}(t)$. Mallat (1989) montre qu'en multipliant ce vecteur de coefficients par $\sqrt{2^j}$, on peut faire une approximation discrète de $f(t)$, tel qu'illustré par la figure 3.13.

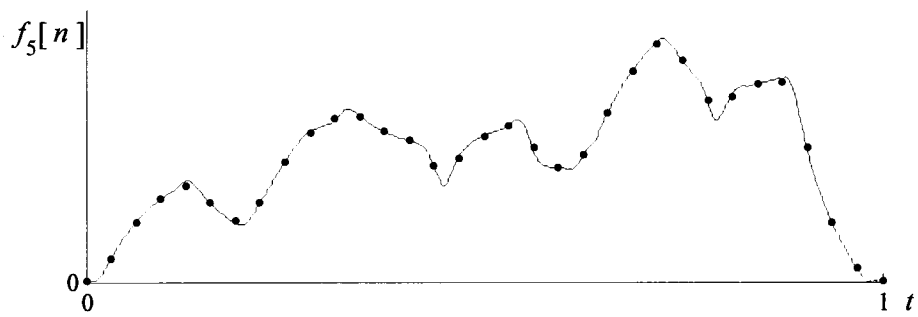


Figure 3.13 : Approximation discrète de P(t) pour $j = 5$.

Notons ici que le maximum global de la fonction peut être trouvé en localisant le coefficient le plus élevé du vecteur et en le multipliant par $\sqrt{2^j}$. Ce maximum est situé à $t = 2^{-j} n$. Dans le cas des figures 3.12 et 3.13, d'après une approximation à $j = 5$, le maximum global de la fonction est 3.9051 et est situé à $t = 0.7188$.

Une des propriétés intéressantes des approximations multirésolutions est la convergence vers une valeur nulle de l'erreur quadratique d'approximation à une résolution j donnée lorsque j tend vers l'infini, c'est à dire :

$$\lim_{j\to\infty} f_j = f \quad \Leftrightarrow \quad \lim_{j\to\infty} \| f - f_j \|^2 = 0. \tag{3.40}$$

Comme nous n'avons pas une approximation multirésolution, l'erreur de l'approximation multiéchelle à une résolution j ne tend pas nécessairement vers zéro. Au contraire, le prochain théorème indique qu'avec une fonction d'erreur quadratique, pour une valeur de j quelconque définie par :

$$\varepsilon[\,j\,] = \left\| f(t) - f_j(t) \right\|^2 = \int_0^1 \left(f(t) - f_j(t) \right)^2 dt\,, \tag{3.41}$$

l'erreur tend vers une constante.

La méthodologie derrière la preuve du prochain théorème (3.3) consiste tout d'abord à approximer une fonction avec une série de Fourier, où chacune des sinusoïdes a été approximée par des paraboles. Comme la décomposition par séries de Fourier possède la propriété d'approximation, si l'erreur quadratique de l'approximation d'une sinusoïde tend vers une valeur nulle, la méthode multiéchelle possède également la propriété d'approximation. C'est pourquoi la preuve de convergence sera faite en approximant la fonction $f(t)$ positive :

$$f(t) = \sin(2\pi\, t) + 1. \tag{3.42}$$

Pour conserver la continuité de l'approximation du sinus, l'approximation sera faite sur l'ensemble du domaine. Par contre, comme le montre la figure 3.14, la fonction d'erreur sera seulement considérée sur l'espace normalisé $L^2[0,1]$.

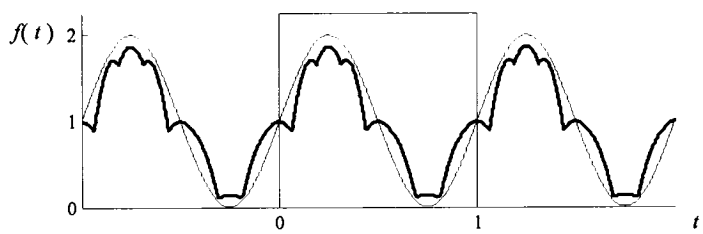


Figure 3.14 : Intervalle défini pour la fonction d'erreur.

Avant d'énoncer le théorème, nous devons établir quelques lemmes concernant l'approximation de fonctions périodiques sur un espace normalisé $L^2[0,1]$. Pour simplifier la notation, le résultat du produit scalaire entre $f(t)$ et la $n^{\text{ième}}$ parabole est identifié par :

$$c_{j,n} = \left\langle f(t), \theta_{j,n}(t) \right\rangle. \tag{3.43}$$

Lemme 3.4 : Pour n'importe quelles fonctions périodiques :

$$c_{j,0} \int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j} \int_0^1 f(t)\theta_{j,2^j}(t)\,dt = {c_{j,0}}^2 \tag{3.44}$$

Preuve : La périodicité de $f(t)$ fait en sorte que la section de cette dernière recoupant la parabole à $n = 0$ est identique à celle recouvrant la parabole à $n = 2^j$. Ceci implique que $c_{j,0} = c_{j,2^j}$ et que $\theta_{j,0}(t)$ est identique à $\theta_{j,2^j}(t)$. Alors :

$$c_{j,0} \int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j} \int_0^1 f(t)\theta_{j,2^j}(t)\,dt = c_{j,0}\left(\int_0^1 f(t)\theta_{j,0}(t)\,dt + \int_0^1 f(t)\theta_{j,2^j}(t)\,dt \right).$$

Avec la propriété d'invariance à la translation, mettre ensemble les demi-paraboles formées par $\theta_{j,0}(t)$ et $\theta_{j,2^j}(t)$ revient à décomposer la fonction $f(t)$ sur la parabole entière $\theta_{j,0}(t)$ sur tout le domaine :

$$c_{j,0}\int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j}\int_0^1 f(t)\theta_{j,2^j}(t)\,dt = c_{j,0}\left(\int_{-\infty}^{\infty} f(t)\theta_{j,0}(t)\,dt\right)$$

$$= c_{j,0}\left\langle f(t), \theta_{j,0}(t)\right\rangle$$

$$= {c_{j,0}}^2.$$

■

Lemme 3.5 : Pour n'importe quelles fonctions périodiques :

$${c_{j,0}}^2\int_0^1 \theta_{j,0}(t)^2\,dt + {c_{j,2^j}}^2\int_0^1 \theta_{j,2^j}(t)^2\,dt = {c_{j,0}}^2\int_{-\infty}^{\infty}\theta_{j,0}(t)^2\,dt \qquad (3.45)$$

Preuve : De la même manière que dans le lemme 3.4 : $c_{j,0} = c_{j,2^j}$; ce qui implique que ${c_{j,0}}^2 = {c_{j,2^j}}^2$. Alors :

$${c_{j,0}}^2\int_0^1 \theta_{j,0}(t)^2\,dt + {c_{j,2^j}}^2\int_0^1 \theta_{j,2^j}(t)^2\,dt = {c_{j,0}}^2\int_{-\infty}^{\infty}\frac{\theta_{j,0}(t)^2}{2}dt + {c_{j,2^j}}^2\int_{-\infty}^{\infty}\frac{\theta_{j,2^j}(t)^2}{2}dt$$

$$= {c_{j,0}}^2\left(\int_{-\infty}^{\infty}\frac{\theta_{j,0}(t)^2}{2}dt + \int_{-\infty}^{\infty}\frac{\theta_{j,2^j}(t)^2}{2}dt\right)$$

Pour les paraboles aux extrémités, l'aire sous la courbe dans l'intervalle [0,1] correspond à la moitié de l'aire totale et grâce à la propriété d'invariance à la translation :

$$c_{j,0}{}^2 \int_0^1 \theta_{j,0}(t)^2\,dt + c_{j,2^j}{}^2 \int_0^1 \theta_{j,2^j}(t)^2\,dt = c_{j,0}{}^2 \left(\int_{-\infty}^{\infty} \frac{\theta_{j,0}(t)^2}{2} dt + \int_{-\infty}^{\infty} \frac{\theta_{j,2^j}(t)^2}{2} dt \right)$$

$$= c_{j,0}{}^2 \left(\int_{-\infty}^{\infty} \frac{\theta_{j,0}(t)^2}{2} dt + \int_{-\infty}^{\infty} \frac{\theta_{j,0}(t)^2}{2} dt \right)$$

$$= c_{j,0}{}^2 \int_{-\infty}^{\infty} \theta_{j,0}(t)^2\,dt$$

■

Lemme 3.6 : Pour n'importe quelles fonctions périodiques :

$$\lim_{j\to\infty} \sum_{n=0}^{2^j-1} c_{j,n}\, c_{j,n+1} = \lim_{j\to\infty} \sum_{n=0}^{2^j-1} c_{j,n}{}^2 . \tag{3.46}$$

Preuve : Puisque l'on considère les coefficients $c_{j,n}$ comme étant des vecteurs, l'intervalle entre chaque vecteur est de 2^{-j}. Lorsque j tend vers l'infini, les vecteurs de la représentation de la fonction $f(t)$ sont de moins en moins espacés et la valeur de $c_{j,n+1}$ se rapproche graduellement de $c_{j,n}$. Alors, à l'infini, nous avons $c_{j,n+1} = c_{j,n}$.

■

Tout est mis en place pour le théorème de la convergence de l'erreur quadratique.

Théorème 3.3 : Pour une erreur quadratique d'une approximation multiéchelle telle que définie par l'équation 3.41 où $f(t)$ est une sinusoïde définie selon l'équation 3.42,

alors l'erreur sur l'espace normalisé $L^2[0,1]$ à une résolution j qui tend vers l'infini est de 1/162.

Preuve : Le développement de la fonction de l'erreur quadratique donne :

$$\begin{aligned}\varepsilon[j] &= \int_0^1 \left(f(t) - f_j(t)\right)^2 dt \\ &= \int_0^1 \left(f(t)^2 - 2f(t)f_j(t) + f_j(t)^2\right) dt \\ &= \underbrace{\int_0^1 f(t)^2 dt}_{1} - 2\underbrace{\int_0^1 f(t)f_j(t)\,dt}_{2} + \underbrace{\int_0^1 f_j(t)^2 dt}_{3}. \end{aligned} \tag{3.47}$$

Le premier terme de l'équation 3.47 se résout facilement :

$$\int_0^1 f(t)^2 dt = \int_0^1 \left(\sin(2\pi t) + 1\right)^2 dt = 3/2. \tag{3.48}$$

En développant le deuxième terme de l'équation 3.47 :

$$\begin{aligned}\int_0^1 f(t)f_j(t)\,dt &= \int_0^1 f(t)\sum_{n=0}^{2^j} \left\langle f(t), \theta_{j,n}(t)\right\rangle \theta_{j,n}(t)\,dt \\ &= \int_0^1 f(t)\sum_{n=0}^{2^j} c_{j,n}\,\theta_{j,n}(t)\,dt \\ &= \sum_{n=1}^{2^j-1} c_{j,n} \int_0^1 f(t)\theta_{j,n}(t)\,dt + c_{j,0}\int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j}\int_0^1 f(t)\theta_{j,2^j}(t)\,dt\,. \end{aligned}$$

Comme les extrémités de chacune des paraboles sont situées dans l'intervalle [0, 1] pour les indices n compris dans l'ensemble $\{1, 2, ..., 2^j - 1\}$, les intégrales concernant ces paraboles peuvent être interprétées comme des produits scalaires. Alors :

$$\begin{aligned}\int_0^1 f(t) f_j(t)\,dt &= \sum_{n=1}^{2^j-1} c_{j,n} \int_{-\infty}^{\infty} f(t)\theta_{j,n}(t)\,dt + c_{j,0}\int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j}\int_0^1 f(t)\theta_{j,2^j}(t)\,dt \\ &= \sum_{n=1}^{2^j-1} c_{j,n}\left\langle f(t),\theta_{j,n}(t)\right\rangle + c_{j,0}\int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j}\int_0^1 f(t)\theta_{j,2^j}(t)\,dt \\ &= \sum_{n=1}^{2^j-1} {c_{j,n}}^2 + c_{j,0}\int_0^1 f(t)\theta_{j,0}(t)\,dt + c_{j,2^j}\int_0^1 f(t)\theta_{j,2^j}(t)\,dt.\end{aligned}$$

Avec le résultat du lemme 3.4 :

$$\int_0^1 f(t) f_j(t)\,dt = \sum_{n=0}^{2^j-1} {c_{j,n}}^2 . \tag{3.49}$$

Pour le troisième terme de l'équation 3.47 :

$$\begin{aligned}\int_0^1 f_j(t)^2\,dt &= \int_0^1 \left(\sum_{n=0}^{2^j} \left\langle f(t),\theta_{j,n}(t)\right\rangle \theta_{j,n}(t) \right)^2 dt \\ &= \int_0^1 \begin{pmatrix} {c_{j,0}}^2 \theta_{j,0}(t)^2 + 2c_{j,0}c_{j,1}\theta_{j,0}(t)\theta_{j,1}(t) + 2c_{j,0}c_{j,2}\theta_{j,0}(t)\theta_{j,2}(t) + \ldots + \\ {c_{j,1}}^2 \theta_{j,1}(t)^2 + 2c_{j,1}c_{j,2}\theta_{j,1}(t)\theta_{j,2}(t) + 2c_{j,1}c_{j,3}\theta_{j,1}(t)\theta_{j,3}(t) + \ldots + \\ 2c_{j,2^j-1}c_{j,2^j}\theta_{j,2^j-1}(t)\theta_{j,2^n}(t) + {c_{j,2^j}}^2 \theta_{j,2^j}(t)^2 \end{pmatrix} dt.\end{aligned}$$

Lors de la décomposition, deux paraboles se recouvrent si et seulement si la différence de leur indice n est une valeur de l'ensemble {-1, 0, 1}. Le produit de toutes les autres combinaisons de paraboles est alors nul et l'expression précédente peut s'écrire :

$$\int_0^1 f_j(t)^2\,dt = \int_0^1 \left(\sum_{n=0}^{2^j} c_{j,n}{}^2\, \theta_{j,n}(t)^2 \right) dt \quad + \quad \int_0^1 \left(\sum_{n=0}^{2^j-1} c_{j,n}\, c_{j,n+1}\, \theta_{j,n}(t)\, \theta_{j,n+1}(t) \right) dt$$

$$= \sum_{n=1}^{2^j-1} c_{j,n}{}^2 \int_0^1 \theta_{j,n}(t)^2\,dt \quad + \quad c_{j,0}{}^2 \int_0^1 \theta_{j,0}(t)^2\,dt \quad + \quad c_{j,2^j}{}^2 \int_0^1 \theta_{j,2^j}(t)^2\,dt$$

$$+ \quad 2 \sum_{n=0}^{2^j-1} c_{j,n}\, c_{j,n+1} \int_0^1 \theta_{j,n}(t)\, \theta_{j,n+1}(t)\,dt.$$

Avec le lemme 3.5 :

$$\int_0^1 f_j(t)^2\,dt = \sum_{n=1}^{2^j-1} c_{j,n}{}^2 \int_0^1 \theta_{j,n}(t)^2\,dt \quad + \quad c_{j,0}{}^2 \int_{-\infty}^{\infty} \theta_{j,0}(t)^2\,dt \quad + \quad 2 \sum_{n=0}^{2^j-1} c_{j,n}\, c_{j,n+1} \int_0^1 \theta_{j,n}(t)\, \theta_{j,n+1}(t)\,dt \;;$$

en intégrant sur tout le domaine les fonctions de base dont le support est compris dans l'intervalle [0, 1] :

$$\int_0^1 f_j(t)^2\,dt = \sum_{n=1}^{2^j-1} c_{j,n}{}^2 \int_{-\infty}^{\infty} \theta_{j,n}(t)^2\,dt \quad + \quad c_{j,0}{}^2 \int_{-\infty}^{\infty} \theta_{j,0}(t)^2\,dt \quad + \quad 2 \sum_{n=0}^{2^j-1} c_{j,n}\, c_{j,n+1} \int_0^1 \theta_{j,n}(t)\, \theta_{j,n+1}(t)\,dt \;;$$

en considérant que le support du produit de $\theta_{j,n}(t)$ et $\theta_{j,n+1}(t)$ est la section où les fonctions se recoupent et que ce support est toujours situé dans l'intervalle [0, 1] :

$$\int_0^1 f_j(t)^2\,dt = \sum_{n=1}^{2^j-1} c_{j,n}{}^2 \int_{-\infty}^{\infty} \theta_{j,n}(t)^2\,dt \;+\; c_{j,0}{}^2 \int_{-\infty}^{\infty} \theta_{j,0}(t)^2\,dt \;+\; 2\sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1} \int_{-\infty}^{\infty} \theta_{j,n}(t)\theta_{j,n+1}(t)\,dt.$$

Avec l'invariance à la translation, les fonctions de base sont les mêmes pour un même j, peu importe n. On pose alors $n = 0$ et on laisse tomber les indices en exprimant $\theta_{j,n}(t)$ selon sa définition (éq. 3.17) :

$$\begin{aligned}
\int_{-\infty}^{\infty} f_j(t)^2\,dt &= \sum_{n=1}^{2^j-1} c_{j,n}{}^2 \int_{-\infty}^{\infty} 2^j\theta(2^jt)^2\,dt + \int_{-\infty}^{\infty} c_{j,0}{}^2\, 2^j\theta(2^jt)^2\,dt \;+\; 2\sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1} \int_{-\infty}^{\infty} 2^j\theta(2^jt)\theta(2^jt-1)\,dt \\
&= \sum_{n=0}^{2^j-1} c_{j,n}{}^2 \int_{-\infty}^{\infty} 2^j\theta(2^jt)^2\,dt \;+\; 2\sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1} \int_{-\infty}^{\infty} 2^j\theta(2^jt)\theta(2^jt-1)\,dt \\
&= \sum_{n=0}^{2^j-1} c_{j,n}{}^2 \frac{4}{5} \;+\; 2\sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1} \frac{124}{1215} \\
&= \frac{4}{5}\sum_{n=0}^{2^j-1} c_{j,n}{}^2 \;+\; \frac{248}{1215}\sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1}.
\end{aligned} \tag{3.50}$$

En réunissant les trois termes de l'équation 3.47 (éq. 3.48, 3.49 et 3.50), l'erreur devient :

$$\varepsilon[j] = \frac{3}{2} \;-\; 2\sum_{n=0}^{2^j-1} c_{j,n}{}^2 \;+\; \frac{4}{5}\sum_{n=0}^{2^j-1} c_{j,n}{}^2 \;+\; \frac{248}{1215}\sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1}, \tag{3.51}$$

et la limite donne :

$$\lim_{j\to\infty} \varepsilon[j] = \frac{3}{2} \;-\; \frac{6}{5}\lim_{j\to\infty} \sum_{n=0}^{2^j-1} c_{j,n}{}^2 \;+\; \frac{248}{1215}\lim_{j\to\infty} \sum_{n=0}^{2^j-1} c_{j,n}c_{j,n+1}.$$

Avec le résultat du lemme 3.6 :

$$\lim_{j\to\infty}\varepsilon[j]=\frac{3}{2}\;-\;\frac{6}{5}\lim_{j\to\infty}\sum_{n=0}^{2^j-1}{c_{j,n}}^2\;+\;\frac{248}{1215}\lim_{j\to\infty}\sum_{n=0}^{2^j-1}{c_{j,n}}^2$$

$$=\frac{3}{2}\;-\;\frac{242}{243}\lim_{j\to\infty}\sum_{n=0}^{2^j-1}{c_{j,n}}^2. \qquad (3.52)$$

En tenant compte du rapport de $\sqrt{2^j}$ entre l'approximation numérique formée par les coefficients et $f(t)$ (voir figure 3.13), la somme des coefficients peut être interprétée comme une intégrale de droite selon Riemann, où chaque rectangle a une largeur de 2^{-j} :

$$\lim_{j\to\infty}\sum_{n=0}^{2^j-1}{c_{j,n}}^2 2^{-j}=\int_0^1\left(\frac{f(t)}{\sqrt{2^j}}\right)^2 dt \;\Rightarrow\; \lim_{j\to\infty}\sum_{n=0}^{2^j-1}{c_{j,n}}^2=\int_0^1 f(t)^2\,dt. \qquad (3.53)$$

Finalement, en appliquant 3.53 dans 3.52 :

$$\lim_{j\to\infty}\varepsilon[j]=\frac{3}{2}\;-\;\frac{242}{243}\int_0^1\left(\sin(2\pi t)+1\right)^2 dt, \qquad (3.54)$$

qui donne

$$\lim_{j\to\infty}\varepsilon[j]=\frac{1}{162}.$$

■

Bien que le théorème précédent nous donne une convergence pour l'approximation d'une période d'une sinusoïde, l'équation 3.54 est très intéressante car la valeur 3/2 est en fait l'intégrale du carré de la sinusoïde approximée. Nous pouvons donc substituer ce terme par l'égalité 3.48 et généraliser le théorème 3.3 en utilisant cette fois-ci une erreur quadratique absolue définie par

$$\varepsilon_{abs}[\,j\,] = \frac{\left\| f(t) - f_j(t) \right\|^2}{\left\| f(t) \right\|^2} = \frac{\int_0^1 \left(f(t) - f_j(t) \right)^2 dt}{\int_0^1 \left(f(t) \right)^2 dt}. \qquad (3.55)$$

Théorème 3.4 : L'erreur quadratique absolue d'approximation à une résolution j de n'importe quelle fonction positive et périodique sur l'espace normalisé $L^2[0,1]$ tend vers 1/243, lorsque j tend vers l'infini.

Preuve : Sans remplacer le premier terme de l'équation 3.48 par sa valeur, l'équation 3.54 devient :

$$\lim_{j\to\infty} \varepsilon[j] = \int_0^1 (\sin(2\pi t) + 1)^2 dt \quad - \quad \frac{242}{243} \int_0^1 (\sin(2\pi t) + 1)^2 dt. \qquad (3.56)$$

Si nous généralisons maintenant la sinusoïde par n'importe quelle fonction périodique sur l'espace $L^2[0,1]$, nous pouvons toujours appliquer les résultats des lemme 3.4, 3.5 et 3.6. L'équation 3.56 devient alors

$$\lim_{j\to\infty}\varepsilon[j]=\int_0^1\left(f(t)\right)^2dt \quad - \quad \frac{242}{243}\int_0^1\left(f(t)\right)^2dt. \tag{3.57}$$

Finalement, en normalisant l'erreur quadratique :

$$\lim_{j\to\infty}\varepsilon_{abs}[j]=\frac{\int_0^1\left(f(t)\right)^2dt \quad - \quad \frac{242}{243}\int_0^1\left(f(t)\right)^2dt}{\int_0^1\left(f(t)\right)^2dt}$$

$$=1-\frac{242}{243}$$

$$=\frac{1}{243}.$$

■

Avec le théorème 3.4, nous pouvons effectuer une démarche semblable à une approximation de Fourier; soit rendre une fonction périodique et en faire la projection sur notre espace d'approximation. En fait, la périodicité est nécessaire uniquement à cause des paraboles à chaque extrémité de l'approximation. Si le support de la fonction est compris dans l'intervalle $]0,1[$, et que j tend vers l'infini, les paraboles d'indice $n=0$ et $n=2^j$ vont éventuellement devenir nulles. À ce moment-ci, nous pouvons éviter de rendre la fonction périodique et véritablement approximer une fonction non stationnaire. Cet élément est très intéressant car la très grande majorité des signaux réels, y compris ceux issus du quantron, sont de nature non stationnaire.

3.2 Approximation multiéchelle

Bien que nous ayons utilisé l'appellation «approximation multiéchelle» jusqu'à présent, nous n'avons travaillé qu'avec une seule échelle (ou résolution) à la fois. Dans cette section, nous appliquerons notre démarche à plusieurs résolutions simultanément pour avoir une approximation multiéchelle.

Avec un réseau de neurones nommé HRBF, Ferrari et al. (2004) démontrent qu'une approche multiéchelle utilisant une distribution gaussienne comme fonction de base peut posséder la propriété d'approximation sans toutefois être une analyse multirésolution. Dans le même ordre d'idées que le HBRF, notre extension à plusieurs résolutions sera faite en effectuant une suite d'approximations d'erreurs dans le but d'avoir des sous-espaces organisés selon le schéma de la figure 3.15.

Dans un premier temps, si la fonction n'est pas positive, on lui additionne une constante k_0 de manière à la rendre positive et à pouvoir l'approximer avec des paraboles positives. Ensuite, cette fonction positive $(f(t)+k_0)$ est approximée à une résolution de 1 (2^0) pour obtenir la fonction $a_0(t)$. L'erreur générée par cette approximation forme ce que nous appelons une fonction résiduelle $r_0(t)$ et la fonction originale peut être exprimée par la relation $f(t) = a_0(t) + r_0(t) - k_0$.

Forcément, cette fonction résiduelle n'est pas positive; elle est donc ajustée par une constante k_1 pour obtenir une fonction positive qui est alors approximée à une résolution plus grande pour obtenir $a_1(t)$. De cette approximation résulte une fonction

résiduelle $r_1(t)$ qui est ajustée et approximée à une plus grande résolution, et ainsi de suite.

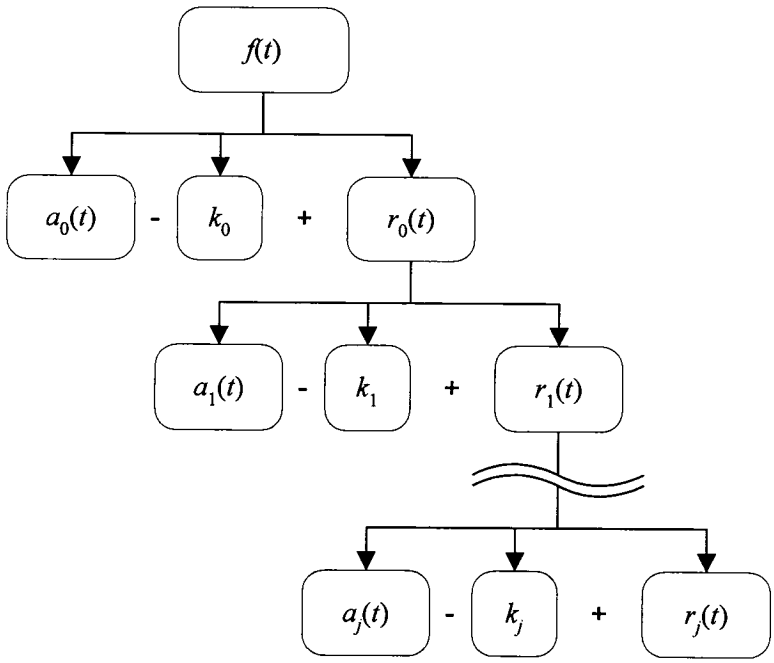


Figure 3.15 : Organisation des sous-espaces V_j pour une approximation multiéchelle.

La reconstruction de la fonction originale $f(t)$ jusqu'à une résolution j est notée $g_j(t)$. Cette fonction ($g_j(t)$) est obtenue en additionnant les différentes approximations $a_j(t)$ et en soustrayant les différentes constantes qui ont servi à ajuster $f(t)$ et les fonctions résiduelles $r_j(t)$ (voir figure 3.16).

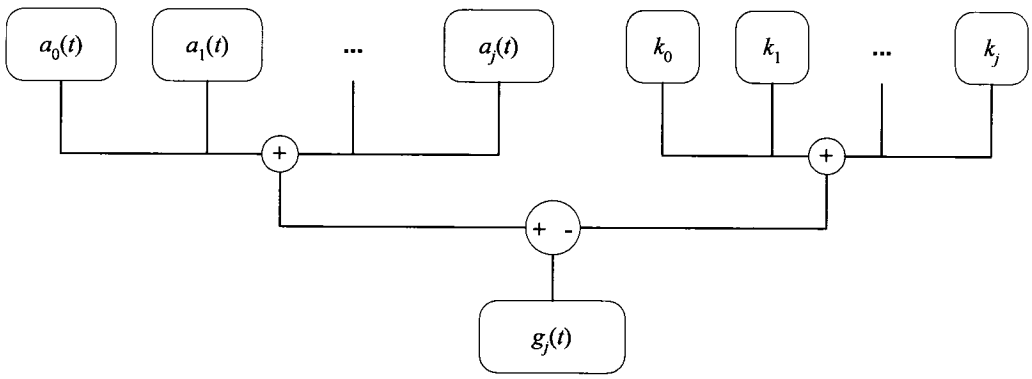


Figure 3.16 : Reconstruction de $g_j(t)$.

À titre d'exemple, les figures 3.17 à 3.20 montrent le résultat d'une telle approximation sur une fonction sinusoïdale de période 2π pour des résolutions allant jusqu'à $j = 3$, c'est-à-dire les fonctions $g_0(t)$ à $g_3(t)$. La sinusoïde ainsi que toutes les fonctions résiduelles ne sont pas positives et elles sont toutes ajustées d'une constante $k_j = 1$.

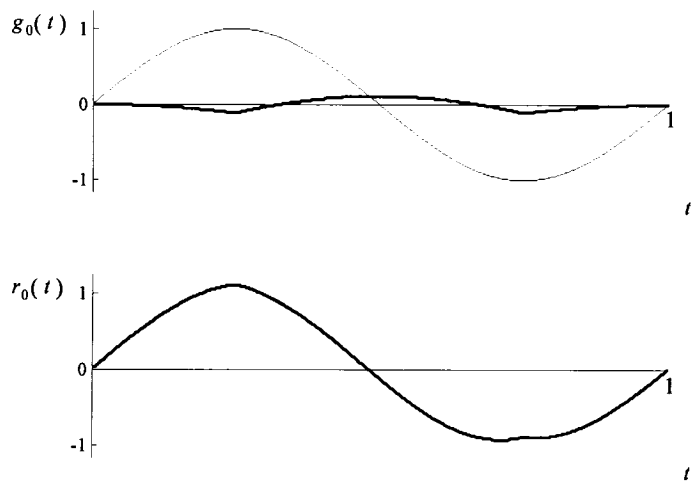


Figure 3.17 : Approximation $g_0(t)$ et fonction résiduelle $r_0(t)$.

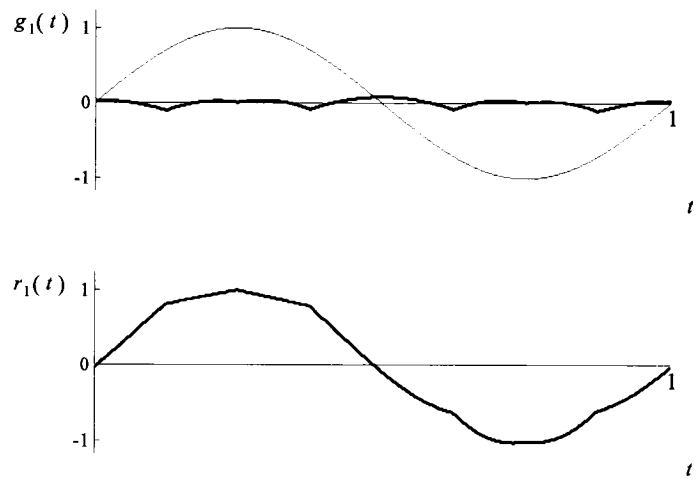


Figure 3.18 : Approximation $g_1(t)$ et fonction résiduelle $r_1(t)$.

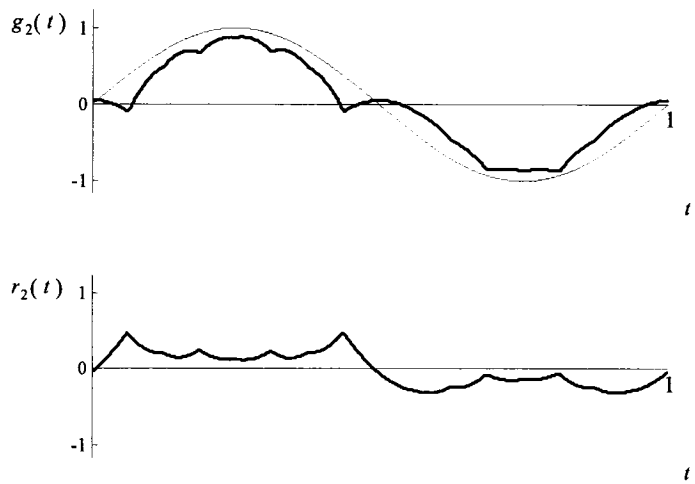


Figure 3.19 : Approximation $g_2(t)$ et fonction résiduelle $r_2(t)$.

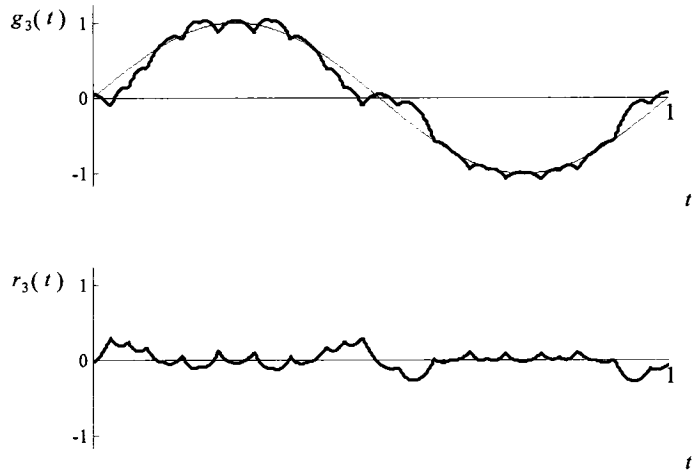


Figure 3.20 : Approximation $g_3(t)$ et fonction résiduelle $r_3(t)$.

Avec cette approche, il est possible de trouver l'erreur totale d'approximation à partir de l'erreur d'approximation des fonctions résiduelles. Ce résultat, présenté dans le lemme 3.7, nous sera utile pour le théorème 3.5 qui montre que la pertinence de cette méthode réside dans la convergence vers zéro de son erreur d'approximation.

Lemme 3.7 : L'erreur quadratique absolue de la fonction reconstruite est le produit des erreurs quadratiques absolues de l'approximation de $f(t)$ et de l'approximation de chacune des fonctions résiduelles $r_j(t)$:

$$\varepsilon_{abs}[j] = \frac{\left\| \left(f(t)+k_0\right) - a_0(t) \right\|^2}{\left\| f(t) \right\|^2} \prod_{n=0}^{j} \frac{\left\| \left(r_n(t)+k_{n+1}\right) - a_{n+1}(t) \right\|^2}{\left\| r_n(t) \right\|^2}. \qquad (3.58)$$

Preuve : En suivant l'algorithme de reconstruction, l'erreur quadratique absolue de l'approximation multiéchelle est définie par :

$$\varepsilon_{abs}[j] = \frac{\left\| f(t) - \left(a_0(t) + a_1(t) + \ldots + a_j(t) - k_0 - k_1 - \ldots - k_j\right)\right\|^2}{\left\| f(t)\right\|^2}. \qquad (3.59)$$

Pour la première approximation seulement:

$$\varepsilon_{abs}[0] = \frac{\left\| f(t) - \left(a_0(t) - k_0\right)\right\|^2}{\left\| f(t)\right\|^2}. \qquad (3.60)$$

Une fois ajustée par k_1 et approximée par $a_1(t)$, l'erreur d'approximation de la fonction résiduelle $r_0(t)$ est :

$$\varepsilon_{r_0} = \frac{\left\| r_0(t) - \left(a_1(t) - k_1\right)\right\|^2}{\left\| r_0(t)\right\|^2}. \qquad (3.61)$$

D'après le schéma de la figure 3.15, nous pouvons remplacer $r_0(t)$ par son équivalent et réécrire l'équation 3.61 :

$$\varepsilon_{r_0} = \frac{\left\| f(t) - a_0(t) + k_0 - \left(a_1(t) - k_1\right)\right\|^2}{\left\| f(t) - a_0(t) + k_0\right\|^2}. \qquad (3.62)$$

En généralisant 3.62, l'erreur d'approximation de $r_j(t)$ donne :

$$\varepsilon_{r_j} = \frac{\left\| r_{j-1}(t) - \left(a_j(t) - k_j\right) \right\|^2}{\left\| r_{j-1}(t) \right\|^2}$$

$$= \frac{\left\| f(t) - \sum_{n=0}^{j} a_n(t) - k_n \right\|^2}{\left\| f(t) - \sum_{n=0}^{j-1} a_n(t) - k_n \right\|^2}. \qquad (3.63)$$

À ce moment-ci, l'erreur totale est :

$$\varepsilon_{abs}[j] = \varepsilon_{abs}[0] \prod_{n=0}^{j} \varepsilon_{r_n}$$

$$= \frac{\left\| f(t) - \left(a_0(t) - k_0\right) \right\|^2}{\left\| f(t) \right\|^2} \frac{\left\| f(t) - \left(a_0(t) + a_1(t) - k_0 - k_1\right) \right\|^2}{\left\| f(t) - \left(a_0(t) - k_0\right) \right\|^2} \cdots$$

$$\frac{\left\| f(t) - \sum_{n=0}^{j-2} a_n(t) - k_n \right\|^2}{\left\| f(t) - \sum_{n=0}^{j-3} a_n(t) - k_n \right\|^2} \frac{\left\| f(t) - \sum_{n=0}^{j-1} a_n(t) - k_n \right\|^2}{\left\| f(t) - \sum_{n=0}^{j-2} a_n(t) - k_n \right\|^2} \frac{\left\| f(t) - \sum_{n=0}^{j} a_n(t) - k_n \right\|^2}{\left\| f(t) - \sum_{n=0}^{j-1} a_n(t) - k_n \right\|^2}$$

$$= \frac{\left\| f(t) - \sum_{n=0}^{j} a_n(t) - k_n \right\|^2}{\left\| f(t) \right\|^2}$$

$$= \frac{\left\| f(t) - \left(a_0(t) + a_1(t) + \ldots + a_j(t) - k_0 - k_1 - \ldots - k_j\right) \right\|^2}{\left\| f(t) \right\|^2}.$$

■

Théorème 3.5 : Pour n'importe quelle valeur de $\varepsilon > 0$, nous pouvons avoir une approximation $a_j(t)$ pour un j suffisamment grand où $\varepsilon_{abs}[j] < \varepsilon$.

Preuve : Si on considère une fonction $f(t)$ dans l'espace $L^2[0,1]$, l'erreur de la première approximation de $f(t)$ donne une certaine constante définie par l'équation 3.60. Si l'approximation est poursuivie jusqu'à une résolution j, nous aurons alors effectué l'approximation de j-1 fonctions résiduelles.

Étant donné que la fonction résiduelle $r_j(t)$ est l'erreur d'approximation à une résolution j et que la résolution double pour chaque changement d'échelle, cette fonction résiduelle $r_j(t)$ peut toujours être approximée avec une certaine précision à une résolution j+1. Ainsi, l'erreur d'approximation des fonctions résiduelles qui correspondent aux termes à l'intérieur du produit de l'équation 3.58 sera toujours inférieure à 1.

Nous savons que $\lim_{n\to\infty} r^n$ tend vers zéro pour tout r compris entre 0 et 1. Comme chacun des termes du produit de l'équation 3.58 est entre 0 et 1, ce produit tend vers zéro lorsque j tend vers l'infini. Du même coup, l'erreur quadratique absolue de la fonction reconstruite tend également vers zéro lorsque j tend vers l'infini.

■

CHAPITRE 4 – APPLICATIONS ET DISCUSSION

Dans ce chapitre, nous proposons des applications utilisant des approximations multiéchelles et présentons les résultats des celles-ci. Nous présentons une étude d'approximations à une échelle, suivie d'une étude d'approximation à plusieurs échelles.

Dans les deux cas, les résultats obtenus avec la fonction de base formée d'une parabole sont comparés avec les résultats obtenus en utilisant la fonction d'échelle de Haar. La comparaison est faite en utilisant l'erreur quadratique absolue. Pour quantifier l'erreur en pourcentage, nous utilisons également l'erreur absolue (non quadratique). Cette dernière est obtenue en effectuant la racine carrée de l'erreur quadratique de manière à obtenir l'équation suivante :

$$\varepsilon_{app} = \frac{\| f - f_{app} \|}{\| f \|}. \tag{4.1}$$

Cette erreur est alors convertie en pourcentage en utilisant la relation :

$$\varepsilon_{pourcent} = \varepsilon_{app} \cdot 100\%. \tag{4.2}$$

Pour l'approximation à une résolution, le théorème 3.4 impose une limite d'erreur quadratique de 1/243, ce qui, en pourcentage, correspond à une erreur limite de 6.4%.

C'est pourquoi nous chercherons à atteindre un seuil de 10% (une erreur quadratique de 0.01), ce qui donne une approximation juste.

Pour obtenir ces applications, les équations du chapitre précédent ont été implantées dans des logiciels de mathématique. Un des nos objectifs étant d'obtenir une fonction analytique, le choix logique fut d'aborder le problème avec un logiciel de calcul symbolique : Maple. Bien qu'utile pour les développements d'équations ayant servi à la compréhension du problème et à faire les différentes preuves mathématiques, Maple se révèle inadéquat pour développer des applications. En effet, la lourdeur des expressions symboliques avec lesquelles travaille Maple fait en sorte que le temps de calcul devient rapidement long. C'est pourquoi les applications ont été développées en utilisant des méthodes numériques implantées dans Matlab. De cette manière, les calculs s'effectuent rapidement et la reconstruction de $f_j(t)$ et de $g_j(t)$ à partir des coefficients obtenus par ces calculs nous permet de fournir une fonction analytique comme solution.

4.1 Approximation à une échelle

Dans cette section, nous considérons l'approximation de trois types de fonctions : un changement de polarisation selon une seule fonction $\rho(t)$, selon une fonction P(t) positive représentant un changement de polarisation pour dix fonctions $\rho(t)$ et une fonction P(t) partiellement positive formée encore une fois de dix fonctions $\rho(t)$.

Le cas d'une variation de polarisation exprimé par une seule fonction $\rho(t)$ revient à considérer un quantron avec une seule entrée dont les paramètres sont (dans l'exemple de la figure 4.1) : $S = 0.23$, $a = 2$, $w = 1$ et $\theta = 0.20$.

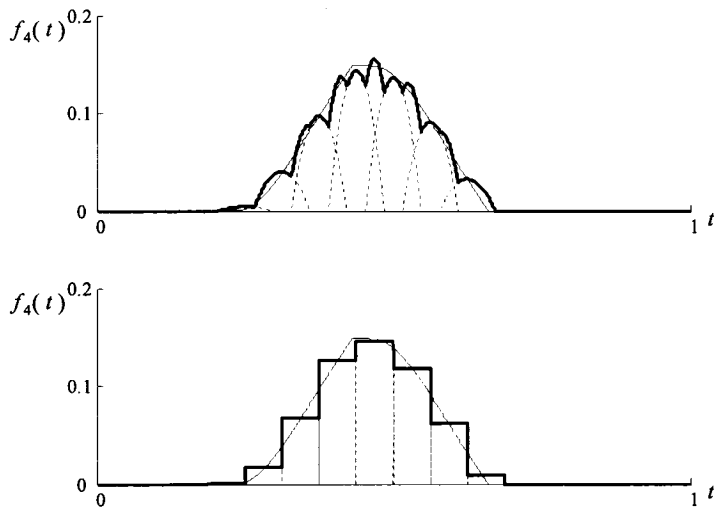


Figure 4.1 : Approximation d'une fonction $\rho(t)$ avec des paraboles et avec des fonctions de Haar.

Dans cet exemple, le paramètre de résolution a dû être ajusté à $j = 4$ de façon à avoir une erreur quadratique plus petite que 0.01 (voir tableau 4.1). En comparant la fonction formée de paraboles et celle formée avec des fonctions de Haar, on voit que les paraboles permettent d'aller chercher les courbures de $\rho(t)$, ce que les fonctions de Haar ne permettent pas. Les données du tableau 4.1 montrent en effet que l'approximation est meilleure lorsqu'on utilise les paraboles jusqu'à ce que l'on approche de l'erreur quadratique limite de 0.00412 (1/243). À partir de ce moment, l'erreur d'approximation avec les fonctions de Haar diminue vers une valeur nulle.

Tableau 4.1 : Erreur quadratique absolue d'approximation d'une fonction $\rho(t)$ pour des valeurs de j variant de 0 à 7.

	Erreur quadratique absolue	
j	Paraboles	Haar
0	0,7292	0,6963
1	0,3553	0,6672
2	0,1113	0,3348
3	0,0427	0,0716
4	0,0093	0,0208
5	0,0050	0,0528
6	0,0043	0,0013
7	0,0042	0,0003

Un point important à souligner est qu'avec une approximation à une seule résolution, nous pouvons facilement retrouver le maximum global de la fonction $\rho(t)$. En observant les coefficients non nuls de la décomposition du tableau 4.2 et en multipliant ce vecteur par $\sqrt{2^j}$, tel que discuté à la section 3.1.3, nous obtenons la valeur du sommet de chacune des paraboles (tableau 4.3).

Tableau 4.2 : Coefficients non nuls d'une approximation pour $j = 4$.

n	4	5	6	7	8	9	10
Coefficient	0,0011	0,01	0,0243	0,0358	0,0342	0,0227	0,0081

Tableau 4.3 : Sommet des paraboles d'une approximation pour $j = 4$.

n	4	5	6	7	8	9	10
Sommet	0,0044	0,04	0,0972	0,1432	0,1368	0,0908	0,0324

Comme l'indice n nous permet de localiser chacune des paraboles, nous pouvons déterminer approximativement le maximum global d'une fonction à partir de l'équation suivante :

$$\max_{global}(f(t)) \simeq f_j(n \cdot 2^{-j}) \tag{4.3}$$

où n est l'indice du coefficient le plus élevé. Dans le cas de la fonction $\rho(t)$ de la figure 4.1, le maximum global est d'environ $f_j(7 \cdot 2^{-4}) = 0.1432$ et il est situé à $t = 7 \cdot 2^{-4} = 0.4375$. En comparant avec le maximum global déterminé numériquement et valant 0.1484, nous avons une erreur de 3.5%. Comme le démontre le théorème 3.1, plus la résolution augmente, plus le sommet de chacune des paraboles tend vers une valeur de la fonction approximée. Ainsi, l'erreur d'approximation du maximum global tend vers zéro lorsque j tend vers l'infini. Le fait que nous puissions travailler avec des données numériques pour obtenir une fonction analytique nous permet de faire la comparaison entre notre maximum réel et approximatif.

À ce moment-ci, nous pouvons approximer une fonction issue du quantron avec une fonction dérivable, trouver son maximum global et localiser ce dernier sans utiliser la méthode de Laplace.

Dans l'exemple illustré à la figure 4.2, un quantron à quatre entrées est considéré (voir figure 1.2). Dans cette exemple, l'erreur d'approximation est de 10.7%, ce qui est juste au-delà de notre seuil de 10%. Ceci est un bon compromis étant donné qu'en

augmentant la résolution de 1, nous doublons le nombre de paraboles. Ceci étant dit, rien ne nous oblige à restreindre la résolution à un nombre entier. Dans l'exemple de la figure 4.2, nous obtenons notre seuil de 10% avec une valeur de $j = 4.3$.

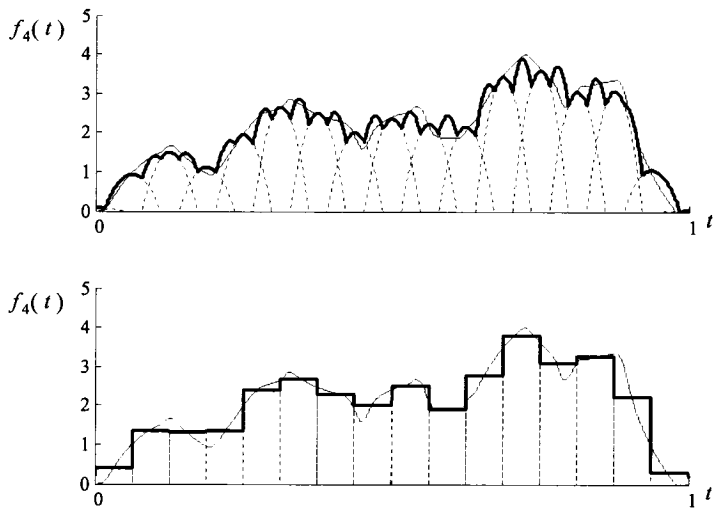


Figure 4.2 : Approximation d'une fonction P(t) positive avec des paraboles et des fonctions de Haar.

Bien que la fonction soit continue (donc avec une certaine courbure), la fonction P(t) combinant plusieurs entrées avec des paramètres différents vers un seul quantron est généralement moins régulière qu'une seule fonction $\rho(t)$. C'est pourquoi les paraboles performent mieux que les fonctions de Haar. Toutefois, en excluant les extrémités, la fonction P(t) a moins de variation que la fonction $\rho(t)$ de l'exemple précédent, donc moins de courbure. Cette plus grande similarité avec une droite de pente nulle, qui n'est pas l'idéal à approximer avec des paraboles, fait en sorte que la différence entre les erreurs d'approximation est moins grande et par toujours de

même signe. Cette similarité avec une fonction plus adaptée à l'approximation avec des fonctions de Haar fait en sorte que pour des petites résolutions, celles-ci performent mieux que les paraboles. Et encore une fois, alors que nous sommes limités à une erreur de 6.4% en utilisant les paraboles, l'erreur converge vers zéro lorsque l'on utilise les fonctions de Haar.

Tableau 4.4 : Erreur quadratique absolue d'approximation d'une fonction P(t) pour des valeurs de j variant de 0 à 7.

	Erreur quadratique absolue	
j	Paraboles	Haar
0	0,3737	0,1822
1	0,1602	0,1558
2	0,0577	0,1134
3	0,0332	0,0534
4	0,0115	0,0178
5	0,0064	0,0043
6	0,0044	0,0013
7	0,0042	0,0003

Pour ce qui est du maximum global de la fonction P(t), le maximum approximé est de 3.585 avec une erreur de 10.3% par rapport au maximum numérique de 4.0002. À ce moment-ci, il pourrait être judicieux d'augmenter la résolution de 1. En effet, avec une résolution de 5, l'approximation du maximum global est alors 3.905, ce qui représente une erreur de 2.3%.

Le dernier cas à considérer est le cas illustré à la figure 4.3, où une partie de la fonction P(t) est négative. Dans cet exemple, les contraintes entourant l'approximation de cette fonction nous obligent à avoir une fonction dérivable autour

du maximum. Ainsi, nous ne gardons que la partie positive de façon à ce que notre approximation ne soit formée que de paraboles positives.

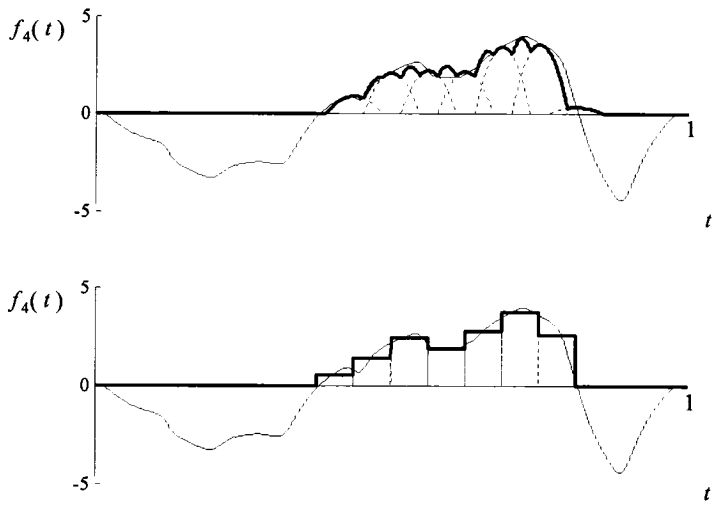


Figure 4.3 : Approximation d'une fonction P(t) partiellement positive avec des paraboles et des fonctions de Haar.

Si la fonction est entièrement négative, tous les coefficients sont nuls et il n'y a pas d'approximation. Ceci ne nous importe pas puisque l'objectif de cette recherche est de repérer si la polarisation de la membrane post-synaptique dépasse un certain seuil positif Γ. Si la fonction est négative, il est clair que ce seuil ne sera pas atteint.

4.2 Approximation multiéchelle

Pour les approximations multiéchelles, nous allons procéder avec les mêmes cas que les approximations à une résolution : une fonction $\rho(t)$ issue d'un seul

quantron, une fonction P(t) positive issue de plusieurs quantrons et une fonction P(t) partiellement positive. Mais avant d'aborder ces exemples, certaines conditions d'application doivent être spécifiées pour effectuer une approximation multiéchelle. Ces conditions sont la normalisation des fonctions à approximer et la définition des constantes k_j pour ajuster les fonctions résiduelles.

4.2.1 Conditions d'application

Premièrement, rien ne nous oblige à commencer nos approximations à $j = 0$. C'est pourquoi une première approximation sera faite à une résolution convenable et l'algorithme étendra l'approximation à des échelles subséquentes jusqu'à ce que l'erreur désirée soit atteinte.

Deuxièmement, dans le cas de l'approximation d'une fonction $f(t)$ à support compact, pour appliquer convenablement la méthode multiéchelle, nous normalisons $f(t)$ pour avoir un support sur l'ensemble de l'espace $L^2[0,1]$. Nous normalisons $f(t)$ dans le but de limiter le domaine où $f(t) = 0$.

Bien qu'une première approximation d'une fonction où $f(t) = 0$ pour une partie du domaine utilise un nombre limité de paraboles ($g_3(t)$ dans l'exemple de la figure 4.4), nous avons une fonction résiduelle où $r_3(t) = 0$ pour une partie du domaine. Lorsque que $r_3(t)$ est ajustée pour effectuer une approximation à $j = 4$, cela équivaut à approximer deux segments de droite de pente nulle avec des paraboles. Comme

approximer ces deux segments consiste à approximer $f(t) = 0$, nous introduisons des paraboles qui n'ont à priori aucune utilité.

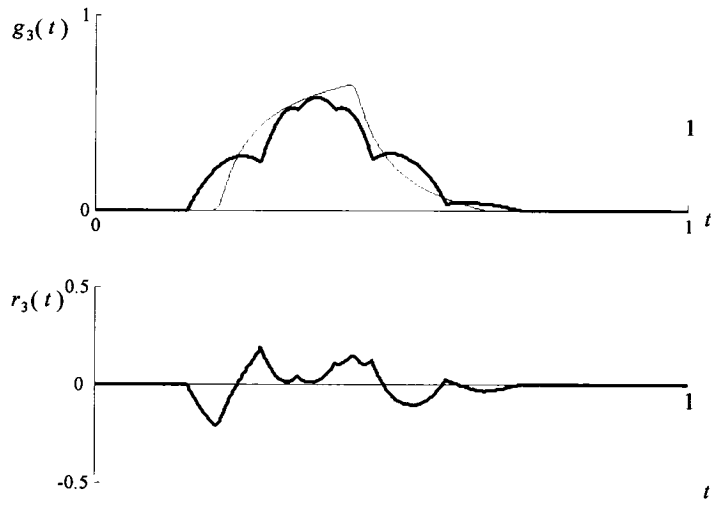


Figure 4.4 : Approximation $g_3(t)$ d'une fonction $\rho(t)$ et fonction résiduelle $r_3(t)$.

La troisième condition d'application consiste à définir une constante k_j appropriée pour ajuster chacune des fonctions $r_j(t)$ et les rendre positives. Pour ce faire, nous profitons du fait que l'implantation de notre méthode nous permet de passer d'une fonction numérique à une fonction analytique et nous posons comme valeur de k_j :

$$k_j = -\min_{numérique}\left(r_j(t)\right), \tag{4.4}$$

qui est le minimum global de la fonction discrétisée.

4.2.2 Exemples

Dans le prochain exemple, illustré par les figures 4.5 à 4.7, une fonction $\rho(t)$ est approximée. Cette fonction du quantron est normalisée en posant $S = 0.5$ et $\theta = 0$, alors que les autres paramètres sont $a = 1.25$ et $w = 1$. Dans cet exemple, nous commençons notre approximation à une résolution donnant une erreur de 21%, et nous poursuivons cette approximation pour avoir une erreur sous les 10%. Notons ici que contrairement aux approximations à une résolution, l'erreur tend vers zéro, alors le seuil de tolérance pourrait être encore moins élevé.

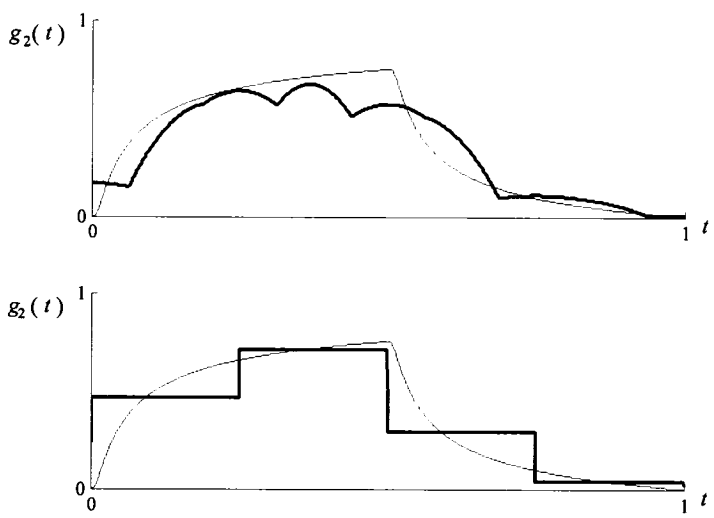


Figure 4.5 : Approximation $g_2(t)$ d'une fonction $\rho(t)$ pour des paraboles et des fonctions de Haar.

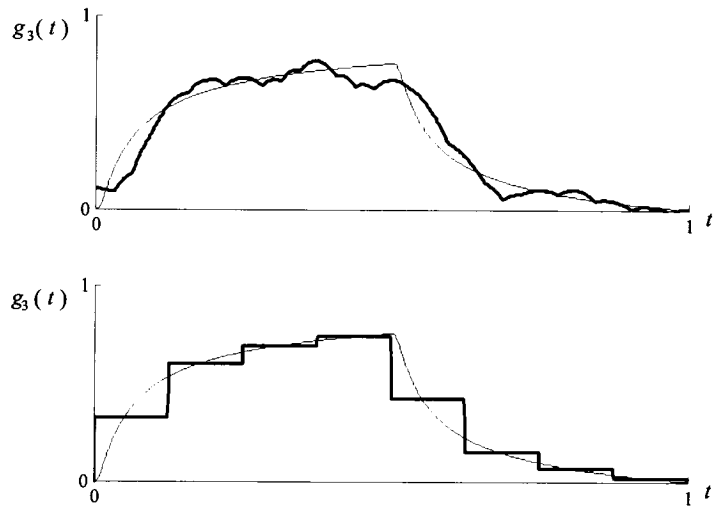


Figure 4.6 : Approximation $g_3(t)$ d'une fonction $\rho(t)$ pour des paraboles et des fonctions de Haar.

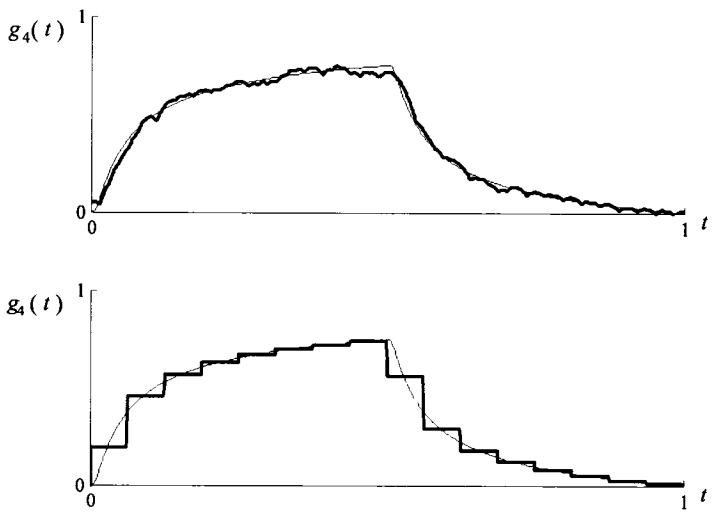


Figure 4.7 : Approximation $g_4(t)$ d'une fonction $\rho(t)$ pour des paraboles et des fonctions de Haar.

Avec les données de l'exemple précédent (tableau 4.5), nous constatons que l'erreur de l'approximation utilisant des paraboles converge vers zéro plus rapidement que l'erreur de l'approximation utilisant des fonctions de Haar. Ceci est dû au fait que les paraboles permettent d'aller chercher la courbure de $\rho(t)$, ainsi qu'à la différence entre une approximation multiéchelle et une approximation multirésolution.

Tableau 4.5 : Erreur quadratique absolue d'approximation d'une fonction $\rho(t)$ pour des valeurs de j variant de 2 à 4.

	Erreur quadratique absolue	
j	Paraboles	Haar
2	0,0443	0,0769
3	0,0182	0,0322
4	0,0021	0,0106

De façon à mieux illustrer ce propos, nous introduisons le prochain exemple (illustré par les figures 4.8 à 4.10), soit l'approximation de la même fonction $P(t)$ que celle de la figure 4.2. Notons ici que cette fonction n'a pas été normalisée étant donné que le support de $P(t)$ s'étend sur presque tout le domaine [0,1].

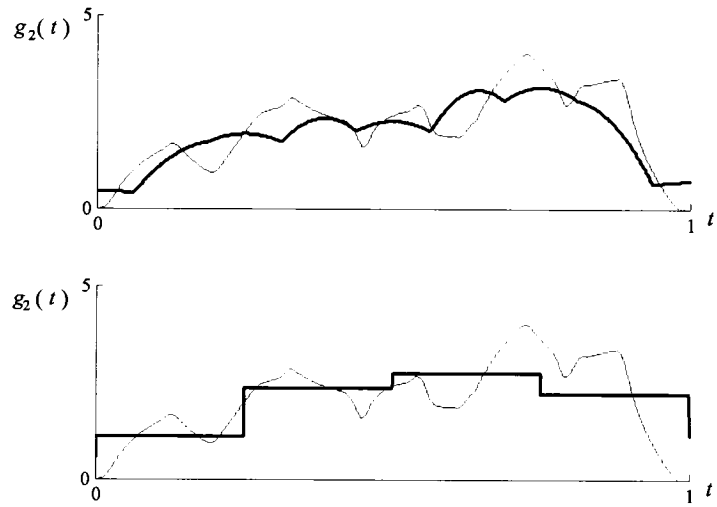


Figure 4.8 : Approximation $g_2(t)$ d'une fonction P(t) pour des paraboles et des fonctions de Haar.

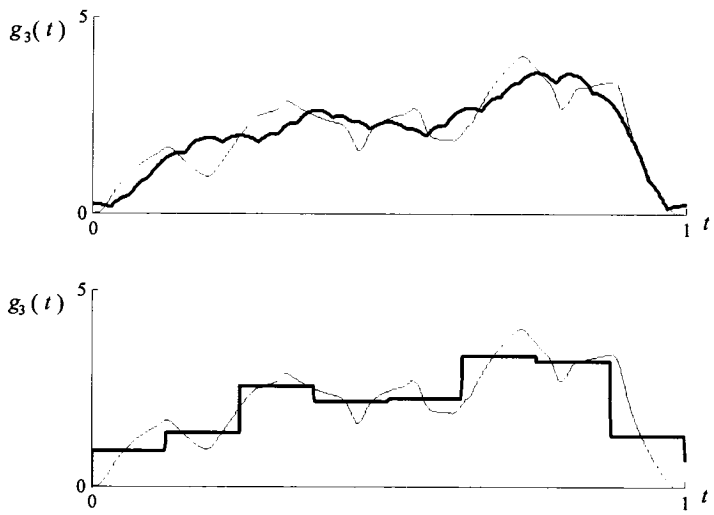


Figure 4.9 : Approximation $g_3(t)$ d'une fonction P(t) pour des paraboles et des fonctions de Haar.

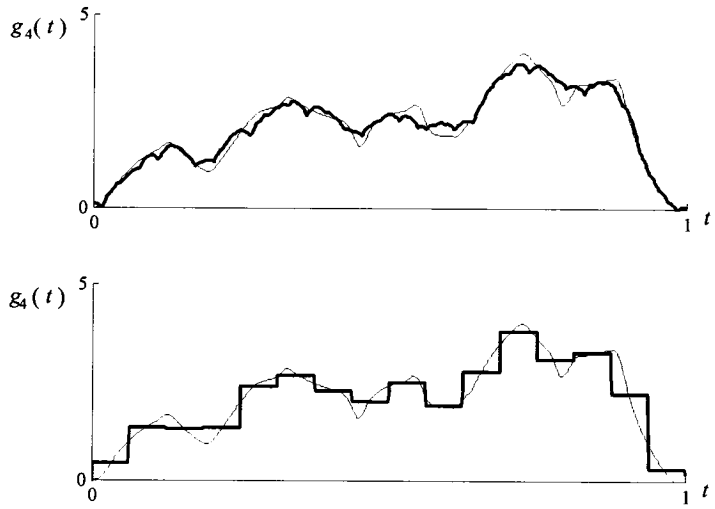


Figure 4.10 : Approximation $g_4(t)$ d'une fonction P(t) pour des paraboles et des fonctions de Haar.

L'erreur des approximations successives donnée dans le tableau 4.6 montre encore une fois que l'approximation multiéchelle avec des paraboles donne de meilleurs résultats que les fonctions de Haar.

Tableau 4.6 : Erreur quadratique absolue d'approximation d'une fonction P(t) positive pour des valeurs de j variant de 2 à 4.

	Erreur quadratique absolue	
j	Paraboles	Haar
2	0,0578	0,1131
3	0,0318	0,0531
4	0,0061	0,0181

Pour expliquer ceci, rappelons que l'absence de la propriété d'inclusion définie par l'équation 2.18 fait en sorte que nous n'avons pas une analyse multirésolution. Alors,

si on fait une approximation avec des paraboles à une résolution j, cette fonction $f_j(t)$ ne peut pas être entièrement reconstituée à une résolution $j + 1$. C'est pourquoi, dans le but d'augmenter la précision, nous décidons d'approximer l'erreur de celle-ci à une résolution $j + 1$. Ainsi, la fonction $g_4(t)$ (figure 4.10) est plus précise que la fonction $f_4(t)$ (figure 4.2).

Par contre, dans le cas des analyses multirésolutions (que permettent les fonctions de Haar), nous avons la propriété d'inclusion. Comme nous pouvons reconstituer l'approximation d'une résolution j à une résolution $j + 1$, au lieu d'approximer une fonction résiduelle $r_j(t)$ pour obtenir $g_{j+1}(t)$, nous pouvons procéder comme dans la section précédente et effectuer une approximation simple pour obtenir $f_{j+1}(t)$. Ainsi, lorsque l'on compare les approximations faites avec les fonctions de Haar des figures 4.2 et 4.10, la fonction $g_4(t)$ est sensiblement la même que la fonction $f_4(t)$. Cette similitude est d'ailleurs confirmée par le fait que les erreurs d'approximation des fonctions de Haar dans les tableaux 4.4 et 4.6 soient pratiquement identiques pour une même résolution. Comme il y a quelques différences au niveau de l'implantation des méthodes à simple échelle et multiéchelle, il est normal que les erreurs quadratiques ne soit pas exactement les mêmes.

Ceci nous permet de dire que la méthode multiéchelle devient futile lorsqu'elle est employée avec une fonction d'échelle pouvant effectuer une analyse multirésolution (les fonctions de Haar par exemple). C'est d'ailleurs la raison pour laquelle les analyses multirésolutions utilisent les fonctions d'ondelettes de manière à récupérer l'erreur d'approximation entre $f_j(t)$ et $f_{j+1}(t)$ dans un sous-espace orthogonal à V_j et

V_{j+1}. Cette capacité de l'analyse multirésolution à générer des sous-espaces d'approximation orthogonaux entre eux est un avantage important de celle-ci sur notre méthode multiéchelle. Mais puisque que la fonction de base formée d'une parabole ne peut effectuer une analyse multirésolution, notre méthode multiéchelle est efficace.

Finalement, observons l'approximation d'une fonction P(t) partiellement négative illustrée par les figure 4.11 à 4.14. Pour ce cas, il existe deux solutions possibles : soit nous approximons la fonction au complet, soit nous redéfinissons $f(t)$ en ne gardant que la partie positive et en normalisant celle-ci. À titre d'exemple supplémentaire, nous avons choisi d'approximer P(t) dans son ensemble.

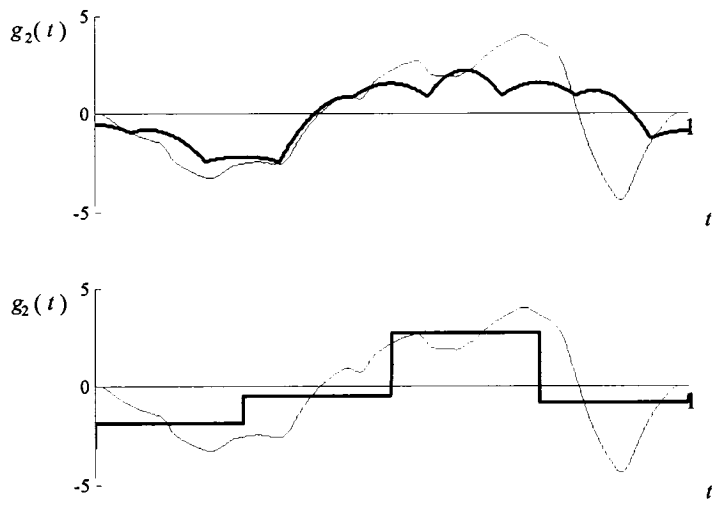


Figure 4.11 : Approximation $g_2(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.

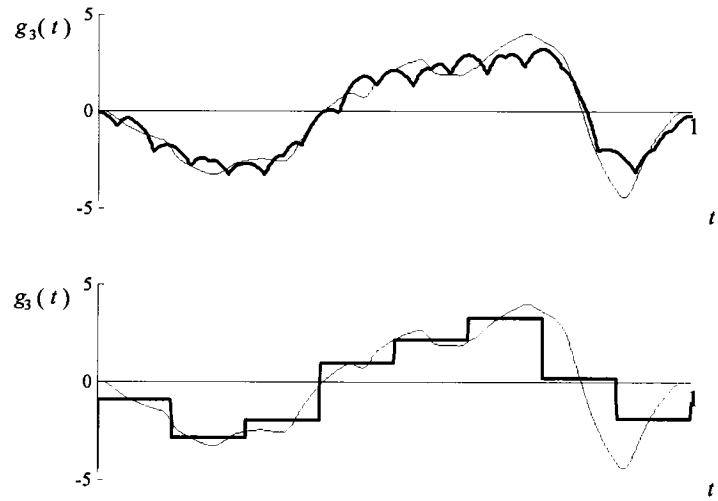


Figure 4.12 : Approximation $g_3(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.

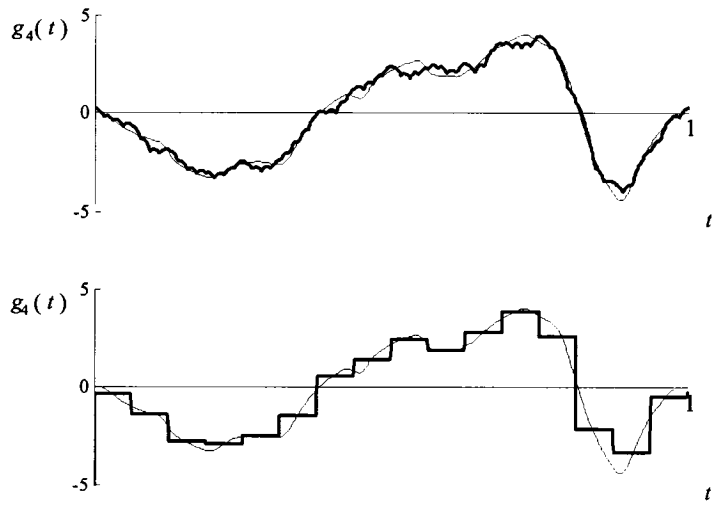


Figure 4.13 : Approximation $g_4(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.

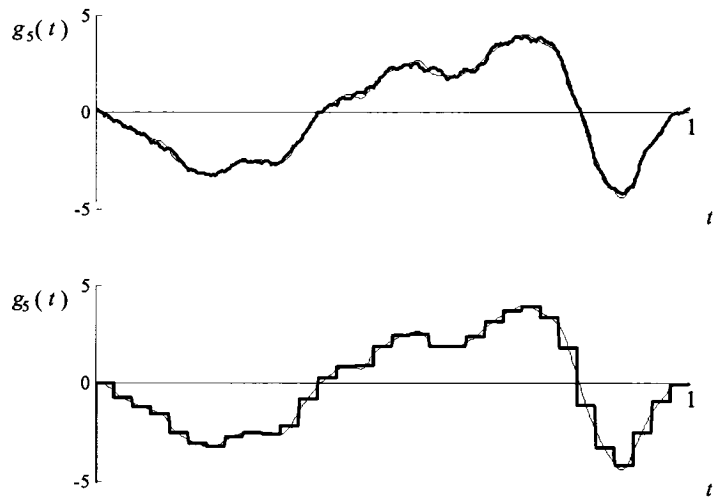


Figure 4.14 : Approximation $g_5(t)$ d'une fonction P(t) partiellement positive pour des paraboles et des fonctions de Haar.

Encore une fois, les données du tableau 4.7 nous montrent que l'approche multiéchelle fonctionne mieux avec les paraboles.

Tableau 4.7 : Erreur quadratique absolue d'approximation d'une fonction P(t) pour des valeurs de j variant de 2 à 5.

	Erreur quadratique absolue	
j	Paraboles	Haar
2	0,4491	0,4624
3	0,0730	0,2594
4	0,0152	0,0664
5	0,0039	0,0251

Nous venons de montrer que l'approche multiéchelle est capable d'effectuer des approximations de qualité en approximant à une seule résolution. De plus, en approximant une fonction avec un minimum de fonctions de base, nous remarquons

que les paraboles peuvent aller chercher la courbure des fonctions issues du quantron, ce que ne peuvent faire les fonctions de Haar. C'est pourquoi nous obtenons de meilleurs résultats avec les paraboles. Nous sommes aussi en mesure d'approximer facilement le maximum global d'une fonction à partir de la décomposition de celle-ci à l'aide de paraboles.

Nous avons également vu que sous certaines conditions (choix des résolutions d'approximation, normalisation et ajustement des fonctions), nous pouvons effectuer une approximation multiéchelle plus performante en utilisant des paraboles comme fonctions de base qu'en utilisant les fonctions de Haar. Par contre, en utilisant l'approximation multiéchelle, nous ne pouvons trouver facilement le maximum global et nous devons nous en remettre à la méthode de Laplace.

CONCLUSION

L'objectif de cette recherche était de développer une méthode d'approximation analytique dont la forme devait nous permettre d'optimiser celle-ci en utilisant la méthode de Laplace. Pour ce faire, nous nous sommes inspiré des analyses mutirésolutions et nous avons développé une méthode d'approximation multiéchelle utilisant des paraboles positives à support compact. Dans un premier temps, nous avons observé une approximation faite à une seule échelle pour ensuite passer à une approximation multiéchelle.

Dans le cas de l'approximation à une échelle donnée, il a été démontré que l'approximation était stable et que l'erreur quadratique absolue de celle-ci convergeait vers 1/243, alors que dans le cas d'une approximation multiéchelle, l'erreur convergeait vers 0. Ces deux approches théoriques ont ensuite été appliquées à des fonctions issues du quantron. À l'aide de ces exemples, nous avons pu comparer les performances de notre approche à une analyse mutlirésolution utilisant des fonctions de Haar. Dans les deux cas (approximation à simple échelle et multiéchelle), nous remarquons que les fonctions de base formée de paraboles donnent de meilleurs résultats que les fonctions de Haar. De plus, nous avons pu voir qu'en approximant à une seule résolution, nous pouvions facilement trouver le maximum global des fonctions $\rho(t)$ et $P(t)$ sans utiliser la méthode de Laplace

Cette approximation analytique servira au développement de réseaux de neurones artificiels utilisant le quantron de deux manières : premièrement, en permettant l'élimination des neurones qui ne transmettent pas d'information (ceux dont la polarisation n'atteint pas le seuil Γ) et deuxièmement, en donnant une fonction dérivable qui permet l'apprentissage du réseau.

Les analyses multirésolutions ont permis non seulement d'atteindre les objectifs énoncés précédemment, mais elles ouvrent également la porte à un outil puissant de traitement de signal : les ondelettes. En effet, il serait intéressant de prendre une fonction $\rho(t)$ du quantron, de démontrer que cette fonction peut effectuer une analyse multirésolution, de trouver l'ondelette associée à $\rho(t)$ et de voir les possibilités d'application de celle-ci.

BIBLIOGRAPHIE

DAUBECHIES, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.

FAGGIN, F. (1991), « VLSI Implementation of Neural Networks », *International joint Conference on Neural Networks*, Seattle.

FERRARI, S., MAGGIONI, M. And Borghese N.A. (2004), « Multiscale Approximation With Hierarchical Radial Basis Functions Networks », *IEEE Transactions on Neural Networks*, Vol 15, No. 1 (January), pp. 178-188.

HAAR, A. (1910), « Zur Theorie der orthogonalen Funktionnensysteme », *Math. Anal.*, 69, pp. 331-371.

HAYKIN S. (1999), *Neural Networks : A Comprehensive Foundation*, second edition, Prentice Hall, Inc., New Jersey.

HEBB, D.O. (1943), *The Organization of Behavior : A Neuropsychological Theory*, Wiley, New-York.

LABIB R. (1999), « New single neuron structure for solving nonlinear problems », *Proc. Of the 1999 international Joint Conference on Neural Networks*, manuscript #0476, 4 pages.

LAPLACE, P. S. (1774). « Mémoire sur la probabilité des causes par les événemens », *Savants étranges* 6, pp. 621-656, *Oeuvres* 8, pp. 27-65.

MALLAT, S. (1989), « Multiresolution Approximation and Wavelet Orthonormal Bases of L^2(R) », *Transaction of the American Mathematical Society*, Vol. 315, No. 1 (Sep.) pp. 69-87.

MCCULLOCH, W.S. and W. PITTS. (1943), « A logical calculus of the ideas immanent in nervous activity », *Bulletin of mathematical Biophysics*, vol. 5, pp. 115-133.

MEYER, Y. (1992), *Wavelets and Operator*, Advanced mathematics, Cambridge University Press.

MINSKY, M.L. and PAPERT, S.A. (1969), *Perceptrons*, Cambridge, MA : MIT Press.

ROSENBLAT F. (1958), « The Perceptron : A Probabilistic model for information storage and organization in the brain », *Psychological Review*, vol. 65, pp. 386-408.

SHEPHERD, G.M. and KOCH, C. (1990), « Introduction to synaptic circuits », *The synaptic Organization of the Brain*, G.M. Shepherd, ed., pp.3-31. Oxford University Press.

SWOKOWSKI (1993), *Calculus 5th ed.*, DeBoeck-Wesmael S.A, Buxelles.

WIDROW B. and HOFF, Jr., M.E. (1960), « Adaptive switching circuits », *IRE WESCON Convention Record*, pp. 96-104.