

Khalil Ahmad · Abdullah

Wavelet packets and Their Statistical Applications



Springer

Forum for Interdisciplinary Mathematics

Editor-in-chief

P. V. Subrahmanyam, Indian Institute of Technology Madras, Chennai, India

Editorial Board

Yogendra Prasad Chaubey, Concordia University, Montreal (Québec), Canada

Janusz Matkowski, University of Zielona Góra, Poland

Thiruvenkatachari Parthasarathy, Chennai Mathematical Institute, Kelambakkam, India

Mathieu Dutour Sikirić, Rudjer Bošković Institute, Zagreb, Croatia

Bhu Dev Sharma, Jaypee Institute of Information Technology, Noida, India

The *Forum for Interdisciplinary Mathematics* series publishes high-quality monographs and lecture notes in mathematics and interdisciplinary areas where mathematics has a fundamental role, such as statistics, operations research, computer science, financial mathematics, industrial mathematics, and bio-mathematics. It reflects the increasing demand of researchers working at the interface between mathematics and other scientific disciplines.

More information about this series at <http://www.springer.com/series/13386>

Khalil Ahmad · Abdullah

Wavelet Packets and Their Statistical Applications



Springer

Khalil Ahmad
Department of Mathematics
Al-Falah University
Faridabad
India

Abdullah
Department of Mathematics
Zakir Husain Delhi College
New Delhi
India

ISSN 2364-6748

ISSN 2364-6756 (electronic)

Forum for Interdisciplinary Mathematics

ISBN 978-981-13-0267-1

ISBN 978-981-13-0268-8 (eBook)

<https://doi.org/10.1007/978-981-13-0268-8>

Library of Congress Control Number: 2018939017

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
part of Springer Nature

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721,
Singapore

To

Prof. H. P. Dikshit

Foreword

It has been our endeavor, to motivate and encourage teaching and research in emerging areas of mathematics and its applications in our country. One of the important initiatives taken by some of us including Prof. Khalil Ahmad in this direction was to organize an International Conference on Advances in Computational Mathematics in 1993 at the Indira Gandhi National Open University, New Delhi, which was supported by the National Board of Higher Mathematics (NBHM), Department of Atomic Energy, Government of India, and the National Science Foundation, USA. Wavelet and its applications were some of the main areas covered in the Conference, proceedings of which were published with Prof. C. A. Micchelli of T. J. Watson Research Center of IBM at Yorktown Heights, USA, and the undersigned as the editors. Since then, a few good schools of teaching and research emerged in the areas of wavelet and its applications in our country, like the one led by Prof. Khalil Ahmad, former Dean Faculty of Natural Sciences, Jamia Millia Islamia University, Delhi. Thus, Prof. Khalil Ahmad and several research students motivated by him have significantly contributed to the important areas of wavelet and its applications. Besides a large number of research contributions to his credit, Prof. Khalil Ahmad in his typically lucid and clear style of expression published jointly with F. A. Shah, a beautiful and exhaustive book *Introduction to Wavelet Analysis with Applications*, Real World Education Publishers, New Delhi (2013), which was well received by students and researchers alike. The present monograph entitled *Wavelet Packets and Their Statistical Applications* are jointly written by him and his coauthor Abdullah is an important contribution to wavelet packets which are a simple but powerful extension of wavelets and multiresolution analysis. A distinctive feature of the monograph is that two separate chapters are devoted to applications of wavelets and wavelet packets. The wavelet packets allow more flexibility in adapting the basis to the frequency contents of a signal, and it is easy to develop a fast wavelet packet transform. The power of wavelet packets lies on the fact that we have much more freedom in deciding which basis function we should use to represent the given function.

Generally, the contributions in the areas of wavelets and their extensions lay varying degree of emphasis on the theory, application, and computation aspects, but this monograph is different as starting from the essential mathematical tools, to a fairly complete and clear development of the theory and its potential areas of applications to computational implementation, all have been treated equally well with clarity. I am confident that this approach will be especially useful for interdisciplinary research in a variety of fields including the computational harmonic analysis and its applications to physical, biological, and medical sciences.

Starting with some basic results in functional analysis, wavelet analysis, and thresholding, construction of wavelet packets and band-limited wavelet packets and some of their important properties are presented in Chap. 2. Chapter 3 deals with the pointwise convergence of wavelet packet series, convolution bounds and convergence of wavelet packet series. Chapter 4 is devoted to characterizations of certain Lebesgue spaces, Hardy space, and Sobolev function spaces by using wavelet packets.

The last two Chaps. 5 and 6 provide a comprehensive study of applications of wavelets and wavelet packets to the important areas of signal and image processing. Speech denoising methods based on wavelets and wavelet packet decompositions of speech signals have been given. The proposed method of wavelet decomposition of speech signals and Wiener filter as post-filtering gives better results in comparison with Donoho's thresholding method. Similarly, the proposed method of wavelet packet decomposition of speech signals gives better results in comparison with those presented in J. P. Areanas (Combining adaptive filtering and wavelet techniques for vibration signal enhancement, *Acustica*, Paper ID-99, (2004), 1–8). A novel wavelet packet denoising method based on optimal decomposition and global threshold value has been also proposed for speech denoising in this work. To check the performance of this method, the signal-to-noise ratio is computed for the denoised speech signal.

Applications of wavelets in biomedical signals related to cardiac problems have been presented with clarity, especially in reference to ECG. An ECG signal denoising method based on wavelet transform is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy ECG signal and the original noise-free ECG signal. In comparison with the method used in A. Mikhled and D. Khaled (ECG signal denoising by wavelet transform thresholding, *American Journal of Applied Sciences*, 5(3) (2008), 276–281), the proposed method gives better result. Similarly, an ECG signal denoising method based on wavelet packet decomposition is proposed which gives a better result in comparison with the method proposed in M. Chendeb, K. Mohamad, and D. Jacques (Methodology of wavelet packet selection for event detection, *Signal Processing Archive*, 86(12) (2006), 3826–3841). For correct estimation of baseline drift in ECG signal, wavelet packet transform has been used in the present work.

Simulation result shows that level 8 is best for correct estimation of baseline drift in ECG signal. Applications of wavelets and wavelet packets to image processing are given in the last chapter.

Jabalpur (M.P.), India

H. P. Dikshit
Ex Vice Chancellor IGNOU

Preface

The theory of wavelets is the latest comer to the world of signal processing (more than 30 years now). It grew and brewed in different areas of science. Harmonic analysts had developed powerful time–frequency tools, electrical engineers were busy with subband coding, and quantum physicists were trying to understand coherent states. They were not aware of each other's progress until the late 1980's when a synthesis of all these ideas came to be, and what is now called wavelet theory contains all those ideas and much more. Wavelet is not one magical transform that solves all problems. It is a library of bases that is appropriate for a large number of situations where the traditional tools, for example, Fourier analysis, are not so good. There are many other problems which cannot be optimally treated with either of the known tools; therefore, new ones have to be designed.

A simple, but a powerful extension of wavelets and multiresolution analysis is wavelet packets, pioneered by Coifman, Meyer, Wickerhauser, and other researchers. The wavelet transform is generalized to produce a library of an orthonormal basis of modulated wavelet packets, where each basis comes with a fast transform. By generalizing the method of multiresolution decomposition, it is possible to construct an orthonormal basis for $L^2(\mathbb{R})$. Discrete wavelet packets have been thoroughly studied by Wickerhauser, who has also developed computer programs and implemented them. The wavelet packets allow more flexibility in adapting the basis to the frequency contents of a signal, and it is easy to develop a fast wavelet packet transform. The power of wavelet packet lies on the fact that we have much more freedom in deciding which basis function we use to represent the given function.

Wavelet packet functions are generated by scaling and translating a family of basic function shapes, which include father wavelet ϕ and mother wavelet ψ . In addition to ϕ , and ψ there is a whole range of wavelet packet functions ω_n . These functions are parametrized by an oscillation or frequency index n . A father wavelet corresponds to $n = 0$, so $\phi = \omega_0$. A mother wavelet corresponds to $n = 1$, so $\psi = \omega_1$. Larger values of n correspond to wavelet packets with more oscillations and higher frequency. Wavelet packets are particular linear combinations or

superpositions of wavelets. They form bases which retain many of the orthogonality, smoothness, and localization properties of their parent wavelets.

The text begins with an elementary chapter on preliminaries such as basic concepts of functional analysis, a short tour of the wavelet transform, Haar scaling functions and function space, Lebesgue spaces $L^p(\mathbb{R})$, Hardy space, Sobolev spaces, Besov spaces, wavelets, Symlets wavelets, and Coiflets wavelets and thresholding.

Chapters 2 and 3 are devoted to the construction of wavelet packets, certain results on wavelet packets, band-limited wavelet packets, characterizations of wavelet packets, MRA wavelet packets, pointwise convergence, the convergence of wavelet packet series, and convolution bounds. Characterizations of function spaces like Lebesgue spaces $L^p(\mathbb{R})$, Hardy space $\mathcal{H}^1(\mathbb{R})$ and Sobolev spaces $L^{p,s}(\mathbb{R})$ in terms of wavelet packets are given in Chap. 4.

A signal can be defined as a function that conveys information, generally about the state or behaviour of the physical system. In almost every area of science and technology, signals must be processed to assist the extraction of information. Thus, the development of signal processing techniques and systems is of great importance. The presence of noise in speech signal can significantly reduce the intelligibility of speech and degrade automatic speech recognition performance. These noises may be due to the background noise of the environment in which the speaker is speaking, or it may be introduced by the transmission media during transmission of the speech signal. It is often necessary to perform speech denoising as the presence of noise, which severely degrades the speech signal. Chapter 5 is devoted to applications of wavelets and wavelet packets in speech denoising and biomedical signals.

The growth of media communication industry and demand of the high quality of visual information in the modern age has an interest to researchers to develop various image denoising techniques. In recent years, there has been a plethora of work on using wavelet thresholding techniques for removing noise in both signal and image processing. Chapter 6 is devoted to applications of wavelets and wavelet packets in image denoising. An exhaustive list of references is given at the end of the monograph.

The present book is intended to serve as a reference book for those working in the area of wavelet packets and their applications in different branches of mathematics and engineering, in particular in signal and image processing. It is also useful for statisticians and to those working in the industrial sector.

New Delhi, India
January 2018

Khalil Ahmad
Abdullah

Contents

1 Preliminaries	1
1.1 Introduction	1
1.2 Basic Concepts of Functional Analysis	1
1.3 A Short Tour of Wavelet Transform	6
1.3.1 Fourier Transform	6
1.3.2 Wavelet Transform	8
1.3.3 Continuous Wavelet Transform	8
1.3.4 Discrete Wavelet Transform	10
1.4 Haar Scaling Functions and Function Space	10
1.4.1 Translation and Scaling of $\varphi(t)$	11
1.4.2 Orthogonality of Translates of $\varphi(t)$	11
1.4.3 Function Space V_0	12
1.4.4 Finer Haar Scaling Functions	13
1.4.5 Nested Space	15
1.4.6 Haar Wavelet Function	16
1.4.7 Scaled Haar Wavelet Functions	17
1.4.8 Orthogonality of $\varphi(t)$ and $\psi(t)$	21
1.4.9 Normalization of Haar Bases at Different Scales	22
1.5 Function Spaces	23
1.5.1 Lebesgue Spaces	23
1.5.2 Hardy Space	24
1.5.3 Sobolev Spaces	24
1.5.4 Besov Spaces	25
1.6 Wavelets	26
1.7 Thresholding	34
2 Wavelet Packets	37
2.1 Introduction	37
2.2 Construction of Wavelet Packets	38
2.3 Certain Results on Wavelet Packets	46

2.4	Band-Limited Wavelet Packets	53
2.5	Characterization of Wavelet Packets	59
2.6	Characterizations of MRA Wavelet Packets	70
3	Convergence of Wavelet Packet Series	79
3.1	Introduction	79
3.2	Pointwise Convergence of Wavelet Packet Series	79
3.3	Convolution Bounds and Convergence of Wavelet Packet Series	85
4	Function Spaces and Wavelet Packets	93
4.1	Introduction	93
4.2	Lebesgue Spaces $L^p(\mathbb{R})$ and Wavelet Packets	93
4.3	Hardy Space $\mathcal{H}^1(\mathbb{R})$ and Wavelet Packets	106
4.4	Sobolev Spaces $L^{p,s}(\mathbb{R})$ and Wavelet Packets	116
5	Applications in Signal Processing	131
5.1	Introduction	131
5.2	Applications of Wavelets in Speech Denoising	132
5.2.1	Speech Signal Processing	132
5.2.2	Noise Reduction in Speech Signal Using Thresholding in Wavelet Transform Domain	137
5.2.3	Estimation of Speech Signal in the Presence of White Noise Using Wavelet Transform	142
5.3	Applications of Wavelet Packets in Speech Denoising	147
5.3.1	Denoising of Speech Signal by Using Wavelet Packet Transform	147
5.3.2	Selection of Optimal Decomposition Level Based on Entropy for Speech Denoising Using Wavelet Packets	154
5.3.3	Thresholding Method for Denoising Speech Signal Using Wavelet Packet Transform	164
5.4	Applications of Wavelets in Biomedical Signals	170
5.4.1	ECG Signal Characteristics	170
5.4.2	Performance Comparison of Wavelet Threshold Estimators for ECG Signal Denoising	177
5.4.3	Denoising of ECG Signal Using Wavelet Transform	182
5.5	Applications of Wavelet Packets in ECG Signal Denoising	189
5.5.1	Wavelet Packet Thresholding Method for Denoising ECG Signal	191
5.5.2	Baseline Drift Correction in ECG Signal Using Wavelet Packets Transform	194

6 Applications in Image Processing	203
6.1 Introduction	203
6.2 Applications of Wavelets in Image Denoising	215
6.3 Applications of Wavelet Packets in Image Denoising	220
References	225
Index	237

About the Authors

Khalil Ahmad has been a Professor at the Department of Mathematics, Al-Falah University, India, since 2015. He previously served as a Professor at the Department of Mathematics, Jamia Millia Islamia, New Delhi, from 1975 to 2014 and worked at Aligarh Muslim University and DAV College, Kanpur. He received his Ph.D. in Mathematics from the University of Delhi in 1982. His major fields of research are functional analysis and applicable mathematics. He is a life member of several academic societies such as the Indian Mathematical Society, Indian Society of Industrial and Applied Mathematics, Jammu Mathematical Society, and Gwalior Academy of Mathematical Sciences. He has over 50 research papers to his credit, published in national and international journals of repute. He is the author of 12 books and has delivered lectures at over 60 mathematical conferences.

Abdullah is an Assistant Professor at the Department of Mathematics, Zakir Husain Delhi College, University of Delhi, New Delhi, India. His main fields of research are wavelet analysis and frame theory. He received his Ph.D. in Mathematics from Jamia Millia Islamia, New Delhi, in 2007. He has over 25 research papers to his credit and has attended more than 20 mathematical conferences.

Chapter 1

Preliminaries



1.1 Introduction

In this chapter we enlist those concepts and results concerning functional analysis, wavelet analysis and thresholding which are already known in the literature and we require in the subsequent chapters.

1.2 Basic Concepts of Functional Analysis

Throughout, the functions f , g , φ , ψ , and ω_n will stand for $f(x)$, $g(x)$, $\varphi(x)$, $\psi(x)$, and $\omega_n(x)$, respectively.

Let \mathbb{Z} and \mathbb{R} denote the set of integers and real numbers, respectively, and \mathbb{T} denote the unit circle in the complex plane which can be identified with the interval $[-\pi, \pi]$. The inner product of two functions $f \in L^2(\mathbb{R})$ and $g \in L^2(\mathbb{R})$ is denoted by $\langle f, g \rangle$ and is defined as

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x) \overline{g(x)} dx.$$

The norm of $f \in L^2(\mathbb{R})$ is written as $\|f\|$. The Fourier transform of any function $f \in L^2(\mathbb{R})$ is denoted by \hat{f} and is defined as

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i\xi x} dx.$$

The inverse Fourier transform of any function $g \in L^2(\mathbb{R})$ is denoted by \check{g} and is defined as

$$\check{g} = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\xi) e^{i\xi x} d\xi$$

and if we apply it to $g = \hat{f}$ we obtain $\check{g} = f$, that is $(\hat{f})^\vee = f$. With this definition, the Plancherel theorem asserts that

$$\langle f, g \rangle = \frac{1}{2\pi} \langle \hat{f}, \hat{g} \rangle.$$

The set $l^2(\mathbb{Z})$ is the vector space of square-summable sequences, i.e.,

$$l^2(\mathbb{Z}) = \left\{ \{h_k\}_{k \in \mathbb{Z}} : \sum_{k=-\infty}^{\infty} |h_k|^2 < \infty \right\}.$$

Throughout we shall denote \mathbb{R}^0 , S for the regularity class and Schwartz class, respectively. For the dual of Schwartz class, we denote S' .

Lemma 1.2.1 *Let \mathbb{H} be a Hilbert space and $\{e_j : j = 1, 2, \dots\}$ be a family of elements of \mathbb{H} . Then*

$$(i) \quad \|f\|^2 = \sum_{j=1}^{\infty} |\langle f, e_j \rangle|^2 \text{ holds for all } f \in \mathbb{H}$$

if and only if

$$(ii) \quad f = \sum_{j=1}^{\infty} \langle f, e_j \rangle e_j, \text{ with convergence in } \mathbb{H}, \text{ for all } f \in \mathbb{H}.$$

Lemma 1.2.2 *Suppose $\{e_j : j = 1, 2, \dots\}$ be a family of elements in a Hilbert space \mathbb{H} such that equality (i) in Lemma 1.2.1 holds for all f belonging to a dense subset D of \mathbb{H} , then the equality is valid for all $f \in \mathbb{H}$.*

Lemma 1.2.3 *Let C be a positive integer and let $\{v_j : j \geq 1\}$ be a family of vectors in a Hilbert space \mathbb{H} such that*

$$(i) \quad \sum_{n=1}^{\infty} \|v_n\|^2 = C \text{ and}$$

$$(ii) \quad v_n = \sum_{m=1}^{\infty} \langle v_n, v_m \rangle v_m \text{ for all } n \geq 1.$$

Let $\mathbb{F} = \overline{\text{span}\{v_j : j \geq 1\}}$. Then, $\dim \mathbb{F} = \sum_{j=1}^{\infty} \|v_j\|^2 = C$ (Number of basis elements of \mathbb{F}).

Definition 1.2.4 For a given function g defined on \mathbb{R} , we say that a bounded function $H : [0, \infty) \rightarrow \mathbb{R}^+$ is a radial decreasing L^1 -majorant of g if $|g(x)| \leq H(|x|)$ and H satisfies the following conditions

$$\begin{cases} (i) H \in L^1[0, \infty) \\ (ii) H \text{ is decreasing} \\ (iii) H(0) < \infty \end{cases} \quad (1.2.1)$$

The set of all bounded radially decreasing functions is denoted by RB .

Lemma 1.2.5 *Let H be a function on $[0, \infty)$ satisfying condition (1.2.1). Then*

$$\sum_{k \in \mathbb{Z}} H(|x - k|) H(|y - k|) \leq CH \left[\frac{|x - y|}{2} \right], \quad \forall x, y \in \mathbb{R}$$

where C is a constant depending on H .

Definition 1.2.6 The point $x \in \mathbb{R}$ is said to be a Lebesgue point of a function f on \mathbb{R} if f is integrable in some neighborhood of x and

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{V(B_\varepsilon)} \int_{B_\varepsilon} |f(x) - f(x + y)| dy = 0$$

where B_ε denotes the ball of radius ε about the origin and V denotes volume.

Definition 1.2.7 For a function g defined on \mathbb{R} and for a real number $\lambda > 0$, the maximal function is defined by

$$g_\lambda^*(x) = \sup_{y \in \mathbb{R}} \frac{|g(x - y)|}{(1 + |y|)^\lambda}, \quad x \in \mathbb{R}. \quad (1.2.2)$$

Definition 1.2.8 Hardy–Littlewood maximal function, $\mathcal{M}f(x)$, is defined by

$$\mathcal{M}f(x) = \sup_{r > 0} \frac{1}{2r} \int_{|y-x| \leq r} |f(y)| dy \quad (1.2.3)$$

for a locally integrable function f on \mathbb{R} .

It is well known that \mathcal{M} is bounded on $L^p(\mathbb{R})$, $1 < p \leq \infty$. An important property of \mathcal{M} that we shall need is the following vector-valued inequality:

Lemma 1.2.9 *Suppose $1 < p, q < \infty$; then there exists a constant $C_{p,q}$ such that*

$$\left\| \left\{ \sum_{i=1}^{\infty} (\mathcal{M}f_i)^q \right\}^{\frac{1}{q}} \right\|_{L^p(\mathbb{R})} \leq C_{p,q} \left\| \left\{ \sum_{i=1}^{\infty} |f_i|^q \right\}^{\frac{1}{q}} \right\|_{L^p(\mathbb{R})} \quad (1.2.4)$$

for any sequence $\{f_i : i = 1, 2, \dots\}$ of locally integrable functions.

Lemma 1.2.10 *Let g be a band-limited function on \mathbb{R} such that $g_\lambda^*(x) < \infty$ for all $x \in \mathbb{R}$. Then, there exists a constant C_λ such that*

$$g_\lambda^*(x) \leq C_\lambda \left\{ \mathcal{M}(|g|^{\frac{1}{\lambda}})(x) \right\}^\lambda, \quad x \in \mathbb{R}. \quad (1.2.5)$$

Lemma 1.2.11 *If g is a band-limited function, i.e., support of \hat{g} is contained in a finite interval, defined on \mathbb{R} such that $g \in L^p(\mathbb{R})$, $0 < p \leq \infty$, then we have $g_\lambda^*(x) < \infty$ for all $x \in \mathbb{R}$.*

Lemma 1.2.12 (Hörmander–Mihlin Multiplier Theorem). *Let \mathbb{H}_0 and \mathbb{H}_1 be two Hilbert spaces, and we denote by $\mathcal{L}(\mathbb{H}_0, \mathbb{H}_1)$ the set of all bounded linear operators T from \mathbb{H}_0 to \mathbb{H}_1 . Assume that m is a function defined on \mathbb{R} with values in $\mathcal{L}(\mathbb{H}_0, \mathbb{H}_1)$ such that*

$$\|(D^j m)(\xi)\|_{\mathcal{L}(\mathbb{H}_0, \mathbb{H}_1)} \leq B \frac{1}{|\xi|^j}, \quad j = 0, 1 \quad (1.2.6)$$

for some positive constant $B < \infty$. Then, the operator T_m given by

$$(T_m f)(\xi) = m(\xi) \hat{f}(\xi) \quad \text{for all } f \in S(\mathbb{H}_0)$$

can be extended to a bounded linear operator from $L^p(\mathbb{R}; \mathbb{H}_0)$ to $L^p(\mathbb{R}; \mathbb{H}_1)$, $1 < p < \infty$. That is, there exists a constant C , $0 < C < \infty$ such that

$$\|T_m f\|_{L^p(\mathbb{R}; \mathbb{H}_1)} \leq C \|f\|_{L^p(\mathbb{R}; \mathbb{H}_0)} \quad \text{for all } f \in L^p(\mathbb{R}; \mathbb{H}_0). \quad (1.2.7)$$

Lemma 1.2.13 *Given $\varepsilon > 0$ and $1 \leq r < 1 + \varepsilon$, there exists a constant C such that for all sequences $\{s_{l,k} : l, k \in \mathbb{Z}\}$ of complex numbers and all $x \in I_{l,k}$,*

$$(a) \sum_{k' \in \mathbb{Z}} \frac{|s_{l',k'}|}{(1 + 2^{l'}|2^{-l}k - 2^{-l'}k'|)^{1+\varepsilon}} \leq C \left[\mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |s_{l',k'}|^{\frac{1}{r}} \chi_{I_{l',k'}} \right) (x) \right]^r \text{ if } l' \leq l$$

and

$$(b) \sum_{k' \in \mathbb{Z}} \frac{|s_{l',k'}|}{(1 + 2^l|2^{-l'}k' - 2^{-l}k|)^{1+\varepsilon}} \leq C 2^{(l'-l)r} \times \left[\mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |s_{l',k'}|^{\frac{1}{r}} \chi_{I_{l',k'}} \right) (x) \right]^r \text{ if } l' \geq l$$

where \mathcal{M} is the Hardy–Littlewood maximal function defined in Definition 1.2.6 and $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$.

Definition 1.2.14 We say that a function φ defined on \mathbb{R} belongs to the regularity class \mathbb{R}^0 if there exist constants C_0, C_1, γ , and $\varepsilon > 0$ such that

- (i) $\int_{\mathbb{R}} \varphi(x) dx = 0$
- (ii) $|\varphi(x)| \leq \frac{C_0}{(1 + |\alpha|)^{2+\gamma}}$ for all $x \in \mathbb{R}$
- (iii) $|\varphi'(x)| \leq \frac{C_1}{(1 + |\alpha|)^{1+\varepsilon}}$ for all $x \in \mathbb{R}$.

Lemma 1.2.15 Let $\varepsilon > 0$. Suppose that g and h satisfy

- (a) $|g(x)| \leq \frac{C_1}{(1+|x|)^{1+\varepsilon}}$ for all $x \in \mathbb{R}$ and
- (b) $|h(x)| \leq \frac{C_2}{(1+|x|)^{1+\varepsilon}}$ for all $x \in \mathbb{R}$

with C_1 and C_2 independent of $x \in \mathbb{R}$. Then, there exists a constant C such that for all $l, k, l', k' \in \mathbb{Z}$ and $l \leq l'$, we have

$$|(g_{l,k} * h_{l',k'})(x)| \leq \frac{C 2^{\frac{1}{2}(l-l')}}{(1+2^l|x-2^{-l}k-2^{-l'}k'|)^{1+\varepsilon}} \quad \text{for all } x \in \mathbb{R}.$$

Lemma 1.2.16 Let $r \geq \varepsilon > 0$ and $N \in \mathbb{N}$. Suppose that g and h satisfy

- (a) $\left| \frac{d^n g}{dx^n}(x) \right| \leq \frac{C_{n,1}}{(1+|x|)^{1+\varepsilon}}$ for all $x \in \mathbb{R}$ and $0 \leq n \leq N+1$
- (b) $\int_{\mathbb{R}} x^n h(x) dx = 0$ for all $0 \leq n \leq N$
- (c) $|h(x)| \leq \frac{C_2}{(1+|x|)^{2+N+r}}$ for all $x \in \mathbb{R}$

with $C_{n,1}$, $0 \leq n \leq N+1$, and C_2 independent of $x \in \mathbb{R}$. Then, there exists a constant C such that for all $l, k, l', k' \in \mathbb{Z}$ and $l \leq l'$, we have

$$|(g_{l,k} * h_{l',k'})(x)| \leq \frac{C 2^{(l-l')(\frac{1}{2}+N+1)}}{(1+2^l|x-2^{-l}k-2^{-l'}k'|)^{1+\varepsilon}} \quad \text{for all } x \in \mathbb{R}.$$

For $N \in \mathbb{N} \cup \{-1\}$, let \mathcal{D}^N be the set of all functions f defined on \mathbb{R} for which there exist constants $\varepsilon > 0$ and $C_n < \infty$, $n = 0, 1, \dots, N+1$, such that

$$\left| D^n f(x) \right| \leq \frac{C_n}{(1+|x|)^{1+\varepsilon}} \quad \text{for all } x \in \mathbb{R} \text{ and } 0 \leq n \leq N+1.$$

We write \mathcal{M}^N for the set of all functions f defined on \mathbb{R} for which there exist constants $\gamma > 0$ and $C < \infty$ such that

$$\int_{\mathbb{R}} x^n f(x) dx = 0 \quad \text{for } n = 0, 1, \dots, N$$

$$\text{and} \quad |f(x)| \leq C \frac{1}{(1+|x|)^{2+N+\gamma}} \quad \text{for all } x \in \mathbb{R}.$$

Definition 1.2.17 For a nonnegative integer s , let $\mathbb{R}^s = \mathcal{D}^s \cap \mathcal{M}^s$; that is, $f \in \mathbb{R}^s$ if there exist constants $\varepsilon > 0$, $\gamma > 0$, $C < \infty$, and $C_n < \infty$, $n = 1, 2, \dots, s+1$, such that

- (i) $\int_{\mathbb{R}} x^n f(x) dx = 0 \quad \text{for } n = 0, 1, \dots, s$
- (ii) $|f(x)| \leq \frac{C}{(1+|x|)^{2+s+\gamma}} \quad \text{for all } x \in \mathbb{R}$
- (iii) $|D^n f(x)| \leq \frac{C_n}{(1+|x|)^{1+\varepsilon}} \quad \text{for all } x \in \mathbb{R}, n = 1, 2, \dots, s+1.$

Definition 1.2.18 The Schwartz space S is the subspace of C^∞ (the set of all bounded continuous functions) given

$$S = \bigcap_{N=1}^{\infty} \left\{ f \in C^\infty : \sum_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^n} | \langle x \rangle^N \partial^\alpha f(x) | < \infty \right\}.$$

The topology of S is the weakest one for which the mapping $f \mapsto p_N(f) \in \mathbb{R}$ is continuous for all $N \in \mathbb{N}$, where

$$p_N(f) = \sum_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^n} | \langle x \rangle^N \partial^\alpha f(x) |, \quad \langle x \rangle = \sqrt{1 + |x|^2}.$$

Definition 1.2.19 One defines

$$S' = \{f : S \rightarrow \mathbb{C} : f \text{ is linear and continuous}\}$$

One equips S' with the weakest topology so that the mapping

$$f \in S' \rightarrow \langle f, \varphi \rangle \in \mathbb{C}$$

is continuous for all $\varphi \in S$.

1.3 A Short Tour of Wavelet Transform

The need of simultaneous representation and localizations of both time and frequency for nonstationary signals (e.g., music, speech, images) led toward the advancement of wavelet transform from the popular Fourier transform. Different “time–frequency representations” (TFR) are very informative in understanding and modeling of wavelet transform [68, 124, 143].

1.3.1 Fourier Transform

Fourier transform is a well-known mathematical tool to transform time-domain signal to frequency-domain for efficient extraction of information and it is reversible also. For a signal $x(t)$, the Fourier transform is given by:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt$$

The Fourier transform has a great ability to detain signal's frequency content as long as $x(t)$ is composed of few stationary components (e.g., sine waves). However, any abrupt change in time for nonstationary signal $x(t)$ is spread out over the whole frequency axis in $X(f)$. Hence, the time-domain signal sampled with Dirac delta function is highly localized in time but spills over entire frequency band and vice versa. The limitation of Fourier transform is that it cannot offer both time and frequency localization of a signal at the same time.

To overcome the limitations of the standard Fourier transform, Gabor [117] introduced the initial concept of short-time Fourier transform (STFT). The advantage of STFT is that it uses an arbitrary but fixed-length window $g(t)$ for analysis, over which the actual nonstationary signal is assumed to be approximately stationary. The STFT decomposes such a pseudo-stationary signal $x(t)$ into a two-dimensional time–frequency representation $S(\tau, f)$ using that sliding window $g(t)$ at different times τ . Thus, the Fourier transform of windowed signal $x(t) * (t - \tau)$ yields STFT as:

$$\text{STFT}_x(\tau, f) = \int_{-\infty}^{\infty} x(t)g * (t - \tau)e^{-j2\pi ft} dt$$

Filter bank interpretation is an alternative way of seeing “windowing of the signal” viewpoint of STFT [19, 219]. With the modulated filter bank, a signal can be seen as passing through a bandpass filter centered at frequency f with an impulse response of the window function modulated to that frequency. From this dual interpretation, a possible drawback related to time–frequency resolution of STFT can be shown through “Heisenberg’s uncertainty principle” [41, 256]. For a window $g(t)$ and its Fourier transform $G(f)$, both centered around the origin in time as well as in frequency, i.e., satisfying $\int t|g(t)|^2 dt = 0$ and $\int f|G(f)|^2 df = 0$. Then, the spreads in time and frequency are defined as:

$$\Delta_t^2 = \frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt}, \quad \Delta_f^2 = \frac{\int_{-\infty}^{\infty} f^2 |G(f)|^2 df}{\int_{-\infty}^{\infty} |G(f)|^2 df}$$

Thus, the time–frequency resolution for STFT is lower bounded by their product as:

$$\text{Time-Bandwidth product } \Delta_t \Delta_f \geq \frac{1}{4\pi}.$$

Once a window has been chosen for STFT, the time–frequency resolution is fixed over the entire time–frequency plane because the same window is used at all frequencies. There is always a trade-off between time resolution and frequency resolution in STFT.

1.3.2 Wavelet Transform

Fixed resolution limitation of STFT can be resolved by letting the resolution Δ_t and Δ_f vary in time–frequency plane in order to obtain multiresolution analysis. The wavelet transform (WT) in its continuous form; i.e., CWT provides a flexible time–frequency window, which narrows when observing high-frequency phenomena and widens when analyzing low-frequency behavior. Thus, time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies. This kind of analysis is suitable for signals composed of high-frequency components with short duration and low-frequency components with long duration, which is often the case in practical situations [230].

When analysis is viewed as a filter bank, the wavelet transform, generally termed as standard discrete wavelet transform (DWT), is seen as a composition of bandpass filters with constant relative bandwidth such that Δ_f/f is always constant. As Δ_f changes with frequencies, corresponding time resolution Δ_t also changes so as to satisfy the uncertainty condition. The frequency responses of bandpass filters are logarithmically spread over frequency.

1.3.3 Continuous Wavelet Transform

It is very clear that wavelet means “small wave,” so wavelet analysis is about analyzing signal with short duration finite energy functions. Mathematically, wavelet can be represented as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$$

with some wavelet admissibility conditions:

$$C_\psi = \int_0^\infty \frac{|\hat{\psi}(\omega)|}{\omega} d\omega < \infty$$

and

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$$

Where “ b ” is location parameter, “ a ” is scaling parameter, $\hat{\psi}(\omega)$ is the Fourier transform, which ensures that $\hat{\psi}(\omega)$ goes to zero quickly as $\omega \rightarrow 0$. In fact to guarantee that $C_\psi < \infty$, we must impose $\hat{\psi}(0) = 0$. Wavelet transform is defined as:

$$W(a, b) = \int_t f(t) \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) dt. \quad (1.3.1)$$

According to above equation, for every (a, b) , we have a wavelet transform coefficient, representing how much the scaled wavelet is similar to the function at location $t = \frac{a}{b}$.

Now, we can say that continuous wavelet transform (CWT) is a function of two parameters and, therefore, contains a high amount of extra (redundant) information when analyzing a function. A critical sampling of the CWT

$$W(a, b) = \int_t f(t) \frac{1}{\sqrt{|a|}} \psi \left(\frac{t - b}{a} \right) dt$$

is obtained via $a = 2^{-j}$, where j and k (in next integral) are integers representing the set of discrete translations and discrete dilations. Upon substitution, Eq. (1.3.1) can become

$$\int f(t) 2^{j/2} \psi(2^j t - k) dt$$

which is function of j and k . We denote it by $W(j, k)$. In general, $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$ is wavelet for all integers j and k and produces an orthogonal basis. We call $\psi_{0,0}(t) = \psi(t)$ as mother wavelet. Other wavelets are produced by translation and dilation of the mother wavelet [60].

In earlier time, a phenomenon that is known as Heisenberg's uncertainty principle, which says that a signal cannot be simultaneously localized in time and frequency. Wavelets are an attempt to overcome this shortcoming. They provide a way to do time-frequency analysis. The idea is that one chooses a "mother wavelet," i.e., a function subject to some conditions like mean value by using two variable base (one for the amount to shift and one for the amount of dilation); we are able to introduce enough redundancy to maintain the local properties of the original function.

Overall, we can say that continuous wavelet transform is defined as the sum over all time of the signal multiplied by scaled, shifted version of the wavelet function ψ :

$$C(\text{scale, position}) = \int_{-\infty}^{\infty} f(t) \psi((\text{scale, position})) dt$$

The results of the CWT are many wavelet coefficients C , which are a function of scale and position. Multiplying each coefficient by the appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal.

Now, we are very eager to know that what is continuous in CWT, because in any signal processing real-world data must be performed on a discrete signal. The speciality of CWT is that it can operate at every scale, from that of the original signal up to some maximum scale that we determine by trading off our need for detailed analysis. The CWT is also continuous in term of shifting: during computation, which makes continuous wavelet transform distinguishing from others.

1.3.4 Discrete Wavelet Transform

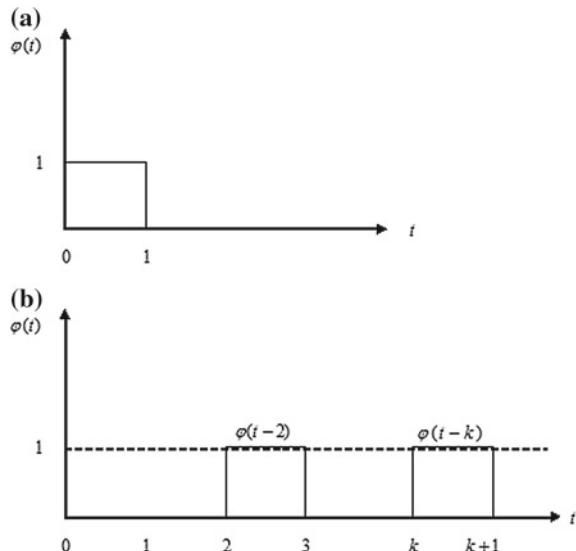
Unlike conventional methods, in wavelet transform, one can use a single function and its dilations and translations to generate a set of orthonormal basis functions to represent a signal. Numbers of such functions are infinite, and we can choose one that suits to the application. Unfortunately, most of the wavelets used in discrete wavelet transform are fractal in nature. They are expressed in terms of a recurrence relation so that to see them we must do several iterations. But fortunately, we have two special functions known as Haar wavelet functions and Haar scaling functions, which have explicit expression. To understand mathematical (geometry) approach insight wavelet transform, Haar function is the only hope. Scaling functions and wavelet functions are just like twins; corresponding to wavelet function there is a scaling function. The details about these will be discussed in next sections of this chapter [18, 257].

1.4 Haar Scaling Functions and Function Space

In discrete wavelet transform, we have to deal with basically two sets of functions—scaling and wavelet functions. Understanding the relation between these two functions, consider Haar scaling function $\varphi(t)$ defined in Eq. (1.4.1) and shown in Fig. 1.1

$$\varphi(t) = \begin{cases} 1, & 0 \leq t \leq 1 \\ 0 & \text{elsewhere} \end{cases} \quad (1.4.1)$$

Fig. 1.1 **a** Haar scaling function and **b** translation of Haar scaling function



The domain of the function is $[0, 1]$. Here, we see that the function is time-limited and has finite energy. That is, $\int_{-\infty}^{\infty} |f(t)|^2 dt$ exists and finite. Now, we will make a set of functions that are translates of $\varphi(t)$. Then, it will be shown that these functions are orthogonal and can form a basis set just like sine and cosine functions form a basis set for Fourier series analysis.

1.4.1 Translation and Scaling of $\varphi(t)$

Consider functions of the type $\varphi(t - 1)$, $\varphi(t - 2)$, $\varphi(t + 1)$, or, in general, $\varphi(t - k)$. These functions are known as translates of $\varphi(t)$. In function $\varphi(t)$, the function exists practically for values of t in the range $[0, 1]$. Beyond this range function value is zero. We say that the domain of the function is $[0, 1]$. What about $\varphi(t - 1)$? Where does it exist? To visualize this, consider $\varphi(t)$ again. According to Eq. 1.4.1, quantity inside the bracket of φ may vary only in the closed interval $[0, 1]$ to get a nonzero value. Once the function is defined, this remains same irrespective of what we write inside the bracket. For example, once we define $\varphi(t)$ as in Eq. (1.4.1), then the quantity inside the bracket, that is, $t - 1$ may vary in the range $[0, 1]$ to have nonzero values for φ . Therefore, t in $\varphi(t - 1)$ when vary between 1 and 2, function value is 1, and for all other values of t , the function value is zero. We can generalize this fact for $\varphi(t - k)$, we set $t - k = 0$ to get $t = k$, and set $t - k = 1$ to get $t = k + 1$. So the function $\varphi(t - k)$ has nonzero value for t in the range $[k, k + 1]$.

If we consider the function $\varphi(10t)$, then it has nonzero value in the range $[0, 1/10]$; here $\varphi(10t)$ is a scaled version of $\varphi(t)$. Now, what about $\varphi(10t - k)$? We set $10t - k = 0$ to get $t = k/10$, and set $10t - k = 1$ to get $t = (k + 1)/10$. Therefore, the function $\varphi(10t - k)$ has nonzero value in the interval $[k/10, (k + 1)/10]$. Thus, $\varphi(10t - k)$ is a scaled and translated version of $\varphi(t)$.

1.4.2 Orthogonality of Translates of $\varphi(t)$

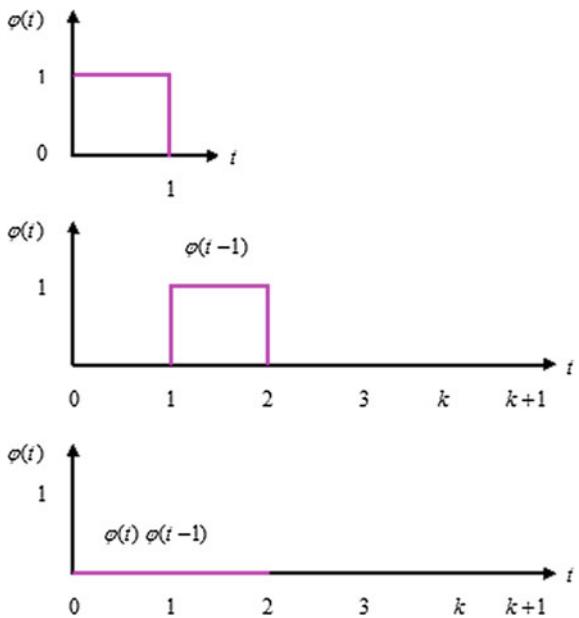
Consider $\int_{-\infty}^{\infty}$. For Haar, $\varphi(t)\varphi(t - 1)$ is identically zero (Fig. 1.2). Now, we are able to tell the location where the function exists. It can be easily seen that pointwise multiplication of these two functions results in zero for every value of t . Therefore, the integral is zero. Here it is quite obvious, since, there is no overlap between the two functions in the sense that wherever $\varphi(t)$ has nonzero value $\varphi(t - 1)$ is zero.

So, for Haar scaling function

$$\int_{-\infty}^{\infty} \varphi(t - m)\varphi(t - n)dt = \delta_{m-n}. \quad (1.4.2)$$

When m and n are equal, the functions are identical and perfectly overlap so that the area is equal to

Fig. 1.2 Orthogonality of translates of $\varphi(t)$



$$\int_{-\infty}^{\infty} \varphi(t-m)\varphi(t-n)dt = \delta_{m-n} = \int_{-\infty}^{\infty} \varphi(t)\varphi(t)dt = \int_0^1 1^2 dt = 1.$$

Equation (1.4.2) implies that the functions $\varphi(t - m)$ and $\varphi(t - n)$ are orthonormal.

1.4.3 Function Space V_0

We now have a set of orthonormal functions $\{\dots, \varphi(t+1), \varphi(t), \varphi(t-1), \dots\}$, which are translates of a single function $\varphi(t)$. These orthonormal functions can now represent certain type of signal. For example, consider Fig. 1.3, a sort of “random staircase” type signal, which can be expressed using the bases we have just defined.

Let V_0 be the space spanned by the set of bases $\{\dots, \varphi(t+1), \varphi(t), \varphi(t-1), \dots\}$, and we denote this as:

$$V_0 = \text{Span}\{\overline{\varphi(t-k)}\}. \quad (1.4.3)$$

Consider a function $f(t) = \sum_{k=-\infty}^{\infty} a_k \varphi(t-k)$ where a_k 's are real numbers (scalars) which we call as coefficients of $\varphi(t-k)$'s. For, one set of a_k 's, we have one particular signal. But assume that we are continuously changing a_k 's to generate continuously new functions or signals. The set of all such signals constitute the function space V_0 . Note that, all types of signals cannot be elements of V_0 . For a signal $f(t)$ to be an element of V_0 , we must be able to express $f(t)$ using the bases of V_0 . What is the

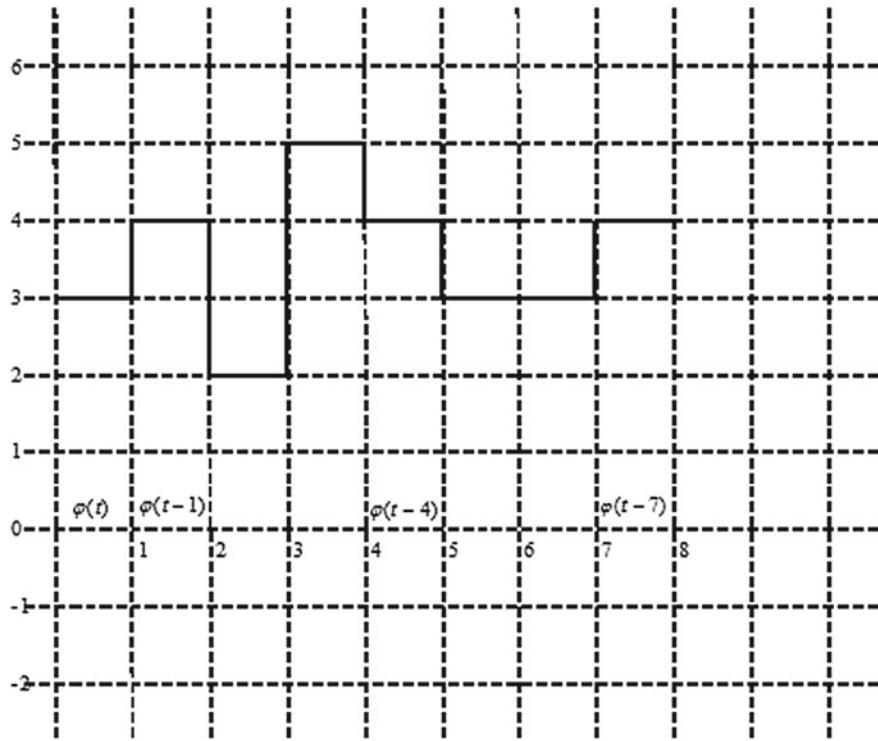


Fig. 1.3 A representative signal in V_0 , spanned by Haar scaling functions

speciality of a signal in V_0 which is spanned by the Haar scaling functions? They are piecewise constant in the unit interval. Figure 1.3 shows such a signal. Here, $f(t)$ is given by

$$\begin{aligned} f(t) = & 3\varphi(t) + 4\varphi(t-1) + 2\varphi(t-2) + 5\varphi(t-3) + 4\varphi(t-4) \\ & + 3\varphi(t-5) + 3\varphi(t-6) + 4\varphi(t-7) \end{aligned} \quad (1.4.4)$$

Here, the set of all such signals constitute the signal space V_0 . Can the signals given in Fig. 1.4 belong to V_0 (spanned by Haar scaling function)?

No, both $f(t)$ and $f_1(t)$ cannot be represented by the bases of V_0 . The signal in V_0 space must be piecewise constant in each unit interval [257].

1.4.4 Finer Haar Scaling Functions

Let us now scale the Haar basis function and form a new basis set. We scale $\varphi(t)$ by 2 to form functions of the type $\varphi(2t+1)$, $\varphi(2t)$, $\varphi(2t-2)$ or in general $\varphi(2t-k)$. These functions are again nonoverlapping and are, therefore, orthogonal among themselves.

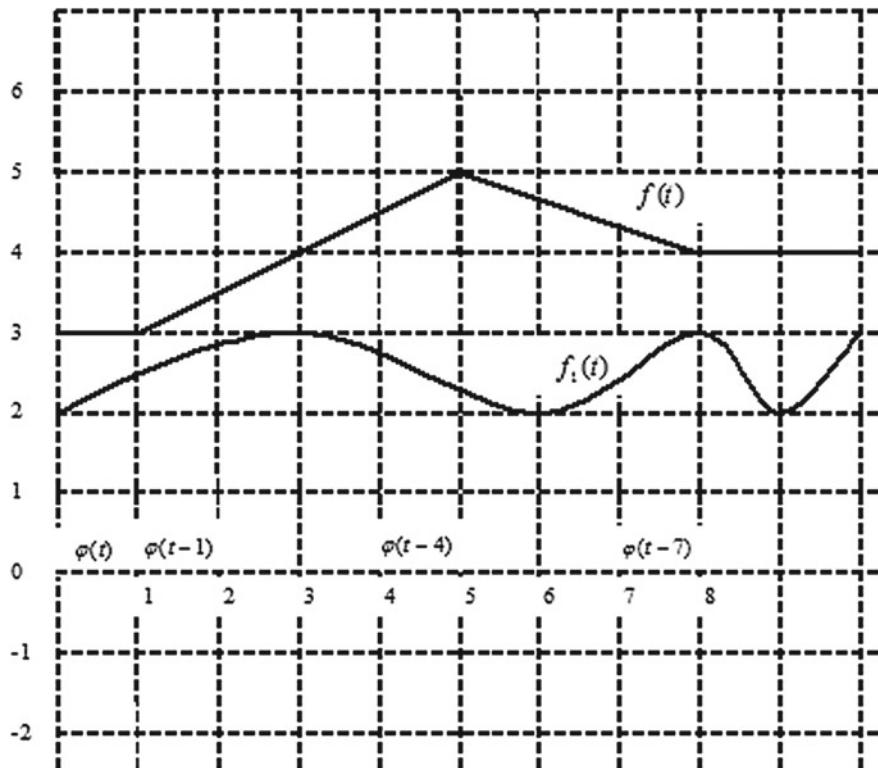


Fig. 1.4 Signals which are not in V_0

We call the space spanned by this set of functions $\{\varphi(2t - k) : k \in \mathbb{N}\}$ as V_1 . Figure 1.5 shows the new set of bases. Formally,

$$V_1 = \text{Span}_k \{\overline{\varphi(2t - k)}\}$$

Any signal in such span can be written as:

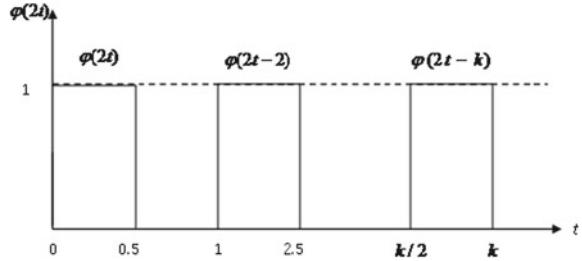
$$f_1(t) = \sum_{k=-\infty}^{\infty} a_k \varphi(2t - k). \quad (1.4.5)$$

By varying a_k 's in Eq. (1.4.5), we can generate new functions and set of all such possible functions which constitute the space V_1 . Similarly, V_2 is the space spanned by $\varphi(2^2 t - k)$, that is

$$V_2 = \text{Span}_k \{\overline{\varphi(2^2 t - k)}\}.$$

Generalizing in this way, V_j is the space spanned by $\varphi(2^j t - k)$, that is

Fig. 1.5 Haar scaling functions which form the basis for V_1



$$V_j = \text{Span}_k \{\overline{\varphi(2^j t - k)}\}.$$

Let us now move on to the concept of nested spaces, which is one of the cornerstones of wavelet theory [257].

1.4.5 Nested Space

If we see Figs. 1.2 and 1.4 closely, we have a signal in V_0 space which basically means that the signal can be represented using the basis of V_0 . Our signal is given by Eq. (1.4.4), i.e.,

$$\begin{aligned} f(t) = & 3\varphi(t) + 4\varphi(t-1) + 2\varphi(t-2) + 5\varphi(t-3) + 4\varphi(t-4) \\ & + 3\varphi(t-5) + 3\varphi(t-6) + 4\varphi(t-7) \end{aligned}$$

The same signal can also be represented by using the basis of V_1 , i.e.,

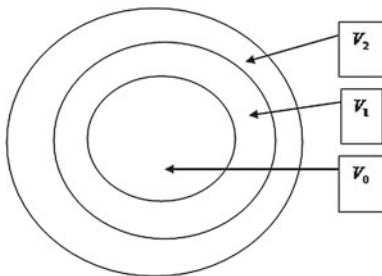
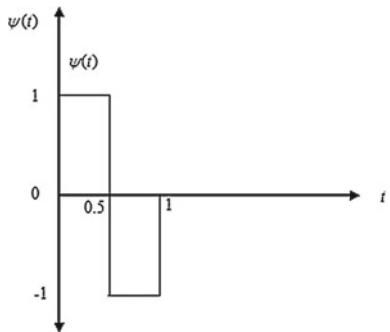
$$\begin{aligned} f(t) = & 3\varphi(2t) + 3\varphi(2t-1) + 4\varphi(2t-2) + 4\varphi(2t-3) + 2\varphi(2t-4) \\ & + 2\varphi(2t-5) + \dots \end{aligned} \quad (1.4.6)$$

We have substituted $3\varphi(t)$ by $3\varphi(2t) + 3\varphi(2t-1)$, $4\varphi(t-1)$ by $4\varphi(2t-2) + 4\varphi(2t-3)$, and so on. This is true for any signal in space V_0 . Therefore, we say that V_0 is contained in V_1 or $V_0 \subset V_1$. Moreover, V_1 is a finer space than V_0 and contains all the signals in V_0 . This is possible because bases of V_0 itself can be represented by using bases of V_1 . For example,

$$\begin{aligned} \varphi(t) &= \varphi(2t) + \varphi(2t-1) \\ \varphi(t-1) &= \varphi(2t-2) + \varphi(2t-3) \end{aligned} \quad (1.4.7)$$

The above relation is called scaling relation or refinement relation or dilation equation. Similarly, $V_1 \subset V_2$; i.e., any signal in V_1 can be represented by using the basis of V_2 . Again this is possible because any basis of V_1 can be represented by using the basis of V_2 , i.e.,

$$\varphi(2t) = \varphi(4t) + \varphi(4t-1)$$

Fig. 1.6 Nested spaces**Fig. 1.7** Haar wavelet function

In general, we can write

$$\dots V_{-1} \subset V_0 \subset V_1 \subset V_2 \dots \subset V_\infty \quad (1.4.8)$$

Pictorially, we visualize this as exhibited in Fig. 1.6.

Based on the relation $V_0 \subset V_1$, it is natural to ask, what is missing in V_0 that makes V_0 a subset of V_1 ? To answer this question, let us explore another function called Haar wavelet function.

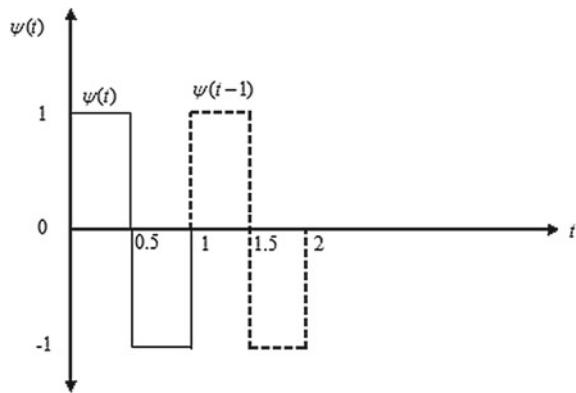
1.4.6 Haar Wavelet Function

Haar wavelet function $\psi(t)$ (Fig. 1.7) is given by

$$\psi(t) = \begin{cases} 1, & 0 \leq t \leq 1/2 \\ -1, & 1/2 \leq t \leq 1 \\ 0, & \text{elsewhere} \end{cases} \quad (1.4.9)$$

It is quite easy to see that the set of functions $\{\psi(t-k) : k \in \mathbb{N}\}$ form an orthonormal set of basis functions. Like Haar scaling function, there is no overlap between the translates of $\psi(t)$ (Fig. 1.8).

Fig. 1.8 Haar wavelet function and its translates



Therefore

$$\int_{-\infty}^{\infty} \psi(t)\psi(t-1)dt = 0.$$

In general

$$\int_{-\infty}^{\infty} \psi(t-m)\psi(t-n)dt = \delta_{m-n}. \quad (1.4.10)$$

That is, the integral is 1 if $m = n$ and 0 if $m \neq n$. Figure 1.9 shows that

$$\int_{-\infty}^{\infty} \psi(t)\psi(t)dt = 1$$

Let W_0 be the space spanned by the orthonormal set of bases $\{\psi(t-k) : k \in \mathbb{N}\}$.

Formally, we define

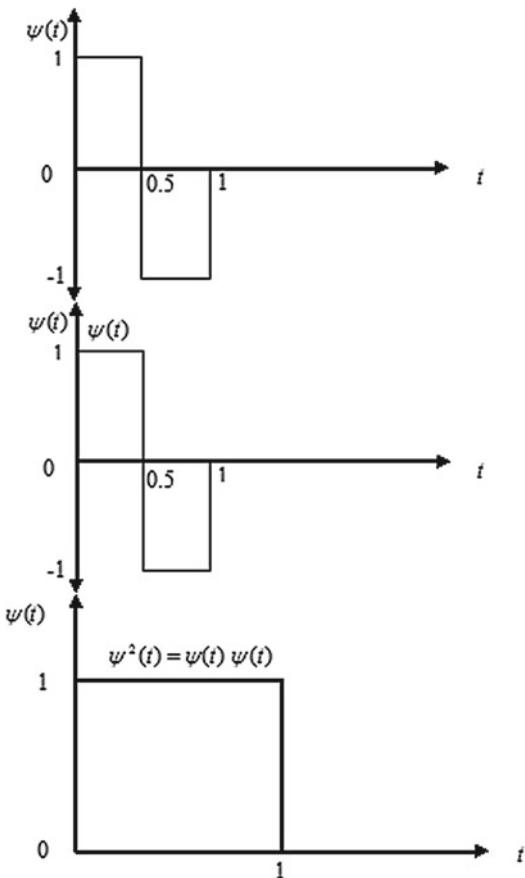
$$W_0 = \text{Span}_k \{\overline{\psi(t-k)}\}. \quad (1.4.11)$$

The speciality of signal in W_0 (spanned by Haar wavelet function) is that in every unit interval, if the function value in first half is $\pm m$ units, then in the next half it must necessarily be $\mp m$ units. Otherwise, we cannot express the signal using the basis of W_0 , i.e., $\psi(t)$ and its translates. Thus, the space W_0 is a highly restricted space [257].

1.4.7 Scaled Haar Wavelet Functions

Consider a scale version of $\psi(t)$. We will go only for dyadic scaling, i.e., scaling by the integer power of two. Consider at first $\psi(2t)$. Given $\psi(t)$ as defined in Eq. (1.4.9), $\psi(2t)$ can easily be shown as

Fig. 1.9 Norm of Haar wavelet function



$$\psi(2t) = \begin{cases} 1, & 0 \leq t \leq 1/4 \\ -1, & 1/4 \leq t \leq 1/2 \\ 0, & \text{elsewhere} \end{cases}$$

We denote the space spanned by the set of orthogonal bases $\{\psi(2t - k) : k \in \mathbb{N}\}$ as W_1 . Formally, we define

$$W_1 = \text{Span}_k \overline{\{\psi(2t - k)\}}.$$

Is $W_0 \subset W_1$? or in other words, can we represent a signal in W_0 by using the bases of W_1 . The answer is “No” (Fig. 1.10).

Since linear combination of $\psi(2t)$ and its translates cannot make a $\psi(t)$, no signal in W_0 can be expressed using bases of W_1 . Therefore, $W_0 \not\subset W_1$. However, we shall see that $W_0 \perp W_1$. This means that bases of W_0 are orthonormal to bases of W_1 . Figure 1.11 shows that

Fig. 1.10 $\psi(2t)$ and its translates cannot express a $\psi(t)$

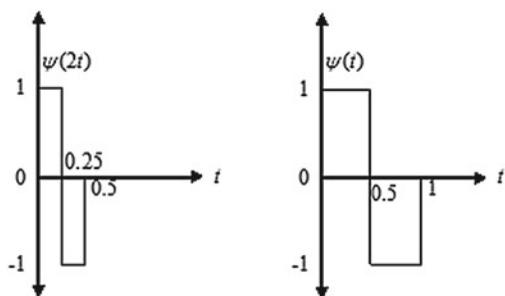
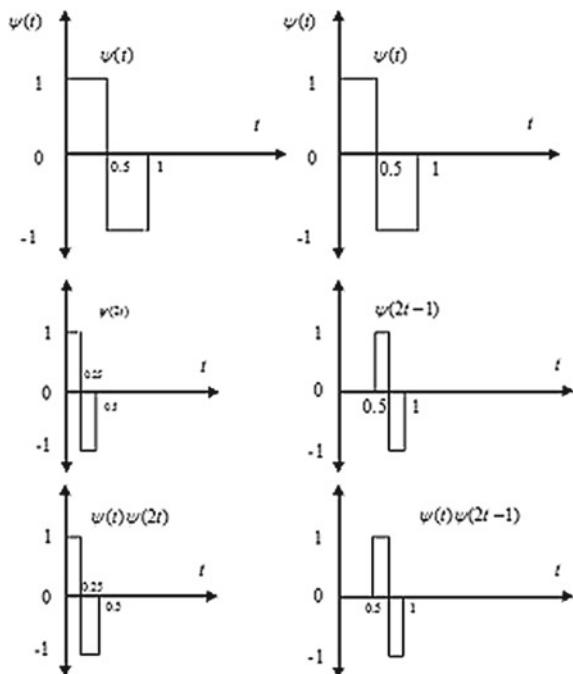


Fig. 1.11 Product function $\psi(t)\psi(2t)$ and $\psi(t)\psi(2t-1)$



$$\int \psi(t)\psi(2t)dt = 0 \text{ and } \int \psi(t)\psi(2t-1)dt = 0$$

We note that the area under the function, which is the product of $\psi(t)$ and $\psi(2t)$, is zero. Similarly, the area under the function which is the product of $\psi(t)$ and $\psi(2t-1)$ is also zero. This implies that space W_0 is orthogonal to the space W_1 . Again it can be proved that $W_1 \perp W_2$. Thus, we have another most important relation in wavelet theory

$$\dots W_{-1} \perp W_0 \perp W_1 \perp W_2 \perp \dots \quad (1.4.12)$$

So far we have discussed two very important concepts in wavelet theory with respect to function spaces, which are:

1. Space spanned by scaling function bases are nested, i.e.,

$$\dots V_{-1} \subset V_0 \subset V_1 \subset \dots \subset V_\infty.$$

2. Spaces spanned by wavelet function bases are orthogonal among themselves.
Thus

$$\dots W_{-1} \perp W_0 \perp W_1 \perp W_2 \dots$$

We are now ready to answer the question we have posed earlier. What is that missing in V_0 to make it a subset of V_1 . To answer the question, we consider a signal in V_1 . Since it is a signal in V_1 , we can express it by using the basis of V_1 . Thus

$$\begin{aligned} f(t) = & 4\varphi(2t) + 2\varphi(2t - 1) + 6\varphi(2t - 2) - 2\varphi(2t - 3) + 4\varphi(2t - 4) \\ & + 6\varphi(2t - 5) + 2\varphi(2t - 6) + 2\varphi(2t - 7). \end{aligned}$$

It may be observed that the signal segment in the first unit interval can be represented by using the basis of V_0 and W_0 , i.e.,

$$4\varphi(2t) + 2\varphi(2t - 1) = \frac{4+2}{2}\varphi(t) + \frac{4-2}{2}\psi(t) = 3\varphi(t) + \psi(t).$$

Similarly

$$6\varphi(2t - 2) - 2\varphi(2t - 3) = \frac{6-2}{2}\varphi(t - 2) + \frac{6+2}{2}\psi(t) = 2\varphi(t - 1) + 4\psi(t - 1).$$

Again

$$4\varphi(2t - 4) + 6\varphi(2t - 5) = \frac{4+6}{2}\varphi(t - 3) + \frac{4-6}{2}\psi(t - 3) = 5\varphi(t - 2) - \psi(t - 2).$$

Lastly

$$2\varphi(2t - 6) + 2\varphi(2t - 7) = \frac{2+2}{2}\varphi(t - 4) + \frac{2-2}{2}\psi(t - 4) = 2\varphi(t - 3) + 0\psi(t - 3).$$

Thus

$$\begin{aligned} f(t) = & 4\varphi(2t) + 2\varphi(2t - 1) + 6\varphi(2t - 2) - 2\varphi(2t - 3) + 4\varphi(2t - 4) \\ & + 5\varphi(2t - 5) + 2\varphi(2t - 6) + 2\varphi(2t - 7). \end{aligned} \tag{1.4.13a}$$

$$\begin{aligned} f(t) = & 3\varphi(t) + 2\varphi(t - 1) + 6\varphi(t - 2) + 2\varphi(t - 3) + \varphi(t) \\ & + 4\varphi(t - 4) - \varphi(t - 5) + 0\varphi(t - 6). \end{aligned} \tag{1.4.13b}$$

We could express a signal in V_1 space, in terms of bases of V_0 and W_0 spaces. If we combine the bases of V_0 and W_0 spaces, we can express any signal in V_1 space. We express this fact mathematically as:

$$V_1 = V_0 \oplus W_0.$$

We call W_0 as the complementary space of V_0 . Now, we found out the “thing” that was missing in V_0 that made V_0 a proper subset of V_1 . We call V_0 and W_0 spaces as complementary spaces because V_0 and W_0 spaces are orthogonal and their bases together can represent any signal in the next “higher” or finer space V_1 . We call this as the decomposition of a finer signal into two coarser signals.

1.4.8 Orthogonality of $\varphi(t)$ and $\psi(t)$

To prove that $V_0 \perp W_0$, we must show that the bases of V_0 are orthogonal to the bases of W_0 . Now, we need to show the orthogonality only in the case where both the bases overlap (exist in the same interval). If the bases do not overlap, naturally they are orthogonal because the product of the two functions will identically be zero at every point. We know that $\varphi(t)$ and $\psi(t)$ are defined in the same interval $[0, 1]$ and hence overlap. So we will try to show that $\varphi(t)$ is orthogonal to $\psi(t)$. We have

$$\langle \varphi(t), \psi(t) \rangle = \int_t^1 \varphi(t)\psi(t)dt = \int_0^{0.5} 1.1dt + \int_{0.5}^1 1.(-1)dt = 0.5 - 0.5 = 0.$$

Product of $\varphi(t)$ and $\psi(t)$ gives $\psi(t)$ whose net area is zero (Fig. 1.12).

We know that the bases for V_0 and W_0 together span V_1 , i.e., $V_1 = V_0 \oplus W_0$.

Using the same case, we may write

$$V_2 = V_1 \oplus W_1.$$

In general

$$V_j = V_{j-1} \oplus W_{j-1}. \quad (1.4.14)$$

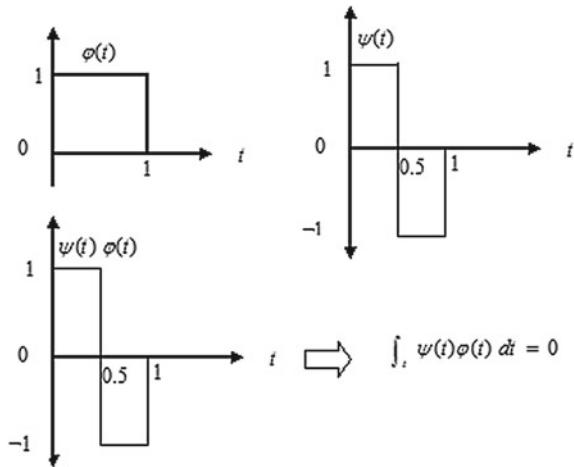
Let us now express V_j using wavelet spaces alone. (Except the last term)

$$V_j = V_{j-1} \oplus W_{j-1}.$$

But

$$V_{j-1} = V_{j-2} \oplus W_{j-2}.$$

Fig. 1.12 Orthogonality of $\varphi(t)$ and $\psi(t)$



Therefore

$$\begin{aligned} V_j &= W_{j-1} \oplus W_{j-2} \oplus V_{j-2} \\ &\quad \dots \quad \dots \quad \dots \\ &\quad \dots \quad \dots \quad \dots \\ &V_j = W_{j-1} \oplus W_{j-2} \oplus W_{j-3} \oplus \dots \oplus W_0 \oplus V_0. \end{aligned} \tag{1.4.15}$$

Equation (1.4.15) tells us that any signal in V_j can be expressed by using the bases of $W_{j-1}, W_{j-2}, \dots, W_0$ and V_0 . Equation (1.4.15) is the foundation of wavelet base decomposition of the signal.

1.4.9 Normalization of Haar Bases at Different Scales

We know that the vectors in \mathbb{R}^3 are orthonormal. In addition to orthogonality, they satisfy the relation $\hat{i} \cdot \hat{i} = \hat{j} \cdot \hat{j} = \hat{k} \cdot \hat{k} = 1$, i.e., orthonormality. In function space, corresponding to dot product we have multiplication and integration.

Consider bases $\varphi(t - k)$'s of V_0 . We know that they are orthogonal. In addition, we have

$$\int \varphi(t)\varphi(t) = 1$$

or

$$\int \varphi(t - k)\varphi(t - k) = 1$$

making the bases of V_0 orthonormal. We say that bases of Haar at scale level '0' are orthonormal. The actual value of scale is given by 2^j where j is the scale level.

What about orthonormality of bases of V_1 , where scales level is 1? The bases of space V_1 are translates of $\varphi(2t)$'s. Let us check the value of integral $\int \varphi(2t)\varphi(2t)dt$. Its value is 1/2. So to normalize $\varphi(2t)$ and translates of $\varphi(2t)$, we multiply each by $\sqrt{2}$. Now, we redefine V_1 as:

$$V_1 = \text{Span}_k\{\overline{\sqrt{2}\varphi(2t - k)}\}.$$

Consider the bases for V_2 . Here the scale level is 2. Hence, we call bases for V_2 as level 2 bases. Since $\int \varphi(4t)\varphi(4t)dt = 1/4$, we should multiply each base by $\sqrt{4} = 2$.

We redefine V_2 as:

$$V_2 = \text{Span}_k\{\overline{2\varphi(4t - k)}\} = \text{Span}_k\{\overline{2\varphi(2^2t - k)}\}.$$

Let us now generalize, by considering unnormalized bases for V_j . The unnormalized bases for V_j are translates of $\varphi(2^j t)$. But

$$\int \varphi(2^j t)\varphi(2^j t)dt = \frac{1}{2^j}.$$

Therefore, the normalizing constant is $\sqrt{2^j} = 2^{j/2}$. Now, V_j is redefined as:

$$V_j = \text{Span}_k\{\overline{2^{j/2}\varphi(2^j t - k)}\}. \quad (1.4.16)$$

1.5 Function Spaces

In this section, we discuss certain important spaces which have significance with the matter of the book.

1.5.1 Lebesgue Spaces

Many of the classical spaces in analysis consist of measurable functions, and most of the important norms on such spaces have been defined by integrals. The Lebesgue L^p -spaces are among such important classes.

Let f be a Lebesgue measurable function defined on a measurable set E and $1 \leq p < \infty$. Designate by $L^p(E)$, the class of all Lebesgue p -integrable functions on E , i.e.,

$$L^p(E) = \left\{ f : \int_E |f|^p < \infty \right\}.$$

Then, it is easy to verify that $L^p(E)$, $1 \leq p < \infty$, is a linear space over \mathbb{R} .

Now, we define a function $\|\cdot\|_p : L^p(E) \rightarrow \mathbb{R}$, $0 < p < \infty$ as follows:

$$\|f\|_p = \left(\int_E |f|^p \right)^{\frac{1}{p}}, \quad 0 < p < \infty.$$

Then, the spaces $L^p(E)$, $1 \leq p < \infty$ is a Banach space.

1.5.2 Hardy Space

The Hardy space can be defined in the various ways (see [121]). Here, we shall use the following simple definition:

$$\mathcal{H}^1(\mathbb{R}) = \{f \in L^1(\mathbb{R}) : Hf \in L^1(\mathbb{R})\}$$

where H is the Hilbert transform

$$Hf(x) = p\nu \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(y)}{x-y} dy.$$

Furthermore

$$\|f\|_{\mathcal{H}^1} = \|f\|_{L^1} + \|Hf\|_{L^1}$$

defines a Banach space norm on $\mathcal{H}^1(\mathbb{R})$.

The definition we have just given shows that $\mathcal{H}^1(\mathbb{R})$ consists of functions in $L^1(\mathbb{R})$. It is our goal to obtain characterization of $\mathcal{H}^1(\mathbb{R})$ using wavelet packets. For this purpose, we need a simple characterization of functions in $\mathcal{H}^1(\mathbb{R})$. It is more convenient to consider the atomic decomposition of $\mathcal{H}^1(\mathbb{R})$ in terms of atoms that are in $L^2(\mathbb{R})$. A 2-atom is a measurable function “ a ” defined on \mathbb{R} that satisfies

$$\begin{cases} (i) \text{the support ‘}a\text{’ is contained in a finite interval } I \subset \mathbb{R} \\ (ii) \|a\|_{L^2(\mathbb{R})} \leq \frac{1}{|I|^{\frac{1}{2}}} \quad \left[\text{Equivalently } \left(\int_I |a(x)|^2 \frac{dx}{|I|} \right)^{\frac{1}{2}} \leq \frac{1}{|I|} \right] \\ (iii) \int_{\mathbb{R}} a(x) dx = 0 \end{cases} \quad (1.5.1)$$

For more details, see [121].

1.5.3 Sobolev Spaces

Definition 1.5.3.1 For $1 < p < \infty$, $s = 1, 2, 3, \dots$, we define the Sobolev space $L^{p,s}(\mathbb{R}) \equiv L^{p,s}$ to be the space of all functions $f \in L^p(\mathbb{R})$ such that,

$\forall n = 1, 2, 3, \dots, s$, the n th derivative of f also belongs to $L^p(\mathbb{R})$. The n th derivative of a function $f \in L^p(\mathbb{R})$ is considered here in the sense of distributions; i.e., it is a function $D^n f$ such that

$$\int_{\mathbb{R}} (D^n f)(x) \varphi(x) dx = (-1)^n \int_{\mathbb{R}} f(x) D^n \varphi(x) dx$$

for every test function $\varphi \in S$. The quantity

$$\|f\|_{L^{p,s}} = \|f\|_{L^p} + \sum_{n=1}^s \|D^n f\|_{L^p} \quad (1.5.2)$$

is a norm on the space $L^{p,s}$, with respect to which it is a Banach space. There are other equivalent definitions of the space $L^{p,s}(\mathbb{R})$. One of them involves multiplier $(1 + |\xi|^2)^{s/2}$.

For a function $f \in L^p(\mathbb{R})$, we consider the quantity

$$\|f\|_{W^{p,s}} = \left\| \left[(1 + |\cdot|^2)^{\frac{s}{2}} \hat{f}(\cdot) \right] \right\|_{L^p}, \quad 1 < p < \infty, \quad s = 1, 2, \dots,$$

where $\hat{\cdot}$ and $\check{\cdot}$ denote the Fourier transform and the inverse Fourier transform.

1.5.4 Besov Spaces

Definition For $0 < \alpha < r$, $0 < p \leq \infty$, Besov spaces denoted by $B_{pq}^\alpha(\mathbb{R})$ are defined as

$$\begin{aligned} B_{pq}^\alpha(\mathbb{R}) &= \left\{ f \in L_p : \left(\int_0^\infty [t^{-\alpha} \omega_r(f, t)_p]^q \frac{dt}{t} \right)^{\frac{1}{q}} < \infty \right\} \\ &= \left\{ f \in L_p : \sup_{t \geq 0} t^{-\alpha} \omega_r(f, t)_p < \infty \right\}, \quad q = \infty. \end{aligned}$$

On $B_{pq}^\alpha(\mathbb{R})$ norm is defined as

$$\begin{aligned} \|f\|_{B_{pq}^\alpha} &= \|f\|_{L_p} + \left(\int_0^\infty [t^{-\alpha} \omega_r(f, t)_p]^q \frac{dt}{t} \right)^{\frac{1}{q}}, \quad 0 < q < \infty \\ &= \|f\|_{L_p} + \sup_{t \geq 0} t^{-\alpha} \omega_r(f, t)_p, \quad q = \infty. \end{aligned}$$

It is a quasi-norm on $B_{pq}^\alpha(\mathbb{R})$ and for $1 \leq p < \infty$, $1 \leq q < \infty$ it is a norm.

1.6 Wavelets

The theory of wavelets is the latest comer to the world of signal processing (more than 30 years now). It grew and brewed in different areas of science. Harmonic analysts had developed powerful time–frequency tools, electrical engineers were busy with subband coding, and quantum physicists were trying to understand coherent states. They were not aware of each others’ progress until the late 80’s when a synthesis of all these ideas came to be, and what is now called wavelet theory contains all those ideas and much more. Wavelet is not one magical transform that solves all problems. It is a library of bases that is appropriate for a large number of situations where the traditional tools, for example Fourier analysis, are not so good. There are many other problems which cannot be optimally treated with either of the known tools; therefore, new ones have to be designed.

Wavelets have gone from what some thought would be a short-lived fashion, to come part of the standard curriculum in electrical engineering, statistics, physics, mathematical analysis, and applied mathematics. It has become part of the toolbox of statisticians, signal and image processors, medical imaging, geophysicists, speech recognition, video coding, Internet communications, etc. Among the most spectacular and well-known applications are the wavelet-based FBI standard for storing and retrieving fingerprints.

Since the days of Fourier, scientists and engineers, besides mathematicians themselves, have made vigorous effort to represent square integrable functions (signals having finite energy) as a linear combination of functions having some nice properties. The discovery of wavelets (small waves) was an attempt to search a function which will generate the space of square integrable functions over the real line.

Definition 1.6.1 A function $\psi \in L^p(\mathbb{R})$ is called a (basic) wavelet of class m , $m \in \mathbb{N}$ be an integer, if the following four properties hold:

- (a) if $m = 0$, $\psi \in L^\infty(\mathbb{R})$; if $m \geq 1$, ψ and all its derivatives up to order m belong to $L^\infty(\mathbb{R})$;
- (b) ψ and all its derivatives upto order m decrease rapidly as $x \rightarrow \pm\infty$;
- (c) $\int_{\mathbb{R}} x^k \psi(x) dx = 0$, for $0 \leq k \leq m$;
- (d) the collection of functions $\{2^{j/2}\psi(2^j x - k) : j, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$.

The functions $2^{j/2}\psi(2^j x - k)$; $j, k \in \mathbb{Z}$ are the wavelets (generated by the “mother wavelet” ψ).

Let I denote the dyadic interval $[k 2^{-j}, (k+1) 2^{-j}]$, and $\psi_I = 2^{j/2}\psi(2^j x - k)$ are the functions generated by ψ , and the representation of the function f on the real line by a wavelet series is the identity:

$$f(x) = \sum_I \alpha(I)\psi_I(x), \quad (1.6.1)$$

where $\alpha(I) = \langle f, \psi_I \rangle$ are wavelet coefficients and the wavelet transform of f be

$$L_\psi f(a, b) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} f(t) \psi \left\{ \frac{(t-b)}{a} \right\} dt. \quad (1.6.2)$$

The identity (1.6.1) works if $f \in L^p(\mathbb{R})$, for $1 < p < \infty$, but does not hold when f belongs to $L^\infty(\mathbb{R})$ or $L^1(\mathbb{R})$. All difficulties disappear if we have a second function φ , called the “father wavelet,” satisfying the conditions (a) and (b) just like ψ , and having two other properties:

(c') φ satisfies $\int_{-\infty}^{\infty} \varphi(x) dx = 1$;

(d') the functions $\varphi(x-k)$, $k \in \mathbb{Z}$, and ψ_I , I a dyadic interval with $|I| \leq 1$, together form an orthonormal basis of $L^2(\mathbb{R})$.

Definition 1.6.2 A multiresolution analysis (MRA) consists of a sequence $\{V_j : j \in \mathbb{Z}\}$ of closed subspaces of $L^2(\mathbb{R})$ satisfying

- (i) $V_j \subset V_{j+1}$, $\forall j \in \mathbb{Z}$;
- (ii) $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$ and $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$;
- (iii) For every $f \in L^2(\mathbb{R})$, $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$, $\forall j \in \mathbb{Z}$;
- (iv) $\begin{cases} \text{There exists a function } \varphi \in V_0 \text{ such that } \{\varphi(x-k) : k \in \mathbb{Z}\} \\ \text{is an orthonormal basis for } V_0; \end{cases}$
- (v) $\begin{cases} \text{There exists an isomorphism } I \text{ from } V_0 \text{ onto } l^2(\mathbb{Z}) \text{ which} \\ \text{commutes with the action of } \mathbb{Z}. \end{cases}$

The function φ whose existence is asserted in (iv) is called a scaling function of the given MRA. For more details, see [18, 189, 190].

Now, we construct wavelets from multiresolution analysis. Let W_0 be the orthogonal complement of V_0 in V_1 , i.e., $V_1 = V_0 \oplus W_0$. Then, if we dilate the elements of W_0 by 2^j , we obtain a closed subspace W_j of V_{j+1} as

$$V_{j+1} = V_j \oplus W_j, \quad j \in \mathbb{Z}. \quad (1.6.3)$$

Consider $V_0 = W_{-1} \oplus V_{-1}$, and observe that $\frac{1}{\sqrt{2}} \varphi \left(\frac{x}{2} \right) \in V_{-1} \subset V_0$. By (iv), we can express this function in terms of the basis $\{\varphi(x-k) : k \in \mathbb{Z}\}$ to obtain

$$\varphi(x/2) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \varphi(x-k), \quad (1.6.4)$$

where $h_k = \frac{1}{\sqrt{2}} \int_{\mathbb{R}} \varphi(x/2) \overline{\varphi(x-k)} dx$; the convergence in (1.6.4) is in $L^2(\mathbb{R})$ and $\sum_{k \in \mathbb{Z}} |h_k|^2 < \infty$. Taking Fourier transform of (1.6.4), we obtain

$$\hat{\varphi}(2\xi) = m_0(\xi) \hat{\varphi}(\xi), \quad (1.6.5)$$

where

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbb{Z}} h_k e^{-i\xi k}. \quad (1.6.6)$$

The sequence $\{h_k\}$ and its Fourier series $m_0(\xi)$ need to satisfy specific constraints that are related to the properties of φ :

- (i) If $\varphi \in L^1(\mathbb{R})$, then $\sum_{k \in \mathbb{Z}} h_k = \sqrt{2}$;
- (ii) φ is compactly supported in $[p, q]$ if and only if $h_k = 0$ for $k < p$ or $k > q$;
- (iii) When the two previous constraints are satisfied, the orthogonality of $\{\varphi(x - k) : k \in \mathbb{Z}\}$ is ensured if and only if

$$|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1 \quad (1.6.7)$$

and there exists a compact set K , congruent to $[-\pi, \pi]$ modulo 2π , such that $m_0(2^{-k}\xi) \neq 0, \forall \xi \in K$, and $k > 0$.

We then define

$$m_1(\xi) = e^{-i\xi} \overline{m_0(\xi + \pi)} \quad (1.6.8)$$

and

$$\hat{\psi}(\xi) = m_1(\xi/2)\hat{\varphi}(\xi/2), \quad (1.6.9)$$

where

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k) \quad (1.6.10)$$

and

$$g_k = (-1)^{k+1} \bar{h}_{1-k}. \quad (1.6.11)$$

Now, we consider a function $\psi \in W_0$ such that $\{\psi(x - k) : k \in \mathbb{Z}\}$ is an orthonormal basis for W_0 . In fact, if this is the case, then $\{2^{j/2}\psi(2^j x - k) : k \in \mathbb{Z}\}$ is an orthonormal basis for W_j for all $j \in \mathbb{Z}$ due to (iv) of Definition 1.6.2 and the definition of W_j . Hence, $\{\psi_{j,k} : j, k \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(\mathbb{R})$, which shows that ψ is an orthonormal wavelet on \mathbb{R} .

Lemma 1.6.3 *Let $\varphi \in L^2(\mathbb{R})$ be a scaling function. Then, $\{\varphi(\cdot - k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis of V_0 if and only if*

$$\sum_{k \in \mathbb{Z}} |\hat{\varphi}(\xi + 2k\pi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}.$$

Lemma 1.6.4 *Let $\psi \in L^2(\mathbb{R})$ be a wavelet associated with the scaling function φ . Then*

$$|\hat{\psi}(\xi)|^2 = \sum_{j=1}^{\infty} \left| \hat{\psi}(2^j \xi) \right|^2 \text{ for a.e. } \xi \in \mathbb{R}.$$

Lemma 1.6.5 Let ψ be a band-limited orthonormal wavelet. Then

$$\sum_{j \in \mathbb{Z}} \left| \hat{\psi}(2^j \xi) \right|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R} - \{0\}.$$

Lemma 1.6.6 Let ψ be a band-limited orthonormal wavelet such that $|\hat{\psi}|$ is continuous at 0. Then, $\hat{\psi}(0) = 0$ in an open neighborhood of the origin.

Now, we discuss few wavelets which are useful in applications.

(a) Daubechies Wavelets

General characteristics of the wavelets generated by I. Daubechies are that these are compactly supported. These wavelets are having extremal phase and highest number of vanishing moments for a given support width. Associated scaling filters are minimum-phase filters. The short name of these wavelets is “db”. Different orders of this family are present, for example db1 or Haar, db2, db4, db10. These different orders of wavelets are orthogonal as well as biorthogonal characteristics. Both kinds of transformation, i.e., continuous wavelet transform and discrete wavelet transform, are possible for these wavelets. If filters length is $2N$, then support width is $2N - 1$ and regularity is about $0.2N$ for large N . Symmetry is very far from. Number of vanishing moments for wavelet function is N . In Figs. 1.13, 1.14, and 1.15, db2 wavelet function, db2 scaling function, and db2 with central frequency-based approximation are shown.

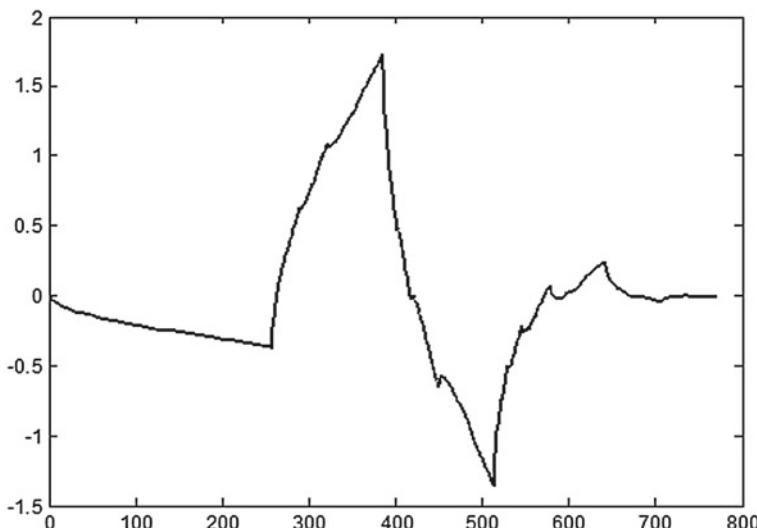


Fig. 1.13 db2 wavelet function

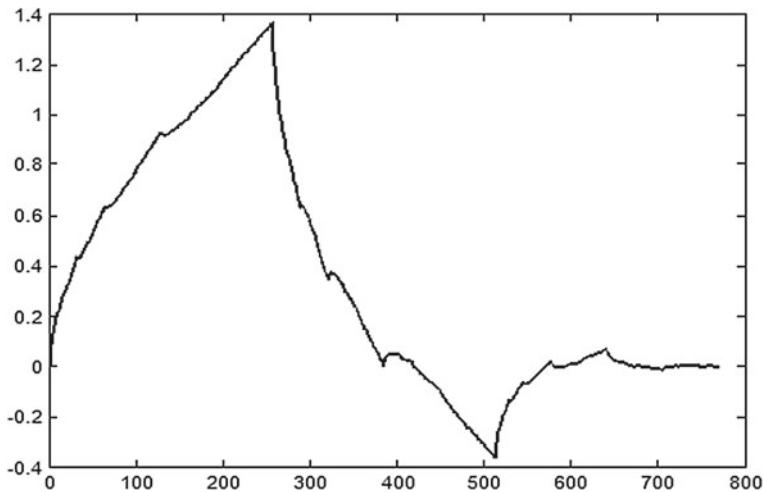


Fig. 1.14 db2 scaling function

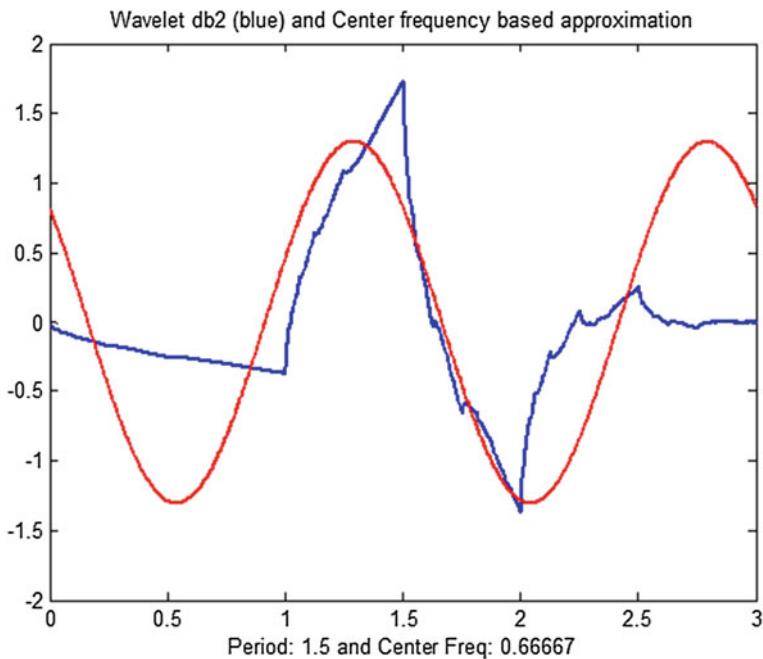


Fig. 1.15 db2 with central frequency-based approximation

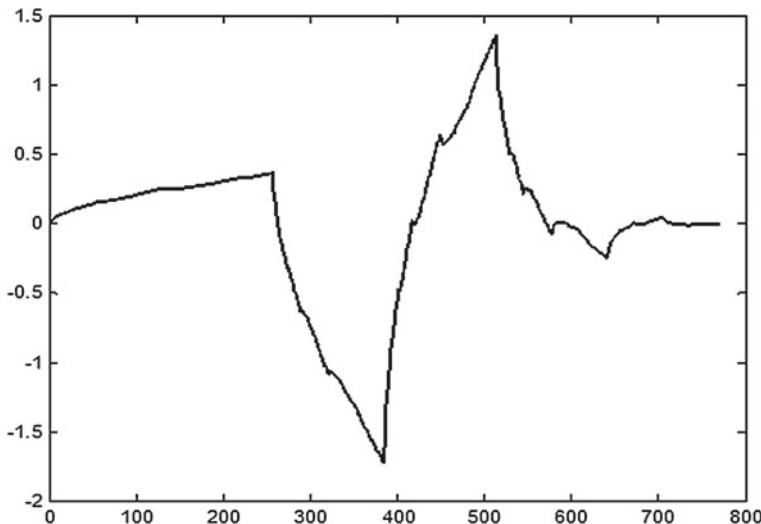


Fig. 1.16 sym2 wavelet function

(b) Symlets Wavelets

General characteristics of the wavelets are that these wavelets are also compactly supported wavelets with least asymmetry and highest number of vanishing moments for a given support width. Associated scaling filters are near linear-phase filters. The short name of these wavelets is “sym”. Different orders of this family are present, for example sym1, sym2, sym4, db8. These different orders of wavelets are orthogonal as well as biorthogonal characteristics. Both kinds of transformation, i.e., continuous wavelet transform and discrete wavelet transform, are possible for these wavelets. If filters length is $2N$, then support width is $2N - 1$. Symmetry is very near from. Number of vanishing moments for wavelet function is N . In Figs. 1.16, 1.17, and 1.18, sym2 wavelet function, sym2 scaling function, and sym2 with central frequency-based approximation are shown.

(c) Coiflets Wavelets

General characteristics of the wavelets are that these wavelets are also compactly supported wavelets with highest number of vanishing moments for both scaling functions and wavelet functions for a given support width. The short name of these wavelets is “coif”. Different orders of this family are present, for example coif1, coif2, coif3, coif4, and coif5. These different orders of wavelets are orthogonal as well as biorthogonal characteristics. Both kinds of transformation, i.e., continuous wavelet transform and discrete wavelet transform, are possible for these wavelets. If filters length is $6N$, then support width is $6N - 1$. Symmetry is near from. Number of vanishing moments for wavelet function is $2N$, and scaling function is $2N - 1$. In Figs. 1.19, 1.20, and 1.21, coif2 wavelet function, coif2 scaling function, and coif2 with central frequency-based approximation are shown.

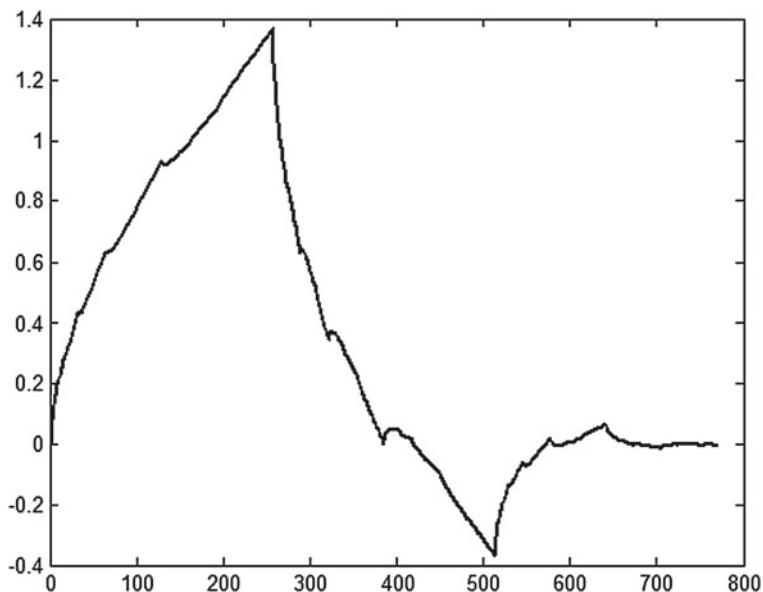


Fig. 1.17 sym2 scaling function

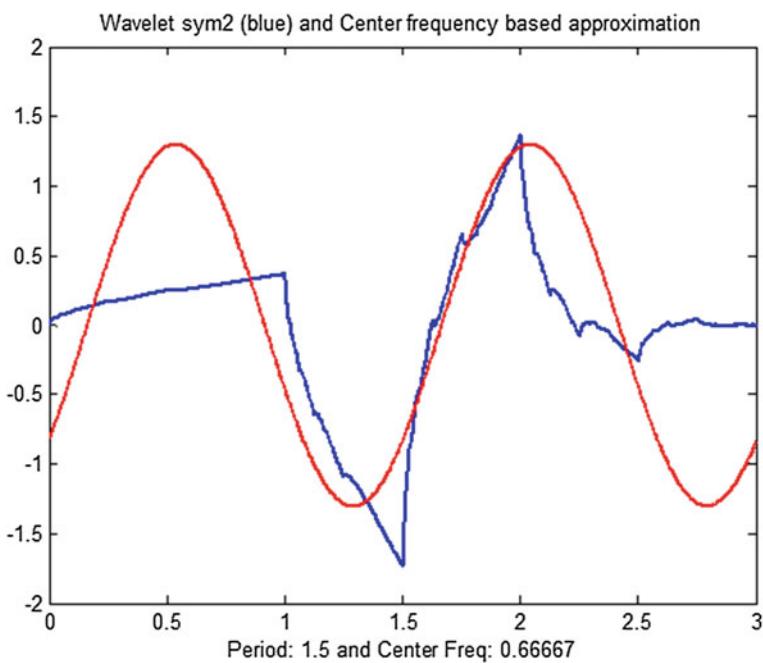


Fig. 1.18 sym2 with central frequency-based approximation

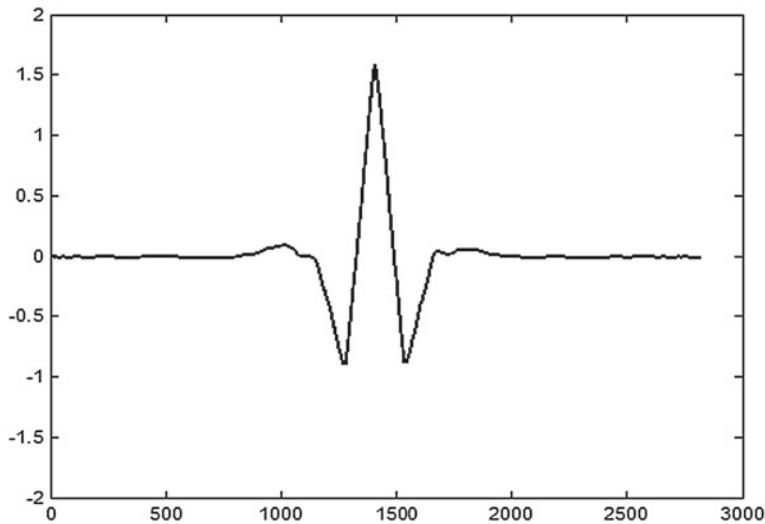


Fig. 1.19 coif2 wavelet function

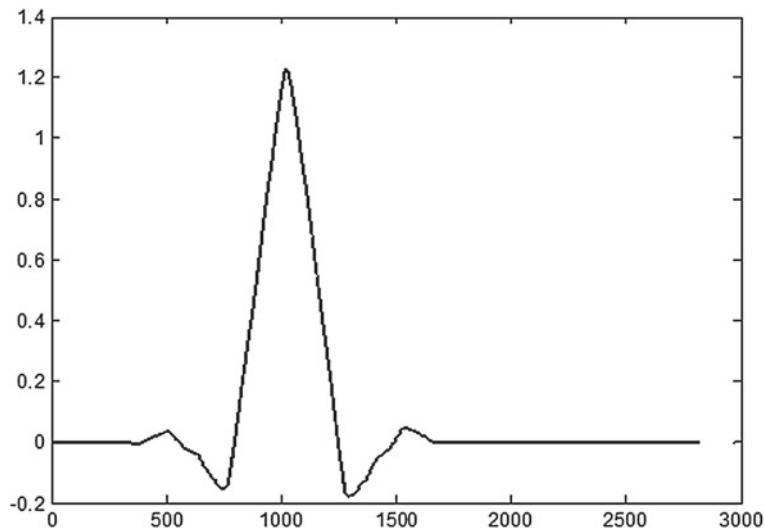


Fig. 1.20 coif2 scaling function

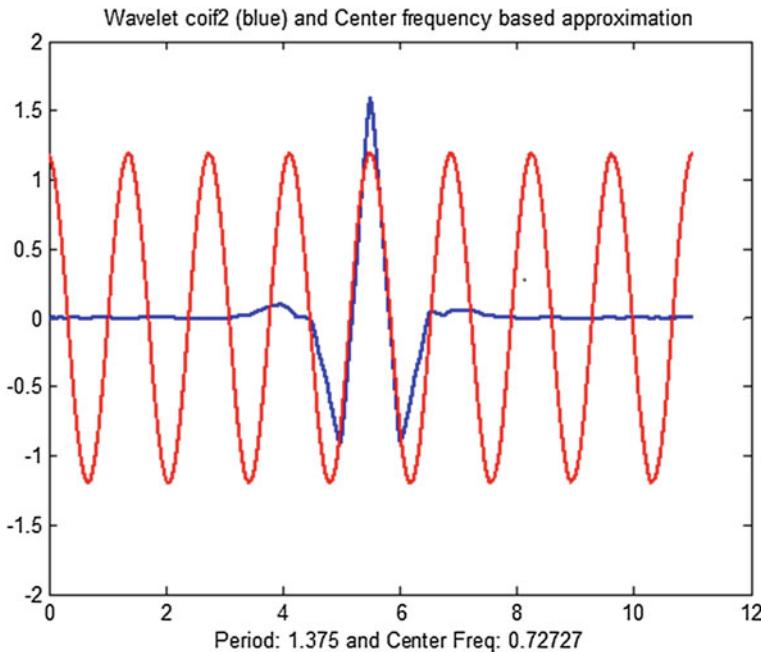


Fig. 1.21 coif2 with central frequency-based approximation

1.7 Thresholding

Thresholding is used in wavelet domain to smooth out or to remove some coefficients of wavelet transform subsignals of the measured signal. This reduces the noise content of the signal under the nonstationary environment. There are two common ways to threshold the resulting wavelet coefficients. In the first case, the coefficients values are set to zero whose absolute value is below a threshold value. This is commonly known as hard thresholding. The second one, known as soft thresholding, goes one step further and reduces the magnitude of the remaining coefficients by the threshold value. Hard thresholding maintains the scale of the signal but introduces ringing and artifacts after reconstruction due to discontinuity in the wavelet coefficients. Soft thresholding eliminates this problem resulting in smoother signal.

Let t denote the threshold. The hard threshold signal is x if $|x| > t$, and is 0 if $|x| \leq t$. The soft threshold signal is $\text{sign}(x)(|x| - t)$ if $|x| > t$, and is 0 if $|x| \leq t$.

Hard thresholding can be described as the usual process of setting to zero the elements whose absolute values are lower than the threshold. Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients toward zero as shown in Fig. 1.22a–c.

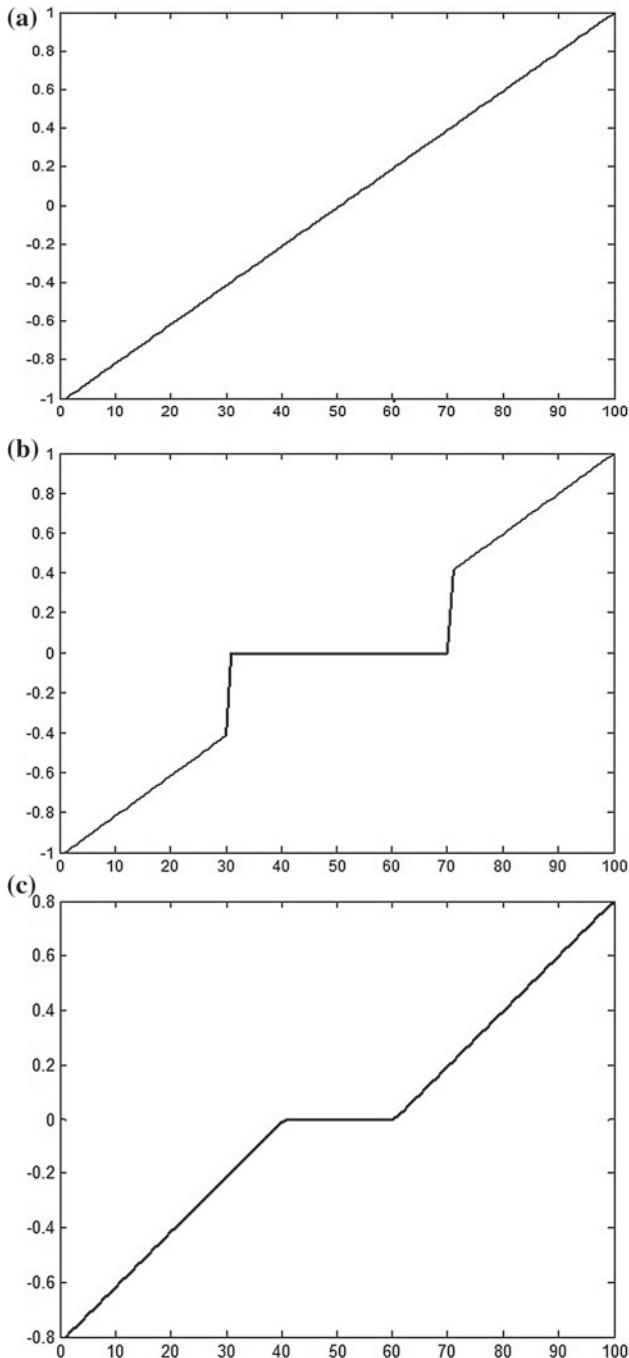


Fig. 1.22 **a** A simple signal $y = x$. **b** Applying hard threshold. **c** Applying soft threshold

As can be seen from the figures that the hard thresholding procedure creates discontinuities at $x = \pm t$, while the soft thresholding procedure does not.

Different Rules for Threshold Selection

According to the basic characteristics of noise, four threshold selection rules are implemented in the literature. These are classified below:

1. First option for the soft threshold estimator is a threshold selection rule, which is based on Stein's unbiased estimate of risk (quadratic loss function). We get an estimate of the risk for a particular threshold value t , and minimizing the risks in t gives a selection of the threshold value.
2. Second option uses a fixed form threshold yielding minimax performance multiplied by a small factor proportional to $\log(\text{length}(\text{signal}))$.
3. Third option is a mixture of the two previous options. As a result, if the signal-to-noise ratio is very small, Stein's unbiased estimate of risk estimate is very noisy. So if such a situation is detected, the fixed form threshold is used.
4. Last option uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics to design estimators. Since the denoised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the option that realizes the minimum, over a given set of functions, of the maximum mean square error.

In general, one can suppress the noise level if it is estimated. The detailed coefficients (the finest scale) are essentially noise coefficients with standard deviation equal to σ . The median absolute deviation of the coefficients is a robust estimate of σ . The use of a robust estimate is crucial for two reasons. The first one is that if the first-level coefficients contain details of the function, then these details are concentrated in a few coefficients if the function f is sufficiently regular. The second reason is to avoid signal end effects, which are pure artifacts due to computations on the edges. Often rescaling using a single estimation of level noise is based on the first-level coefficients. When we suspect a nonwhite noise, thresholds must be rescaled by a level-dependent estimation of the level noise. The same kind of strategy as in the previous option is used by estimating level by level. This estimation is implemented directly in the wavelet decomposition structure of the original signal. Another option is also possible, which handles threshold rescaling using a level-dependent estimation of the level noise.

Chapter 2

Wavelet Packets



2.1 Introduction

A simple but powerful extension of wavelets and multiresolution analysis is wavelet packets, pioneered by Coifman, Meyer, Wickerhauser, and other researchers [74, 75, 78, 293]. The wavelet transform is generalized to produce a library of orthonormal basis of modulated wavelet packets, where each basis comes with a fast transform. By generalizing the method of multiresolution decomposition, it is possible to construct orthonormal basis for $L^2(\mathbb{R})$ (see [60, 64, 71, 74, 77–79, 87, 140, 195, 293]). Discrete wavelet packets have been thoroughly studied by Wickerhauser [293], who has also developed computer programs and implemented them. The wavelet packets allow more flexibility in adapting the basis to the frequency contents of a signal, and it is easy to develop a fast wavelet packet transform. The power of wavelet packet lies on the fact that we have much more freedom in deciding which basis function we use to represent the given function. The best basis selection criteria and application to image processing can be found in [79, 293].

Wavelet packet functions are generated by scaling and translating a family of basic function shapes, which include father wavelet φ and mother wavelet ψ . In addition to φ and ψ there is a whole range of wavelet packet functions ω_n . These functions are parametrized by an oscillation or frequency index n . A father wavelet corresponds to $n = 0$, so $\varphi = \omega_0$. A mother wavelet corresponds to $n = 1$, so $\psi = \omega_1$. Larger values of n correspond to wavelet packets with more oscillations and higher frequency.

Wavelet packets are particular linear combinations or superpositions of wavelets. They form bases which retain many of the orthogonality, smoothness, and localization properties of their parent wavelets.

2.2 Construction of Wavelet Packets

If $\{h_k\}$ and $\{g_k\}$ be two sequences of $l^2(\mathbb{Z})$ such that

$$\sum_{n \in \mathbb{Z}} h_{n-2k} h_{n-2l} = \delta_{k,l}, \quad (2.2.1)$$

$$\sum_{n \in \mathbb{Z}} h_n = \sqrt{2}, \quad (2.2.2)$$

$$g_k = (-1)^k \bar{h}_{1-k}. \quad (2.2.3)$$

Further, let φ be a continuous and compactly supported real-valued function on \mathbb{R} that solves the equation

$$\varphi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2x - k) \quad (2.2.4)$$

with $\hat{\varphi}(0) = 1$.

Let ψ be an associated function defined by

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k). \quad (2.2.5)$$

Now, we construct wavelet packets from multiresolution analysis. In general, we consider two sequences $\{\alpha_n\}_{n \in \mathbb{Z}}$ and $\{\beta_n\}_{n \in \mathbb{Z}}$ in $l^2(\mathbb{Z})$. Let \mathbb{H} be a Hilbert space with orthonormal basis $\{e_k\}_{k \in \mathbb{Z}}$. Then, the sequences

$$\left. \begin{aligned} f_{2n} &= \sqrt{2} \sum_{k \in \mathbb{Z}} \alpha_{2n-k} e_k \\ f_{2n+1} &= \sqrt{2} \sum_{k \in \mathbb{Z}} \beta_{2n-k} e_k \end{aligned} \right\}. \quad (2.2.6)$$

are orthonormal bases of two orthogonal closed subspaces \mathbb{H}_1 and \mathbb{H}_0 , respectively, such that

$$\mathbb{H} = \mathbb{H}_1 \oplus \mathbb{H}_0. \quad (2.2.7)$$

Using this “splitting trick,” we now define the basic wavelet packets associated with a scaling function φ as defined in MRA.

Let $\omega_0 = \varphi$. The basic wavelet packets ω_n , $n = 0, 1, 2, \dots$, associated with the scaling function φ are defined recursively by

$$\left. \begin{aligned} \omega_{2n}(x) &= \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \omega_n(2x - k) \\ \omega_{2n+1}(x) &= \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \omega_n(2x - k) \end{aligned} \right\} \quad (2.2.8)$$

where ω_1 and ω_0 are mother and father wavelets, respectively.

The collection $\{\omega_n(x - k) : k \in \mathbb{Z}, n = 0, 1, 2, \dots\}$ is an orthonormal basis of $L^2(\mathbb{R})$, where

$$\omega_n(x - k) = \frac{1}{\sqrt{2}} \sum_i h_{k-2i} \omega_{2n} \left(\left(\frac{x}{2} \right) - i \right) + \frac{1}{\sqrt{2}} \sum_i g_{k-2i} \omega_{2n+1} \left(\left(\frac{x}{2} \right) - i \right).$$

The families $\{\omega_n(x - k)\}_{k \in \mathbb{Z}} = \{\omega_{n,k}\}$ for $2^j \leq n \leq 2^{j+1}$ are the results of j -splitting of the space W_j more or less than j -times. For a family of wavelet packets $\{\omega_n\}$ corresponding to some orthogonal scaling function $\varphi = \omega_0$, define the family of subspaces of $L^2(\mathbb{R})$ given by

$$U_j^n = \overline{\text{span}\{\omega_n(2^j x - k) : k \in \mathbb{Z}\}}, \quad j \in \mathbb{Z}, \quad n = 0, 1, 2, \dots \quad (2.2.9)$$

Here we observe that

$$U_j^0 = V_j \text{ and } U_j^1 = W_j. \quad (2.2.10)$$

This result can be generalized for other values of $n = 1, 2, 3, \dots$ as given below

$$U_{j+1}^n = U_j^{2n} \oplus U_j^{2n+1}, \quad j \in \mathbb{Z}. \quad (2.2.11)$$

Lemma 2.2.1 For each $j = 1, 2, 3, \dots$, decomposition trick (2.2.11) gives

$$\left. \begin{aligned} W_j &= U_j^1 = U_{j-1}^2 \oplus U_{j-1}^3 \\ W_j &= U_j^1 = U_{j-2}^4 \oplus U_{j-2}^5 \oplus U_{j-2}^6 \oplus U_{j-2}^7 \\ &\vdots \\ &\vdots \\ W_j &= U_j^1 = U_{j-k}^{2^k} \oplus U_{j-k}^{2^k+1} \oplus \dots \oplus U_{j-k}^{2^{k+1}-1} \\ &\vdots \\ &\vdots \\ W_j &= U_j^1 = U_0^{2^j} \oplus U_0^{2^j+1} \oplus \dots \oplus U_0^{2^{j+1}-1} \end{aligned} \right\} \quad (2.2.12)$$

where U_j^n is defined in (2.2.9). Moreover, for each $j = 1, 2, 3, \dots$; $k = 1, 2, 3, \dots, j$ and $m = 0, 1, \dots, 2^k - 1$, the set $\left\{ 2^{\frac{j-k}{2}} \omega_p(2^{j-k}x - l) : l \in \mathbb{Z} \right\}$ is an orthonormal basis of U_{j-k}^p where $p = 2^k + m$.

The families $\{\omega_n(x - k)\}_{k \in \mathbb{Z}} = \{\omega_{n,k}\}$ for $2^j \leq n \leq 2^{j+1}$ are the results of j -splitting of the space W_j more or less than j -times. However, all the elements of this basis have the general form

$$\omega_{j,n,k}(x) = 2^{j/2} \omega_n(2^j x - k). \quad (2.2.13)$$

If a function $f \in L^2(\mathbb{R})$, then

$$f(x) \sim \sum_{j \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} C_{l,n,k} \omega_{l,n,k}(x) \quad (2.2.14)$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$; will be a wavelet packet expansion of f and $C_{l,n,k}$ the wavelet packet coefficients defined as

$$C_{l,n,k} = \langle f, \omega_{l,n,k} \rangle. \quad (2.2.15)$$

[Lemma 2.2.1](#) allows us to obtain many orthogonal bases of $L^2(\mathbb{R})$. Since

$$L^2(\mathbb{R}) = V_0 \oplus W_0 \oplus W_1 \oplus W_2 \oplus \dots$$

we can choose whether to decompose a W_j , $j = 0, 1, 2, \dots$ further with any of the decompositions that appeared in [\(2.2.12\)](#). If we choose not to decompose any of the W_j , we obtain the wavelet decomposition of $L^2(\mathbb{R})$. If we choose the last formula in [\(2.2.12\)](#), for each W_j we obtain the wavelet packets. In between these two decompositions, there are a denumerable number of ways to choose a decomposition of $L^2(\mathbb{R})$, and hence, produce new orthonormal bases.

Wavelet packets have been introduced as a flexible method for time-frequency analysis of signals combining the advantages of windowed Fourier and wavelet analysis. Similarly, periodic wavelet packets provide an interesting alternative to Fourier series.

Lemma 2.2.2 Let $\omega_n \in L^2(\mathbb{R})$ be the wavelet packets relative to the orthonormal scaling function φ ($\varphi = \omega_0$). Then

$$\langle \omega_n(x - r), \omega_n(x - k) \rangle = \delta_{r-k}$$

for all $r, k \in \mathbb{Z}$.

Lemma 2.2.3 Let $P = P_\varphi$ and $G^* = G_\varphi^*$ be two admissible two-scale symbols such that

$$P(z) = \left(\frac{1+z}{2} \right)^N S(z) \text{ and } G^*(z) = \left(\frac{1+z}{2} \right)^{\tilde{N}} \tilde{S}(z), |z| = 1 \quad (2.2.16)$$

are in $L^2(\mathbb{R})$. Also, let φ and $\tilde{\varphi}$ be the corresponding scaling functions whose Fourier transforms are given by

$$\hat{\varphi}(\omega) = \prod_{k=1}^{\infty} P(e^{-i\omega/2^k}) \text{ and } \hat{\tilde{\varphi}}(\omega) = \prod_{k=1}^{\infty} G^*(e^{-i\omega/2^k}). \quad (2.2.17)$$

If φ and $\tilde{\varphi}$ are dual scaling functions, then P and G^* are dual to each other in the sense that

$$P(z)G(z) + P(-z)G(-z) = 1, \quad |z| = 1. \quad (2.2.18)$$

Conversely, if P and G^* are dual to each other and satisfy the conditions

$$B = \max_{|z|=1} |S(z)| < 2^{N-(1/2)} \text{ and } \tilde{B} = \max_{|z|=1} |\tilde{S}(z)| < 2^{\tilde{N}-(1/2)} \quad (2.2.19)$$

then φ and $\tilde{\varphi}$ are dual scaling functions.

Let P^0 and G^0 be Laurentz series in the Wiener class \mathcal{W} satisfying

$$P^0(z) = \left(\frac{1+z}{2}\right)^N S(z) \text{ and } G^0(z) = \left(\frac{1+z}{2}\right)^{\tilde{N}} \tilde{S}(z) \quad (2.2.20)$$

$$S(1) = \tilde{S}(1) = 1 \quad (2.2.21)$$

and

$$\begin{cases} \inf_{j>0} \max_{\xi} \prod_{k=1}^j |S(e^{i2^{-k}\xi})|^{1/j} < 2^{N-\frac{1}{2}} \\ \inf_{j>0} \max_{\xi} \prod_{k=1}^j |\tilde{S}(e^{i2^{-k}\xi})|^{1/j} < 2^{\tilde{N}-\frac{1}{2}} \end{cases} \quad (2.2.22)$$

where N and \tilde{N} are positive integers, and ξ is variable. Then, the infinite products

$$\hat{\varphi}(\xi) = \prod_{k=1}^{\infty} P^0(e^{-i\xi/2^k}) \text{ and } \hat{\tilde{\varphi}}(\xi) = \prod_{k=1}^{\infty} \overline{G^0(e^{-i\xi/2^k})} \quad (2.2.23)$$

converge in L^p , and the limit functions are Fourier transforms of some functions $\varphi, \tilde{\varphi} \in L^p$, that generate two (possibly different) multiresolution analyses (MRA) $\{V_n\}$ and $\{\tilde{V}_n\}$, respectively, of L^p (see Lemma 2.2.3). In addition, from (2.2.23) it follows that $\hat{\varphi}$ and $\hat{\tilde{\varphi}}$ satisfy

$$\hat{\varphi}(\xi) = P^0(z)\hat{\varphi}(\xi/2) \text{ and } \hat{\tilde{\varphi}}(\xi) = \overline{G^0(z)}\hat{\tilde{\varphi}}(\xi/2) \quad (2.2.24)$$

where $z = e^{-i\xi/2}$.

Now, writing

$$P^0(z) = \frac{1}{2} \sum_{n \in \mathbb{Z}} p_n^0 z^n, \quad G^0(z) = \frac{1}{2} \sum_{n \in \mathbb{Z}} g_n^0 z^n \quad (2.2.25)$$

we see that (2.2.24) is equivalent to the “two-scale relations”

$$\varphi(x) = \sum_{k \in \mathbb{Z}} p_k^0 \varphi(2x - k), \quad \tilde{\varphi}(x) = \sum_{k \in \mathbb{Z}} \overline{g_{-k}^0} \tilde{\varphi}(2x - k) \quad (2.2.26)$$

of the “scaling functions” φ and $\tilde{\varphi}$. Hence, the sequences $\{p_k^0\}$ and $\{g_{-k}^0\}$ are the corresponding “two-scale sequences,” and $P^0(z)$ and $\overline{G^0(z)}$, the corresponding “two-scale symbols,” of the scaling functions φ and $\tilde{\varphi}$, respectively.

We say that φ and $\tilde{\varphi}$ are dual to each other if they satisfy

$$\langle \varphi(\cdot - j), \tilde{\varphi}(\cdot - k) \rangle = \delta_{j,k}, \quad j, k \in \mathbb{Z}. \quad (2.2.27)$$

Under the assumption (2.2.20)–(2.2.22), a necessary and sufficient condition for the duality relationship (2.2.27) is that P^0 and G^0 are “dual two-scale symbols,” in the sense that

$$P^0(z)G^0(z) + P^0(-z)G^0(-z) = 1, \quad |z| = 1 \quad (2.2.28)$$

A proof of this statement is given in Lemma 2.2.3 for the case of polynomial symbols. It is to be noted that the assumption in (2.2.20)–(2.2.22) for P^0 and G^0 can be somewhat weakened. Hence, for more generality, we will drop this requirement, but simply assume that $\hat{\varphi}$ and $\hat{\tilde{\varphi}}$ defined in (2.2.23) are in L^p , $\{V_n\}$ and $\{\tilde{V}_n\}$ are MRA of L^p , such that (2.2.27) and (2.2.28) are satisfied.

Next, let us consider an arbitrary Laurent series $R(z)$ of class \mathcal{W} , which never vanishes on the unit circle $|z| = 1$. We also have that $\frac{1}{R(z)} \in \mathcal{W}$, and this yields two other Laurent series in \mathcal{W} , namely

$$P^1(z) = -zG^0(-z)R(z^2) \quad \text{and} \quad G^1(z) = -z^{-1}P^0(-z)/R(z^2). \quad (2.2.29)$$

The reason for introducing P^1 and G^1 is that the two matrices

$$M(z) = \begin{bmatrix} P^0(z) & P^1(z) \\ P^0(-z) & P^1(-z) \end{bmatrix} \quad \text{and} \quad \tilde{M}(z) = \begin{bmatrix} G^0(z) & G^1(z) \\ G^0(-z) & G^1(-z) \end{bmatrix} \quad (2.2.30)$$

are nonsingular for $|z| = 1$, independent of the choice of R , and that $M^T(z)$ and $\tilde{M}(z)$ are inverse of each other

$$P^1(z) = \frac{1}{2} \sum_{n \in \mathbb{Z}} p_n^1 z^n, \quad G^1(z) = \frac{1}{2} \sum_{n \in \mathbb{Z}} g_n^1 z^n, \quad (2.2.31)$$

and set

$$\omega_0 = \varphi \text{ and } \tilde{\omega}_0 = \tilde{\varphi}. \quad (2.2.32)$$

Then, in view of (2.2.26) we may introduce two sequences of L^p functions $\{\omega_n\}$ and $\{\tilde{\omega}_n\}$ defined by

$$\begin{cases} \omega_{2n+\lambda}(x) = \sum_{k \in \mathbb{Z}} p_k^\lambda \omega_n(2x - k), & \lambda = 0, 1 \\ \tilde{\omega}_{2n+\lambda}(x) = \sum_{k \in \mathbb{Z}} \overline{g_{-k}^\lambda} \tilde{\omega}_n(2x - k), & \lambda = 0, 1 \end{cases} \quad (2.2.33)$$

where $n = 0, 1, 2, \dots$

Of course (2.2.33) reduces to (2.2.26) when $\lambda = 0$ and $n = 0$. The functions

$$\psi = \omega_1 \text{ and } \tilde{\psi} = \tilde{\omega}_1 \quad (2.2.34)$$

obtained by setting $\lambda = 1$ and $n = 0$ in (2.2.33) are dual wavelet packets in the sense that

$$\langle \omega_{j,1,k}, \tilde{\omega}_{l,1,m} \rangle = \delta_{j,l} \delta_{k,m}, \quad j, k, l, m \in \mathbb{Z}. \quad (2.2.35)$$

We will document this statement in a moment. In general, we call $\{\omega_n\}$ and $\{\tilde{\omega}_n\}$ sequences of the wavelet packets, and $\{\tilde{\omega}_n\}$ the dual of $\{\omega_n\}$. In the orthonormal setting, where $\tilde{\omega}_n = \omega_n$, $n = 0, 1, \dots$ (obtain by setting $G^\lambda(z) = P^\lambda(z)$ and $|R(z)| = 1$ for $\lambda = 0, 1$ into (2.2.29)) the functions ω_n become the orthonormal wavelet packets introduced by Coifman and Meyer [75].

We now observe that the Laurent series P^λ and G^λ , $\lambda = 0, 1$ that satisfy (2.2.28) and (2.2.29) for some $R \in \mathcal{W}$ with $R(z) \neq 0$ on $|z| = 1$ also satisfy

$$\begin{cases} P^0(z)G^0(z) + P^1(z)G^1(z) = 1 \\ P^0(-z)G^0(z) + P^1(-z)G^1(z) = 0, \quad |z| = 1 \end{cases} \quad (2.2.36)$$

$$P^\lambda(z)G^\mu(z) + P^\lambda(-z)G^\mu(-z) = \delta_{\lambda, \mu}, \quad |z| = 1, \quad \lambda, \mu = 0, 1. \quad (2.2.37)$$

We also remark that if $P^\lambda, G^\lambda \in \mathcal{W}$ satisfy (2.2.36) or (2.2.37), then there exists some $R \in \mathcal{W}$, with $R(z) \neq 0$ on $|z| = 1$, such that

$$P^0(z)G^0(z) + P^0(-z)G^0(-z) = 1, \quad P^1(z) = -zG^0(-z)R(z^2) \quad (2.2.38)$$

$$G^1(z) = \frac{-z^{-1}P^0(-z)}{R(z^2)}, \quad |z| = 1.$$

Now, we assume that P^0 and G^0 are dual symbols as in (2.2.28), it follows that P^1 and G^1 are defined by (2.2.29), and the series expressions (2.2.25) and (2.2.31) are used. With these (two-scale) coefficient sequences $\{p_n^\lambda\}$ and $\{g_{-n}^\lambda\}$, $\lambda = 0, 1$ define linear operators P_λ and G_λ , $\lambda = 0, 1$ as introduced in [64] on

$$l^2 = \left\{ v = \{v_k\} : \sum_{k \in \mathbb{Z}} |v_k|^2 < \infty \right\} \quad (2.2.39)$$

by

$$(P_\lambda(v))_l = \sum_{k \in \mathbb{Z}} p_{k-2l}^\lambda v_k, \quad l \in \mathbb{Z}, \lambda = 0, 1 \quad (2.2.40)$$

and

$$(G_\lambda(v))_\lambda = \sum_{k \in \mathbb{Z}} \overline{g_{-k+2l}^\lambda} v_k, \quad l \in \mathbb{Z}, \lambda = 0, 1. \quad (2.2.41)$$

Then, the adjoints P_λ^* and G_λ^* of P_λ and G_λ , respectively, are given by

$$(P_\lambda^* v)_l = \sum_{k \in \mathbb{Z}} p_{l-2k}^\lambda v_k, \quad l \in \mathbb{Z}, \lambda = 0, 1 \quad (2.2.42)$$

and

$$(G_\lambda^* v)_l = \sum_{k \in \mathbb{Z}} g_{-l+2k}^\lambda v_k, \quad l \in \mathbb{Z}, \lambda = 0, 1. \quad (2.2.43)$$

Indeed, by adopting the notation

$$\langle u, v \rangle_{l^2} = \sum_{k \in \mathbb{Z}} u_k \overline{v_k} \quad (2.2.44)$$

where $u = \{u_k\}$ and $v = \{v_k\}$, for the inner product of the l^2 -space, we have

$$\langle u, P_\lambda v \rangle = \sum_{l \in \mathbb{Z}} u_l (\overline{P_\lambda v})_l = \sum_{k \in \mathbb{Z}} \left(\sum_{l \in \mathbb{Z}} \overline{P_{k-2l}^\lambda} u_l \right) \overline{v_k} = \langle P_\lambda^* u, v \rangle \quad (2.2.45)$$

and similarly

$$\langle u, G_\lambda v \rangle = \langle G_\lambda^* u, v \rangle. \quad (2.2.46)$$

Remark 2.2.4 Let P_λ and G_λ be as defined in (2.2.40) and (2.2.41) with adjoints P_λ^* and G_λ^* , respectively. Then

$$P_\lambda G_\mu^* = 2\delta_{\lambda,\mu} I, \quad \lambda, \mu = 0, 1 \quad (2.2.47)$$

and

$$G_0^* P_0 + G_1^* P_1 = 2I \quad (2.2.48)$$

where I is the identity operator on l^2 .

By taking the adjoint of both (2.2.47) and (2.2.48), we also have

$$G_\mu P_\lambda^* = 2\delta_{\lambda,\mu} I, \quad \lambda, \mu = 0, 1 \quad (2.2.49)$$

and

$$P_0^* G_0 + P_1^* G_1 = 2I. \quad (2.2.50)$$

That is, the identities (2.2.47) and (2.2.48) are equivalent to (2.2.49) and (2.2.50), respectively.

Remark 2.2.5 We can view P_λ and G_λ , $\lambda = 0, 1$, as bi-infinite matrices, namely

$$P_\lambda = [p_{k-2l}^\lambda]_{l,k \in \mathbb{Z}} \quad \text{and} \quad G_\lambda = [\overline{g_{k-2l}^\lambda}]_{l,k \in \mathbb{Z}}$$

with $P_\lambda^* = \overline{P_\lambda}^T$ and $G_\lambda^* = \overline{P_\lambda}^T$ being the conjugate transposes of P_λ and G_λ , respectively, and I being the identity matrix. Then, the equalities (2.2.47)–(2.2.50) can be viewed as matrix identities.

As a consequence of Remark 2.2.4, we see that

$$\left(\frac{1}{2} G_\lambda^* P_\lambda\right) \left(\frac{1}{2} G_\lambda^* P_\lambda\right) = \frac{1}{2} G_\lambda^* \left(\frac{1}{2} P_\lambda G_\lambda^*\right) P_\lambda = \frac{1}{2} G_\lambda^* P_\lambda$$

and

$$\left(\frac{1}{2} G_\lambda^* P_\lambda\right) \left(\frac{1}{2} G_\mu^* P_\mu\right) = \frac{1}{2} G_\lambda^* \left(\frac{1}{2} P_\lambda G_\mu^*\right) P_\mu = 0, \quad \lambda \neq \mu.$$

This shows that

- (i) $\frac{1}{2} G_\lambda^* P_\lambda$ is a projection on l^2 , $\lambda = 0, 1$;
- (ii) The range of $\frac{1}{2} P_0^* G_0$ and $\frac{1}{2} G_1^* P_1$ is orthogonal (since $\langle P_0^* G_0 u, G_1^* P_1 v \rangle_{l^2} = \langle u, G_0^* P_0 G_1^* P_1 v \rangle_{l^2} = 0$); while
- (iii) The range of $\frac{1}{2} G_0^* P_0$ and $\frac{1}{2} G_1^* P_1$ forms a direct sum decomposition of l^2 .

Now, using the notation of (2.2.33) in (2.2.40), we have

$$\omega_{2n+\lambda}(x-l) = (P_\lambda \{\omega_n(2x-.)\})_l, \quad l \in \mathbb{Z}, \quad \lambda = 0, 1. \quad (2.2.51)$$

Similarly, we also have, from (2.2.33) and (2.2.41),

$$\tilde{\omega}_{2n+\lambda}(x-l) = (G_\lambda \{\tilde{\omega}_n(2x-.)\})_l, \quad l \in \mathbb{Z}, \quad \lambda = 0, 1. \quad (2.2.52)$$

As an application of (2.2.33) in Remark 2.2.5 by putting (2.2.51) in (2.2.43), we obtain for all $k \in \mathbb{Z}$

$$\begin{aligned}\omega_n(2x - k) &= \frac{1}{2}[G_0^*(P_0\{\omega_0(2x - .)\})]_k + \frac{1}{2}[G_1^*(P_1\{\omega_n(2x - .)\})]_k \\ &= \frac{1}{2} \sum_{l \in \mathbb{Z}} g_{2l-k}^0 \omega_{2n}(x - l) + \frac{1}{2} \sum_{l \in \mathbb{Z}} g_{2l-k}^1 \omega_{2n+1}(x - l).\end{aligned}\quad (2.2.53)$$

In the same manner, from (2.2.52), (2.2.50), and (2.2.42), we also have for $k \in \mathbb{Z}$

$$\begin{aligned}\tilde{\omega}_n(2x - k) &= \frac{1}{2}[P_0^*(G_0\{\tilde{\omega}_n(2x - .)\})]_k + \frac{1}{2}[P_1^*(G_1\{\tilde{\omega}_n(2x - .)\})]_k \\ &= \frac{1}{2} \sum_{l \in \mathbb{Z}} \overline{p_{k-2l}^0} \tilde{\omega}_{2n}(x - l) + \frac{1}{2} \sum_{l \in \mathbb{Z}} p_{k-2l}^1 \tilde{\omega}_{2n+1}(x - l).\end{aligned}\quad (2.2.54)$$

For $n = 0$, (2.2.53) and (2.2.54) give the so-called decomposition formulae in wavelet packet analysis. In general, we call (2.2.53) and (2.2.54) “decomposition formulae” for wavelet packets $\{\omega_n\}$ and $\{\tilde{\omega}_n\}$, $n \in \mathbb{Z}^+$.

2.3 Certain Results on Wavelet Packets

Let P_j and Q_j , respectively, be the orthogonal projections onto the spaces V_j and W_j with kernels $P_j(x, y)$ and $Q_j(x, y)$, defined as follows:

$$P_j(x, y) = \sum_k \varphi_{j,k}(x) \overline{\varphi_{j,k}(y)}, \quad (2.3.1)$$

where $\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k)$ and

$$Q_j(x, y) = \sum_k \psi_{j,k}(x) \overline{\psi_{j,k}(y)}. \quad (2.3.2)$$

In light of $V_{j+1} = V_j \oplus W_j$, $P_j(x, y)$ can be written as

$$P_j(x, y) = \sum_{m < j} Q_m(x, y) = \sum_{m < j; k} \psi_{m,k}(x) \overline{\psi_{m,k}(y)}. \quad (2.3.3)$$

Now, we consider a projection Q_j^n onto U_j^n with kernel $Q_j^n(x, y)$ defined as

$$Q_j^n(x, y) = \sum_{k \in \mathbb{Z}} \omega_{j,n,k}(x) \overline{\omega_{j,n,k}(y)} \quad (2.3.4)$$

where $\omega_{j,n,k}$ are the wavelet packets and $\omega_{j,n,k}(x) = 2^{j/2}\omega_n(2^j x - k)$. Thus, we observe that $Q_j^0 = P_j$ and $Q_j^1 = Q_j$.

Theorem 2.3.1 Let $\omega_n \in L^2(\mathbb{R})$ be orthonormal wavelet packets for all $n = 0, 1, 2, \dots$. Then

$$|\hat{\omega}_n(\xi)|^2 = \sum_{s=0}^{2^u-1} |\hat{\omega}_{2^u n+s}(2^u \xi)|^2 \quad \text{a.e. } \xi \in \mathbb{R}$$

where $u = 0, 1, 2, \dots$

Proof From the definition of wavelet packets, we have (2.2.8). The Fourier transform of (2.2.8) gives

$$\left. \begin{aligned} \hat{\omega}_{2n}(\xi) &= m_0\left(\frac{\xi}{2}\right) \hat{\omega}_{2n}\left(\frac{\xi}{2}\right) \\ \hat{\omega}_{2n+1}(\xi) &= m_1\left(\frac{\xi}{2}\right) \hat{\omega}_{2n}\left(\frac{\xi}{2}\right) \end{aligned} \right\} \quad (2.3.5)$$

Using (1.6.7)–(1.6.8) in (2.3.5), it gives

$$\begin{aligned} |\hat{\omega}_{2n}(2\xi)|^2 + |\hat{\omega}_{2n+1}(2\xi)|^2 &= |m_0(\xi)\hat{\omega}_n(\xi)|^2 + |m_1(\xi)\hat{\omega}_n(\xi)|^2 \\ &= \left\{ |m_0(\xi)|^2 + |e^{i\xi}m_0(\xi + \pi)|^2 \right\} |\hat{\omega}_n(\xi)|^2. \end{aligned}$$

That is, $|\hat{\omega}_{2n}(2\xi)|^2 + |\hat{\omega}_{2n+1}(2\xi)|^2 = |\hat{\omega}_n(\xi)|^2$. A simple iteration of this result yields $|\hat{\omega}_n(\xi)|^2 = \sum_{s=0}^{2^u-1} |\hat{\omega}_{2^u n+s}(2^u \xi)|^2$, and the statement is proved.

Theorem 2.3.2 If $\omega_n \in L^2(\mathbb{R})$ are wavelet packets associated with the scaling function $\varphi = \omega_0$, then

$$\sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 = 1 \quad \text{for a.e. } \xi \in \mathbb{R} \quad (2.3.6)$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$.

Proof From Theorem 2.3.1, we have

$$|\hat{\omega}_1(\xi)|^2 = \sum_{s=0}^{2^u-1} |\hat{\omega}_{2^u+s}(2^u \xi)|^2 = \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^u \xi)|^2 \quad (2.3.7)$$

since $\omega_1 = \psi$. Lemma 1.6.3 together with Lemma 1.6.4 gives

$$\sum_{k \in \mathbb{Z}} |\hat{\varphi}(\xi + 2k\pi)|^2 = \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_1(2^l(\xi + 2k\pi))|^2.$$

Therefore

$$\sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_1(2^l(\xi + 2k\pi))|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}. \quad (2.3.8)$$

From (2.3.7) and (2.3.8), we obtain the equality (2.3.6).

Theorem 2.3.3 *For orthonormal wavelet packets $\omega_n \in L^2(\mathbb{R})$, the expression*

$$D(\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2,$$

where $l = j - u$, $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}^+$ is well defined and finite for almost every $\xi \in \mathbb{R}$. Moreover

$$\int_I D(\xi) d\xi = 2\pi \quad (2.3.9)$$

for any interval I of length 2π in \mathbb{R} .

Proof $D(\xi)$ is well defined and cannot be infinite on a set of positive measure if the last part of the theorem is true. The second part will follow if we prove the result for $I = [0, 2\pi]$ since $D(\xi)$ is 2π -periodic. Now

$$\begin{aligned} \int_I D(\xi) d\xi &= \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} \int_0^{2\pi} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 d\xi \\ &= \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} \int_{2k\pi}^{2(k+1)\pi} |\hat{\omega}_n(2^l\xi)|^2 d\xi \\ &= 2\pi \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \int_{-\infty}^{\infty} |\hat{\omega}_n(2^l\xi)|^2 d\xi \\ &= 2\pi \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} 2^{-l} \int_{-\infty}^{\infty} |\hat{\omega}_n(\xi)|^2 d\xi \\ &= 2\pi \sum_{l=1}^{\infty} 2^{-l} \sum_{n=2^u}^{2^{u+1}-1} \|\hat{\omega}_n\|_2^2 \\ &= 2\pi \sum_{l=1}^{\infty} 2^{-l} 2^u \|\hat{\omega}_n\|_2^2 \\ &= 2\pi \|\hat{\omega}_n\|_2^2 = 2\pi \end{aligned}$$

and the statement is proved.

Theorem 2.3.4 Let $\omega_n \in L^2(\mathbb{R})$ be the wavelet packets. Then, the necessary and sufficient conditions for the orthonormality of the system $\{\omega_{l,n,k} : n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1, l = j - u, j, k \in \mathbb{Z}\}$, where $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$ are

$$\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(\xi + 2k\pi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R} \quad (2.3.10)$$

and

$$\sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^l(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} = 0 \text{ for a.e. } \xi \in \mathbb{R}, l \geq 1 \quad (2.3.11)$$

where $t = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$.

Proof Since $\omega_n \in L^2(\mathbb{R})$ are wavelet packets from periodization argument, we know that

$$\begin{aligned} \delta_{k,0} &= \int_{\mathbb{R}} \omega_n(x) \overline{\omega_n(x-k)} dx \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{\omega}_n(\mu)|^2 e^{ik\mu} d\mu \quad [\text{By Plancherel theorem}] \\ &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \int_{2s\pi}^{2(s+1)\pi} |\hat{\omega}_n(\mu)|^2 e^{ik\mu} d\mu \\ &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \int_0^{2\pi} |\hat{\omega}_n(\xi + 2s\pi)|^2 e^{ik\xi} d\xi \\ &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \int_0^{2\pi} \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(\xi + 2s\pi)|^2 \right) e^{ik\xi} d\xi. \end{aligned}$$

This tells us that the 2π -periodic function $\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(\xi + 2k\pi)|^2$ equals 1, i.e., since it has Fourier coefficient 1 at the frequency $k = 0$ and all the other coefficients are zero. The converse is immediate. Thus, $\{\omega_n(x-k) : k \in \mathbb{Z}\}$ is an orthonormal system if and only if

$$\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(\xi + 2k\pi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}.$$

Now, performing a change of variables, we see that

$$\langle \omega_{l,n,k}, \omega_{l,t,s} \rangle = \langle \omega_{0,n,k}, \omega_{0,t,s} \rangle.$$

This implies that the system $\{\omega_{l,n,k} : n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1, k \in \mathbb{Z}\}$, where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$, is orthonormal for each fixed j when (2.3.10) is satisfied.

If, say, $l > m$, then the change of variables $x = 2^{-m}(y + r)$ shows that

$$\langle \omega_{l,n,k}, \omega_{m,t,r} \rangle = \langle \omega_{s,n,p}, \omega_{0,t,0} \rangle,$$

where $s = l - m$ and $p = k - 2^{l-m}r$. Relabeling this, we see that the orthogonality between $\omega_{l,n,k}$ and $\omega_{m,t,r}$ for $l > m$ and $k, r \in \mathbb{Z}$, can be reduced to the orthogonality between $\omega_{l,n,k}$ and ω_t , when $l > 0$ and $k \in \mathbb{Z}$. By the Plancherel theorem, it follows that for $l \geq 1$ and $k \in \mathbb{Z}$,

$$\begin{aligned} 0 &= \langle \omega_t, \omega_{l,n,k} \rangle \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \hat{\omega}_t(\xi) 2^{-\frac{l}{2}} \overline{\hat{\omega}_n(2^{-l}\xi)} e^{i2^{-l}\xi k} d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} 2^{\frac{l}{2}} \hat{\omega}_t(2^l \mu) \overline{\hat{\omega}_n(\mu)} e^{ik\mu} d\xi. \end{aligned}$$

Thus

$$\begin{aligned} 0 &= \sum_{s \in \mathbb{Z}} \int_{2s\pi}^{2(s+1)\pi} \hat{\omega}_t(2^l \xi) \overline{\hat{\omega}_n(\xi)} e^{ik\xi} d\xi \\ &= \int_0^{2\pi} \left\{ \sum_{s \in \mathbb{Z}} \hat{\omega}_t(2^l(\xi + 2s\pi)) \overline{\hat{\omega}_n(\xi + 2s\pi)} \right\} e^{ik\xi} d\xi \end{aligned}$$

for all $k \in \mathbb{Z}$ when $l \geq 1$. This shows that

$$\sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^l(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} = 0 \quad \text{a.e. on } \mathbb{R}, \quad l \geq 1.$$

Converse is immediate.

Theorem 2.3.5 *Let $\omega_n \in L^2(\mathbb{R})$ be the wavelet packets for all n . Then, the two series (2.3.10) and (2.3.11) converge for a.e. $\xi \in \mathbb{R}$.*

Proof Let us define

$$\zeta_j(\xi) = \sum_{k \in \mathbb{Z}} \hat{\omega}_n(2^j(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)}, \quad \text{a.e. } \xi \in \mathbb{R}, \quad j \in \mathbb{Z}. \quad (2.3.12)$$

Then, the change of variables combined with the Schwartz inequality gives us

$$\begin{aligned}
\int_T |\zeta_j(\xi)| d\xi &= \int_T \left| \sum_{k \in \mathbb{Z}} \hat{\omega}_n(2^j(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} \right| d\xi \\
&\leq \int_{\mathbb{R}} \left| \hat{\omega}_n(2^j \xi) \hat{\omega}_n(\xi) \right| d\xi \\
&\leq \left(\int_{\mathbb{R}} |\hat{\omega}_n(2^j \xi)|^2 d\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} |\hat{\omega}_n(\xi)|^2 d\xi \right)^{\frac{1}{2}} \\
&= 2^{-\frac{j}{2}} \int_{\mathbb{R}} |\hat{\omega}_n(\xi)|^2 d\xi = 2^{-\frac{j}{2}} 2\pi \int_{\mathbb{R}} |\omega_n(\xi)|^2 d\xi \\
&= 2^{-\frac{j}{2}} 2\pi \|\omega_n(\xi)\| < \infty.
\end{aligned}$$

This shows that $\zeta_j \in L^1(T)$ for all $j \in \mathbb{Z}$ and gives us the a.e. absolute convergence of the series (2.3.10) and (2.3.11) when $j = 0$ and $j \neq 0$, respectively.

Theorem 2.3.6 *Let $f \in L^2(\mathbb{R})$. Then*

(A) *f is orthogonal to U_j^n if and only if*

$$\sum_{k \in \mathbb{Z}} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)} = 0 \text{ for a.e. } \xi \in \mathbb{R}.$$

(B) *For the projection operator Q_j^n , we have*

$$(Q_j^n f) \hat{=} (\hat{\omega}_n(2^{-j}\xi) \sum_{k \in \mathbb{Z}} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)}) \text{ for a.e. } \xi \in \mathbb{R}.$$

Proof To prove this, we examine the orthogonal projection Q_j^n of $L^2(\mathbb{R})$ onto U_j^n , where

$$U_j^n = \overline{\text{span}\{\omega_{j,n,k} : k \in \mathbb{Z}\}}, \quad j \in \mathbb{Z}, \quad n = 0, 1, 2, \dots$$

We know that $\{\omega_{j,n,k} : k \in \mathbb{Z}\}$ is an orthonormal basis for U_j^n . Therefore

$$Q_j^n f = \sum_{k \in \mathbb{Z}} \langle f, \omega_{j,n,k} \rangle \omega_{j,n,k} \text{ for all } f \in L^2(\mathbb{R}). \quad (2.3.13)$$

Since $(\omega_{j,n,k}) \hat{=} 2^{-j/2} \hat{\omega}_n(2^{-j}\xi) e^{-i2^{-j}k\xi}$ for all $j, k \in \mathbb{Z}$, we see that $\{\gamma_{j,n,k} : k \in \mathbb{Z}\}$ is an orthonormal basis for $(U_j^n) \hat{=}$, where

$$\gamma_{j,n,k}(\xi) = \frac{2^{-j/2}}{\sqrt{2\pi}} e^{-i2^{-j}k\xi} \hat{\omega}_n(2^{-j}\xi), \quad j, k \in \mathbb{Z}. \quad (2.3.14)$$

The factor $\frac{1}{\sqrt{2\pi}}$ comes from the Plancherel theorem, and it normalizes the $\gamma_{j,n,k}$'s. Hence, we can write

$$(\mathcal{Q}_j^n f) \hat{=} \sum_{k \in \mathbb{Z}} \langle \hat{f}, \gamma_{j,n,k} \rangle \gamma_{j,n,k} \text{ for all } f \in L^2(\mathbb{R}). \quad (2.3.15)$$

Let $g_j(\mu) = \hat{f}(\mu) \overline{\hat{\omega}_n(2^{-j}\mu)}$ and $F_j(\xi) = \sum_{l \in \mathbb{Z}} g_j(\xi + 2^{j+1}l\pi)$. The last function is $2^{j+1}\pi$ periodic and so are the functions

$$E_k^{(j)}(\xi) = \frac{2^{-j/2}}{\sqrt{2\pi}} e^{-i2^{-j}k\xi}, \quad k \in \mathbb{Z}. \quad (2.3.16)$$

In fact, the system $\{E_k^{(j)}(\xi) : k \in \mathbb{Z}\}$ is an orthonormal basis for $L^2([0, 2^{j+1}\pi])$. Using again a periodization argument, we obtain

$$\begin{aligned} \langle \hat{f}, \gamma_{j,n,k} \rangle &= \int_{\mathbb{R}} \hat{f}(\xi) \frac{2^{-j/2}}{\sqrt{2\pi}} e^{i2^{-j}k\xi} \overline{\hat{\omega}_n(2^{-j}\xi)} d\xi \\ &= \int_0^{2^{j+1}\pi} \sum_{l \in \mathbb{Z}} g_j(\xi + 2^{j+1}l\pi) \frac{2^{-j/2}}{\sqrt{2\pi}} e^{i2^{-j}k\xi} d\xi. \end{aligned}$$

This gives

$$\langle \hat{f}, \gamma_{j,n,k} \rangle = \langle F_j, E_k^{(j)} \rangle \quad (2.3.17)$$

where the last inner product is one associated with $L^2([0, 2^{j+1}\pi])$. Hence

$$F_j(\xi) = \sum_{k \in \mathbb{Z}} \langle \hat{f}, \gamma_{j,n,k} \rangle E_k^{(j)}(\xi) \quad (2.3.18)$$

with convergence in $L^2([0, 2^{j+1}\pi])$. More explicitly, we have obtained

$$\sum_{l \in \mathbb{Z}} \hat{f}(\xi + 2^{j+1}l\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2l\pi)} = \sum_{k \in \mathbb{Z}} \langle \hat{f}, \gamma_{j,n,k} \rangle E_k^{(j)}(\xi) \quad (2.3.19)$$

Now, we prove the required results.

(A) Let f be orthogonal to U_j^n . Then, by Plancherel theorem

$$\begin{aligned} 0 &= \langle f, \omega_{j,n,k} \rangle = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\xi) 2^{-j/2} \overline{\hat{\omega}_n(2^{-j}\xi)} e^{-i2^{-j}k\xi} d\xi \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\xi) \overline{\gamma_{j,n,k}(\xi)} d\xi. \end{aligned}$$

Thus

$$\langle \hat{f}, \gamma_{j,n,k} \rangle = 0 \text{ for a.e. } \xi \in \mathbb{R} \text{ and } k \in \mathbb{Z}. \quad (2.3.20)$$

Equation (2.3.19) together with (2.3.20) gives the required result (A).

(B) On multiplying (2.3.19) by $\hat{\omega}_n(2^{-j}\xi)$, we get

$$\begin{aligned} & \hat{\omega}_n(2^{-j}\xi) \sum_{k \in \mathbb{Z}} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)} \\ &= \sum_{k \in \mathbb{Z}} \langle \hat{f}, \gamma_{j,n,k} \rangle \hat{\omega}_n(2^{-j}\xi) E_k^{(j)}(\xi). \end{aligned}$$

From (2.3.14) and (2.3.16), we observe that $\hat{\omega}_n(2^{-j}\xi) E_k^{(j)}(\xi) = \gamma_{j,n,k}(\xi)$. Thus

$$\hat{\omega}_n(2^{-j}\xi) \sum_{k \in \mathbb{Z}} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)} = \sum_{k \in \mathbb{Z}} \langle \hat{f}, \gamma_{j,n,k} \rangle \gamma_{j,n,k}(\xi).$$

Relation (2.3.15) together with the above result gives

$$(Q_j^n f)(\xi) = \hat{\omega}_n(2^{-j}\xi) \sum_{k \in \mathbb{Z}} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)} \text{ for a.e. } \xi \in \mathbb{R}.$$

2.4 Band-Limited Wavelet Packets

A function $f \in L^2(\mathbb{R})$ is said to be band-limited if the support of \hat{f} is contained in a finite interval. Suppose that $\omega_n \in L^2(\mathbb{R})$ are band-limited wavelet packets for all $n = 0, 1, 2, \dots$. Then, we can always find an integer J such that $\text{supp}(\hat{\omega}_n) \subset [-2^J\pi, 2^J\pi]$. We will assume this inclusion throughout this section.

Theorem 2.4.1 Suppose that $f \in L^2(\mathbb{R})$ and \hat{f} has a support contained in $I = (a, b)$, where $b - a \leq 2^{-j}\pi$ and $I \cap [-\pi, \pi] = \emptyset$. Then, for all $j \in \mathbb{Z}$,

$$(Q_j^n f)(\xi) = \hat{f}(\xi) |\hat{\omega}_n(2^{-j}\xi)|^2 \text{ for a.e. } \xi \in I.$$

Proof If $-j \leq J$, $k \neq 0$ and $\xi \in I$, we have that $\xi + 2^{j+1}k\pi$ lies outside the support of \hat{f} . In fact, if $k > 0$,

$$b = a + (b - a) \leq a + 2^{-j}\pi \leq a + 2^{j+1}k\pi < \xi + 2^{j+1}k\pi$$

so that, in this case $\xi + 2^{j+1}k\pi$ is to the right of I . Further, if $k < 0$,

$$a = b + (a - b) \geq b - 2^{-J}\pi \geq b + 2^{-J}k\pi \geq b + 2^{j+1}k\pi > \xi + 2^{j+1}k\pi$$

which proves that $\xi + 2^{j+1}k\pi$ is to the left of I . Therefore, in the formula for $(Q_j^n f)^\wedge$ given in Theorem 2.3.6(B), all terms are zero except the one corresponding to $k = 0$ and this proves the theorem if $-j \leq J$. Note that the condition $I \cap [-\pi, \pi] = \emptyset$ has not been used in this part.

If $-j > J$, we have

$$|2^{-j}\xi| \geq 2^J|\xi| \geq 2^J\pi$$

when $\xi \in I$ since $I \cap [-\pi, \pi] = \emptyset$. Hence $\hat{\omega}_n(2^{-j}\xi) = 0$, which proves that $\hat{f}(\xi)$ $|\hat{\omega}_n(2^{-j}\xi)|^2 = 0$. On the other hand, if $\xi + 2^{j+1}k\pi \in I$, then

$$|2^{-j}\xi + 2k\pi| \geq 2^J|\xi + 2^{j+1}k\pi| \geq 2^J\pi$$

and hence $\hat{\omega}_n(2^{-j}\xi + 2k\pi) = 0$. On using Theorem 2.3.6(B), we deduce that $(Q_j^n f)^\wedge(\xi) = 0$. Thus

$$(Q_j^n f)^\wedge(\xi) = \hat{f}(\xi)|\hat{\omega}_n(2^{-j}\xi)|^2$$

when $-j > J$ and $\xi \in I$ (since both terms are zero).

Theorem 2.4.2 *Let ω_n be band-limited orthonormal wavelet packets for all $n = 0, 1, 2, \dots$. Then*

$$\sum_{j \leq 0} |\hat{\omega}_1(2^{-j}\xi)|^2 + \sum_{j > 0} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^{-l}\xi)|^2 = 1$$

for a.e. $\xi \in \mathbb{R} - \{0\}$, where $u = 1, 2, \dots, j$, $l = j - u$.

Proof From Lemma 1.6.5, we can write

$$\sum_{j \in \mathbb{Z}} |\hat{\omega}_1(2^{-j}\xi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R} - \{0\} \quad (2.4.1)$$

which gives

$$\sum_{j \leq 0} |\hat{\omega}_1(2^{-j}\xi)|^2 + \sum_{j > 0} |\hat{\omega}_1(2^{-j}\xi)|^2 = 1 \quad (2.4.2)$$

for a.e. $\xi \in \mathbb{R} - \{0\}$.

Now if we take $n = 1$ in Theorem 2.3.1, we get

$$|\hat{\omega}_1(\xi)|^2 = \sum_{s=0}^{2^u-1} |\hat{\omega}_{2^u+s}(2^u\xi)|^2 = \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^u\xi)|^2$$

where $u = 1, 2, \dots$

The above result together with (2.4.2) gives

$$\sum_{j \leq 0} |\hat{\omega}_1(2^{-j}\xi)|^2 + \sum_{j > 0} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^{-j+u}\xi)|^2 = 1$$

$$\sum_{j \leq 0} |\hat{\omega}_1(2^{-j}\xi)|^2 + \sum_{j > 0} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^{-(j-u)}\xi)|^2 = 1$$

for a.e. $\xi \in \mathbb{R} - \{0\}$. Clearly, u can take values from $1, 2, \dots, j$ since it is impossible to decompose any wavelet space W_j more than j -times. This completes the proof.

Remark 2.4.3 If we choose the last formula in (2.2.12) to decompose each W_j , then due to $j - u = 0$, we get

$$\sum_{j \leq 0} |\hat{\omega}_1(2^{-j}\xi)|^2 + \sum_{n=2}^{\infty} |\omega_n(\xi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R} - \{0\}.$$

But

$$|\hat{\omega}_0(\xi)|^2 = \sum_{j=1}^{\infty} |\hat{\omega}_1(2^j\xi)|^2, \quad \forall \xi \in \mathbb{R}.$$

Hence (2.4.2) reduces to

$$\sum_{n=0}^{\infty} |\omega_n(\xi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}. \quad (2.4.3)$$

Theorem 2.4.4 Let ω_n be a band-limited orthonormal wavelet packet, $|\hat{\omega}_n|$ be continuous at zero and $\{\omega_{l,n,k} : n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1, l = j - u; j, k \in \mathbb{Z}\}$ be orthonormal system, where $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$. Then, $\hat{\omega}_n(0) = 0$.

Proof Since $\omega_n \perp U_j^n$ for all $j \neq 0$, Theorem 2.3.6(A) shows that

$$\sum_{k \in \mathbb{Z}} \hat{\omega}_n(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)} = 0 \text{ for a.e. } \xi \in \mathbb{R}$$

where $j \neq 0$.

Recall that $\text{supp}(\hat{\omega}_n) \subset (-2^J\pi, 2^J\pi)$, hence the point $\xi + 2^{j+1}k\pi$ lies outside of $\hat{\omega}_n$ when $k \neq 0$, $\xi \in \text{supp}(\hat{\omega}_n)$ and $j \geq J$. Therefore, the above formula reduces to

$$\hat{\omega}_n(\xi) \overline{\hat{\omega}_n(2^{-j}\xi)} = 0 \text{ a.e. on } \text{supp}(\hat{\omega}_n)$$

where $j \geq J$. But this equality is trivially true if $\xi \notin \text{supp}(\hat{\omega}_n)$, so that

$$|\hat{\omega}_n(\xi) \hat{\omega}_n(2^{-j}\xi)| = 0 \text{ for a.e. } \xi \in \mathbb{R}$$

where $j \geq J$. Taking $j \rightarrow \infty$ and using the continuity of $|\hat{\omega}_n|$ at zero, we obtain

$$|\hat{\omega}_n(\xi) \hat{\omega}_n(0)| = 0 \text{ for a.e. } \xi \in \mathbb{R}.$$

Since $\hat{\omega}_n$ is not identically zero, we deduce that $\hat{\omega}_n(0) = 0$.

Theorem 2.4.5 *Let ω_n be the band-limited orthonormal wavelet packets such that $|\hat{\omega}_n|$ is continuous at zero, for all $n = 1, 2, 3, \dots$. Then, $\hat{\omega}_n(0) = 0$ in an open neighborhood of the origin.*

Proof Since ω_n are band-limited orthonormal wavelet packets and $|\hat{\omega}_n|$ is continuous at zero, by Theorem 2.3.1 and Lemma 1.6.6, it follows that

$$\sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(0)|^2 = 0, \text{ for } u = 1, 2, 3 \dots$$

in an open neighborhood of the origin. Moreover, $\hat{\omega}_n(0) = 0$ in an open neighborhood of the origin for $n = 1, 2, \dots$

Theorem 2.4.6 *Let ω_n be the band-limited orthonormal wavelet packets such that $|\hat{\omega}_n|$ is continuous at zero, for all $n = 1, 2, 3, \dots$. Then, for each odd integer q , we have*

$$\sum_{j=0}^{\infty} \hat{\omega}_1(2^j\xi) \overline{\hat{\omega}_1(2^j(\xi + 2q\pi))} = 0 \text{ for a.e. } \xi \in \mathbb{R}. \quad (2.4.4)$$

Proof Isolating the term $k = 0$ in Theorem 2.3.6(B), we have

$$(Q_j^n f)(\xi) = |\hat{\omega}_n(2^{-j}\xi)|^2 \hat{f}(\xi) + \hat{\omega}_n(2^{-j}\xi) \sum_{k \neq 0} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)}. \quad (2.4.5)$$

Since $|\hat{\omega}_n|$ is continuous at 0, by Theorem 2.4.5 we can find $J \in \mathbb{N}$ such that $\text{supp}(\hat{\omega}_n) \subset [-2^J\pi, -2^{-J}\pi] \cup [-2^J\pi, 2^J\pi]$. If $0 < a < |\xi| < b$, only a finite number of j 's, say, $|j| \leq M$, gives us nonzero terms in the second summand of (2.4.5). Let us fix such a j . Then, there exists at most a finite number of k 's such that $\hat{\omega}_n(2^{-j}\xi + 2k\pi) \neq 0$. Thus, there exists at most a finite number of values of the form $v = 2^j k$ that occur in the nonzero summands. When $k \neq 0$, they are of the (unique) form $v = 2^m q$ for a finite number of $m, q \in \mathbb{Z}$, $m \geq j$ and q odd. Thus

$$\begin{aligned} (Q_j^n f)(\xi) &= |\hat{\omega}_n(2^{-j}\xi)|^2 \hat{f}(\xi) \\ &\quad + \hat{\omega}_n(2^{-j}\xi) \sum_{m \geq j} \sum_{q \in 2\mathbb{Z}+1} \hat{f}(\xi + 2^{m+1}q\pi) \overline{\hat{\omega}_n(2^{-j}(\xi + 2^{m+1}q\pi))}. \end{aligned} \quad (2.4.6)$$

Summing over all $j \in \mathbb{Z}$, $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ for all $j \geq 0$ and using Theorem 2.4.1, we obtain

$$\hat{f}(\xi) = \sum_{j \leq 0} (\mathcal{Q}_j^1 f) \hat{\gamma}(\xi) + \sum_{j > 0} \sum_{n=2^u}^{2^{u+1}-1} (\mathcal{Q}_l^n f) \hat{\gamma}(\xi)$$

where $l = j - u$. Hence

$$\begin{aligned} \hat{f}(\xi) &= \hat{f}(\xi) + \sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \sum_{j \leq 0} \hat{\omega}_1(2^{-j}\xi) \overline{\hat{\omega}_1(2^{-j}(\xi + 2^{m+1}q\pi))} \\ &\quad + \sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \sum_{j=0}^{m-1} \sum_{n=2^u}^{2^{u+1}-1} \hat{\omega}_n(2^{-l}\xi) \overline{\hat{\omega}_n(2^{-l}(\xi + 2^{m+1}q\pi))} \end{aligned} \quad (2.4.7)$$

where we have changed the order of summation, which we can do since all the sums involved are finite and $q \in 2\mathbb{Z} + 1$. Also, if we choose $m \leq 0$, the last term in (2.4.7) will not appear and in second term of (2.4.7) summation over $j \leq 0$ will become summation over $j < m$. Thus, for a.e. $\xi \in (-b, -a) \cup (a, b)$,

$$\begin{aligned} &\sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \sum_{j \leq 0} \hat{\omega}_1(2^{-j}\xi) \overline{\hat{\omega}_1(2^{-j}(\xi + 2^{m+1}q\pi))} \\ &+ \sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \sum_{0 < j < m} \sum_{n=2^u}^{2^{u+1}-1} \hat{\omega}_n(2^{-l}\xi) \overline{\hat{\omega}_n(2^{-l}(\xi + 2^{m+1}q\pi))} = 0 \end{aligned}$$

where $q \in 2\mathbb{Z} + 1$. Let us define

$$h_k^n(\xi) = \sum_{j=0}^{\infty} \hat{\omega}_n(2^{-j}\xi) \overline{\hat{\omega}_n(2^j(\xi + 2k\pi))}, \quad k \in 2\mathbb{Z} + 1. \quad (2.4.8)$$

Then, the change of indices $m - j = r$ give us, for a.e. $\xi \in (-b, -a) \cup (a, b)$

$$\begin{aligned} &\sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \sum_{r \geq 0} \hat{\omega}_1(2^r 2^{-m}\xi) \overline{\hat{\omega}_1(2^r(2^{-m}\xi + 2q\pi))} \\ &+ \sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \sum_{r \geq 0} \sum_{n=2^u}^{2^{u+1}-1} \hat{\omega}_n(2^r 2^{-s}) \overline{\hat{\omega}_n(2^r(2^{-s}\xi + 2q\pi))} = 0 \end{aligned}$$

where $s = m - u$. This gives

$$\sum_{q \in 2\mathbb{Z}+1} \sum_{m \in \mathbb{Z}} \hat{f}(\xi + 2^{m+1}q\pi) \left\{ h_q^1(2^{-m}\xi) + \sum_{n=2^u}^{2^{u+1}-1} h_q^n(2^{-s}\xi) \right\} = 0. \quad (2.4.9)$$

If we fix $q_0 \in 2\mathbb{Z}+1$ and $m_0 = 0$, the second term will not appear in (2.4.7) and for each $\xi_0 \in (-b, -a) \cup (a, b)$ there exists a $\delta > 0$ such that the interval

$$U = (\xi_0 + 2q_0\pi - \delta, \xi_0 + 2q_0\pi + \delta)$$

contains no point of the form $\xi_0 + 2^{m+1}q\pi$ if $(m, q) \neq (0, q_0)$. With $\hat{f} = \chi_U$ in the above equality we obtain, for a.e. $\xi \in (\xi_0 - \delta, \xi_0 + \delta) \cap \{(-b, -a) \cup (a, b)\}$,

$$0 = \sum_{q \in 2\mathbb{Z}+1} \hat{f}(\xi + 2^{m+1}q\pi) h_{q_0}^1(2^{-m}\xi) = h_{q_0}(\xi).$$

Since ξ_0 is an arbitrary point of $(-b, -a) \cup (a, b)$, the above equality is true for almost every $\xi \in (-b, -a) \cup (a, b)$. Letting $a \rightarrow 0$ and $b \rightarrow \infty$, we see that $h_{q_0}(\xi) = 0$ for almost every $\xi \in \mathbb{R} - \{0\}$. Also, observe that $h_{q_0}(\xi) = 0$ is true if $\xi = 0$ since $\hat{\omega}_1(0) = 0$. Hence, the desired result follows since q_0 is an arbitrary odd integer.

Theorem 2.4.7 *Let $\omega_n \in L^2(\mathbb{R})$ be the band-limited orthonormal wavelet packets such that $\hat{\omega}_n$ is zero on an open neighborhood of the origin and let $\{\omega_{l,n,k} : l = j - u, n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1, j, k \in \mathbb{Z}\}$ be an orthonormal system that satisfies (2.4.1) and (2.4.4), where $u = 0$, if $j < 0$ and $u = 0, 1, 2, \dots, j$, if $j \geq 0$. Then, ω_n is an orthonormal wavelet packet for all n .*

Proof It is enough to show that

$$\sum_{j < 0} (Q_j^1 f)(\xi) + \sum_{j \geq 0} \sum_{n=2^u}^{2^{u+1}-1} (Q_j^n f)(\xi) = \hat{f}(\xi)$$

a.e. on \mathbb{R} for $f \in L^2(\mathbb{R})$, where $u = 0, 1, 2, \dots, j$, $l = j - u$ and Q_j^n is an orthonormal projection onto Q_j^n as defined in (2.3.13). By Theorem 2.3.6(B), we can write

$$(Q_j^n f)(\xi) = \hat{f}(\xi) |\hat{\omega}_n(2^{-j}\xi)|^2 + \hat{\omega}_n(2^{-j}\xi) \sum_{k \neq 0} \hat{f}(\xi + 2^{j+1}k\pi) \overline{\hat{\omega}_n(2^{-j}\xi + 2k\pi)}.$$

For j fixed and $k \neq 0$, we can write $2^j k = 2^j 2^m q = 2^n q$ for $m \geq 0$ and $q \in 2\mathbb{Z}+1$. Since ω_n is band limited and $\hat{\omega}_n$ is zero in an open neighborhood of the origin, we can interchange the order of summation when we sum the above expression over n and j . Thus using (2.4.1), we obtain

$$\begin{aligned} \sum_{j<0} (\mathcal{Q}_j^1 f) \hat{\xi} + \sum_{j \geq 0} \sum_{n=2^u}^{2^{u+1}-1} (\mathcal{Q}_l^n f) \hat{\xi} &= \hat{f}(\hat{\xi}) \\ &+ \sum_{q \in 2\mathbb{Z}+1} \sum_{r \in \mathbb{Z}} \hat{f}(\hat{\xi} + 2^{r+1}q\pi) \sum_{j \leq r} \hat{\omega}_1(2^{-j}\hat{\xi}) \overline{\hat{\omega}_1(2^{-j}(\hat{\xi} + 2^{r+1}q\pi))}. \end{aligned}$$

Writing $\hat{\xi} = 2^s \mu$ and using the change of indices $t = r - s$, we see that the last sum coincides with the left-hand side of (2.4.4), which we assume to be zero. Hence

$$\sum_{j<0} (\mathcal{Q}_j^1 f) \hat{\xi} + \sum_{j \geq 0} \sum_{n=2^u}^{2^{u+1}-1} (\mathcal{Q}_l^n f) \hat{\xi} = \hat{f}(\hat{\xi}) \text{ a.e. on } \mathbb{R}$$

which proves the completeness of the system.

Now, we summarize the above results.

Theorem 2.4.8 *Let $\omega_n \in L^2(\mathbb{R})$ be the band-limited orthonormal wavelet packets such that $\hat{\omega}_n$ is zero in a neighborhood of the origin and let $\{\omega_{l,n,k} : l = j - u, n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1, j, k \in \mathbb{Z}\}$ be an orthonormal system, where $u = 0$, if $j < 0$ and $u = 0, 1, 2, \dots, j$, if $j \geq 0$. Then, this system is complete if and only if*

- (i) $\sum_{j<0} |\hat{\omega}_1(2^{-j}\hat{\xi})|^2 + \sum_{j \geq 0} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^{-l}\hat{\xi})|^2 = 1 \text{ a.e. } \hat{\xi} \in \mathbb{R} - \{0\} \text{ where } u = 0, 1, 2, \dots, j \text{ and } l = j - u.$
- (ii) $\sum_{j=0}^{\infty} \hat{\omega}_1(2^j\hat{\xi}) \overline{\hat{\omega}_1(2^j(\hat{\xi} + 2k\pi))} = 0 \text{ for a.e. } \hat{\xi} \in \mathbb{R}, k \in 2\mathbb{Z} + 1.$

2.5 Characterization of Wavelet Packets

Theorem 2.5.1 *Let $\{\omega_n\}_{n \in \mathbb{N}}$ be a orthonormal sequence of wavelet packets of $L^2(\mathbb{R})$ and $\omega_{l,n,k}$ are given by Eq. (2.2.13). Then, the function ω_n is an orthonormal wavelet packet if and only if*

$$\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l\hat{\xi})|^2 = 1, \text{ for a.e. } \hat{\xi} \in \mathbb{R} \quad (2.5.1)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and

$$\sum_{l=0}^{\infty} \hat{\omega}_n(2^l\hat{\xi}) \overline{\hat{\omega}_n(2^l(\hat{\xi} + 2m\pi))} = 0, \text{ for a.e. } \hat{\xi} \in \mathbb{R}, m \in 2\mathbb{Z} + 1. \quad (2.5.2)$$

Proof In order to prove the theorem, first of all we prove the almost everywhere convergence of the following series

$$\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(\xi + 2k\pi)|^2 = 1, \quad \text{for a.e. } \xi \in \mathbb{R} \quad (2.5.3)$$

and

$$\sum_{k \in \mathbb{Z}} \hat{\omega}_n(2^l(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} = 0, \quad \text{for a.e. } \xi \in \mathbb{R}, \quad l \geq 1. \quad (2.5.4)$$

The convergence of the series in (2.5.3) follows by integrating the absolute values of its terms over the interval $[0, 2\pi]$. Since the left-hand side of Eq. (2.5.3) is 2π -periodic, we have

$$\sum_{k \in \mathbb{Z}} \int_0^{2\pi} |\hat{\omega}_n(\xi + 2k\pi)|^2 d\xi = \sum_{k \in \mathbb{Z}} \int_{2k\pi}^{2(k+1)\pi} |\hat{\omega}_n(\xi)|^2 d\xi = \|\hat{\omega}_n\|_2^2 < \infty.$$

The almost everywhere convergence now follows from Beppo Levi's theorem. For the series in Eq. (2.5.4), we have

$$\begin{aligned} & \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi)) \hat{\omega}_n(\xi + 2k\pi)| \\ & \leq \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(\xi + 2k\pi)|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Thus, the almost everywhere convergence in Eq. (2.5.4) follows from the same result we have just proved for (2.5.3). For the convergence of the series in Eq. (2.5.2), we make use of the fact that the sum ranges over the nonnegative integers by a change of variables and Schwartz's inequality

$$\begin{aligned} & \sum_{l=0}^{\infty} \int_{-\infty}^{\infty} |\hat{\omega}_n(2^l \xi) \hat{\omega}_n(2^l(\xi + 2m\pi))| d\xi \\ & = \sum_{l=0}^{\infty} \frac{1}{2^l} \int_{\mathbb{R}} |\hat{\omega}_n(\eta) \hat{\omega}_n(\eta + 2^{l+1}m\pi))| d\eta \\ & \leq \sum_{l=0}^{\infty} \frac{1}{2^l} \|\hat{\omega}_n\|_2^2 < \infty. \end{aligned}$$

Again, the almost everywhere convergence follows by Beppo Levi's theorem.

One of the difficulties with the series that appears in Eq. (2.5.1) is that the sum is extended over all the integers and, as a consequence, the almost everywhere convergence is not established easily. It will follow, however, from the argument we shall present below.

An immediate consequence of Theorem 2.5.1 is that Eqs. (2.5.3) and (2.5.4) follow from Eqs. (2.5.1) and (2.5.2) provided $\|\omega_n\|_2 = 1$. The hypothesis $\|\omega_n\|_2 = 1$ in Theorem 2.5.1 is very natural, the system $\{\omega_{l,n,k} : l = j - u, j, k \in \mathbb{Z}; n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1\}$ where $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, cannot be normal without this assumption. It is important to point out, however, that the two Eqs. (2.5.1) and (2.5.2) without the hypothesis $\|\omega_n\|_2 = 1$ characterize systems that are not necessarily orthonormal bases for $L^2(\mathbb{R})$, but they do satisfy the basis “analyzing” and “reconstruction” properties of these bases. We shall show, in fact, that the following result is true:

Theorem 2.5.2 *If $\omega_n \in L^2(\mathbb{R})$, the following statements are equivalent:*

- $$(A) \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2 = \|f\|_2^2 \text{ for all } f \in L^2(\mathbb{R})$$
- with $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$;
- (B) ω_n satisfies Eqs. (2.5.1) and (2.5.2).

Remark 2.5.3 The proof of above theorem will be given in the form of lemmas and propositions.

In the case of $\mathbb{H} = L^2(\mathbb{R})$ and the system we are studying is $\{\omega_{l,n,k} : l = j - u, j, k \in \mathbb{Z}; n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1\}$ where $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $\omega_{l,n,k}(x) = 2^{l/2} \hat{\omega}_n(2^l x - k)$, $l, k \in \mathbb{Z}$. Let D be the class of all those functions $f \in L^2(\mathbb{R})$ such that $\hat{f} \in L^\infty(\mathbb{R})$ and \hat{f} is compactly supported in $\mathbb{R} \setminus \{0\}$. We begin by examining the expression

$$I = \sum_{l \in \mathbb{Z}} \sum_{n=2^p}^{2^{p+1}-1} \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2, \quad f \in D \quad (2.5.5)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$. Since

$$(\omega_{l,n,k})(\xi) = 2^{-l/2} \hat{\omega}_n(2^{-l} \xi) e^{-i2^{-l} k \xi}$$

an application of Plancherel’s theorem gives us

$$I = \frac{1}{(2\pi)^2} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} 2^{-l} \left| \int_{\mathbb{R}} \hat{f}(2^{-l} \xi) \overline{\hat{\omega}_n(\xi)} e^{ik\xi} d\xi \right|^2$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$. Let

$$F_l^n(\xi) = \hat{f}(2^{-l} \xi) \overline{\hat{\omega}_n(\xi)}, \quad l \in \mathbb{Z}.$$

Each F_l^n is compactly supported in $\mathbb{R} \setminus \{0\}$ and belongs to $L^2(\mathbb{R})$. If F is such a function, then $\sum_{k \in \mathbb{Z}} F(\xi + 2k\pi)$ is a 2π -periodic function in $L^2(\mathbb{T})$ whose Fourier coefficients are $\frac{1}{2\pi} \hat{F}(k)$, $k \in \mathbb{Z}$. Thus

$$\sum_{k \in \mathbb{Z}} F(\xi + 2k\pi) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{F}(k) e^{ik\xi}$$

with convergence in $L^2(\mathbb{T})$. Hence

$$\frac{1}{2\pi} \int_{\mathbb{R}} \overline{F(\xi)} \sum_{k \in \mathbb{Z}} \hat{F}(k) e^{ik\xi} d\xi = \int_{\mathbb{R}} \overline{F(\xi)} \sum_{k \in \mathbb{Z}} F(\xi + 2k\pi) d\xi$$

with each series converging in $L^2(\text{supp}(F))$. But the left side of the last equality equals

$$\frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{F}(k) \int_{\mathbb{R}} \overline{F(\xi) e^{-ik\xi}} d\xi = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} |\hat{F}(k)|^2.$$

It follows that

$$\frac{1}{2\pi} \sum_{k \in \mathbb{Z}} |\hat{F}(k)|^2 = \int_{\mathbb{R}} \overline{F(\xi)} \sum_{k \in \mathbb{Z}} F(\xi + 2k\pi) d\xi. \quad (2.5.6)$$

Applying (2.5.6) when $F = F_l^n$ in (2.5.5), we obtain

$$I = \frac{1}{2\pi} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{-l} \int_{\mathbb{R}} \hat{f}(2^{-l}\xi) \hat{\omega}_n(\xi) \sum_{k \in \mathbb{Z}} \hat{f}(2^{-l}(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} d\xi.$$

In the sum with respect to k , we separate the term $k = 0$ from the others and write $I = I_0 + I_1$ where

$$\begin{aligned} I_0 &= \frac{1}{2\pi} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{-l} \int_{\mathbb{R}} |\hat{f}(2^{-l}\xi)|^2 |\hat{\omega}_n(\xi)|^2 d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2 d\xi \end{aligned}$$

and

$$I_1 = \frac{1}{2\pi} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{-l} \int_{\mathbb{R}} \overline{\hat{f}(2^{-l}\xi)} \hat{\omega}_n(\xi) \sum_{k \neq 0} \hat{f}(2^{-l}(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} d\xi.$$

We can commute integration and summation in I_0 due to positivity of summands. Now, to find I_1 we state and prove the following lemmas.

Lemma 2.5.4 Let $\omega_n \in L^2(\mathbb{R})$ be wavelet packets. Then

$$\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{-l} \int_{\mathbb{R}} \left| \hat{f}(2^{-l}\xi) \hat{\omega}_n(\xi) \right| \sum_{k \neq 0} \left| \hat{f}(2^{-l}(\xi + 2k\pi)) \hat{\omega}_n(\xi + 2k\pi) \right| d\xi < \infty$$

whenever $f \in D$.

Remark 2.5.5 A consequence of this lemma is that I is finite for all $f \in D$ if and only if I_0 is finite and later is true if and only if

$$\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2$$

with $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$; is locally integrable in $\mathbb{R} \setminus \{0\}$. As

$$2|\hat{\omega}_n(\xi)||\hat{\omega}_n(\xi + 2k\pi)| \leq |\hat{\omega}_n(\xi)|^2 + |\hat{\omega}_n(\xi + 2k\pi)|^2$$

it is sufficient, for the proof of the Lemma 2.5.4, to show that

$$\int_{\mathbb{R}} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \neq 0} 2^{-l} \left| \hat{f}(2^{-l}\xi) \right| \left| \hat{f}(2^{-l}(\xi + 2k\pi)) \right| |\hat{\omega}_n(\xi)|^2 d\xi < \infty \quad (2.5.7)$$

since the term involving $|\hat{\omega}_n(\xi + 2k\pi)|^2$ equals the integral in (2.5.7) (after the change of variable $\eta = \xi + 2k\pi$ and replacing $-k$ by k in the sum over k). But, inequality (2.5.7) and thus, Lemma 2.5.4 is an immediate consequence of the following lemma.

Lemma 2.5.6 Suppose $0 < a < b < \infty$, $f \in D$ with

$$\text{supp}(\hat{f}) \subseteq \{\xi : a < |\xi| < b\}$$

and $\delta = \text{diam}(\text{supp}(\hat{f}))$. Then

$$\begin{aligned} \alpha(\xi) &= \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \neq 0} 2^{-l} \left| \hat{f}(2^{-l}\xi) \right| \left| \hat{f}(2^{-l}(\xi + 2k\pi)) \right| \\ &\leq \frac{\delta}{\pi} \left(1 + \log_2 \frac{b}{a} \right) \|\hat{f}\|_{L^\infty}^2 \text{ for all } \xi \in \mathbb{R}. \end{aligned}$$

Now, we consider

$$t_q^n(\xi) = \sum_{l=0}^{\infty} \hat{\omega}_n(2^l \xi) \overline{\hat{\omega}_n(2^l(\xi + 2q\pi))}, \quad q \in \mathbb{Z}. \quad (2.5.8)$$

The function t_q^n belongs to $L^1(\mathbb{R})$ since

$$\begin{aligned} \int_{\mathbb{R}} |t_q^n(\xi)| d\xi &\leq \left(\int_{\mathbb{R}} \sum_{l=0}^{\infty} |\hat{\omega}_n(2^l \xi)|^2 d\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} \sum_{l=0}^{\infty} |\hat{\omega}_n(2^l(\xi + 2q\pi))|^2 d\xi \right)^{\frac{1}{2}} \\ &= \left(\sum_{l=0}^{\infty} \int_{\mathbb{R}} 2^{-l} |\hat{\omega}_n(\xi)|^2 d\xi \right)^{\frac{1}{2}} \left(\sum_{l=0}^{\infty} \int_{\mathbb{R}} 2^{-l} |\hat{\omega}_n(\xi + 2q\pi)|^2 d\xi \right)^{\frac{1}{2}} \\ &= 2 \|\hat{\omega}_n\|_2^2. \end{aligned}$$

It will be useful to express I_1 as a sum involving these functions t_q^n with q ranging throughout the odd integers $2\mathbb{Z} + 1$.

Lemma 2.5.7 If $f \in D$, then

$$I_1 = \frac{1}{2\pi} \int_{\mathbb{R}} \sum_{s \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{q \in 2\mathbb{Z}+1} \overline{\hat{f}(\xi)} \hat{f}(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi.$$

Proof For each k in the expression for I , there exists a unique nonnegative integer r and a unique odd integer q such that $k = 2^r q$. But the sum defining I is absolutely convergent by Lemma 2.5.4. Therefore, the following equalities that lead to an expression involving t_q^n are valid:

$$\begin{aligned} 2\pi I_1 &= \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \int_{\mathbb{R}} \overline{\hat{f}(\xi)} \hat{\omega}_n(2^l \xi) \sum_{k \neq 0} \hat{f}(\xi + 2^{-l}(2k\pi)) \overline{\hat{\omega}_n(2^l \xi + 2k\pi)} d\xi \\ &= \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \int_{\mathbb{R}} \overline{\hat{f}(\xi)} \hat{\omega}_n(2^l \xi) \sum_{r=0}^{\infty} \sum_{q \in 2\mathbb{Z}+1} \hat{f}(\xi + 2^{r-l}(2q\pi)) d\xi \\ &= \int_{\mathbb{R}} \overline{\hat{f}(\xi)} \sum_{s \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{q \in 2\mathbb{Z}+1} \sum_{r=0}^{\infty} \hat{f}(\xi + 2^s(2q\pi)) \hat{\omega}_n(2^r(2^{-s}\xi)) d\xi \\ &= \int_{\mathbb{R}} \overline{\hat{f}(\xi)} \sum_{s \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{q \in 2\mathbb{Z}+1} \hat{f}(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi. \end{aligned}$$

Now, let us collect the results we have obtained about the sum of squares I :

Proposition 2.5.8 Suppose $\omega_n \in L^2(\mathbb{R})$ and $f \in D$, then

$$\begin{aligned} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2 &= \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2 d\xi \\ &+ \frac{1}{2\pi} \int_{\mathbb{R}} \overline{\hat{f}(\xi)} \sum_{s \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{q \in 2\mathbb{Z}+1} \hat{f}(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi. \end{aligned} \quad (2.5.9)$$

The last integrand is integrable and so is the first, when $\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2$ is locally integrable in $\mathbb{R} \setminus \{0\}$.

Proof of Theorem 2.5.2. In view of Eq. (2.5.8), $t_q^n(2^{-s}\xi)$ in the last expression of Eq. (2.5.9) can be written as

$$t_q^n(2^{-s}\xi) = \sum_{l=0}^{\infty} \hat{\omega}_n(2^{l-s}\xi) \overline{\hat{\omega}_n(2^l(2^{-s}\xi + 2q\pi))}, \quad q \in 2\mathbb{Z}+1 \quad s \in \mathbb{Z}.$$

Now, using Eq. (2.5.2), we get

$$t_q^n(2^{-s}\xi) = 0.$$

Hence the Eq. (2.5.9) becomes

$$\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2 = \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 d\xi = \|\hat{f}(\xi)\|_2^2 = \|f(\xi)\|_2^2. \quad (\text{by Plancheral's theorem})$$

Clearly, condition (B) guarantees condition (A) for all $f \in D$. Therefore, by Lemma 1.2.2, shows that (A) is valid for all $f \in L^2(\mathbb{R})$. In order to complete the proof of Theorem 2.5.2, we need to show that if (A) is satisfied for all $f \in D$, then the two Eqs. (2.5.1) and (2.5.2) are satisfied.

If condition (A) is satisfied for all $f \in D$, then, by the Remark 2.5.5

$$\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2$$

is locally integrable in $\mathbb{R} \setminus \{0\}$. Let $\xi_0 \neq 0$ be a Lebesgue point for this function. Then

$$\lim_{\delta \rightarrow 0^+} \frac{1}{2\delta} \int_{\xi_0-\delta}^{\xi_0+\delta} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2 d\xi = \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2$$

where $[\xi_0 - \delta, \xi_0 + \delta] \subset \mathbb{R} \setminus \{0\}$. Let I^δ , I_0^δ , and I_1^δ be the quantities I , I_0 , and I_1 when $f = f_\delta$, where

$$\hat{f}_\delta(\xi) = \frac{1}{\sqrt{2\delta}} \chi_{[\xi_0 - \delta, \xi_0 + \delta]}(\xi).$$

Using Proposition 2.5.8 together with (A), we obtain

$$I^\delta = \|f_\delta\|_2^2 = \frac{1}{2\pi} \|\hat{f}_\delta\|_2^2 = \frac{1}{2\pi} \int_{\xi_0 - \delta}^{\xi_0 + \delta} \frac{1}{2\delta} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2 d\xi + I_1^\delta.$$

Since $\|\hat{f}_\delta\|_2 = 1$, we must have

$$1 = \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi)|^2 + 2\pi \lim_{\delta \rightarrow 0^+} I_1^\delta. \quad (2.5.10)$$

If we can show that the last limit is 0, we then have

$$\sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^l \xi_0)|^2 = 1$$

with $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, at each such Lebesgue point ξ_0 and the equality (2.5.1) must hold. By using the arguments of the proof of Lemma 2.5.4, the estimate I_1^δ is obtained as

$$|I_1^\delta| \leq \int_{\mathbb{R}} \sum_{l \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \neq 0} 2^{-l} \left| \hat{f}_\delta(2^{-l} \xi) \right| \left| \hat{f}_\delta(2^{-l}(\xi + 2k\pi)) \right| |\hat{\omega}_n(\xi)|^2 d\xi. \quad (2.5.11)$$

Now, we apply Lemma 2.5.6 for this choice of f . The diameter of the support of $\hat{f} = \hat{f}_\delta$ is 2δ . Since $k \neq 0$, we have

$$\left| \hat{f}_\delta(2^{-l} \xi) \right| \left| \hat{f}_\delta(2^{-l}(\xi + 2k\pi)) \right| = 0 \text{ if } 2^{-l}(2\pi) > 2\delta.$$

Thus, if l_0 is the least integer such that $2^{l_0}\delta \geq \pi$ we need only to consider $l \geq l_0$ in the sum in (2.5.11). Also, if $\hat{f}_\delta(2^{-l}\xi) \neq 0$ we must have $\xi_0 - \delta < 2^{-l}\xi$ and, without loss of generality, we can assume $0 < \xi_0 - \delta$. Thus, the integral in (2.5.11) can be assumed to be over the region $\{\xi : 2^{l_0}(\xi_0 - \delta) < \xi\} \subseteq \{\xi : \frac{\pi}{\delta}(\xi_0 - \delta) < \xi\}$. Therefore, using the notation of Lemma 2.5.6, we have

$$\begin{aligned} |I_1^\delta| &\leq \int_{\frac{\pi}{\delta}(\xi_0-\delta)}^{\infty} \sum_{n=2^u}^{2^{u+1}-1} \alpha_\delta(\xi) |\hat{\omega}_n(\xi)|^2 d\xi \\ &\leq \frac{\pi}{\delta} \left(1 + \log_2 \frac{\xi_0 + \delta}{\xi_0 - \delta}\right) \frac{1}{2\delta} \sum_{n=2^u}^{2^{u+1}-1} \int_{\frac{\pi}{\delta}(\xi_0-\delta)}^{\infty} |\hat{\omega}_n(\xi)|^2 d\xi \end{aligned}$$

since $\|\hat{f}_\delta\|_{L^\infty}^2 = \frac{1}{2\delta}$. It is now clear that

$$\lim_{\delta \rightarrow 0^+} I_1^\delta = 0.$$

Finally, we show that if (A) holds for all $f \in D$, then Eq. (2.5.2) is true. But it follows from the equality (2.5.9) and just-established fact that (2.5.1) holds, that (A) implies

$$\int_{\mathbb{R}} \sum_{s \in \mathbb{Z}} \sum_{q \in 2\mathbb{Z}+1} \overline{\hat{f}(\xi)} \hat{f}(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi = 0$$

for all $f \in D$. By polarization identity, we have

$$\int_{\mathbb{R}} \sum_{s \in \mathbb{Z}} \sum_{q \in 2\mathbb{Z}+1} \overline{\hat{f}(\xi)} \hat{g}(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi = 0 \quad (2.5.12)$$

for all $f, g \in D$.

Let us fix an odd integer q_0 and let ξ_0 be a Lebesgue point of $t_{q_0}^n$ such that ξ_0 and $\xi_0 + 2q_0\pi$ are not 0. We can also assume δ to be sufficiently small and positive so that neither $[\xi_0 - \delta, \xi_0 + \delta]$ nor $[\xi_0 + 2q_0\pi - \delta, \xi_0 + 2q_0\pi + \delta]$ contains 0. In fact, without loss of generality, we can assume $\xi_0 > 0$ and $0 < \delta < \frac{1}{3}\xi_0$.

Let $f = f_\delta$ be the function satisfying

$$\hat{f}_\delta = \frac{1}{\sqrt{2\delta}} \chi_{[\xi_0-\delta, \xi_0+\delta]}$$

and $g = \hat{g}_\delta$ such that $\hat{g}_\delta(\xi) = \hat{f}_\delta(\xi - 2q_0\pi)$. Then

$$\hat{f}_\delta(\xi) \hat{g}_\delta(\xi + 2q_0\pi) = \frac{1}{2\delta} \chi_{[\xi_0-\delta, \xi_0+\delta]}.$$

Equality (2.5.12) then can be written as

$$\begin{aligned} 0 &= \frac{1}{2\delta} \int_{\xi_0-\delta}^{\xi_0+\delta} t_{q_0}^n(\xi) d\xi + \sum_{s \in \mathbb{Z}} \sum_{\substack{q \in 2\mathbb{Z}+1 \\ (s,q) \neq (0,q_0)}} \int_{\mathbb{R}} \overline{\hat{f}_\delta(\xi)} \\ &\quad \times \hat{g}_\delta(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi \end{aligned}$$

$$= \frac{1}{2\delta} \int_{\xi_0-\delta}^{\xi_0+\delta} t_{q_0}^n(\xi) + J_\delta.$$

But the first summand tends to $t_{q_0}(\xi_0)$ as $\delta = 0^+$, and it is sufficient to prove that

$$\lim_{\delta \rightarrow 0^+} J_\delta = 0.$$

If $\overline{\hat{f}_\delta(\xi)} \hat{g}_\delta(\xi + 2^s(2q\pi)) \neq 0$, we must have

$$|\xi - \xi_0| \leq \delta \text{ and } |\xi - 2q_0\pi + 2^s(2q\pi) - \xi_0| \leq \delta.$$

Thus

$$|q_0 - 2^s q| \leq \frac{1}{2\pi} |\xi - 2q_0\pi + 2^s(2q\pi) - \xi_0| + |\xi - \xi_0| \leq \frac{\delta}{\pi}. \quad (2.5.13)$$

Since we are interested in the behavior of these quantities $\delta \rightarrow 0^+$ as we can assume that $\delta < \pi$. Now, if $s > 0$, since q_0 is odd, $|q_0 - 2^s q|$ must be odd and therefore cannot be less than 1. If $s = 0$, in which case $q \neq q_0$ we cannot have $|q_0 - 2^s q| = |q_0 - q| < 1$. Finally, if $s < 0$, then $|q_0 - 2^s q| = 2^s |2^{-s} q_0 - q| \geq 2^s$ since q is odd. Inequality (2.5.13), therefore, implies $2^s \leq \frac{\delta}{\pi}$. This shows that if $\delta < \pi$ and l_0 is the largest integer satisfying $2^{l_0} \leq \frac{\delta}{\pi}$,

$$\begin{aligned} J_\delta &= \sum_{s \leq l_0} \sum_{q \in 2\mathbb{Z}+1} \int_{\mathbb{R}} \overline{\hat{f}_\delta(\xi)} \hat{g}_\delta(\xi + 2^s(2q\pi)) t_q^n(2^{-s}\xi) d\xi \\ &= \sum_{s \leq l_0} \sum_{q \in 2\mathbb{Z}+1} 2^s \int_{\mathbb{R}} \overline{\hat{f}_\delta(2^s \xi)} \hat{g}_\delta(2^s(\xi + 2q\pi)) t_q^n(\xi) d\xi. \end{aligned}$$

Since

$$2|t_q^n(\xi)| \leq \sum_{l=0}^{\infty} |\hat{\omega}_n(2^l \xi)|^2 + \sum_{l=0}^{\infty} |\hat{\omega}_n(2^l(\xi + 2q\pi))|^2$$

we have

$$|J_\delta| \leq J_\delta^{(1)} + J_\delta^{(2)}$$

where

$$J_\delta^{(1)} = \sum_{s \leq l_0} \sum_{q \in 2\mathbb{Z}+1} 2^s \int_{\mathbb{R}} |\hat{f}_\delta(2^s \xi)| |\hat{g}_\delta(2^s(\xi + 2q\pi))| [\tau(\xi)]^2 d\xi$$

with

$$\int_{\mathbb{R}} [\tau(\xi)]^2 d\xi = \sum_{l=0}^{\infty} \int_{\mathbb{R}} |\hat{\omega}_n(2^l \xi)|^2 d\xi = 2 \|\hat{\omega}_n\|_2^2 < \infty$$

and

$$\begin{aligned} J_{\delta}^{(2)} &= \sum_{s \leq l_0} \sum_{q \in 2\mathbb{Z}+1} 2^s \int_{\mathbb{R}} |\hat{f}_{\delta}(2^s \xi)| |\hat{g}_{\delta}(2^s(\xi + 2q\pi))| [\tau(\xi + 2q\pi)]^2 d\xi \\ &= \sum_{s \leq l_0} \sum_{q \in 2\mathbb{Z}+1} 2^s \int_{\mathbb{R}} |\hat{f}_{\delta}(2^s(\eta - 2q\pi))| |\hat{g}_{\delta}(2^s\eta)| [\tau(\eta)]^2 d\eta. \end{aligned}$$

Thus, $J_{\delta}^{(2)}$ has the same form as $J_{\delta}^{(1)}$ with the roles of \hat{f}_{δ} and \hat{g}_{δ} interchanged. Since

$$\hat{f}_{\delta} = \frac{1}{\sqrt{2\delta}} \chi_{[\xi_0 - \delta, \xi_0 + \delta]}$$

we deduce that

$$J_{\delta}^{(1)} = \sum_{s \leq l_0} \sum_{q \in 2\mathbb{Z}+1} \frac{2^s}{\sqrt{2\delta}} \int_{2^{-s}(\xi_0 - \delta)}^{2^{-s}(\xi_0 + \delta)} |\hat{g}(2^s(\xi + 2q\pi))| [\tau(\xi)]^2 d\xi. \quad (2.5.14)$$

Fix s , for $\hat{g}(2^s(\xi + 2q\pi))$ not to be 0, we must have

$$2^s(\xi + 2q\pi) \in [\xi_0 + 2q_0\pi - \delta, \xi_0 + 2q_0\pi + \delta].$$

Since $2^s \xi \in [\xi_0 - \delta, \xi_0 + \delta]$ a simple calculation shows that

$$-2^{-s} \frac{\delta}{\pi} \leq q - 2^{-s} q_0 \leq 2^{-s} \frac{\delta}{\pi} \quad (2.5.15)$$

if $\hat{g}(2^s(\xi + 2q\pi)) \neq 0$. But q is odd and $2^{-s} q_0$ is even. Thus $q - 2^{-s} q_0$ is odd. We claim that there are at most $2^{-s} \frac{(2\delta)}{\pi}$ odd integers in the interval $[-2^{-s} \frac{\delta}{\pi}, 2^{-s} \frac{\delta}{\pi}]$ and this gives us an upper bound for the number of integers of the form $q - 2^{-s} q_0$ that satisfy (2.5.15).

Thus by Eq. (2.5.14), we have

$$\begin{aligned} J_{\delta}^{(1)} &= \sum_{s \leq l_0} \frac{2^s}{\sqrt{2\delta}} \int_{2^{-s}(\xi_0 - \delta)}^{2^{-s}(\xi_0 + \delta)} [\tau(\xi)]^2 \sum_{q \in 2\mathbb{Z}+1} |\hat{g}_{\delta}(2^s(\xi + 2q\pi))| d\xi \\ &= \sum_{s \leq l_0} \frac{2^s}{\sqrt{2\delta}} \int_{2^{-s}(\xi_0 - \delta)}^{2^{-s}(\xi_0 + \delta)} 2^{-s} \frac{2\delta}{\pi} \frac{1}{\sqrt{2\delta}} [\tau(\xi)]^2 d\xi \\ &= \frac{1}{\pi} \sum_{s \leq l_0} \int_{2^{-s}(\xi_0 - \delta)}^{2^{-s}(\xi_0 + \delta)} [\tau(\xi)]^2 d\xi. \end{aligned}$$

Recall that we are assuming $0 < \delta < \frac{1}{3}\xi_0$. This implies that the intervals $[-2^{-s}(\xi_0 - \delta), 2^{-s}(\xi_0 + \delta)]$, $s = l_0, l_0 - 1, l_0 - 2, \dots$, are disjoint. Therefore,

$$\begin{aligned}
J_\delta^{(1)} &\leq \frac{1}{\pi} \sum_{s \leq l_0} \int_{2^{-s}(\xi_0 - \delta)}^{2^{-s}(\xi_0 + \delta)} [\tau(\xi)]^2 d\xi \\
&\leq \frac{1}{\pi} \int_{2^{-s}(\xi_0 - \delta)}^{2^{-s}(\xi_0 + \delta)} [\tau(\xi)]^2 d\xi \\
&\leq \frac{1}{\pi} \int_{\frac{\pi}{\delta}(\xi_0 - \delta)}^{\infty} [\tau(\xi)]^2 d\xi.
\end{aligned}$$

The last expression tends to 0 as $\delta \rightarrow 0^+$ since $\tau^2 \in L^2(\mathbb{R})$. In the case $\xi_0 < 0$ this argument can be modified in an obvious way to obtain this same result

$$\lim_{\delta \rightarrow 0^+} J_\delta^{(1)} = 0.$$

The case involving $J_\delta^{(2)}$ is similar, with the roles of \hat{f}_δ and \hat{g}_δ interchanged and the point $\xi_0 + 2q_0\pi$ replacing ξ_0 in the above argument. This completes the proof of Theorem 2.5.2.

2.6 Characterizations of MRA Wavelet Packets

The dimension function of wavelet packets $\omega_n \in L^2(\mathbb{R})$ associated with a dilation by 2 is the function D_{ω_β} given by

$$D_{\omega_\beta}(\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \quad (2.6.1)$$

where $\beta = u + 1$ and $l = j - u$, $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}^+$.

Theorem 2.6.1 *The functions $\omega_n \in L^2(\mathbb{R})$ are MRA wavelet packets if and only if $D_{\omega_\beta}(\xi) = 1$ for almost every $\xi \in \mathbb{R}$.*

Proof From Theorem 2.3.2, we have

$$\sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 = 1 \Rightarrow D_{\omega_\beta}(\xi) = 1, \text{ a.e.}$$

where ω_n are MRA wavelet packets. Now, to complete the proof of Theorem 2.6.1 it is sufficient to show that $D_{\omega_\beta}(\xi) = 1$ a.e., so that the wavelet packets are MRA wavelet packets. We shall break up the proof of this into several lemmas.

Lemma 2.6.2 *If ω_n are orthonormal wavelet packets, then*

$$\hat{\omega}_t(2^m\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^m(\xi + 2k\pi)) \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \hat{\omega}_n(2^l\xi) \quad (2.6.2)$$

a.e. for all $m \geq 1$ and $t = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$, where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Proof First of all we show that the series in (2.6.2) is well defined. By using Schwartz's inequality and (2.3.6), we obtain

$$\begin{aligned} & \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \left| \hat{\omega}_t(2^m(\xi + 2k\pi)) \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \right| \\ & \leq \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_t(2^m(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} \sum_{n=2^u}^{2^{u+1}-1} \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} \\ & \leq \left(\sum_{s \in \mathbb{Z}} |\hat{\omega}_t(2^m\xi + 2s\pi)|^2 \right)^{\frac{1}{2}} \left(\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} \\ & = \left(\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Taking the sum over all $l \geq 1$, using Schwartz's inequality and (2.5.1), we obtain

$$\begin{aligned} & \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} \left| \hat{\omega}_t(2^m(\xi + 2k\pi)) \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \hat{\omega}_n(2^l\xi) \right| \\ & \leq \sum_{l=1}^{\infty} \sum_{n=2^u}^{2^{u+1}-1} \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} (|\hat{\omega}_n(2^l\xi)|) \\ & \leq \left(\sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} \left(\sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} |\hat{\omega}_n(2^l\xi)|^2 \right)^{\frac{1}{2}} \\ & \leq \left(\sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} .1 = \sqrt{D_{\omega_\beta}(\xi)}. \end{aligned}$$

But all the above inequalities are true for a.e. $\xi \in \mathbb{R}$. Therefore, the series in (2.6.2) is well defined almost everywhere. Let $G_m(\xi)$ be the RHS of (2.6.2). Then, we show that $G_m(\xi) = \hat{\omega}_t(2^m\xi)$ for a.e. $\xi \in \mathbb{R}$. First of all we show that $G_m(\xi) = G_{m-1}(2\xi)$, and, then, that $G_1(\xi) = \hat{\omega}_t(2\xi)$. Clearly, this gives (2.6.2).

Replacing l by m in (2.5.4) and then using it, we get

$$\begin{aligned} G_m(\xi) &= \sum_{\substack{k \in \mathbb{Z} \\ 2^{u+1}-1}} \hat{\omega}_t(2^m(\xi + 2k\pi)) \sum_{n=2^u}^{\infty} \sum_{l=1}^{\infty} \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \hat{\omega}_n(2^l\xi) \\ &= \sum_{n=2^u}^{\infty} \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^m(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} \hat{\omega}_n(\xi) \\ &\quad + \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^m(\xi + 2k\pi)) \sum_{n=2^u}^{\infty} \sum_{l=1}^{\infty} \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \hat{\omega}_n(2^l\xi) \\ &= \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^m(\xi + 2k\pi)) \sum_{n=2^u}^{\infty} \sum_{l=0}^{\infty} \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \hat{\omega}_n(2^l\xi). \end{aligned}$$

By (2.5.2), the terms in the summation over l where k is odd are zero a.e. Therefore, on replacing k by $2s$, we get

$$\begin{aligned} G_m(\xi) &= \sum_{s \in \mathbb{Z}} \hat{\omega}_t(2^m(\xi + 4s\pi)) \sum_{n=2^u}^{\infty} \sum_{l=0}^{\infty} \overline{\hat{\omega}_n(2^l(\xi + 4s\pi))} \hat{\omega}_n(2^l\xi) \\ &= \sum_{s \in \mathbb{Z}} \hat{\omega}_t\left(2^{m+1}\left(\frac{\xi}{2} + 2s\pi\right)\right) \sum_{n=2^u}^{\infty} \sum_{l=0}^{\infty} \overline{\hat{\omega}_n\left(2^{l+1}\left(\frac{\xi}{2} + 2s\pi\right)\right)} \\ &\quad \times \hat{\omega}_n\left(2^{l+1}\frac{\xi}{2}\right) \\ &= \sum_{s \in \mathbb{Z}} \hat{\omega}_t\left(2^{m+1}\left(\frac{\xi}{2} + 2s\pi\right)\right) \sum_{n=2^u}^{\infty} \sum_{l=1}^{\infty} \overline{\hat{\omega}_n\left(2^l\left(\frac{\xi}{2} + 2s\pi\right)\right)} \\ &\quad \times \hat{\omega}_n\left(2^l\frac{\xi}{2}\right) \\ &= G_{m+1}\left(\frac{\xi}{2}\right). \end{aligned}$$

This shows that $G_m(\xi) = G_{m-1}(2\xi)$ almost everywhere.

We, now, calculate $G_1(\xi)$. Changing variables in the sum over l , we obtain

$$\begin{aligned} G_1(\xi) &= \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2(\xi + 2k\pi)) \sum_{n=2^u}^{\infty} \sum_{l=1}^{\infty} \overline{\hat{\omega}_n(2^l(\xi + 2k\pi))} \hat{\omega}_n(2^l\xi) \\ &= \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2\xi + 4k\pi) \sum_{n=2^u}^{\infty} \sum_{l=0}^{\infty} \overline{\hat{\omega}_n(2^l(2\xi + 4k\pi))} \hat{\omega}_n(2^l2\xi). \end{aligned}$$

Further, in the last sum over k we add all the corresponding terms with $2k$ replaced by $2k + 1$, which are zero by (2.5.2). This gives us

$$G_1(\xi) = \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2\xi + 2k\pi) \sum_{n=2^u}^{\infty} \sum_{l=0}^{\infty} \overline{\hat{\omega}_n(2^l(2\xi + 2k\pi))} \hat{\omega}_n(2^l2\xi).$$

Interchanging the order of summation, using (2.5.4) when $l \geq 1$ and (2.5.3) when $l = 0$, we obtain $G_1(\xi) = \hat{\omega}_t(2\xi)$. This completes the proof of Lemma 2.6.2.

Now, consider $l^2(\mathbb{Z})$, and denote its (usual) norm by $\|\cdot\|_{l^2}$ and the inner product by $\langle \cdot, \cdot \rangle_{l^2}$. If ω_n are orthonormal wavelet packets, we define the vector

$$\Psi_{n,l}(\xi) = \left\{ \hat{\omega}_n(2^l(\xi + 2k\pi)) : n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1; l = j - u; j, k \in \mathbb{Z}; \right\}, l \geq 1.$$

But (2.5.3) implies that

$$\begin{aligned} \|\Psi_{n,l}(\xi)\|_{l^2} &= \left(\sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{s \in \mathbb{Z}} |\hat{\omega}_n(2^l\xi + 2s\pi)|^2 \right)^{\frac{1}{2}} \\ &= 1 \text{ for a.e. } \xi \in \mathbb{R}. \end{aligned}$$

Therefore, for almost every ξ the vector $\Psi_{n,l}(\xi) \in l^2(\mathbb{Z})$.

Let $\mathbb{F}_{\omega_\beta}(\xi)$ be the closure of the span of the set of vectors $\{\Psi_{n,l}(\xi) : l \geq 1; n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1\}$. Then, $\mathbb{F}_{\omega_\beta}(\xi)$ is a well-defined subspace of $l^2(\mathbb{Z})$ for almost every $\xi \in \mathbb{R}$. We can rewrite (2.6.2) in terms of the above notation as

$$\hat{\omega}_t(2^m\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \langle \Psi_{t,m}(\xi), \Psi_{n,l}(\xi) \rangle_{l^2} \hat{\omega}_n(2^l\xi) \text{ for a.e. } \xi \in \mathbb{R}$$

for all $t = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$. In the above replacing ξ by $\xi + 2s\pi$ we obtain, for $n \geq 1$,

$$\hat{\omega}_t(2^m(\xi + 2s\pi)) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \langle \Psi_{t,m}(\xi), \Psi_{n,l}(\xi) \rangle_{l^2} \hat{\omega}_n(2^l(\xi + 2s\pi)) \text{ a.e.}$$

But $\langle \Psi_{t,m}(\xi), \Psi_{n,l}(\xi) \rangle_{l^2}$ is 2π -periodic. Therefore, we can write this equality vectorially as

$$\Psi_{t,m}(\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \langle \Psi_{t,m}(\xi), \Psi_{n,l}(\xi) \rangle_{l^2} \Psi_{n,l}(\xi). \quad (2.6.3)$$

Further, since $D_{\omega_\beta}(\xi)$ is finite a.e., simple calculation shows that

$$D_{\omega_\beta}(\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \|\Psi_{n,l}(\xi)\|_{l^2}^2. \quad (2.6.4)$$

Let S be the subset of \mathbb{T} on which $D_{\omega_\beta}(\xi) < \infty$. Then, the vectors $\Psi_{n,l}(\xi)$, $l \geq 1$, are well defined on S (observe that $|S| = 2\pi$). For $\xi \in S$, let $\mathbb{F}_{\omega_\beta}(\xi)$ be closure, in $l^2(\mathbb{Z})$, of the span of $\{\Psi_{n,l}(\xi) : l \geq 1; n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1\}$. Then, the hypothesis of Lemma 1.2.3 are satisfied if $v_l = \Psi_{n,l}(\xi)$. This gives us

$$\dim \mathbb{F}_{\omega_\beta}(\xi) = D_{\omega_\beta}(\xi) \text{ on } S. \quad (2.6.5)$$

Now, we are ready to prove the sufficient part of Theorem 2.6.1. Let $\omega_n \in L^2(\mathbb{R})$ be wavelet packets for which $D_{\omega_\beta}(\xi) = 1$ for a.e. $\xi \in \mathbb{R}$. Then, by (2.6.5), $\dim \mathbb{F}_{\omega_\beta}(\xi) = 1$ for a.e. $\xi \in \mathbb{T}$. This shows that, for each $\xi \in S$, $\mathbb{F}_{\omega_\beta}(\xi)$ is generated by a single unit vector $U(\xi)$. We now choose a particular one. For $l \geq 1$, let

$$E_l = \{\xi \in S : \Psi_{n,l}(\xi) \neq \mathbf{0} \text{ and } \Psi_{n,r}(\xi) = \mathbf{0} \text{ for all } r < l\}.$$

The sets E_l , for $l \geq 1$, are mutually disjoint and together with

$$E_0 = \{\xi \in \mathbb{T} : D_{\omega_\beta}(\xi) = 0\}$$

form a partition of S . Hence, for $\xi \in S \setminus E_0$, there exists a unique $l \geq 1$ such that $\xi \in E_l$. But E_0 has measure 0. Therefore

$$U(\xi) = \frac{1}{\|\Psi_{t,l}(\xi)\|_{l^2}} \Psi_{t,l}(\xi), \quad \xi \in E_l \text{ for some } l \geq 1$$

is well defined and $\|U(\xi)\|_{l^2} = 1$ for almost every $\xi \in \mathbb{T}$. Write

$$U(\xi) = \{u_k : k \in \mathbb{Z}\}.$$

If we find the scaling function $\varphi = \omega_0$, in view of Lemma 1.6.3, we hope that $u_k(\xi) = \hat{\omega}_0(\xi + 2k\pi)$. Thus, we let

$$\hat{\omega}_0(\xi) = u_k(\xi - 2k\pi) \text{ if } \xi \in \mathbb{T} + 2k\pi \text{ for some } k \in \mathbb{Z}.$$

This defines $\hat{\omega}_0$ on \mathbb{R} . We claim that $\hat{\omega}_0 \in L^2(\mathbb{R})$:

$$\|\hat{\omega}_0\|_2^2 = \sum_{k \in \mathbb{Z}} \int_{\mathbb{T}} |\hat{\omega}_0(\xi + 2k\pi)|^2 d\xi = \sum_{k \in \mathbb{Z}} \int_{\mathbb{T}} |u_k(\xi)|^2 d\xi = \int_{\mathbb{T}} \|U(\xi)\|_{l^2}^2 d\xi = 2\pi$$

since $U(\xi)$ is a unit vector. We also have

$$\sum_{k \in \mathbb{Z}} |\hat{\omega}_0(\xi + 2k\pi)|^2 = \sum_{k \in \mathbb{Z}} |u_k(\xi)|^2 = \|U(\xi)\|_{l^2}^2 = 1 \text{ for a.e. } \xi \in \mathbb{R} \quad (2.6.6)$$

which is equivalent to the fact that $\{\omega_0(\cdot - k) : k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(\mathbb{R})$. Define $V_0^\#$ as the closed subspace of $L^2(\mathbb{R})$ generated by $\{\varphi(\cdot - k) : k \in \mathbb{Z}\}$. We claim that

$$V_0^\# = V_0 = \bigoplus_{j < 0} W_j. \quad (2.6.7)$$

From this, it follows that $\{V_j : j \in \mathbb{Z}\}$ is the desired MRA.

For each $l \geq 1$, there exists a measurable function $v_{t,l}$, defined on \mathbb{T} , such that

$$\Psi_{t,l}(\xi) = v_{t,l}(\xi)U(\xi) \text{ for a.e. } \xi \in \mathbb{T}.$$

Componentwise,

$$\hat{\omega}_t(2^l(\xi + 2k\pi)) = v_{t,l}(\xi)\hat{\omega}_0(\xi + 2k\pi) \text{ for a.e. } \xi \in \mathbb{T}, \quad k \in \mathbb{Z}.$$

Hence, by (2.6.6), for a.e. $\xi \in \mathbb{T}$,

$$\sum_{k \in \mathbb{Z}} |\hat{\omega}_t(2^l(\xi + 2k\pi))|^2 = \sum_{k \in \mathbb{Z}} |v_{t,l}(\xi)|^2 |\hat{\omega}_0(\xi + 2k\pi)|^2 = |v_{t,l}(\xi)|^2, \quad (2.6.8)$$

which shows that $v_{t,l} \in L^2(\mathbb{T})$ with $\|v_{t,l}\|_{L^2(\mathbb{T})}^2 = 2^{-l}(2\pi)$. Write the Fourier series of $v_{t,l}$, $l \geq 1$, as

$$v_{t,l}(\xi) = \sum_{k \in \mathbb{Z}} a_k^{t,l} e^{-ik\xi} \text{ for a.e. } \xi \in \mathbb{T}$$

with convergence in the $L^2(\mathbb{T})$ -norm, and $\{a_k^{t,l}\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$. Extending $v_{t,l}$, 2π -periodically, we obtain

$$\hat{\omega}_t(2^l\xi) = v_{t,l}(\xi)\hat{\omega}_0(\xi) \text{ for a.e. } \xi \in \mathbb{R}, \quad l \geq 1. \quad (2.6.9)$$

Taking inverse Fourier transform on both sides, we obtain

$$\omega_{-l,t,0}(x) = 2^{-l/2}\omega_t(2^{-l}x) = 2^{l/2} \sum_{k \in \mathbb{Z}} a_k^{t,l} \omega_0(x - k), \quad l \geq 1.$$

Hence, $\omega_{-l,t,0} \in V_0^\#$ for $l \geq 1$. Since $V_0^\#$ is invariant under integral translations and $\omega_{-l,t,k}(x) = 2^{-l/2}\omega_t(2^{-l}(x - 2^l k))$, we have $\omega_{-l,t,k} \in V_0^\#$ for all $t = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$, $k \in \mathbb{Z}$ and $l \geq 1$. Thus, $W_{-l} \subseteq V_0^\#$ for all $l \geq 1$, and hence, $V_0 \subseteq V_0^\#$.

Now, we need to show that $V_0^\# \subseteq V_0$. We do this by showing that ω_0 is perpendicular to W_j for all $j \geq 0$. For $j \geq 0$ and $s \in \mathbb{Z}$, the Plancherel theorem, a change of variables and a periodization argument allow us to write

$$\begin{aligned}
2\pi \sum_{n=2^u}^{2^{u+1}-1} \langle \omega_0, \omega_{l,n,s} \rangle &= \sum_{n=2^u}^{2^{u+1}-1} \langle \hat{\omega}_0, (\omega_{l,n,s})^\wedge \rangle \\
&= \sum_{n=2^u}^{2^{u+1}-1} 2^{-\frac{l}{2}} \int_{\mathbb{R}} \hat{\omega}_0(\xi) \overline{\hat{\omega}_n(2^{-l}\xi)} e^{i2^{-l}s\xi} d\xi \\
&= \sum_{n=2^u}^{2^{u+1}-1} 2^{\frac{l}{2}} \int_{\mathbb{R}} \hat{\omega}_0(2^l\xi) \overline{\hat{\omega}_n(\xi)} e^{is\xi} d\xi \\
&= 2^{\frac{l}{2}} \int_{\mathbb{T}} \left(\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \hat{\omega}_0(2^l(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} \right) \\
&\quad \times e^{is\xi} d\xi. \tag{2.6.10}
\end{aligned}$$

The convergence of the last series in $L^2(\mathbb{T})$ is guaranteed by the fact that $\omega_n \in L^2(\mathbb{R})$, $n = 0, 1, 2, \dots$. From (2.6.8) and our assumption $D_{\omega_\beta}(\xi) = 1$ a.e., we obtain

$$\begin{aligned}
\sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} |\nu_{n,l}(\xi)|^2 &= \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} \sum_{k \in \mathbb{Z}} |\hat{\omega}_n(2^l(\xi + 2k\pi))|^2 \\
&= 1 \text{ for a.e. } \xi \in \mathbb{R}.
\end{aligned}$$

Hence, for such ξ and for each $l \geq 0$, there exists $l_0 \equiv l_0(2^l\xi) \geq 1$ such that $\nu_{t,l_0}(2^l\xi) \neq 0$. This and (2.6.9) imply, for such ξ ,

$$\hat{\omega}_0(2^l(\xi + 2k\pi)) = \frac{1}{\nu_{t,l_0}(2^l\xi)} \hat{\omega}_t(2^{l+l_0}(\xi + 2k\pi)), \quad k \in \mathbb{Z}.$$

We now use (2.5.4) to obtain (observe that $l + l_0 \geq 1$)

$$\begin{aligned}
\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \hat{\omega}_0(2^l(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} \\
= \frac{1}{\nu_{t,l_0}(2^l\xi)} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \hat{\omega}_t(2^{l+l_0}(\xi + 2k\pi)) \overline{\hat{\omega}_n(\xi + 2k\pi)} = 0,
\end{aligned}$$

for a.e. $\xi \in \mathbb{T}$ and for all $t = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$, $l \geq 0$. Therefore, from this result and (2.6.10), we obtain

$$\langle \omega_0, \omega_{l,n,s} \rangle = 0 \text{ for all } n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1, s \in \mathbb{Z} \text{ and } l \geq 0.$$

This shows that ω_0 is orthogonal to W_j which is invariant under integral translation, and we deduce that $V_0^\# \perp W_j$ for all $j \geq 0$. Hence $V_0^\# \subseteq V_0$, and the proof of the Theorem 2.6.1 is finished.

Theorem 2.6.3 *For orthonormal wavelet packets $\omega_n \in L^2(\mathbb{R})$, the following statements are equivalent:*

- (1) ω_n are MRA wavelet packets;
- (2) $D_{\omega_\beta}(\xi) = 1$ for a.e. $\xi \in \mathbb{T}$;
- (3) $D_{\omega_\beta}(\xi) > 0$ for a.e. $\xi \in \mathbb{T}$;
- (4) $\dim \mathbb{F}_{\omega_\beta}(\xi) = 1$ for a.e. $\xi \in \mathbb{T}$,

where $\mathbb{F}_{\omega_\beta}(\xi)$ is the closure, in $l^2(\mathbb{Z})$, of the span of $\{\Psi_{n,l}(\xi) : l \geq 1\}$ and $\Psi_{n,l}(\xi)$ is the vector $\{\hat{\omega}_n(2^l(\xi + 2k\pi)) : k \in \mathbb{Z}; n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1; l = j - u\}$.

Proof The equivalence between (1), (2), and (4) has already been proved (see Theorem 2.6.1 and Equality (2.6.5)). But (2) implies (3), it is sufficient to prove that (3) implies (2). If $D_{\omega_\beta}(\xi) > 0$ for a.e. $\xi \in \mathbb{T}$, the fact that $D_{\omega_\beta}(\xi)$ is an integer a.e. implies that $D_{\omega_\beta}(\xi) \geq 1$ almost everywhere. But this and the equality $\int_{\mathbb{T}} D_{\omega_\beta}(\xi) d\xi = 2\pi$ clearly imply (2).

Corollary 2.6.4 *If ω_n are orthonormal wavelet packets such that $|\hat{\omega}_n|$ is continuous and $|\hat{\omega}_n(\xi)| = O(|\xi|^{-\frac{1}{2}-\alpha})$ at ∞ for some $\alpha > 0$, then ω_n are MRA wavelet packets.*

Proof The behavior at infinity of $|\hat{\omega}_n|$ tells us that the series

$$s(\xi) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l=1}^{\infty} |\hat{\omega}_n(2^l \xi)|^2$$

where $l = j - u$, $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}^+$, converges uniformly on compact subsets of $\mathbb{R} \setminus \{0\}$. Moreover, an easy calculation shows that $s(\xi) = O(|\xi|^{-1-2\alpha})$ at ∞ . It follows that

$$\sum_{k \in \mathbb{Z}} s(\xi + 2k\pi) = D_{\omega_\beta}(\xi)$$

converges uniformly on compact subsets of \mathbb{T} . Thus, $D_{\omega_\beta}(\xi)$ is continuous on $(0, 2\pi)$. Since $D_{\omega_\beta}(\xi)$ is integer-valued and $\int_{\mathbb{T}} D_{\omega_\beta}(\xi) d\xi = 2\pi$ (see Theorem 2.3.3), we must have $D_{\omega_\beta}(\xi) = 1$ a.e. on \mathbb{T} .

Corollary 2.6.5 *If ω_n are band-limited wavelet packets such that $|\hat{\omega}_n|$ are continuous, then ω_n are MRA wavelet packets*

Chapter 3

Convergence of Wavelet Packet Series



3.1 Introduction

This chapter deals with the pointwise convergence of wavelet packet series, convolution bounds, and convergence of wavelet packet series.

3.2 Pointwise Convergence of Wavelet Packet Series

In light of Eq. (2.2.12), Q_j can be expressed as

$$Q_j(x, y) = \sum_{n=2^u}^{2^{u+1}-1} Q_l^n(x, y) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{j, k \in \mathbb{Z}} \omega_{l, n, k}(x) \overline{\omega_{l, n, k}(y)} \quad (3.2.1)$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$.

Thus (2.3.3) gives

$$\begin{aligned} P_j(x, y) &= \sum_{m < j} Q_m(x, y) = \sum_{m < j} \sum_{n=2^u}^{2^{u+1}-1} Q_l^n(x, y) \\ &= \sum_{m < j} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \omega_{l, n, k}(x) \overline{\omega_{l, n, k}(y)} \end{aligned} \quad (3.2.2)$$

where $l = m - u$, $u = 0$ if $m < 0$ and $u = 0, 1, 2, \dots, m$ if $0 \leq m < j$.

Now, we consider a natural operator $S_{j, k}^\sigma$ associated with wavelet packets ω_n with kernel

$$S_{j,k}^\sigma(x, y) = \sum_{m < j} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \omega_{l,n,k}(x) \overline{\omega_{l,n,k}(y)} + R_{j,k}^\sigma(x, y), \quad (3.2.3)$$

where $l = m - u$, $u = 0$ if $m < 0$ and $u = 0, 1, 2, \dots, m$ if $0 \leq m < j$ and

$$R_{j,k}^\sigma(x, y) = \sum_{m=1}^k \sum_{n=2^u}^{2^{u+1}-1} \omega_{j,n,\sigma(m)}(x) \overline{\omega_{j,n,\sigma(m)}(y)},$$

with $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$, $j, k \in \mathbb{Z}$ and σ is a permutation of \mathbb{Z} . This operator is a partial sum of wavelet packet expansion of f .

Convergence Results

Let ω_n , $n = 0, 1, 2, \dots$, be a wavelet packet. Associated with the increasing sequence of subspaces $\{V_j : j \in \mathbb{Z}\}$, we have the orthogonal projections of $L^2(\mathbb{R})$ onto V_j given by

$$P_j f(x) = \sum_{m < j} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \langle f, \omega_{l,n,k} \rangle \omega_{l,n,k}(x) \text{ for } f \in L^2(\mathbb{R}) \quad (3.2.4)$$

where $l = m - u$, $u = 0$ if $m < 0$ and $u = 0, 1, 2, \dots, m$ if $0 \leq m < j$.

We can also consider the projections Q_j^n from $L^2(\mathbb{R})$ onto U_j^n given by

$$Q_j^n f(x) = \sum_{k \in \mathbb{Z}} \langle f, \omega_{j,n,k} \rangle \omega_{j,n,k}(x), \text{ for all } f \in L^2(\mathbb{R}). \quad (3.2.5)$$

There is also a natural operator, associated with a wavelet packet ω_n given by

$$(S_{j,k}^\sigma f)(x) = \sum_{m < j} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \langle f, \omega_{l,n,k} \rangle \omega_{l,n,k}(x) + (R_{j,k}^\sigma f)(x), \quad (3.2.6)$$

where $l = m - u$, $u = 0$ if $m < 0$ and $u = 0, 1, 2, \dots, m$ if $0 \leq m < j$ and

$$(R_{j,k}^\sigma f)(x) = \sum_{m=1}^k \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{j,n,\sigma(m)} \rangle \omega_{j,n,\sigma(m)}(x),$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$, $f \in L^2(\mathbb{R})$ and σ is any permutation of \mathbb{Z} . This operator has a partial sum of wavelet packet expansion of f . Since

$$V_j = U_j^0 = \bigoplus_{m < j} \bigoplus_{n=2^u}^{2^{u+1}-1} U_l^n$$

where $l = m - u$, $u = 0$ if $m < 0$ and $u = 0, 1, 2, \dots, m$ if $0 \leq m < j$, we have

$$(S_{j,k}^\sigma f)(x) = P_j f(x) + \sum_{m=1}^k \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{l,n,\sigma(m)} \rangle \omega_{l,n,\sigma(m)}(x) \quad (3.2.7)$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$.

The above definition makes sense for any $f \in L^2(\mathbb{R})$. Writing $\langle f, \omega_{l,m,n} \rangle$ as an integral and interchanging (formally) the order of summation and integration, we obtain

$$(S_{j,k}^\sigma f)(x) = \int_{-\infty}^{\infty} 2^j P_0(2^j x, 2^j y) f(y) dy + \int_{-\infty}^{\infty} 2^j Q_0^{\sigma,k}(2^j x, 2^j y) f(y) dy \quad (3.2.8)$$

where

$$Q_0^{\sigma,k}(x, y) = \sum_{m=1}^k \omega_n(x - \sigma(m)) \overline{\omega_n(y - \sigma(m))}, \text{ where } n = 1 \quad (3.2.9)$$

and

$$\begin{aligned} P_0(x, y) &= \sum_{j < 0; k \in \mathbb{Z}} \omega_{j,n,k}(x) \overline{\omega_{j,n,k}(y)}, \text{ where } n = 1 \\ &= \sum_{k \in \mathbb{Z}} \omega_0(x - k) \overline{\omega_0(y - k)}, \text{ as } \omega_0(x) = \varphi(x). \end{aligned} \quad (3.2.10)$$

In order to study the operators P_j and $S_{j,k}^\sigma$ simultaneously, we consider the operator

$$(T_j f)(x) = \int_{\mathbb{R}} 2^j P_0(2^j x, 2^j y) \{f(y) - f(x)\} dy \quad (3.2.11)$$

where $|P_0(x, y)| \leq C H \left(\frac{|x-y|}{2} \right)$ and H satisfies (1.2.1) and C is a constant.

Theorem 3.2.1 *Let $\{T_j : j \in \mathbb{Z}\}$ be the family of operators defined by (3.2.11). If $f \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$, then $\lim_{j \rightarrow \infty} T_j f(x) = 0$ for every x in the Lebesgue set of f .*

Proof From (3.2.11), we have

$$\begin{aligned} |T_j f(x)| &\leq \int_{\mathbb{R}} 2^j |P_0(2^j x, 2^j y)| |f(y) - f(x)| dy \\ &\leq C \int_{\mathbb{R}} 2^j H(2^{j-1}|x-y|) |f(y) - f(x)| dy \\ &= C \int_{\mathbb{R}} 2^j H(2^{j-1}|t|) |f(x-t) - f(x)| dt, \text{ (replacing } y \text{ by } x-t). \end{aligned}$$

If x is a point in the Lebesgue set of f and $\delta > 0$, \exists an $\eta > 0$ such that

$$\frac{1}{r} \int_{|t| \leq r} |f(x-t) - f(x)| dt \leq \delta, \quad 0 < r \leq \eta. \quad (3.2.12)$$

Thus,

$$\begin{aligned} C^{-1}|T_j f(x)| &\leq \int_{|t| < \eta} 2^j H(2^{j-1}|t|) |f(x-t) - f(x)| dt \\ &\quad + \int_{|t| \geq \eta} 2^j H(2^{j-1}|t|) |f(x-t) - f(x)| dt = I + II. \end{aligned}$$

But $H(|x|)$ decreases to zero as $|x| \rightarrow \infty$ and $H \in L^1([0, \infty))$. Therefore,

$$rH(r) \leq \int_{r/2 \leq |x| \leq r} H(|x|) dx \rightarrow 0 \quad \text{as } r \rightarrow \infty.$$

Further, since H is continuous at zero, $rH(r) \rightarrow 0$ as $r \rightarrow 0$. Let $g(r) = |f(x-r) - f(x)|$ and $G(r) = \int_0^r g(s) ds$. Then, by (3.2.12)

$$G(r) \leq r\delta \quad \text{when } 0 < r \leq \eta. \quad (3.2.13)$$

On integrating by parts, we obtain

$$\begin{aligned} I &= \int_{|t| < \eta} 2^j H(2^{j-1}|t|) |f(x-t) - f(x)| dt = 2 \int_0^\eta 2^j H(2^{j-1}r) g(r) dr \\ &= 2G(r) 2^j H(2^{j-1}r) \Big|_0^\eta - 2 \int_0^\eta G(r) 2^j 2^{j-1} H'(2^{j-1}r) dr \\ &\leq 2r\delta 2^j H(2^{j-1}r) \Big|_0^\eta - 2 \int_0^{2^j \eta} G(2^{-j}r) 2^{j-1} H' \left(\frac{1}{2}r \right) dr. \end{aligned}$$

Since H is decreasing, $H' \left(\frac{r}{2} \right)$ is negative and by (3.2.13) and the boundedness of $rH(r)$, we obtain

$$\begin{aligned} I &\leq C\delta - 2\delta \int_0^{2^j \eta} 2^{-j} r 2^{j-1} H' \left(\frac{1}{2}r \right) dr \\ &= C\delta - \delta r 2H \left(\frac{1}{2}r \right) \Big|_0^{2^j \eta} + \delta \int_0^{2^j \eta} 2H \left(\frac{1}{2}r \right) dr \\ &= C\delta - 2\delta A H \left(\frac{1}{2}A \right) + 2\delta \int_0^{A/2} 2H(v) dv, \end{aligned}$$

where $A = 2^j \eta$ and $r = 2\nu$. As $j \rightarrow \infty$ (or, equivalently, $A \rightarrow \infty$), the last expression increases to $4\|H\|_{L^1(\mathbb{R})}$. This gives the estimate

$$I \leq C\delta + 4\|H\|_{L^1(\mathbb{R})}\delta = a\delta. \quad (3.2.14)$$

Clearly, the constant a depends on H only.

In order to estimate II , let χ_η be the characteristic function of the set $\{t \in \mathbb{R} : |t| \geq \eta\}$. Let q denote the conjugate exponent to p , i.e., $\frac{1}{p} + \frac{1}{q} = 1$. On using Hölder's inequality, we get

$$II \leq \|f\|_{L^p(\mathbb{R})} \left(\int_{\mathbb{R}} |\chi_\eta(t) 2^j H(2^{j-1}|t|)|^q dt \right)^{\frac{1}{q}} + |f(x)| \int_{\mathbb{R}} |\chi_\eta(t) 2^j H(2^{j-1}|t|)| dt. \quad (3.2.15)$$

But

$$\int_{\mathbb{R}} |\chi_\eta(t) 2^j H(2^{j-1}|t|)| dt = 2 \int_{|s| \geq 2^{j-1}\eta} H(|s|) ds$$

which tends to zero as $j \rightarrow \infty$. Also, the same is true for the first summand of the above inequality since

$$\begin{aligned} \left(\int_{\mathbb{R}} |\chi_\eta(t) 2^j H(2^{j-1}|t|)|^q dt \right)^{\frac{1}{q}} &= \left(\int_{|t| \geq \eta} |2^j H(2^{j-1}|t|)|^{\frac{q}{p}} |2^j H(2^{j-1}|t|)| dt \right)^{\frac{1}{q}} \\ &\leq \left(\sup_{|t| \geq \eta} |2^j H(2^{j-1}|t|)| \right)^{\frac{1}{p}} \left(\int_{|t| \geq \eta} |2^j H(2^{j-1}|t|)| dt \right)^{\frac{1}{q}} \\ &\leq C (\|H\|_{L^1(0,1)})^{\frac{1}{q}} \left(\sup_{|t| \geq \eta} |2^j H(2^{j-1}|t|)| \right)^{\frac{1}{p}}. \end{aligned}$$

But $rH(r) \rightarrow 0$ as $r \rightarrow \infty$. Therefore,

$$\sup_{|t| \geq \eta} |2^j H(2^{j-1}|t|)| = 2^j H(2^{j-1}\eta)$$

tends to zero as $j \rightarrow \infty$. Hence, choosing j large enough we deduce from (3.2.15) that II can be made smaller than δ . This together with inequality (3.2.14) proves the result.

Theorem 3.2.2 Suppose that ω_n are wavelet packets and ω_n has a radial decreasing L^1 -majorant. If $f \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$, then

$$\lim_{j \rightarrow \infty} P_j f(x) = f(x) \quad \text{for every } x \text{ in the Lebesgue set of } f.$$

In particular, $\lim_{j \rightarrow \infty} P_j f(x) = f(x)$ for almost every $x \in \mathbb{R}$.

Proof It is easy to show that $\int_{\mathbb{R}} P_0(2^j x, 2^j y) dy = 1$. Now

$$P_j f(x) - f(x) = \int_{\mathbb{R}} 2^j P_0(2^j x, 2^j y) \{f(y) - f(x)\} dy = T_j f(x).$$

Now, taking limit $j \rightarrow \infty$ and applying Theorem 3.2.1, we have

$$\lim_{j \rightarrow \infty} P_j f(x) - f(x) = 0 \Rightarrow \lim_{j \rightarrow \infty} P_j f(x) = f(x)$$

for every x in the Lebesgue set of f .

Theorem 3.2.3 Let ω_n , $n \geq 0$ be wavelet packets having the radial decreasing L^1 -majorant. If $f \in L^p(\mathbb{R})$, $1 \leq p \leq \infty$, then, for $k = 1, 2, \dots$

$$\lim_{j \rightarrow \infty} S_{j,k}^\sigma f(x) = f(x) \quad \text{for every } x \text{ in the Lebesgue set of } f.$$

In particular, $\lim_{j \rightarrow \infty} S_{j,k}^\sigma f(x) = f(x)$ for almost every $x \in \mathbb{R}$.

Proof Since

$$\int_{\mathbb{R}} \omega_n(y) dy = \hat{\omega}_n(0) = 0$$

for all $n > 0$, by (3.2.7) we obtain

$$(S_{j,k}^\sigma f)(x) - f(x) = [P_j f(x) - f(x)] + R_{j,k}^\sigma f(x)$$

where

$$R_{j,k}^\sigma f(x) = \int_{\mathbb{R}} \left\{ \sum_{m=1}^k \sum_{n=2^u}^{2^{u+1}-1} \omega_{l,n,\sigma(m)}(x) \overline{\omega_{l,n,\sigma(m)}(y)} \right\} \{f(y) - f(x)\} dy,$$

with $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$.

Thus

$$R_{j,k}^\sigma f(x) = \int_{\mathbb{R}} 2^j Q_0^{\sigma,k}(2^j x, 2^j y) \{f(y) - f(x)\} dy.$$

Hence,

$$(S_{j,k}^\sigma f)(x) - f(x) = [P_j f(x) - f(x)] + \int_{\mathbb{R}} 2^j Q_0^{\sigma,k}(2^j x, 2^j y) \{f(y) - f(x)\} dy.$$

But by Theorem 3.2.2, $\lim_{j \rightarrow \infty} (P_j f(x) - f(x)) = 0$. Therefore, by Lemma 1.2.5

$$\left| Q_0^{\sigma, k}(x, y) \right| \leq CH \left(\frac{1}{2} |x - y| \right)$$

(independent of σ and k), and hence, the proof of the theorem follows from Theorem 3.2.1.

3.3 Convolution Bounds and Convergence of Wavelet Packet Series

This section deals with convolution bounds of the kernel $Q_j^n(x, y)$ and convergence of wavelet packet series.

Convolution Bounds of the Kernel $Q_j^n(x, y)$

Theorem 3.3.1 *If $\omega_n(x)$, $n \in \mathbb{Z}$ are wavelet packets constructed with Daubechies filters, then $|\omega_n(x)| \leq H(|x|)$ where H is nonincreasing and belongs to $L^1[0, \infty)$.*

Proof A radial majorant $H(|x|)$ of any function φ is defined as

$$H(|x|) = \text{ess.supp}|\varphi(t)|, \quad \text{where } |t| \geq |x|, \text{ and belongs to } L^1[0, \infty). \quad (3.3.1)$$

From Eq. (2.2.8), we have

$$\omega_1(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k)$$

so that

$$\begin{aligned} |\omega_1(x)| &\leq \sqrt{2} \sum_{k \in \mathbb{Z}} |g_k| |\varphi(2x - k)| \\ &= \sqrt{2} \sum_{k < 0} |g_k| |\varphi(2x - k)| + \sqrt{2} \sum_{k \geq 0} |g_k| |\varphi(2x - k)| \\ &\leq \sqrt{2} \sum_{k < 0} |g_k| H(|2x|) + \sqrt{2} \sum_{k \geq 0} |g_k| H(|2x - k|). \end{aligned}$$

Let $H_1(|x|) = \max\{H(|2x|), H(|2x - k|)\}$, $\forall k \in \mathbb{Z}$. Then

$$\begin{aligned} |\omega_1(x)| &\leq \sqrt{2} \sum_{k < 0} |g_k| H_1(|x|) + \sqrt{2} \sum_{k \geq 0} |g_k| H_1(|x|) \\ &= \sqrt{2} \sum_{k \in \mathbb{Z}} |g_k| H_1(|x|). \end{aligned}$$

But $\sum_{k \in \mathbb{Z}} |g_k| = \sqrt{2}$. Therefore,

$$|\omega_1(x)| \leq 2H_1(|x|) = H'_1(|x|). \quad (3.3.2)$$

Further

$$\begin{aligned} |\omega_2(x)| &\leq \sqrt{2} \sum_{k \in \mathbb{Z}} |h_k| |\omega_1(2x - k)| \\ &= \sqrt{2} \sum_{k < 0} |h_k| |\omega_1(2x - k)| + \sqrt{2} \sum_{k \geq 0} |h_k| |\omega_1(2x - k)| \\ &\leq \sqrt{2} \sum_{k < 0} |h_k| 2H_1(|2x|) + \sqrt{2} \sum_{k \geq 0} |h_k| 2H_1(|2x - k|) \end{aligned}$$

Let $H_2(|x|) = \max\{H_1(|2x|), H_1(|2x - k|)\}, \forall k \in \mathbb{Z}$. Then

$$\begin{aligned} |\omega_2(x)| &\leq 2\sqrt{2} \sum_{k < 0} |h_k| H_2(|x|) + 2\sqrt{2} \sum_{k \geq 0} |h_k| H_2(|x|) \\ &= 2\sqrt{2} \sum_{k \in \mathbb{Z}} |h_k| H_2(|x|). \end{aligned}$$

But $\sum_{k \in \mathbb{Z}} |h_k| = \sqrt{2}$. Therefore,

$$|\omega_2(x)| \leq 2^2 H_2(|x|) = H'_2(|x|).$$

Proceeding in this way, we obtain

$$|\omega_n(x)| \leq H'_n(|x|), \quad x \in \mathbb{Z}.$$

Hence, the result follows.

Theorem 3.3.2 *Let $\omega_n \in RB$ for all $n \geq 0$. Then*

$$|Q_0^n(x, y)| = \sum_{k \in \mathbb{Z}} \omega_{n,k}(x) \overline{\omega_{n,k}(y)}$$

converges uniformly and absolutely on \mathbb{R}^2 and is bounded. Further, if $\omega_n(x) \log(2 + |x|) \in RB$, then

$$|Q_0^n(x, y)| \leq H_1 \left\{ \frac{(|x - y|)}{\log(2 + |x - y|)} \right\} \quad (3.3.3)$$

where $H_1(|.|)$ is a radial decreasing L^1 -majorant of $\omega_n(x) \log(2 + |x|)$. The sum $Q_0^n(x, y)$ is the kernel of the orthogonal projection Q_0^n onto U_0^n .

Proof Since, we have

$$|\omega_n(x)| \leq H(|x|)$$

where $H(|x|)$ is a radial decreasing L^1 -majorant of $\omega_n(x)$, for all $n \geq 0$, define

$$M = \sup_{x \in \mathbb{Z}} |\omega_n(x)|, \quad \text{for all } n \geq 0. \quad (3.3.4)$$

Now

$$\begin{aligned} |Q_0^n(x, y)| &\leq \sum_{k \in \mathbb{Z}} |\omega_{n,k}(x) \overline{\omega_{n,k}(y)}| \\ &\leq M \sum_{k \in \mathbb{Z}} |\omega_{n,k}(x)| \\ &\leq M \sum_{k \in \mathbb{Z}} |\omega_n(x - k)| \\ &\leq M \int H(|x|) dx < \infty. \end{aligned}$$

The last inequality follows from the fact that the sum can be easily estimated by the integral, since $\omega_n(x)$ is bounded, radial, and decreasing. Clearly, the sum converges uniformly and absolutely.

Further, by using Lemma 1.2.5, we obtain

$$\begin{aligned} |Q_0^n(x, y)| &\leq \sum_{k \in \mathbb{Z}} |\omega_{n,k}(x) \overline{\omega_{n,k}(y)}| \\ &\leq \sum_{k \in \mathbb{Z}} H(|x - k|) H(|y - k|) \\ &\leq CH\left(\frac{1}{2}|x - y|\right) = H'(|x - y|). \end{aligned} \quad (3.3.5)$$

Replacing $H'(|x - y|)$ by $H_1\left\{\frac{(|x - y|)}{\log(2 + |x - y|)}\right\}$ in (3.3.5), as $H_1(|.|)$ is a radial decreasing L^1 -majorant of $\omega_n(\log(2 + |x|))$, we get

$$|Q_0^n(x, y)| \leq H_1\left\{\frac{(|x - y|)}{\log(2 + |x - y|)}\right\}.$$

Finally, to prove that the sum $Q_0^n(x, y)$ represents the kernel of the projection Q_0^n , we note that the operators defined by finite partial sums of the series $Q_0^n(x, y)$ converges strongly to Q_0^n and using the standard argument, it follows that the pointwise uniform limit $Q_0^n(x, y)$ of these partial sums is the kernel of the operator Q_0^n . This completes the proof.

Remark 3.3.3 In particular case when $n = 0$, the space $U_0^0 = V_0$, and in this case, Q_0^0 will be P_0 since for $n = 0$, the wavelet packet ω_n is a scaling function φ .

Theorem 3.3.4 Let $\omega_n \in RB$ for all $n \geq 0$. Then, the kernel $Q_j^n(x, y)$ of the projection onto U_j^n satisfy the convolution bound

$$|Q_j^n(x, y)| \leq 2^j H(2^{j-1}|x - y|), \text{ for all } n \geq 0$$

where $H(|.|)$ is a radial decreasing L^1 -majorant.

Proof Since $\omega_{j,n,k}(x) = 2^{j/2}\omega_n(2^j x - k)$, we have

$$|Q_j^n(x, y)| = 2^j \sum_{k \in \mathbb{Z}} \omega_n(2^j x - k) \overline{\omega_n(2^j y - k)}.$$

By using the argument as in (3.3.5), we get

$$|Q_j^n(x, y)| \leq 2^j H(2^{j-1}|x - y|).$$

Remark 3.3.5 In particular case, when $n = 0$, the space $U_j^0 = V_j$, and in this case, Q_j^0 will be P_j , since for $n = 0$ the wavelet packet ω_n is scaling function φ .

Corollary 3.3.6 Let $\omega_n \in RB$ for all $n \geq 0$. Then, the kernel $Q_j(x, y)$ of the projection Q_j onto $U_j^1 = W_j$ satisfy the convolution bound

$$|Q_j(x, y)| \leq 2^j H(2^{j-1}|x - y|)$$

where $H(|.|)$ is a radial decreasing L^1 -majorant.

Proof It follows from Theorem 3.3.4 when $n = 1$.

Theorem 3.3.7 The convolution bound for $P_m (= Q_m^0)$ continues to hold if P_m is replaced by partially complete sum

$$P_m = \sum_{j=m}^{m+M-1} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in K_j} Q_{l,k}^n$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $j \geq 0$. Also, the set K_j is, for each j , an arbitrary collection of k , and M is a fixed constant.

Proof The kernel $P_m(x, y)$ has the form

$$P_m(x, y) = \sum_{j < m} Q_j(x, y)$$

$$\begin{aligned}
&= \sum_{j < m} \sum_{n=2^u}^{2^{u+1}-1} Q_l^n(x, y) \\
&= \sum_{j < m} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \omega_{l,n,k}(x) \overline{\omega_{l,n,k}(y)}
\end{aligned}$$

where $l = j - u$, $u = 0$ if $j < 0$ and $u = 0, 1, 2, \dots, j$ if $0 \leq j < m$.

Now, the proof follows from Theorems 3.3.2 and 3.3.4.

Convergence of Wavelet Packet Series

Theorem 3.3.8 *For the operator T_j defined in (3.2.11), we have*

- (a) *If $1 \leq p < \infty$, then $\lim_{j \rightarrow \infty} \|T_j f\|_{L^p(\mathbb{R})} = 0$ for all $f \in L^p(\mathbb{R})$.*
- (b) *$\lim_{j \rightarrow \infty} \|T_j f\|_\infty = 0$ for all bounded uniformly continuous functions f .*

Proof (a) We choose $f \in L^p(\mathbb{R})$. Now

$$\begin{aligned}
|T_j f(x)| &= \left| \int_{\mathbb{R}} 2^j P_0(2^j x, 2^j(x-t)) \{f(x-t) - f(x)\} dt \right| \\
&\quad \text{(replacing } y \text{ by } x-t) \\
&\leq \int_{\mathbb{R}} 2^j C H \left(\left| \frac{2^j x - 2^j x + 2^j t}{2} \right| \right) |f(x-t) - f(x)| dt \\
&= C \int_{\mathbb{R}} 2^j H(2^{j-1}|t|) |f(x-t) - f(x)| dt.
\end{aligned} \tag{3.3.6}$$

By using Minkowski's inequality for integrals, we obtain

$$\begin{aligned}
\|T_j f\|_{L^p(\mathbb{R})} &\leq C \int_{\mathbb{R}} 2^j H(2^{j-1}|t|) \left(\int_{\mathbb{R}} |f(x-t) - f(x)|^p dx \right)^{\frac{1}{p}} dt \\
&= C \int_{\mathbb{R}} H\left(\frac{1}{2}|t|\right) \left(\int_{\mathbb{R}} |f(x-2^{-j}t) - f(x)|^p dx \right)^{\frac{1}{p}} dt.
\end{aligned}$$

The expression

$$\left(\int_{\mathbb{R}} |f(x-h) - f(x)|^p dx \right)^{\frac{1}{p}} = \alpha_p(h)$$

is the L^p modulus of continuity of f , which satisfies $\alpha_p(h) \rightarrow 0$ as $|h| \rightarrow 0$ if $1 \leq p < \infty$. Now, the proof follows by the Lebesgue dominated convergence theorem.

(b) Let $f \in L^\infty(\mathbb{R})$ be uniformly continuous. Then

$$\alpha_\infty(t) = \sup_{x \in \mathbb{R}} |f(x-t) - f(x)| \rightarrow 0$$

as $t \rightarrow 0$ uniformly on x . By (3.3.6), we obtain

$$\|T_j f\|_\infty \leq C \int_{\mathbb{R}} \alpha_\infty(t) 2^j H(2^{j-1}|t|) dt.$$

The proof follows by applying the Lebesgue dominated convergence theorem.

Theorem 3.3.9 Suppose that ω_n is the wavelet packet with a radial decreasing L^1 -majorant H for all $n \geq 0$. Then:

- (a) If $1 \leq p < \infty$, then $\lim_{j \rightarrow \infty} \|T_j f - f\|_{L^p(\mathbb{R})} = 0$ for all $f \in L^p(\mathbb{R})$.
- (b) $\lim_{j \rightarrow \infty} \|T_j f - f\|_\infty = 0$ for all bounded uniformly continuous functions f .

Proof It is easy to show that

$$\int_{\mathbb{R}} 2^j P_0(2^j x, 2^j y) dy = 1.$$

Now

$$P_j f(x) - f(x) = \int_{\mathbb{R}} 2^j P_0(2^j x, 2^j y) \{f(y) - f(x)\} dy = T_j f(x).$$

The proof follows by application of Theorems 3.3.1 and 3.3.8.

Theorem 3.3.10 Suppose that ω_n is the wavelet packet with a radial decreasing L^1 -majorant H for all $n \geq 0$. Then

- (a) If $1 \leq p < \infty$, then $\lim_{j \rightarrow \infty} \|S_{j,k}^\sigma f - f\|_{L^p(\mathbb{R})} = 0$ for all $k = 1, 2, \dots$, and for all $f \in L^p(\mathbb{R})$.
- (b) $\lim_{j \rightarrow \infty} \|S_{j,k}^\sigma f - f\|_\infty = 0$ for all $k = 1, 2, \dots$, and for all bounded uniformly continuous functions f .

Proof Since

$$\int_{\mathbb{R}} \omega_n(y) dy = \hat{\omega}_n(0) = 0$$

for all $n > 0$, by (3.2.7), we obtain

$$S_{j,k}^\sigma f(x) - f(x) = [P_j f(x) - f(x)] + R_{j,k}^\sigma f(x)$$

where

$$R_{j,k}^\sigma f(x) = \int_{\mathbb{R}} \left\{ \sum_{m=1}^k \sum_{n=2^p}^{2^{p+1}-1} \omega_{l,n,\sigma(m)}(x) \overline{\omega_{l,n,\sigma(m)}(y)} \right\} \{f(y) - f(x)\} dy$$

with $l = j - p$, $p = 0$ if $j < 0$ and $p = 0, 1, 2, \dots, j$ if $j \geq 0$.

Thus,

$$R_{j,k}^\sigma f(x) = \int_{\mathbb{R}} 2^j Q_0^{\sigma,k}(2^j x, 2^j y) \{f(y) - f(x)\} dy.$$

Hence,

$$S_{j,k}^\sigma f(x) - f(x) = [P_j f(x) - f(x)] + \int_{\mathbb{R}} 2^j Q_0^{\sigma,k}(2^j x, 2^j y) \{f(y) - f(x)\} dy.$$

From Theorem 3.3.9, $\lim_{j \rightarrow \infty} \|P_j f - f\|_{L^p(\mathbb{R})} = 0$. By Lemma 1.2.5

$$\left| Q_0^{\sigma,k}(x, y) \right| \leq CH \left(\frac{|x - y|}{2} \right)$$

(independent of σ and k) and hence, the proof of the theorem follows from Theorem 3.3.8.

Chapter 4

Function Spaces and Wavelet Packets



4.1 Introduction

In this chapter, characterizations of function spaces: the Lebesgue spaces $L^p(\mathbb{R})$, $1 < p < \infty$, the Hardy space $\mathcal{H}^1(\mathbb{R})$, and the Sobolev space $L^{p,s}(\mathbb{R})$, $1 < p < \infty$, $s \in \mathbb{Z}$ by using wavelet packets are given.

4.2 Lebesgue Spaces $L^p(\mathbb{R})$ and Wavelet Packets

Theorem 4.2.1 Let ω_n be band-limited wavelet packets, $f \in S'$ and $0 < p \leq \infty$ such that $\omega_{n,2^{-\ell}} * f \in L^p(\mathbb{R})$ for all $\ell \in \mathbb{Z}$. Then, for any real $\lambda > 0$, there exists a constant C_λ such that

$$(\omega_{\ell,n,\lambda}^{**} f)(x) \leq C_\lambda \left\{ \mathcal{M} \left(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}} \right) (x) \right\}^\lambda, \quad x \in \mathbb{R} \quad (4.2.1)$$

where

$$(\omega_{\ell,n,\lambda}^{**} f)(x) \equiv \sup_{y \in \mathbb{R}} \frac{|(\omega_{n,2^{-\ell}} * f)(x-y)|}{(1 + 2^\ell |y|)^\lambda} \quad (4.2.2)$$

for all $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$ and $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Proof Let $g(x) = (\omega_{n,2^{-\ell}} * f)(2^{-\ell}x)$. Then, $g \in L^p(\mathbb{R})$ and by Lemma 1.2.11, $g_\lambda^*(x) < \infty$ for all $x \in \mathbb{R}$. Further, $(\omega_{n,2^{-\ell}})^*(\xi) = \hat{\omega}_n(2^{-\ell}\xi)$ and ω_n is band-limited, so is g . Now, By Lemma 1.2.10, we have

$$g_\lambda^*(t) \leq C_\lambda \left\{ \mathcal{M}(|g|^{\frac{1}{\lambda}})(t) \right\}^\lambda, \quad t \in \mathbb{R}. \quad (4.2.3)$$

But

$$\begin{aligned}
g_\lambda^*(t) &= \sup_{y \in \mathbb{R}} \frac{|g(t-y)|}{(1+|y|)^\lambda} \\
&= \sup_{y \in \mathbb{R}} \frac{|(\omega_{n,2^{-\ell}} * f)(2^{-\ell}t - 2^{-\ell}y)|}{(1+|y|)^\lambda} \\
&= \sup_{z \in \mathbb{R}} \frac{|(\omega_{n,2^{-\ell}} * f)(2^{-\ell}t - z)|}{(1+2^\ell|z|)^\lambda} \\
&= (\omega_{\ell,n,\lambda}^{**} f)(2^{-\ell}t)
\end{aligned} \tag{4.2.4}$$

and

$$\begin{aligned}
\mathcal{M}\left(|g|^{\frac{1}{\lambda}}\right)(t) &= \sup_{r>0} \frac{1}{2r} \int_{t-r}^{t+r} |(\omega_{n,2^{-\ell}} * f)(2^{-\ell}y)|^{\frac{1}{\lambda}} dy \\
&= \sup_{r>0} \frac{2^\ell}{2r} \int_{2^{-\ell}t-2^{-\ell}r}^{2^{-\ell}t+2^{-\ell}r} |(\omega_{n,2^{-\ell}} * f)(z)|^{\frac{1}{\lambda}} dz \\
&= \mathcal{M}\left(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}}\right)(2^{-\ell}t).
\end{aligned} \tag{4.2.5}$$

From Eqs. (4.2.3)–(4.2.5), we get

$$(\omega_{\ell,n,\lambda}^{**} f)(2^{-\ell}t) \leq C_\lambda \left\{ \mathcal{M}\left(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}}\right)(2^{-\ell}t) \right\}^\lambda.$$

By taking $2^{-\ell}t = x$, we get

$$(\omega_{\ell,n,\lambda}^{**} f)(x) \leq C_\lambda \left\{ \mathcal{M}\left(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}}\right)(x) \right\}^\lambda, \quad x \in \mathbb{R}.$$

Theorem 4.2.2 *Let ω_n be an integrable function on \mathbb{R} such that*

$$\hat{\omega}_n(0) = \int_{\mathbb{R}} \omega_n(x) dx = 0$$

and assume that, for some $\alpha > 0$, satisfies

$$\sum_{n=2^u}^{2^{u+1}-1} |\omega_n(x)| \leq C \frac{1}{(1+|x|)^{1+\alpha}}, \quad x \in \mathbb{R} \tag{4.2.6}$$

and

$$\int_{\mathbb{R}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_n(x+h) - \omega_n(x)| dx \leq C|h|^\alpha, \quad h \in \mathbb{R}. \tag{4.2.7}$$

Then, the operator

$$g(f)(x) = \left(\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{n,2^{-\ell}} * f(x)|^2 \right)^{\frac{1}{2}}$$

is bounded in $L^p(\mathbb{R})$, $1 < p < \infty$. This is also bounded operator from $\mathcal{H}^1(\mathbb{R})$ to $L^1(\mathbb{R})$. \square

In fact, we can consider $g(f)$ as $\ell^2(\mathbb{Z})$ -norm of the vector-valued function whose value at x is the sequence

$$(T_{\omega_n} f)(x) = \{\omega_{n,2^{-\ell}} * f(x) : \ell \in \mathbb{Z}, n = 2^u, 2^{u+1}, \dots, 2^{u+1} - 1\}$$

that is

$$\begin{aligned} g(f)(x) &= \|\{\omega_{n,2^{-\ell}} * f(x) : \ell \in \mathbb{Z}, n = 2^u, 2^{u+1}, \dots, 2^{u+1} - 1\}\|_{\ell^2(\mathbb{Z})} \\ &= \|(T_{\omega_n} f)(x)\|_{\ell^2(\mathbb{Z})}. \end{aligned}$$

From Theorem 4.2.2, under appropriate conditions on ω_n , we have

$$\|g(f)\|_{L^p(\mathbb{R})} \leq C \|f\|_{L^p(\mathbb{R})}, \quad 1 < p < \infty. \quad (4.2.8)$$

Theorem 4.2.3 Let $\omega_n \in \mathbb{R}^0$ and suppose that

$$\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^\ell \xi)|^2 = M \quad \text{for a.e. } \xi \in \mathbb{R} \quad (4.2.9)$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$. Then, there exist constants A_p and B_p , $0 < A_p \leq B_p < \infty$, such that

$$A_p \|f\|_{L^p(\mathbb{R})} \leq \|g(f)\|_{L^p(\mathbb{R})} \leq B_p \|f\|_{L^p(\mathbb{R})}$$

for every $f \in L^p(\mathbb{R})$, $1 < p < \infty$, where $g(f)$ is defined in Theorem 4.2.2.

Proof From equality (4.2.9), it can be shown that

$$\|g(f)\|_{L^2(\mathbb{R})} \leq \sqrt{M} \|f\|_{L^2(\mathbb{R})}, \quad f \in L^2(\mathbb{R}). \quad (4.2.10)$$

In fact, using Plancherel's theorem and (4.2.9), we obtain

$$\|g(f)\|_{L^2(\mathbb{R})}^2 = \int_{\mathbb{R}} \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{n,2^{-\ell}} * f(x)|^2 dx$$

$$\begin{aligned}
&= \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \frac{1}{2\pi} \int_{\mathbb{R}} \left| \hat{\omega}_n(2^\ell \xi) \hat{f}(\xi) \right|^2 d\xi \\
&= \frac{M}{2\pi} \int_{\mathbb{R}} \left| \hat{f}(\xi) \right|^2 d\xi = M \|f\|_{L^2(\mathbb{R})}^2.
\end{aligned}$$

We now use the equality (4.2.10) to “reverse” (4.2.8). Applying the polarization identity to (4.2.10), and letting $\langle \cdot, \cdot \rangle$ to denote the inner product in $\ell^2(\mathbb{Z})$, we obtain

$$\begin{aligned}
M \left| \int_{\mathbb{R}} f(x) h(x) dx \right| &= \left| \int_{\mathbb{R}} \langle T_{\omega_n} f(x), T_{\omega_n} h(x) \rangle dx \right| \\
&\leq \int_{\mathbb{R}} g(f)(x) g(h)(x) dx \leq \|g(f)\|_{L^p(\mathbb{R})} \cdot \|g(h)\|_{L^q(\mathbb{R})}
\end{aligned}$$

for $f \in L^p(\mathbb{R}) \cap L^2(\mathbb{R})$, $h \in L^q(\mathbb{R}) \cap L^2(\mathbb{R})$, $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$ (a density argument allows us to drop the assumption that $f, h \in L^2(\mathbb{R})$). By taking the supremum over all h such that $\|h\|_{L^q(\mathbb{R})} \leq 1$ and using (4.2.8) with p replaced by q , we deduce

$$\begin{aligned}
\|f\|_{L^p(\mathbb{R})} &= \sup_{\|h\|_q \leq 1} \left| \int_{\mathbb{R}} f(x) h(x) dx \right| \\
&\leq \frac{1}{M} \|g(f)\|_{L^p(\mathbb{R})} \cdot \|g(h)\|_{L^q(\mathbb{R})} \leq \frac{C_p}{M} \|g(f)\|_{L^p(\mathbb{R})}
\end{aligned}$$

which proves the first equality. The second equality is a consequence of Theorem 4.2.2 (see (4.2.8)).

Theorem 4.2.4 *Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets. Given a real number $\lambda \geq 1$ and $1 < p < \infty$, there exists a constant $A_{p,\lambda} < \infty$ such that*

$$\left\| \left(\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right)^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \leq A_{p,\lambda} \|f\|_{L^p(\mathbb{R})}, \quad \forall f \in L^p(\mathbb{R})$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Proof Let $f \in L^p(\mathbb{R})$, $1 < p < \infty$. But $\omega_{n,2^{-\ell}} * f \in L^p(\mathbb{R})$ for all ℓ . Therefore, we can use Theorem 4.2.1 to obtain

$$\left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})}$$

$$\begin{aligned} &\leq C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left[\mathcal{M} \left(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}} \right) \right]^{2\lambda} \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \\ &= C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left[\mathcal{M} \left(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}} \right) \right]^{2\lambda} \right\}^{\frac{1}{2\lambda}} \right\|_{L^{p\lambda}(\mathbb{R})}^{\lambda}. \end{aligned}$$

Now applying Lemma 1.2.10 to $p\lambda > 1$ and $q = 2\lambda \geq 2 > 1$, we obtain

$$\begin{aligned} \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} &\leq C_{p,\lambda} \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{n,2^{-\ell}} * f|^2 \right\}^{\frac{1}{2\lambda}} \right\|_{L^{p\lambda}(\mathbb{R})}^{\lambda} \\ &= C_{p,\lambda} \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{n,2^{-\ell}} * f|^2 \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \\ &\leq A_{p,\lambda} \|f\|_{L^p(\mathbb{R})} \end{aligned}$$

where the last inequality follows from Theorem 4.2.3.

Remark 4.2.5 Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets and suppose that

$$\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^\ell \xi)|^2 = M \quad \text{for a.e. } \xi \in \mathbb{R}$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$. We have the promised characterization of $L^p(\mathbb{R})$, $1 < p < \infty$, in terms of functions $\omega_{\ell,n,\lambda}^{**} f$. That is

$$A \|f\|_{L^p(\mathbb{R})} \leq \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \leq B \|f\|_{L^p(\mathbb{R})}$$

with A and B depending only on p and λ , and $\lambda \geq 1$. By Theorem 4.2.4, it remains to show the LHS inequality. Clearly, this follows easily from Theorem 4.2.3 and the trivial inequality

$$|(\omega_{n,2^{-\ell}} * f)(x)| \leq (\omega_{\ell,n,\lambda}^{**} f)(x).$$

Theorem 4.2.6 Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets. For p , $1 < p < \infty$, and $f \in L^p(\mathbb{R})$, we have

$$\left\| \left\{ \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \leq C \|f\|_{L^p(\mathbb{R})} \quad (4.2.11)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and C independent of f .

Proof First we shall prove that for each integer n satisfying $2^u \leq n \leq 2^{u+1} - 1$, there exists a constant C_n such that

$$\sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2 \leq C_n \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l+u,1,k} \rangle|^2 \quad (4.2.12)$$

for all $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and $f \in L^p(\mathbb{R})$. For this purpose, we will adopt the dyadic expansion for n , namely $n = \sum_{t=1}^{u+1} \varepsilon_t 2^{t-1}$, where $\varepsilon_t = \{0, 1\}$ and $\varepsilon_{u+1} = 1$. By using the operator P_λ defined in (2.2.40), it follows from (2.2.51) that

$$\omega_{l,n,k} = 2^{-1/2} (P_{\varepsilon_1} \{\omega_{l+1,[n/2],.}\})_k, \quad l, k \in \mathbb{Z} \quad (4.2.13)$$

or equivalently in view of (2.2.51) that

$$\omega_{l,n,k}(x) = 2^{-1/2} \sum_{m \in \mathbb{Z}} P_{m-2k}^{\varepsilon_1} \omega_{l+1,[n/2],m}(x), \quad l, k \in \mathbb{Z}. \quad (4.2.14)$$

Thus, for any $f \in L^p(\mathbb{R})$, we have

$$\begin{aligned} |\langle f, \omega_{l,n,k} \rangle| &\leq 2^{-1/2} \sum_{m \in \mathbb{Z}} |P_{m-2k}^{\varepsilon_1}|^{1/2} |P_{m-2k}^{\varepsilon_1}|^{1/2} |\langle f, \omega_{l+1,[n/2],m} \rangle| \\ &\leq 2^{-1/2} \left(\sum_{m \in \mathbb{Z}} |P_{m-2k}^{\varepsilon_1}| \right)^{1/2} \left(\sum_{m \in \mathbb{Z}} |P_{m-2k}^{\varepsilon_1}| |\langle f, \omega_{l+1,[n/2],m} \rangle|^2 \right)^{1/2} \end{aligned}$$

Moreover, by putting

$$D_\lambda = \max \left\{ \sum_{k \in \mathbb{Z}} |P_{2k}^\lambda|, \sum_{k \in \mathbb{Z}} |P_{2k+1}^\lambda| \right\}, \quad \lambda \in \{0, 1\} \quad (4.2.15)$$

we have

$$\sum_{m \in \mathbb{Z}} |P_{m-2k}^{\varepsilon_1}| = \sum_{m \in \mathbb{Z}} |P_m^{\varepsilon_1}| \leq 2D_{\varepsilon_1}$$

and

$$\sum_{k \in \mathbb{Z}} |P_{m-2k}^{\varepsilon_1}| = \begin{cases} \sum_{k \in \mathbb{Z}} |P_{2k}^{\varepsilon_1}|, & m \text{ is even} \\ \sum_{k \in \mathbb{Z}} |P_{2k+1}^{\varepsilon_1}|, & m \text{ is odd} \end{cases}$$

so that

$$\sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2 \leq D_{\varepsilon_1} \sum_{m \in \mathbb{Z}} \left(\sum_{k \in \mathbb{Z}} |P_{m-2k}^{\varepsilon_1}| \right) |\langle f, \omega_{l+1,[n/2],m} \rangle|^2 \leq D_{\varepsilon_1}^2 \sum_{m \in \mathbb{Z}} |\langle f, \omega_{l+1,[n/2],m} \rangle|^2. \quad (4.2.16)$$

But $\left[\frac{n}{2} \right] = \sum_{t=1}^u \varepsilon_{t+1} 2^{t-1}$, $\left[\frac{n}{2^2} \right] = \sum_{t=1}^{u-1} \varepsilon_{t+2} 2^{t-1}$, etc. Therefore, by repeated application of the inequality (4.2.16) to $\omega_{l+1,[n/2],k}, \omega_{l+2,[n/2^2],k}, \dots, \omega_{l+u,1,k}$ gives

$$\sum_{k \in \mathbb{Z}} |\langle f, \omega_{l,n,k} \rangle|^2 \leq (D_{\varepsilon_1} D_{\varepsilon_2} \dots D_{\varepsilon_u})^2 \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l+u,1,k} \rangle|^2. \quad (4.2.17)$$

Thus, (4.2.12) follows from (4.2.17) by setting

$$C_n = (D_{\varepsilon_1} D_{\varepsilon_2} \dots D_{\varepsilon_u})^2. \quad (4.2.18)$$

Clearly, C_n is a finite positive constant, and it follows from (4.2.15) and $\{P_k^\lambda\} \in l^1$ that D_λ is finite, $\lambda \in \{0, 1\}$. Moreover, by setting $D = \max\{D_0, D_1\}$, (4.2.18) implies that $C_n \leq D^{2u}$. Thus, from (4.2.17), we have

$$\sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 \leq 2^u D^{2u} \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l+u,1,k} \rangle|^2 \quad (4.2.19)$$

$$\Rightarrow \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}} \leq 2^u D^{2u} \sum_{k \in \mathbb{Z}} |\langle f, \omega_{l+u,1,k} \rangle|^2 2^l \chi_{I_{l,k}} \quad (4.2.20)$$

where $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$. It is obvious that for $f \in L^p(\mathbb{R})$ the numbers $\langle f, \omega_{l,n,k} \rangle$ make sense since $\omega_n \in L^p(\mathbb{R})$ (where $\frac{1}{p} + \frac{1}{q} = 1$). In fact, we have

$$\sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 \leq \sum_{n=2^u}^{2^{u+1}-1} 2^{l(\frac{1}{p}-\frac{1}{q})} \|\omega_n\|_{L^q(\mathbb{R})} \|f\|_{L^p(\mathbb{R})}.$$

Also, we have

$$\begin{aligned}
& \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle| \leq \sum_{n=2^u}^{2^{u+1}-1} \left| \int_{\mathbb{R}} f(x) \overline{\omega_{l,n,k}(x)} dx \right| \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} 2^{l/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_n(2^l x - k)} dx \right| \\
& = \sum_{n=2^u}^{2^{u+1}-1} 2^{l/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_n(2^l(x - 2^{-l}k))} dx \right| \\
& = \sum_{n=2^u}^{2^{u+1}-1} 2^{-l/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_{n,2^{-l}}(x - 2^{-l}k)} dx \right| \\
& = \sum_{n=2^u}^{2^{u+1}-1} 2^{-l/2} |(\tilde{\omega}_{n,2^{-l}} * f)(2^{-l}k)| \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} 2^{-l/2} \sup_{y \in I_{l,k}} |(\tilde{\omega}_{n,2^{-l}} * f)(y)|
\end{aligned}$$

where $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$ and $\tilde{\omega}_n(y) = \overline{\omega_n(-y)}$. For each fixed $l \in \mathbb{Z}$, we have

$$\begin{aligned}
& \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}}(x) \leq \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left\{ \sup_{y \in I_{l,k}} |(\tilde{\omega}_{n,2^{-l}} * f)(y)| \right\}^2 \chi_{I_{l,k}}(x) \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} \left\{ \sup_{|z| \leq 2^{-l}} |(\tilde{\omega}_{n,2^{-l}} * f)(x - z)| \right\}^2 \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} \left\{ \sup_{|z| \leq 2^{-l}} \left| \frac{(\tilde{\omega}_{n,2^{-l}} * f)(x - z)}{(1 + 2^l|z|)^{\lambda}} \right| \right\}^2 (1 + 2^l|z|)^{2\lambda} \\
& \leq 2^{2\lambda} [(\omega_{l,n,\lambda}^{**} f)(x)]^2 \text{ for any } \lambda > 0.
\end{aligned}$$

Now, applying Theorems 4.2.4 and 4.2.6 with $\lambda \geq 1$, we obtain

$$\left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \leq A_{p,\lambda} \|f\|_{L^p(\mathbb{R})} \quad (4.2.21)$$

for some constant $A_{p,\lambda}$.

Now, from (4.2.20) and (4.2.21), we have

$$\left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{l,k}(x) \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \leq C \|f\|_{L^p(\mathbb{R})} \quad (4.2.22)$$

This completes the proof of the theorem. \square

To obtain the reverse inequality to (4.2.11), we shall assume that ω_n is orthonormal wavelet packet. We shall use the following notation. Given two functions f and ω_n for which $\langle f, \omega_n \rangle$ makes sense, we define

$$(\mathcal{W}_{\omega_n} f)(x) = \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{l,k}(x) \right\}^{1/2} \quad (4.2.23)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$.

We observe that if T_{ω_n} is the operator which maps f into $l^2(\mathbb{Z} \times \mathbb{Z})$ such that

$$(T_{\omega_n} f)(x) = \{ \langle f, \omega_{l,n,k} \rangle 2^{l/2} \chi_{l,k}(x) : l = j - u; j, l, k \in \mathbb{Z}, n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1 \}$$

where $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, we have

$$(\mathcal{W}_{\omega_n} f)(x) = \sqrt{(T_{\omega_n} f)(x) \cdot (T_{\omega_n} f)(x)}$$

where \cdot denotes the dot product in $l^2(\mathbb{Z} \times \mathbb{Z})$.

Theorem 4.2.7 *Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets. Given $p \in (1, \infty)$, there exist two constants A_p and B_p , $0 < A_p \leq B_p < \infty$, such that*

$$A_p \|f\|_{L^p(\mathbb{R})} \leq \|\mathcal{W}_{\omega_n} f\|_{L^p(\mathbb{R})} \leq B_p \|f\|_{L^p(\mathbb{R})} \quad (4.2.24)$$

for all $f \in L^p(\mathbb{R})$.

Proof Since $\omega_n \in L^q(\mathbb{R})$, where $\frac{1}{p} + \frac{1}{q} = 1$, we observe that the coefficients $\langle f, \omega_{l,n,k} \rangle$ are well defined. The inequality on the right-hand side in (4.2.24) is precisely (4.2.11). Thus, we have already obtained a constant $B_p < \infty$ such that

$$\|\mathcal{W}_{\omega_n} f\|_{L^p(\mathbb{R})} \leq B_p \|f\|_{L^p(\mathbb{R})}. \quad (4.2.25)$$

For $p = 2$, we have equality (with $B_p = 1$) because ω_n is an orthonormal wavelet packet. Now

$$\begin{aligned}
\int_{\mathbb{R}} (T_{\omega_n} f)(x) \cdot (T_{\omega_n} f)(x) dx &= \|\mathcal{W}_{\omega_n} f\|_{L^2(\mathbb{R})}^2 \\
&= \int_{\mathbb{R}} \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{l,k}(x) dx \\
&= \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 \\
&= \|f\|_{L^2(\mathbb{R})}^2.
\end{aligned}$$

From this equality, the polarization identity and a density argument, we obtain

$$\int_{\mathbb{R}} f(x) g(x) dx = \int_{\mathbb{R}} (T_{\omega_n} f)(x) \cdot (T_{\omega_n} g)(x) dx$$

for $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$. Using a duality argument, together with Hölder's inequality and (4.2.25), we deduce

$$\begin{aligned}
\|f\|_{L^p(\mathbb{R})} &= \sup_{\|g\|_{L^q(\mathbb{R})} \leq 1} \left| \int_{\mathbb{R}} f(x) g(x) dx \right| \\
&\leq \sup_{\|g\|_{L^q(\mathbb{R})} \leq 1} \|\mathcal{W}_{\omega_n} f\|_{L^p(\mathbb{R})} \|\mathcal{W}_{\omega_n} g\|_{L^q(\mathbb{R})} \leq B_q \|\mathcal{W}_{\omega_n} f\|_{L^p(\mathbb{R})}.
\end{aligned}$$

This completes the proof of the theorem.

Theorem 4.2.8 *Let $\omega_n \in \mathbb{R}^0$ be wavelet packets. Then:*

(a) *There exist constants $C < \infty$ and $\varepsilon > 0$ such that for $l \geq l'$, for fixed $n \neq 0$, we have*

$$|\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle| \leq \frac{2^{\frac{3}{2}(l'-l)} C}{(1 + 2^{l'} |2^{-l} k - 2^{-l'} m|)^{1+\varepsilon}}$$

(b) *There exist constants $C < \infty$ and $\varepsilon > 0$ such that for $l \leq l'$, we have*

$$|\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle| \leq \frac{2^{\frac{3}{2}(l-l')} C}{(1 + 2^l |2^{-l} m - 2^{-l'} k|)^{1+\varepsilon}}$$

where $k, l', m \in \mathbb{Z}$ and $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$; $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Proof Suppose that ω_n is associated with the constants $C_0^1, C_1^1, \gamma^1 > 0$ and $\varepsilon^1 > 0$ and that ω_0 is associated with the constants $C_0^2, C_1^2, \gamma^2 > 0$ and $\varepsilon^2 > 0$ in the Definition 1.2.14. We choose

$$C = \max\{C_0^1, C_1^1, C_0^2, C_1^2\}, \quad \gamma = \min\{\gamma^1, \gamma^2\} \text{ and } \varepsilon = \min\{\varepsilon^1, \varepsilon^2, \gamma\}.$$

Then, ω_n and ω_0 satisfy the conditions in Definition 1.2.14 with constants C , γ and ε ($\gamma \geq \varepsilon$).

Let g be a function defined on \mathbb{R} , and we write $\tilde{g}(x) = \overline{g(-x)}$. Then, we have

$$\overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle} = \langle \omega_{l',0,m}, \omega_{l,n,k} \rangle = (\omega_{l',0,m} * \tilde{\omega}_{l,n,-k})(0).$$

Now, to prove part (a) we apply Lemma 1.2.16 with $N = 0$ and part (b) can be proved with the help of (a).

Theorem 4.2.9 *Let $\omega_n \in \mathbb{R}^0$ and ω_0 be an orthonormal wavelet packet. Then, there exists a constant C_p , $0 < C_p < \infty$, $1 < p < \infty$, such that*

$$\|\mathcal{W}_{\omega_n} f\|_{L^p(\mathbb{R})} \leq C_p \|\mathcal{W}_{\omega_0} f\|_{L^p(\mathbb{R})}, \quad \forall f \in L^p(\mathbb{R}). \quad (4.2.26)$$

Proof We are given that ω_0 is an orthonormal wavelet packet so

$$\omega_{l,n,k}(x) = \sum_{l',m \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle \omega_{l,n,k}, \omega_{l',0,m} \rangle \omega_{l',0,m}(x)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $j \in \mathbb{Z}$, with convergence in $L^2(\mathbb{R})$ and hence also in the sense of distributions. Then

$$\begin{aligned} (\mathcal{W}_{\omega_n} f)(x) &= \left\{ \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \\ &= \left\{ \sum_{l,k \in \mathbb{Z}} \left| \sum_{l',m \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{l',0,m} \rangle \overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle} \right|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \end{aligned}$$

where $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$. Writing

$$A_1(l, n, k) = \sum_{l' \leq l} \sum_{m \in \mathbb{Z}} \langle f, \omega_{l',0,m} \rangle \overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle}$$

and

$$A_2(l, n, k) = \sum_{l' > l} \sum_{m \in \mathbb{Z}} \langle f, \omega_{l',0,m} \rangle \overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle}$$

we have

$$(\mathcal{W}_{\omega_n} f)(x) \leq \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(l, n, k)|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2}$$

$$+ \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \quad (4.2.27)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $j, k \in \mathbb{Z}$.

To estimate $A_1(l, n, k)$, we use Theorem 4.2.8(a) to obtain

$$\begin{aligned} |A_1(l, n, k)| &\leq \sum_{l' \leq l} \sum_{m \in \mathbb{Z}} |\langle f, \omega_{l', 0, m} \rangle| \frac{2^{\frac{3}{2}(l'-l)} C}{(1 + |2^{(l'-l)}k - m|)^{1+\varepsilon}} \\ &= C \sum_{l' \leq l} 2^{\frac{3}{2}(l'-l)} \left\{ \sum_{m \in \mathbb{Z}} \frac{|\langle f, \omega_{l', 0, m} \rangle|}{(1 + |2^{(l'-l)}k - m|)^{1+\varepsilon}} \right\} \end{aligned}$$

for some $C < \infty$ and $\varepsilon > 0$. By applying Lemma 1.2.13(a) with $r = 1$, we obtain

$$|A_1(l, n, k)| \leq C \sum_{l' \leq l} 2^{\frac{3}{2}(l'-l)} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l', 0, m} \rangle| \chi_{I_{l', m}} \right)(x) \right]$$

for all $x \in I_{l', m}$. Therefore, we have

$$\begin{aligned} &\left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ &\leq C \left\| \left\{ \sum_{l \in \mathbb{Z}} 2^l \left[\sum_{l' \leq l} 2^{\frac{3}{2}(l'-l)} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l', 0, m} \rangle| \chi_{I_{l', m}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ &= C \left\| \left\{ \sum_{l \in \mathbb{Z}} \left[\sum_{l' \leq l} 2^{(l'-l)} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l', 0, m} \rangle| 2^{l'/2} \chi_{I_{l', m}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ &\leq C \left\| \left\{ \sum_{l=0}^{\infty} 2^{-l} \right\} \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l', 0, m} \rangle| 2^{l'/2} \chi_{I_{l', m}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ &\leq C \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l', 0, m} \rangle| 2^{l'/2} \chi_{I_{l', m}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})}, \end{aligned}$$

where we have used Young's inequality for convolutions

$$\| \{a_l\} * \{b_{l'}\} \|_{l^2} \equiv \left\| \left\{ \sum_{l'} a_{l-l'} b_{l'} \right\} \right\|_{l^2} \leq \| \{a_l\} \|_{l^1} \| \{b_{l'}\} \|_{l^2} \quad (4.2.28)$$

with

$$a_l = \begin{cases} 2^{-l} & \text{if } l \geq 0 \\ 0 & \text{if } l < 0 \end{cases}$$

and

$$b_{l'} = \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| 2^{l'/2} \chi_{I_{l',m}} \right) (x).$$

Now, using the vector-valued inequality for the Hardy–Littlewood maximal function with $q = 2$, we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ & \leq C_p \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| 2^{l'/2} \chi_{I_{l',m}} \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ & = C_p \left\| \left\{ \sum_{l' \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^2 2^{l'} \chi_{I_{l',m}} \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ & \leq C_p \|\mathcal{W}_{\omega_0} f\|_{L^p(\mathbb{R})}. \end{aligned} \quad (4.2.29)$$

To estimate $A_2(l, n, k)$, we use Theorem 4.2.8(b) together with Lemma 1.2.13(b) with $r = 1$ to obtain

$$\begin{aligned} |A_2(l, n, k)| & \leq \sum_{l' > l} \sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| \frac{2^{\frac{3}{2}(l-l')} C}{(1 + |2^{(l-l')} m - k|)^{1+\varepsilon}} \\ & \leq C \sum_{l' > l} 2^{\frac{1}{2}(l-l')} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| \chi_{I_{l',m}} \right) (x) \right] \end{aligned}$$

for some $C < \infty$ and $\varepsilon > 0$ and for all $x \in I_{l,k}$. Further, since $\{I_{l,k} : k \in \mathbb{Z}\}$ is a collection of disjoint dyadic intervals, we have

$$\begin{aligned} & \left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \\ & \leq C \left\| \left\{ \sum_{l \in \mathbb{Z}} 2^l \left[\sum_{l' > l} 2^{\frac{1}{2}(l-l')} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| \chi_{I_{l',m}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \end{aligned}$$

$$= C \left\| \left\{ \sum_{l \in \mathbb{Z}} \left[\sum_{l' > l} 2^{(l-l')} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| 2^{l'/2} \chi_{I_{l',m}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p(\mathbb{R})}.$$

As the series $\sum_{l' > l} 2^{(l-l')} = 1$, the estimate for $A_2(l, n, k)$ is now obtained as for the estimate for $A_1(l, n, k)$ so that

$$\left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^p(\mathbb{R})} \leq C_p \| \mathcal{W}_{\omega_0} f \|_{L^p(\mathbb{R})}. \quad (4.2.30)$$

Finally, from (4.2.27), (4.2.29) and (4.2.30), we have

$$\| \mathcal{W}_{\omega_n} f \|_{L^p(\mathbb{R})} \leq C_p \| \mathcal{W}_{\omega_0} f \|_{L^p(\mathbb{R})}.$$

4.3 Hardy Space $\mathcal{H}^1(\mathbb{R})$ and Wavelet Packets

In this section, we shall present the characterization of $\mathcal{H}^1(\mathbb{R})$ using wavelet packet coefficients. In order to achieve the goal, similar methods like characterization of Lebesgue spaces $L^p(\mathbb{R})$ using wavelet packets by Garg, Abdullah, and Ahmad in [119] are used. But here are two difficulties to overcome. Firstly, Lemma 1.2.9 on vector-valued inequalities for Hardy–Littlewood maximal function is not true for $p = 1$ and the second problem is that the space $\mathcal{H}^1(\mathbb{R})$ is not reflexive and hence, the duality argument cannot be used. To achieve characterization, we shall use atomic decomposition of $\mathcal{H}^1(\mathbb{R})$. Throughout $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$.

Following is the precise statement of the atomic characterization of $\mathcal{H}^1 = \mathcal{H}^1(\mathbb{R})$.

Lemma 4.3.1 *A function $f \in \mathcal{H}^1(\mathbb{R})$ if and only if f has a decomposition of the form*

$$f = \sum_{j=1}^{\infty} \lambda_j a_j \quad (4.3.1)$$

where a_j , $j = 1, 2, 3, \dots$ are 2-atoms and $\sum_{j=1}^{\infty} |\lambda_j| < \infty$. Moreover,

$$\| f \|_{\mathcal{H}^1(\mathbb{R})} = \inf \left\{ \sum_{j=1}^{\infty} |\lambda_j| \right\} \quad (4.3.2)$$

where the infimum is taken over all decompositions of f of the form (4.3.1).

Theorem 4.3.2 Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets. Then, there exist constants A and B , $0 < A \leq B < \infty$, such that

$$A \|f\|_{\mathcal{H}^1(\mathbb{R})} \leq \|g(f)\|_{L^1(\mathbb{R})} \leq B \|f\|_{\mathcal{H}^1(\mathbb{R})}, \quad \forall f \in \mathcal{H}^1(\mathbb{R})$$

where $g(f)$ is defined in Theorem 4.2.2.

Theorem 4.3.3 Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets. Given a real number $\lambda > 1$, there exists a constant A_λ , $0 < A_\lambda < \infty$, such that

$$\left\| \left(\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right)^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \leq A_\lambda \|f\|_{\mathcal{H}^1(\mathbb{R})}, \quad \forall f \in \mathcal{H}^1(\mathbb{R})$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and $\omega_{\ell,n,\lambda}^{**}$ is the map defined in Theorem 4.2.1.

Proof Let $f \in \mathcal{H}^1(\mathbb{R})$. By Theorem 4.3.2, $\omega_{n,2^{-\ell}} * f \in L^1(\mathbb{R})$ for all $j \in \mathbb{Z}$ and hence by using Theorem 4.2.1, we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \\ & \leq C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left[\mathcal{M}(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}}) \right]^{2\lambda} \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})}^{\frac{1}{2\lambda}} \\ & = C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left[\mathcal{M}(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}}) \right]^{2\lambda} \right\}^{\frac{1}{2\lambda}} \right\|_{L^\lambda(\mathbb{R})}. \end{aligned}$$

But $\lambda > 1$, therefore we can apply Lemma 1.2.9 with $p = \lambda > 1$ and $q = 2\lambda \geq 2 > 1$, and we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \leq C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{n,2^{-\ell}} * f|^2 \right\}^{\frac{1}{2}} \right\|_{L^\lambda(\mathbb{R})}^{\frac{1}{2\lambda}} \\ & = C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{n,2^{-\ell}} * f|^2 \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \\ & \leq A_\lambda \|f\|_{\mathcal{H}^1(\mathbb{R})} \end{aligned}$$

where the last inequality follows from Theorem 4.3.2.

Remark 4.3.4 We have the promised characterization of $\mathcal{H}^1(\mathbb{R})$, in terms of functions $\omega_{\ell,n,\lambda}^{**} f$. That is

$$A \|f\|_{\mathcal{H}^1(\mathbb{R})} \leq \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\omega_{\ell,n,\lambda}^{**} f|^2 \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \leq B \|f\|_{\mathcal{H}^1(\mathbb{R})}$$

for all $f \in \mathcal{H}^1(\mathbb{R})$, where A and B are depending only on λ , and $\lambda > 1$. By Theorem 4.3.3, it remains to show the LHS inequality. This follows easily from Theorem 4.3.2 and the trivial inequality

$$|(\omega_{n,2^{-\ell}} * f)(x)| \leq (\omega_{\ell,n,\lambda}^{**} f)(x).$$

Theorem 4.3.5 Let $\omega_n \in \mathbb{R}^0$ be band-limited wavelet packets. Then, there exists a constant C , $0 < C < \infty$ such that

$$\left\| \left\{ \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}}(x) \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \leq C \|f\|_{\mathcal{H}^1(\mathbb{R})}, \quad \forall f \in \mathcal{H}^1(\mathbb{R}) \quad (4.3.3)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Proof Let $f \in \mathcal{H}^1(\mathbb{R})$. Then, we have

$$\begin{aligned} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle| &\leq \sum_{n=2^u}^{2^{u+1}-1} 2^{l/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_n(2^l x - k)} dx \right| \\ &= \sum_{n=2^u}^{2^{u+1}-1} 2^{l/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_n(2^l(x - 2^{-l}k))} dx \right| \\ &= \sum_{n=2^u}^{2^{u+1}-1} 2^{-l/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_{n,2^{-l}}(x - 2^{-l}k)} dx \right| \\ &= \sum_{n=2^u}^{2^{u+1}-1} 2^{-l/2} |(\tilde{\omega}_{n,2^{-l}} * f)(2^{-l}k)| \\ &\leq \sum_{n=2^u}^{2^{u+1}-1} 2^{-l/2} \sup_{y \in I_{l,k}} |(\tilde{\omega}_{n,2^{-l}} * f)(y)| \end{aligned}$$

where $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$ and $\tilde{\omega}_n(y) = \overline{\omega_n(-y)}$. For each fixed $l \in \mathbb{Z}$, we have

$$\begin{aligned}
& \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{l,k}(x) \\
& \leq \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left\{ \sup_{y \in I_{l,k}} |(\tilde{\omega}_{n,2^{-l}} * f)(y)| \right\}^2 \chi_{l,k}(x) \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} \left\{ \sup_{|z| \leq 2^{-l}} |(\tilde{\omega}_{n,2^{-l}} * f)(x-z)| \right\}^2 \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} \left\{ \sup_{|z| \leq 2^{-l}} \left| \frac{(\tilde{\omega}_{n,2^{-l}} * f)(x-z)}{(1+2^l|z|)^\lambda} \right| \right\}^2 (1+2^l|z|)^{2\lambda} \\
& \leq \sum_{n=2^u}^{2^{u+1}-1} 2^{2\lambda} [(\omega_{l,n,\lambda}^{**} f)(x)]^2 \text{ for any } \lambda > 0.
\end{aligned}$$

Now, applying Theorem 4.3.3 with $\lambda \geq 1$, we obtain

$$\left\| \left\{ \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{l,k}(x) \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \leq C \|f\|_{\mathcal{H}^1(\mathbb{R})}.$$

Remark 4.3.6 The above theorem is proved for band-limited wavelet packets. But the result is true for more general functions not necessarily band-limited. For this, we state and prove the following theorem:

Theorem 4.3.7 Let $\omega_n \in \mathbb{R}^0$ and ω_0 be an orthonormal wavelet packet. Then, there exists a constant C , $0 < C < \infty$, such that

$$\|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})} \leq C \|\mathcal{W}_{\omega_0} f\|_{L^1(\mathbb{R})}, \quad \forall f \in \mathcal{H}^1(\mathbb{R}) \quad (4.3.4)$$

where

$$(\mathcal{W}_{\omega_n} f)(x) = \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{l,k}(x) \right\}^{1/2} \quad (4.3.5)$$

Proof We are given that ω_0 is an orthonormal wavelet packet. Therefore,

$$\omega_{l,n,k}(x) = \sum_{l',m \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle \omega_{l,n,k}, \omega_{l',0,m} \rangle \omega_{l',0,m}(x)$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $j \in \mathbb{Z}$, with convergence in $L^2(\mathbb{R})$ and hence also in the sense of distributions. Then

$$\begin{aligned}
(\mathcal{W}_{\omega_n} f)(x) &= \left\{ \sum_{l,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{l,n,k} \rangle|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \\
&= \left\{ \sum_{l,k \in \mathbb{Z}} \left| \sum_{l',m \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{l',0,m} \rangle \overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle} \right|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2}
\end{aligned}$$

where $I_{l,k} = [2^{-l}k, 2^{-l}(k+1)]$. Writing

$$A_1(l, n, k) = \sum_{l' \leq l} \sum_{m \in \mathbb{Z}} \langle f, \omega_{l',0,m} \rangle \overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle}$$

and

$$A_2(l, n, k) = \sum_{l' > l} \sum_{m \in \mathbb{Z}} \langle f, \omega_{l',0,m} \rangle \overline{\langle \omega_{l,n,k}, \omega_{l',0,m} \rangle}$$

we have

$$\begin{aligned}
(\mathcal{W}_{\omega_n} f)(x) &\leq \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(l, n, k)|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \\
&\quad + \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}}(x) \right\}^{1/2} \tag{4.3.6}
\end{aligned}$$

where $l = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $j, k \in \mathbb{Z}$.

We observe that in writing (4.3.6), we are using the fact that ω_n is an orthonormal wavelet packet. Hence, using Lemma 1.2.13(a) with $1 < r < 1 + \varepsilon$, we obtain

$$\begin{aligned}
&\left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^1(\mathbb{R})} \\
&\leq C \left\| \left\{ \sum_{l \in \mathbb{Z}} \left[\sum_{l' \leq l} 2^{(l'-l)} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{\frac{l'}{2r}} \chi_{I_{l',m}} \right)^r \right]^2 \right\}^{1/2} \right\|_{L^1(\mathbb{R})} \\
&\leq C \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{\frac{l'}{2r}} \chi_{I_{l',m}} \right)^r \right]^{1/2r} \right\} \right\|_{L^1(\mathbb{R})}
\end{aligned}$$

$$= C \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{\frac{l'}{2r}} \chi_{I_{l',m}} \right) \right]^{2r} \right\}^{1/2r} \right\|_{L^r(\mathbb{R})}^r.$$

Now, applying Lemma 1.2.9 with $p = r > 1$ and $q = 2r > 2 > 1$, we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^1(\mathbb{R})} \\ & \leq C_r \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{\frac{l'}{2r}} \chi_{I_{l',m}} \right]^{2r} \right\}^{1/2r} \right\|_{L^r(\mathbb{R})}^r \\ & = C_r \left\| \left\{ \sum_{l' \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^2 2^{l'} \chi_{I_{l',m}} \right\}^{1/2} \right\|_{L^1(\mathbb{R})} \\ & = C_r \| \mathcal{W}_{\omega_0} f \|_{L^1(\mathbb{R})} \end{aligned} \quad (4.3.7)$$

In order to handle the second summand on the right-hand side of (4.3.6), we proceed as above, but using Lemma 1.2.13(b) with $1 < r < \min\{2, 1 + \varepsilon\}$. Thus, we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \\ & \leq C \left\| \left\{ \sum_{l \in \mathbb{Z}} 2^l \left[\sum_{l' \leq l} 2^{\frac{3}{2}(l-l')} 2^{(l'-l)r} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} \chi_{I_{l',m}} \right)^r \right]^2 \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \\ & = C \left\| \left\{ \sum_{l \in \mathbb{Z}} \left[\sum_{l' \leq l} 2^{(l-l')(2-r)} \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{l'/2r} \chi_{I_{l',m}} \right)^r \right]^2 \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})} \\ & \leq C \left\| \sum_{s=0}^{\infty} 2^{-s(2-r)} \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{l'/2r} \chi_{I_{l',m}} \right)^r \right]^{2r} \right\}^{\frac{1}{2}} \right\|_{L^1(\mathbb{R})}, \end{aligned}$$

where we have used Young's inequality for convolutions

$$\| \{a_l\} * \{b_{l'}\} \|_{l^2} \equiv \left\| \left\{ \sum_{l'} a_{l-l'} b_{l'} \right\} \right\|_{l^2} \leq \| \{a_l\} \|_{l^1} \| \{b_{l'}\} \|_{l^2} \quad (4.3.8)$$

with

$$a_l = \begin{cases} 2^{-l} & \text{if } l \geq 0 \\ 0 & \text{if } l < 0 \end{cases}$$

and

$$b_{l'} = \mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle| 2^{l'/2} \chi_{I_{l',m}} \right) (x).$$

But $2 - r > 0$, the series $\sum_{s=0}^{\infty} 2^{-s(2-r)}$ converges. Therefore,

$$\begin{aligned} & \left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^1(\mathbb{R})} \\ & \leq C_r \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{l'/2r} \chi_{I_{l',m}} \right) \right]^{2r} \right\}^{1/2r} \right\|_{L^1(\mathbb{R})} \\ & = C_r \left\| \left\{ \sum_{l' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{m \in \mathbb{Z}} |\langle f, \omega_{l',0,m} \rangle|^{\frac{1}{r}} 2^{l'/2r} \chi_{I_{l',m}} \right) \right]^{2r} \right\}^{1/2r} \right\|_{L^1(\mathbb{R})}. \end{aligned}$$

From Lemma 1.2.9 with $p = r > 1$ and $q = 2r > 2 > 1$, we obtain

$$\left\| \left\{ \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(l, n, k)|^2 2^l \chi_{I_{l,k}} \right\}^{1/2} \right\|_{L^1(\mathbb{R})} \leq C_r \|\mathcal{W}_{\omega_0} f\|_{L^1(\mathbb{R})}. \quad (4.3.9)$$

On combining (4.3.7), (4.3.6), and (4.3.9), we get

$$\|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})} \leq \|\mathcal{W}_{\omega_0} f\|_{L^1(\mathbb{R})} \quad (4.3.10)$$

for all $f \in \mathcal{H}^1(\mathbb{R})$.

Since the wavelet packets ω_n are not necessarily band-limited, on combining Theorems 4.2.6 and 4.2.8, we obtain the boundedness of \mathcal{W}_{ω_n} from $\mathcal{H}^1(\mathbb{R})$ to $L^1(\mathbb{R})$.

Theorem 4.3.8 *For any $\omega_n \in \mathbb{R}^0$, there exists a constant C , $0 < C < \infty$, such that*

$$\|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})} \leq \|f\|_{\mathcal{H}^1(\mathbb{R})}, \quad \forall f \in \mathcal{H}^1(\mathbb{R}). \quad (4.3.11)$$

In order to complete the characterization of $\mathcal{H}^1(\mathbb{R})$, we need the reverse inequality (4.3.10). But we cannot use duality of $\mathcal{H}^1(\mathbb{R})$ for this purpose. Therefore, we shall use different technique. Further to achieve the required inequality, we shall use atomic concept of $\mathcal{H}^1(\mathbb{R})$. We shall show that for compactly supported wavelet packets ω_n , the representation

$$f \sim \sum_{n=2^u}^{2^{u+1}-1} \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \omega_{l,n,k} \rangle \omega_{l,n,k}$$

can be rearranged to produce the atomic decomposition of $f \in \mathcal{H}^1(\mathbb{R})$. Moreover, if ω_n are not compactly supported, the above representation can be rearranged to produce a molecular decomposition of f which makes the computations more complicated. The desired characterization of $\mathcal{H}^1(\mathbb{R})$ with wavelet packets that belong to \mathbb{R}^0 will be obtained by using above inequality together with Theorem 4.3.7.

Theorem 4.3.9 Let ω_n be compactly supported wavelet packets. Then, there exists a constant B , $0 < B < \infty$, such that

$$\|f\|_{\mathcal{H}^1(\mathbb{R})} \leq B \|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})}, \quad \forall f \in \mathcal{H}^1(\mathbb{R}), \quad (4.3.12)$$

Proof Let $\omega_{n,I} = \omega_{n,I_{j,k}} = \omega_{j,n,k}$ and let \mathcal{D} denote the family of all dyadic intervals on \mathbb{R} . Then

$$(\mathcal{W}_{\omega_n} f)(x) = \left\{ \sum_{n=2^u}^{2^{u+1}-1} \sum_{I \in \mathcal{D}} |\langle f, \omega_{n,I} \rangle|^2 |I|^{-1} \chi_I(x) \right\}^{\frac{1}{2}}.$$

For $k \in \mathbb{Z}$, let $\Omega_k = \{x \in \mathbb{R} : (\mathcal{W}_{\omega_n} f)(x) > 2^k\}$, we have $\Omega_k \supseteq \Omega_{k+1}$ and a summation by parts argument shows

$$\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} 2^k |\Omega_k| \leq 2^{u+1} \|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})}. \quad (4.3.13)$$

Let

$$\mathcal{B}_k = \left\{ I \in \mathcal{D} : |I \cap \Omega_k| \geq \frac{|I|}{2} \text{ and } |I \cap \Omega_{k+1}| < \frac{|I|}{2} \right\}$$

For each dyadic interval I on \mathbb{R} , there exists a unique $k \in \mathbb{Z}$ such that $I \in \mathcal{B}_k$. Moreover, due to nesting property of dyadic intervals $I \in \mathcal{B}_k$, there is a unique maximal dyadic interval $\tilde{I} \in \mathcal{B}_k$ such that $I \subset \tilde{I}$. Denote by $\{\tilde{I}_k^i : i \in \wedge_k\}$ collection of all such maximal dyadic intervals in \mathcal{B}_k , where the index set \wedge_k depends on k . Thus, we have the following partition of \mathcal{D} as

$$\mathcal{D} = \bigcup_{k \in \wedge} \bigcup_{i \in \wedge_k} \{I : I \subset \tilde{I}_k^i, I \in \mathcal{B}_k\}. \quad (4.3.14)$$

Therefore, the representation

$$f(x) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \omega_{l,n,k} \rangle \omega_{l,n,k}(x) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{I \in \mathcal{D}} \langle f, \omega_{n,I} \rangle \omega_{n,I}(x)$$

can be written as

$$f(x) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} \left\{ \sum_{\substack{I \subset \tilde{I}_k^i \\ I \in \mathcal{B}_k}} \langle f, \omega_{n,I} \rangle \omega_{n,I}(x) \right\}. \quad (4.3.15)$$

Now, the compactly supported wavelet packets play an important role. By translations, if necessary, we can always assume that there exists $m \in \mathbb{N}$ such that $\text{supp}(\omega_n) \subset [0, m]$. Thus, the support of $\omega_{n,I_{l,k}} = \omega_{l,n,k}$ is contained in $[2^{-l}k, 2^{-l}(k+m)] = I_{l,k}[m]$, and $|I_{l,k}[m]| = 2^{-l}m = m|I_{l,k}|$. Let

$$\lambda(k, i) = |\tilde{I}_k^i|^{\frac{1}{2}} m^{\frac{1}{2}} \left(\sum_{\substack{I \subset \tilde{I}_k^i \\ I \in \mathcal{B}_k}} |\langle f, \omega_{n,I} \rangle|^2 \right)^{\frac{1}{2}}.$$

Define

$$a_{(k,i)}(x) = \begin{cases} \frac{1}{\lambda(k,i)} \sum_{I \in \mathcal{B}_k} \langle f, \omega_{n,I} \rangle \omega_{n,I}(x), & \text{if } \lambda(k, i) \neq 0 \\ 0, & \text{if } \lambda(k, i) = 0. \end{cases}$$

Then, (4.3.15) can be written as

$$f(x) = \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} \lambda(k, i) a_{(k,i)}(x) \quad (4.3.16)$$

which is the required decomposition of f in terms of 2-atoms. In fact, the support of each $a_{(k,i)}$ is contained in $\tilde{I}_k^i[m]$. Moreover, since ω_n is orthonormal wavelet packet, the above definition gives us

$$\|a_{(k,i)}\|_{L^2(\mathbb{R})}^2 = \frac{1}{|\lambda(k, i)|^2} \sum_{\substack{I \subset \tilde{I}_k^i \\ I \in \mathcal{B}_k}} |\langle f, \omega_{n,I} \rangle|^2 = \frac{1}{m|\tilde{I}_k^i|} = \frac{1}{|\tilde{I}_k^i[m]|}. \quad (4.3.17)$$

The equality (4.3.17) implies that the series which defines $a_{(k,i)}$ also converges in $L^1(\mathbb{R})$. Since $\hat{\omega}_n$ is continuous, we have

$$\int_{\mathbb{R}} \omega_{n,I}(x) dx = 0.$$

Thus, on integrating term by term we deduce that

$$\int_{\mathbb{R}} a_{(k,i)}(x) dx = 0.$$

This shows that all the conditions of (1.5.1) hold for each $a_{(k,i)}$ and, hence, these functions are 2-atoms as promised.

Finally, we need to estimate

$$\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} |\lambda(k, i)|.$$

Since

$$|I \setminus \Omega_{k+1}| = |I| - |I \cap \Omega_{k+1}| > |I| - \frac{1}{2}|I| = \frac{1}{2}|I|$$

when $I \in \mathcal{B}_k$, we have

$$\begin{aligned} \sum_{\substack{I \subset \tilde{I}_k^i \\ I \in \mathcal{B}_k}} |\langle f, \omega_{n,I} \rangle|^2 &\leq 2 \sum_{\substack{I \subset \tilde{I}_k^i \\ I \in \mathcal{B}_k}} |\langle f, \omega_{n,I} \rangle|^2 |I|^{-1} |I \setminus \Omega_{k+1}| \\ &\leq \int_{\tilde{I}_k^i \setminus \Omega_{k+1}} \sum_{I \in \mathcal{D}} |\langle f, \omega_{n,I} \rangle|^2 |I|^{-1} \chi_I(x) dx \\ &\leq \int_{\tilde{I}_k^i \setminus \Omega_{k+1}} [(\mathcal{W}_{\omega_n} f)(x)]^2 dx \leq 2 \cdot 2^{2(k+1)} |\tilde{I}_k^i| \end{aligned}$$

as outside of Ω_{k+1} , $(\mathcal{W}_{\omega_n} f)(x) \leq 2^{k+1}$. Thus,

$$\begin{aligned} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} |\lambda(k, i)| &= \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} |\tilde{I}_k^i|^{\frac{1}{2}} m^{\frac{1}{2}} \left(\sum_{\substack{I \subset \tilde{I}_k^i \\ I \in \mathcal{B}_k}} |\langle f, \omega_{n,I} \rangle|^2 \right)^{\frac{1}{2}} \\ &\leq 2m^{\frac{1}{2}} \sqrt{2} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} 2^k |\tilde{I}_k^i|^{\frac{1}{2}} |\tilde{I}_k^i|^{\frac{1}{2}} \\ &= 2\sqrt{2m} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} 2^k |\tilde{I}_k^i|. \end{aligned}$$

Since $\tilde{I}_k^i \in \mathcal{B}_k$, we have $|\tilde{I}_k^i| \leq 2|\tilde{I}_k^i \cap \Omega_k|$, and, since the \tilde{I}_k^i are disjoint, we obtain

$$\begin{aligned}
\sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} |\lambda(k, i)| &\leq 4\sqrt{2m} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} |\tilde{I}_k^i \cap \Omega_k| \\
&\leq 4\sqrt{2m} \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} 2^k |\Omega_k| \\
&\leq 2^{u+2} \sqrt{2m} \|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})}
\end{aligned}$$

where the last inequality is due to (4.3.13). By Lemma 4.3.1, we have that $f \in \mathcal{H}^1(\mathbb{R})$ and

$$\|f\|_{\mathcal{H}^1(\mathbb{R})} \leq \sum_{n=2^u}^{2^{u+1}-1} \sum_{k \in \mathbb{Z}} \sum_{i \in \wedge_k} |\lambda(k, i)| \leq 2^{u+2} \sqrt{2m} \|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})}.$$

Hence, the theorem is proved by taking $2^{u+2} \sqrt{2m} = B$.

Finally, on combining Theorems 4.3.7, 4.3.8 and Lemma 4.3.1, we obtain the characterization of $\mathcal{H}^1(\mathbb{R})$ in terms of wavelet packet coefficients as given below:

Theorem 4.3.10 *Let ω_n be an orthonormal wavelet packet such that $\omega_n \in \mathbb{R}^0$, $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$ where $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$. Then, there exist two constants A and B , $0 < A \leq B < \infty$, such that*

$$A \|f\|_{\mathcal{H}^1(\mathbb{R})} \leq \|\mathcal{W}_{\omega_n} f\|_{L^1(\mathbb{R})} \leq B \|f\|_{\mathcal{H}^1(\mathbb{R})}$$

for all $f \in \mathcal{H}^1(\mathbb{R})$.

4.4 Sobolev Spaces $L^{p,s}(\mathbb{R})$ and Wavelet Packets

In this section, we consider the Sobolev spaces $L^{p,s}(\mathbb{R})$, $1 < p < \infty$, $s = 1, 2, 3, \dots$, and obtain their characterisations in terms of wavelet packets.

Lemma 4.4.1 *For $1 < p < \infty$ and $s \in \mathbb{N}$, there exist two constants A and B , $0 < A \leq B < \infty$, such that*

$$A \|f\|_{L^{p,s}} \leq \|f\|_{W^{p,s}} \leq B \|f\|_{L^{p,s}} \text{ for all } f \in L^{p,s}(\mathbb{R}). \quad (4.4.1)$$

Thus, $\|\cdot\|_{L^{p,s}}$ and $\|\cdot\|_{W^{p,s}}$ define equivalent norms on the spaces $L^{p,s}(\mathbb{R})$.

The other equivalent norm for the spaces $L^{p,s}(\mathbb{R})$ is given by

$$\|f\|_{L^{p,s}}^* \equiv \|f\|_{L^p} + \|D^s f\|_{L^p}. \quad (4.4.2)$$

Thus, we have

$$\|f\|_{L^{p,s}}^* \leq \|f\|_{L^{p,s}} \leq C \|f\|_{L^{p,s}}^* \quad (4.4.3)$$

for all $f \in L^{p,s}(\mathbb{R})$, $1 < p < \infty$, $s = 1, 2, 3, \dots$.

The version of the Littlewood–Paley function, we need the following:

For $s \in \mathbb{N}$, define

$$g^s(f)(x) = \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} (2^{\ell s} |\omega_{n,2^{-\ell}} * f(x)|)^2 \right\}^{\frac{1}{2}} \quad (4.4.4)$$

where ω_n are band-limited wavelet packets in S with Fourier transform supported in $\{\xi \in \mathbb{R} : 2^{-N} \leq |\xi| \leq 2^N\}$ for some $N \in \mathbb{N}$ and

$$\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^\ell \xi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R}$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Theorem 4.4.2 Let $\omega_n \in S$ be such that

$$\text{supp}(\hat{\omega}_n) \subset \{\xi \in \mathbb{R} : 2^{-N} \leq |\xi| \leq 2^N\} \text{ for some } N \in \mathbb{N},$$

and

$$\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^\ell \xi)|^2 = 1 \text{ for a.e. } \xi \in \mathbb{R} \quad (4.4.5)$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Then, for $1 < p < \infty$ and $s = 1, 2, \dots$, $f \in L^{p,s}(\mathbb{R})$ if and only if $f \in L^p(\mathbb{R})$ and $g^s(f) \in L^p(\mathbb{R})$. Moreover,

$$\|f\|_{L^p} + \|g^s(f)\|_{L^p}$$

defines a norm for $L^{p,s}(\mathbb{R})$ that is equivalent to $\|\cdot\|_{L^{p,s}}$.

Proof Let $f \in L^{p,s}(\mathbb{R})$. Then

$$\begin{aligned} (2^{\ell s} \omega_{n,2^{-\ell}} * f) \hat{(\xi)} &= C 2^{\ell s} (\omega_{n,2^{-\ell}}) \hat{(\xi)} \hat{f}(\xi) \\ &= C 2^{\ell s} \xi^{-s} (\omega_{n,2^{-\ell}}) \hat{(\xi)} \xi^s \hat{f}(\xi). \end{aligned}$$

The function $m : \mathbb{R} \rightarrow \ell^2(\mathbb{Z}) \equiv \mathcal{L}(\mathbb{C}, \ell^2(\mathbb{Z}))$ given by

$$m(\xi) = \{2^{\ell s} \xi^{-s} (\omega_{n,2^{-\ell}}) \hat{(\xi)} : \ell = j - u, j \in \mathbb{Z}; n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1\}$$

where $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, satisfies (1.2.6) with $\mathbb{H}_0 = \mathbb{C}$ and $\mathbb{H}_1 = \ell^2(\mathbb{Z})$. To justify this, we have to estimate $\|m(\xi)\|_{\ell^2(\mathbb{Z})}$ and $\|Dm(\xi)\|_{\ell^2(\mathbb{Z})}$.

Since $(\omega_{n,2^{-\ell}})^\wedge(\xi) = \hat{\omega}_n(2^{-\ell}\xi)$, the assumption we are making about the support of $\hat{\omega}_n$ implies that

$$\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |2^{\ell s} \xi^{-s} (\omega_{n,2^{-\ell}})^\wedge(\xi)|^2 \quad \text{and} \quad \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |2^{\ell s} D(\xi^{-s} (\omega_{n,2^{-\ell}})^\wedge(\xi))|^2$$

have only a finite number (at most 2^{2N}) of nonzero terms. Thus, (1.2.6) will be proved if we show

$$\begin{cases} (i) \ 2^{\ell s} |\xi^{-s} (\omega_{n,2^{-\ell}})^\wedge(\xi)| \leq B_0; \\ (ii) \ 2^{\ell s} |D(\xi^{-s} (\omega_{n,2^{-\ell}})^\wedge(\xi))| \leq B_1 |\xi|^{-1}. \end{cases} \quad (4.4.6)$$

Since on the support of $(\omega_{n,2^{-\ell}})^\wedge$ we have $2^{-N} \leq |2^{-\ell}\xi| \leq 2^N$, we deduce

$$\begin{aligned} 2^{\ell s} |\xi^{-s} (\omega_{n,2^{-\ell}})^\wedge(\xi)| &= 2^{\ell s} |\xi^{-s} \hat{\omega}_n(2^{-\ell}\xi)| \\ &\leq C 2^{\ell s} 2^{(N-\ell)s} = C 2^{Ns} \end{aligned}$$

and

$$\begin{aligned} 2^{\ell s} |D(\xi^{-s} (\omega_{n,2^{-\ell}})^\wedge(\xi))| &= 2^{\ell s} |D(\xi^{-s} \hat{\omega}_n(2^{-\ell}\xi))| \\ &\leq 2^{\ell s} [C_1 |\xi|^{-s-1} \hat{\omega}_n(2^{-\ell}\xi) + C_2 |\xi|^{-s-2} 2^{-\ell} (D\hat{\omega}_n)(2^{-\ell}\xi)] \\ &\leq C 2^{\ell s} \left[2^{(N-\ell)s} \frac{1}{|\xi|} + 2^{N(s+1)} 2^{-s\ell} \frac{1}{|\xi|} \right] \\ &\leq \frac{C}{|\xi|} [2^{Ns} + 2^{N(s+1)}] \quad = \frac{C}{|\xi|} \end{aligned}$$

which proves (4.4.6).

By Lemma 1.2.12, we obtain

$$\begin{aligned} \|g^s(f)\|_{L^p(\mathbb{R})} &= C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left| \left(2^{\ell s} [(\cdot)^{-s} (\omega_{n,2^{-\ell}})^\wedge] (\cdot)^s \hat{f} \right) \right|^2 \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \\ &\leq C \left\| (\cdot)^s \hat{f} \right\|_{L^p(\mathbb{R})} \\ &= C \|D^s f\|_{L^p(\mathbb{R})}. \end{aligned}$$

Thus, $g^s(f) \in L^p(\mathbb{R})$ and

$$\|f\|_{L^p} + \|g^s(f)\|_{L^p} \leq \|f\|_{L^p} + C \|D^s f\|_{L^p} \leq C \|f\|_{L^{p,s}}$$

where the last inequality is due to (4.4.3). Observe that to show this inequality, we have not used (4.4.5). This equality will be needed to prove the converse result.

Further, assume now that $f \in L^p(\mathbb{R})$ and $g^s(f) \in L^p(\mathbb{R})$. By (4.4.3), it suffices to prove that $D^s f \in L^p(\mathbb{R})$. Equality (4.4.5) implies that

$$\hat{f}(\xi) = \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} (\omega_{n,2^{-\ell}})^{\wedge}(\xi) (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\xi) \hat{f}(\xi)$$

where $\tilde{\omega}_n(x) = \overline{\omega_n(-x)}$. Then

$$\begin{aligned} D^s f(x) &= C \left((\cdot)^s \hat{f}(\cdot) \right)^{\wedge}(x) \\ &= C \left(\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} (\cdot)^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\cdot) (\omega_{n,2^{-\ell}})^{\wedge}(\cdot) \hat{f}(\cdot) \right)^{\wedge}(x) \\ &= C \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left(2^{-\ell s} (\cdot)^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\cdot) 2^{s\ell} (\omega_{n,2^{-\ell}})^{\wedge}(\cdot) \hat{f}(\cdot) \right)^{\wedge}(x). \end{aligned}$$

The function $m : \mathbb{R} \rightarrow \mathcal{L}(\ell^2(\mathbb{Z}), \mathbb{C})$ given by

$$m(\xi)(a) = \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{-\ell s} \xi^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\xi) a_\ell \quad \text{for } a = \{a_\ell\}_{\ell \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$$

satisfies (1.2.6) with $\mathbb{H}_0 = \ell^2(\mathbb{Z})$ and $\mathbb{H}_1 = \mathbb{C}$. As before, we only need to show

$$\begin{cases} (i) \quad 2^{-\ell s} |\xi^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\xi)| \leq B_0; \\ (ii) \quad 2^{-\ell s} |D(\xi^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\xi))| \leq B_1 |\xi|^{-1} \end{cases} \quad (4.4.7)$$

taking into account our assumption on the support of $\hat{\omega}_n$. The proof of (4.4.7) is similar to one of (4.4.6). More precisely

$$\begin{aligned} 2^{-\ell s} |\xi^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\xi)| &= 2^{-\ell s} \left| \xi^s \widehat{\omega}_n(2^{-\ell} \xi) \right| \\ &\leq C 2^{-\ell s} 2^{(N+\ell)s} = C 2^{Ns} \end{aligned}$$

and

$$\begin{aligned} 2^{-\ell s} |D(\xi^s (\tilde{\omega}_{n,2^{-\ell}})^{\wedge}(\xi))| &= 2^{-\ell s} \left| D\left(\xi^s \widehat{\omega}_n(2^{-\ell} \xi)\right) \right| \\ &\leq 2^{-\ell s} \left[C_1 \left| \xi^{s-1} \widehat{\omega}_n(2^{-\ell} \xi) \right| + C_2 \left| \xi^s 2^{-\ell} (D\widehat{\omega}_n)(2^{-\ell} \xi) \right| \right] \\ &\leq C 2^{-\ell s} \left[2^{(N+\ell)s} \frac{C_1}{|\xi|} + 2^{(N+\ell)(s+1)} 2^{-\ell} \frac{C_2}{|\xi|} \right] \end{aligned}$$

$$\leq \frac{C}{|\xi|} [2^{Ns} + 2^{N(s+1)}] = \frac{C}{|\xi|}.$$

Further, we use Lemma 1.2.12, with $\mathbb{H}_0 = \ell^2(\mathbb{Z})$ and $\mathbb{H}_1 = \mathbb{C}$ in this case and obtain

$$\begin{aligned} \|D^s f\|_{L^p(\mathbb{R})} &= C \left\| \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left(\{2^{-\ell s} (\cdot)^s (\tilde{\omega}_{n,2^{-\ell}})^\wedge\} 2^{s\ell} (\omega_{n,2^{-\ell}})^\wedge \hat{f} \right)^\vee \right\|_{L^p(\mathbb{R})} \\ &\leq C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \left(2^{s\ell} |(\omega_{n,2^{-\ell}})^\wedge \hat{f}| \right)^\vee \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \\ &= C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} (2^{\ell s} |\omega_{n,2^{-\ell}} * f|)^2 \right\}^{\frac{1}{2}} \right\|_{L^p(\mathbb{R})} \\ &= C \|g^s(f)\|_{L^p(\mathbb{R})}. \end{aligned}$$

This shows that $D^s(f) \in L^p(\mathbb{R})$. Moreover, by (4.4.3)

$$\|f\|_{L^{p,s}} \leq C \left(\|f\|_{L^p} + \|D^s f\|_{L^p} \right) \leq C \left(\|f\|_{L^p} + \|g^s(f)\|_{L^p} \right). \quad \square$$

In order to characterize the Sobolev spaces $L^{p,s}(\mathbb{R})$ using wavelet packets, we denote by “ \mathcal{B} ” the space of all wavelet packets $\omega_n \in S$, $n = 0, 1, 2, \dots$ such that $N \in \mathbb{N}$ for which

$$\text{supp}(\hat{\omega}_n) \subset \{\xi \in \mathbb{R} : 2^{-N} \leq |\xi| \leq 2^N\}$$

and

$$\sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\hat{\omega}_n(2^\ell \xi)|^2 = 1 \quad \text{for a.e. } \xi \in \mathbb{R}$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Theorem 4.4.3 *Let $\omega_n \in \mathcal{B}$, $n = 0, 1, \dots$ be wavelet packets. Given a real number $\lambda \geq 1$, a natural number $s \geq 1$ and $1 < p < \infty$, there exist two constants $A = A_{p,\lambda,s}$ and $B = B_{p,\lambda,s}$, $0 < A \leq B < \infty$, such that*

$$A \|f\|_{L^{p,s}} \leq \|f\|_{L^p} + \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} [2^{\ell s} (\omega_{\ell,n,\lambda}^{**} f)]^2 \right\}^{1/2} \right\|_{L^p} \leq B \|f\|_{L^{p,s}} \quad (4.4.8)$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and $\forall f \in L^{p,s}(\mathbb{R})$, where $\omega_{\ell,n,\lambda}^{**}$ is defined in Theorem 4.2.1 by

$$(\omega_{\ell,n,\lambda}^{**} f)(x) = \sup_{y \in \mathbb{R}} \frac{|(\omega_{n,2^{-\ell}} * f)(x-y)|}{(1+2^\ell|y|)^\lambda} \quad (4.4.9)$$

for all $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$.

Proof Suppose that $f \in L^{p,s}(\mathbb{R})$. Then, $\omega_{n,2^{-\ell}} * f \in L^p(\mathbb{R})$ for all $\ell \in \mathbb{Z}$. Now, using Theorem 4.2.1 and Lemma 1.2.9 with $p = p\lambda > 1$ (since $\lambda \geq 1$) and $q = 2\lambda$, we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |2^{\ell s} (\omega_{\ell,n,\lambda}^{**} f)|^2 \right\}^{\frac{1}{2}} \right\|_{L^p} \\ & \leq C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{2\ell s} [\mathcal{M}(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}})]^{2\lambda} \right\}^{\frac{1}{2}} \right\|_{L^p}^{\frac{1}{2\lambda}} \\ & = C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{2\ell s} [\mathcal{M}(|\omega_{n,2^{-\ell}} * f|^{\frac{1}{\lambda}})]^{2\lambda} \right\}^{\frac{1}{2\lambda}} \right\|_{L^{p\lambda}}^{\frac{1}{2\lambda}} \\ & \leq C_{p,\lambda} \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{2\ell s} |\omega_{n,2^{-\ell}} * f|^2 \right\}^{\frac{1}{2\lambda}} \right\|_{L^{p\lambda}} \\ & = C_{p,\lambda} \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |2^{\ell s} \omega_{n,2^{-\ell}} * f|^2 \right\}^{\frac{1}{2}} \right\|_{L^p} \\ & = C_{p,\lambda} \|g^s(f)\|_{L^p}. \end{aligned}$$

From here, the RHS of the inequalities follows immediately. The LHS inequality follows from the fact

$$|\omega_{n,2^{-\ell}} * f(x)| \leq (\omega_{\ell,n,\lambda}^{**} f)(x) \quad \text{for any } n = 0, 1, 2, \dots$$

and Theorem 4.4.2.

Theorem 4.4.4 Let $\omega_n \in \mathcal{B}$ be band-limited wavelet packets. For $1 < p < \infty$, and $s = 1, 2, \dots$, there exists a constant $C_{p,s}$, $0 < C_{p,s} < \infty$, such that

$$\left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 (1+2^{2\ell s}) 2^\ell \chi_{[2^{-\ell k}, 2^{-\ell(k+1)}]} \right\}^{1/2} \right\|_{L^p} \leq C_{p,s} \|f\|_{L^{p,s}} \quad (4.4.10)$$

for all $f \in L^{p,s}(\mathbb{R})$ and $\ell = j-u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$.

Proof We note that for $f \in L^p(\mathbb{R})$ the numbers $\langle f, \omega_{\ell,n,k} \rangle$ make sense since $\omega_n \in L^q(\mathbb{R})$ (where $\frac{1}{p} + \frac{1}{q} = 1$). In fact

$$\sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle| \leq \sum_{n=2^u}^{2^{u+1}-1} 2^{\ell\left(\frac{1}{p}-\frac{1}{2}\right)} \|\omega_n\|_{L^q(\mathbb{R})} \|f\|_{L^p(\mathbb{R})}.$$

We have

$$\begin{aligned} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle| &\leq \sum_{n=2^u}^{2^{u+1}-1} \left| \int_{\mathbb{R}} f(x) \overline{\omega_{\ell,n,k}(x)} dx \right| \\ &\leq \sum_{n=2^u}^{2^{u+1}-1} 2^{\ell/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_n(2^\ell x - k)} dx \right| \\ &= \sum_{n=2^u}^{2^{u+1}-1} 2^{\ell/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_n(2^\ell(x - 2^{-\ell}k))} dx \right| \\ &= \sum_{n=2^u}^{2^{u+1}-1} 2^{-\ell/2} \left| \int_{\mathbb{R}} f(x) \overline{\omega_{n,2^{-\ell}}(x - 2^{-\ell}k)} dx \right| \\ &= \sum_{n=2^u}^{2^{u+1}-1} 2^{-\ell/2} |(\tilde{\omega}_{n,2^{-\ell}} * f)(2^{-\ell}k)| \\ &\leq \sum_{n=2^u}^{2^{u+1}-1} 2^{-\ell/2} \sup_{y \in I_{\ell,k}} |(\tilde{\omega}_{n,2^{-\ell}} * f)(y)| \end{aligned}$$

where $I_{\ell,k} = [2^{-\ell}k, 2^{-\ell}(k+1)]$ and $\tilde{\omega}_n(y) = \overline{\omega_n(-y)}$. For each fixed $\ell \in \mathbb{Z}$, we have

$$\sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 2^\ell \chi_{I_{\ell,k}}(x) \leq 2^{2\lambda} [(\omega_{\ell,n,\lambda}^{**} f)(x)]^2 \text{ for any } \lambda > 0.$$

Now, applying Theorems 4.2.6 and 4.4.3 with $\lambda \geq 1$, we obtain

$$\begin{aligned} &\left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 (1 + 2^{2\ell s}) 2^\ell \chi_{I_{\ell,k}} \right\}^{1/2} \right\|_{L^p} \\ &\leq \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 2^\ell \chi_{I_{\ell,k}} \right\}^{1/2} \right\|_{L^p} \\ &\quad + \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}} \right\}^{1/2} \right\|_{L^p} \end{aligned}$$

$$\leq C\|f\|_{L^p} + C_\lambda \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} 2^{2\ell s} |(\omega_{\ell,n,\lambda}^{**} f)|^2 \right\}^{1/2} \right\|_{L^p} \leq C\|f\|_{L^{p,s}}.$$

This completes the proof of the theorem. \square

In order to obtain the reverse inequality to (4.4.10), we shall assume that ω_n is orthonormal wavelet packet. We use the following notation. Given two functions f and ω_n for which $\langle f, \omega_n \rangle$ makes sense, we define

$$(\mathcal{W}_{\omega_n}^s f)(x) = \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 (1 + 2^{2\ell s}) 2^\ell \chi_{I_{\ell,k}}(x) \right\}^{1/2} \quad (4.4.11)$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$ and $n = 2^u, 2^u + 1, \dots, 2^{u+1} - 1$.

Theorem 4.4.5 *Let $\omega_n \in S$, for all $n = 0, 1, 2, \dots$, be band-limited orthonormal wavelet packets. Given $p \in (1, \infty)$, and $s = 1, 2, \dots$, there exist two constants $A_{p,s}$ and $B_{p,s}$, $0 < A_{p,s} \leq B_{p,s} < \infty$, such that*

$$A_{p,s} \|f\|_{L^{p,s}(\mathbb{R})} \leq \|\mathcal{W}_{\omega_n}^s f\|_{L^p(\mathbb{R})} \leq B_{p,s} \|f\|_{L^{p,s}(\mathbb{R})} \quad (4.4.12)$$

for all $f \in L^{p,s}(\mathbb{R})$.

Proof The RHS of inequality is proved by using Theorem 4.4.4. Therefore, we need only to prove LHS of inequality. For $f, g \in S$ (where “S” is dense in $L^{p,s}(\mathbb{R})$), we have

$$\begin{aligned} \int_{\mathbb{R}} (D^s f)(x) \cdot g(x) dx &= C \int_{\mathbb{R}} f(x) (D^s g)(x) dx \\ &= C \int_{\mathbb{R}} \left\{ \sum_{\ell,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{\ell,n,k} \rangle \omega_{\ell,n,k}(x) \right\} \\ &\quad \times \left\{ \sum_{\ell',k' \in \mathbb{Z}} \sum_{n'=2^{u'}}^{2^{u'+1}-1} \langle D^s g, \omega_{\ell',n',k'} \rangle \omega_{\ell',n',k'}(x) \right\} dx \\ &= C \sum_{\ell,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{\ell,n,k} \rangle \langle D^s g, \omega_{\ell,n,k} \rangle \\ &= C \int_{\mathbb{R}} \sum_{\ell,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{\ell,n,k} \rangle 2^{\ell s} 2^{\ell/2} \langle D^s g, \omega_{\ell,n,k} \rangle 2^{-\ell s} 2^{\ell/2} \chi_{I_{\ell,k}}(x) dx. \end{aligned}$$

On using the Cauchy–Schwartz inequality for $\ell^2(\mathbb{Z} \times \mathbb{Z})$, we obtain

$$\begin{aligned}
\left| \int_{\mathbb{R}} (D^s f)(x) \cdot g(x) dx \right| &\leq C \int_{\mathbb{R}} \left(\sum_{\ell, k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell, n, k} \rangle|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell, k}}(x) \right)^{1/2} \\
&\quad \times \left(\sum_{\ell, k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle D^s g, \omega_{\ell, n, k} \rangle|^2 2^{-2\ell s} 2^\ell \chi_{I_{\ell, k}}(x) \right)^{1/2} dx \\
&\leq C \int_{\mathbb{R}} (\mathcal{W}_{\omega_n}^s f)(x) \left(\sum_{\ell, k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle D^s g, \omega_{\ell, n, k} \rangle|^2 2^{-\ell s} 2^\ell \chi_{I_{\ell, k}}(x) \right)^{1/2} dx
\end{aligned}$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$. We note that

$$\langle D^s g, \omega_{\ell, n, k} \rangle 2^{-\ell s} = C 2^{-\ell s} \langle g, D^s \omega_{\ell, n, k} \rangle = C \langle g, (D^s \omega_n)_{\ell, k} \rangle.$$

Thus,

$$\left| \int_{\mathbb{R}} (D^s f)(x) \cdot g(x) dx \right| \leq C \int_{\mathbb{R}} (\mathcal{W}_{\omega_n}^s f)(x) (\mathcal{W}_{D^s \omega_n} g)(x) dx.$$

But $\omega_n \in S$ are band-limited orthonormal wavelet packets. Therefore, we can apply Hölder's inequality Theorem 4.2.9 with $\omega_n = D^s \omega_n$ and $\omega_1 = \omega_0$ and Theorem 4.2.7 and obtain

$$\begin{aligned}
\left| \int_{\mathbb{R}} (D^s f)(x) \cdot g(x) dx \right| &\leq C \|\mathcal{W}_{\omega_n}^s f\|_{L^p} \|\mathcal{W}_{D^s \omega_n} g\|_{L^q} \\
&\leq C \|\mathcal{W}_{\omega_n}^s f\|_{L^p} \|\mathcal{W}_{\omega_n} g\|_{L^q} \leq C \|\mathcal{W}_{\omega_n}^s f\|_{L^p} \|g\|_{L^q}.
\end{aligned}$$

Taking the supremum over all $g \in S$ such that $\|g\|_{L^q} \leq 1$, we deduce that

$$\|D^s f\|_{L^p} \leq C \|\mathcal{W}_{\omega_n}^s f\|_{L^p}.$$

Clearly, $(\mathcal{W}_{\omega_n}^s f)(x) \leq (\mathcal{W}_{\omega_n}^s f)(x)$. Since $1 \leq (1 + 2^{2\ell s})$, for all $\ell \in \mathbb{Z}$, by Theorem 4.2.7, we get

$$\|f\|_{L^p} \leq C \|\mathcal{W}_{\omega_n} f\|_{L^p} \leq C \|\mathcal{W}_{\omega_n}^s f\|_{L^p}.$$

Remark 4.4.6 The above Theorem gives a characterization of $L^{p,s}(\mathbb{R})$, $1 < p < \infty$, $s = 1, 2, \dots$, using band-limited orthonormal wavelet packets that belong to the Schwartz class S . This Theorem can be extended to more general wavelet packets.

Theorem 4.4.7 Let $s = 1, 2, \dots$ and $\ell, k, \ell', k' \in \mathbb{Z}$ and $n = 2^u, 2^u + 1, \dots, 2^{u+1}-1$, where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$, $j \in \mathbb{Z}$. Then:

(a) If $\omega_0 \in \mathcal{D}^s$ and $\omega_n \in \mathcal{M}^s$, there exist constants $C < \infty$ and $\varepsilon > 0$ such that

$$|\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle| \leq \frac{C 2^{(\ell'-\ell)(\frac{1}{2}+s+1)}}{(1+2^{\ell'}|2^{-\ell}k - 2^{-\ell'}k'|)^{1+\varepsilon}} \quad \text{for } \ell \geq \ell'.$$

(b) If $\omega_n \in \mathcal{D}^{-1}$, $n = 0, 1, 2, \dots$, there exist constants $C < \infty$ and $\varepsilon > 0$ such that

$$|\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle| \leq \frac{C 2^{\frac{1}{2}(\ell-\ell')}}{(1+2^{\ell}|2^{-\ell}k - 2^{-\ell'}k'|)^{1+\varepsilon}} \quad \text{for } \ell \leq \ell'.$$

Proof Assume that ω_0 is associated with the constants $\varepsilon' > 0$ and C'_m , $m = 0, 1, \dots, N+1$ and ω_n are associated with the constants $\gamma > 0$ and $C' < \infty$. We choose

$$C = \max\{C'_0, \dots, C'_{N+1}, C'\} \text{ and } \varepsilon = \min\{\varepsilon', \gamma\}.$$

Then, $\omega_0 \in \mathcal{D}^s$ with constants C for all $m = 0, 1, \dots, N+1$ and $\varepsilon > 0$ and $\omega_n \in \mathcal{M}^s$ with constant C and ($\gamma \geq \varepsilon$).

For a function g defined on \mathbb{R} , we write $\tilde{g}(x) = \overline{g(-x)}$. Then, we have

$$\overline{\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle} = \langle \omega_{\ell',0,k'}, \omega_{\ell,n,k} \rangle = (\omega_{\ell',0,k'} * \tilde{\omega}_{\ell,n,-k})(0).$$

Now, to prove part (a) apply Lemma 1.2.16 with $N = s$.

To prove part (b), apply Lemma 1.2.15 together with

$$\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle = (\omega_{\ell,n,k} * \tilde{\omega}_{\ell',0,-k'})(0).$$

Theorem 4.4.8 Let $s = 1, 2, 3, \dots$ and $\omega_n \in \mathcal{D}^s \cap \mathcal{M}^s$, for all $n = 0, 1, \dots$. Assume that ω_0 is an orthonormal wavelet packet. Then, for $1 < p < \infty$, there exists a constant $C_{p,s}$, $0 < C_{p,s} < \infty$, such that

$$\|\mathcal{W}_{\omega_n}^s f\|_{L^p(\mathbb{R})} \leq C_{p,s} \left\| \mathcal{W}_{\omega_0}^s f \right\|_{L^p(\mathbb{R})} \quad (4.4.13)$$

for all $f \in L^{p,s}(\mathbb{R})$, where $\mathcal{W}_{\omega_n}^s f$ are defined by

$$(\mathcal{W}_{\omega_n}^s f)(x) = \left\{ \sum_{\ell,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 (1+2^{2\ell s}) 2^\ell \chi_{I_{\ell,k}}(x) \right\}^{1/2} \quad (4.4.14)$$

where $I_{\ell,k} = [2^{-\ell}k, 2^{-\ell}(k+1)]$ and $\ell = j-u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$.

Proof It is sufficient to prove the result for $\widetilde{\mathcal{W}}_{\omega_n}^s f$ instead of $\mathcal{W}_{\omega_n}^s f$, where

$$(\widetilde{\mathcal{W}}_{\omega_n}^s f)(x) = \left\{ \sum_{\ell,k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |\langle f, \omega_{\ell,n,k} \rangle|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}}(x) \right\}^{1/2}$$

that is,

$$\|\widetilde{\mathcal{W}}_{\omega_n}^s f\|_{L^p(\mathbb{R})} \leq C_{p,s} \|\widetilde{\mathcal{W}}_{\omega_0}^s f\|_{L^p(\mathbb{R})} \quad (4.4.15)$$

for all $f \in L^{p,s}(\mathbb{R})$.

We assume that (4.4.15) is true, by Theorem 4.2.9 we have

$$\begin{aligned} \|\mathcal{W}_{\omega_n}^s f\|_{L^p} &\leq \|\mathcal{W}_{\omega_n} f\|_{L^p} + \|\widetilde{\mathcal{W}}_{\omega_n}^s f\|_{L^p} \leq C_1 \|\mathcal{W}_{\omega_0} f\|_{L^p} + C_2 \|\widetilde{\mathcal{W}}_{\omega_0}^s f\|_{L^p} \\ &\leq C \left\{ \|\mathcal{W}_{\omega_0} f\|_{L^p} + \|\mathcal{W}_{\omega_0}^s f\|_{L^p} \right\} = 2C \|\mathcal{W}_{\omega_0}^s f\|_{L^p} \end{aligned}$$

Since ω_0 is an orthonormal wavelet packet, we have

$$\omega_{\ell,n,k}(x) = \sum_{\ell',k' \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle \omega_{\ell',0,k'}(x)$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $j, k \in \mathbb{Z}$.

Thus

$$(\mathcal{W}_{\omega_n}^s f)(x) = \left\{ \sum_{\ell,k \in \mathbb{Z}} \left| \sum_{\ell',k' \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} \langle f, \omega_{\ell',0,k'} \rangle \overline{\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle} \right|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}}(x) \right\}^{1/2}$$

where $I_{\ell,k} = [2^{-\ell}k, 2^{-\ell}(k+1)]$. Writing

$$A_1(\ell, n, k) = \sum_{\ell' \leq \ell} \sum_{k' \in \mathbb{Z}} \langle f, \omega_{\ell',0,k'} \rangle \overline{\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle}$$

and

$$A_2(\ell, n, k) = \sum_{\ell' > \ell} \sum_{k' \in \mathbb{Z}} \langle f, \omega_{\ell',0,k'} \rangle \overline{\langle \omega_{\ell,n,k}, \omega_{\ell',0,k'} \rangle}$$

we have

$$\begin{aligned} (\widetilde{\mathcal{W}}_{\omega_n}^s f)(x) &\leq \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(\ell, n, k)|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}}(x) \right\}^{1/2} \\ &\quad + \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(\ell, n, k)|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}}(x) \right\}^{1/2} \end{aligned} \quad (4.4.16)$$

where $\ell = j - u$, $u = 0$ if $j \leq 0$ and $u = 0, 1, 2, \dots, j$ if $j > 0$ and $j, k \in \mathbb{Z}$.

For estimation of $A_1(\ell, n, k)$, we use Theorem 4.4.7(a) and obtain

$$\begin{aligned}|A_1(\ell, n, k)| &\leq C \sum_{\ell' \leq \ell} \sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle| \frac{2^{(\ell'-\ell)(\frac{1}{2}+s+1)}}{(1+2^{\ell'}|2^{-\ell}k - 2^{-\ell'}k'|)^{1+\varepsilon}} \\&= C \sum_{\ell' \leq \ell} 2^{(\ell'-\ell)(\frac{1}{2}+s+1)} \left\{ \sum_{k' \in \mathbb{Z}} \frac{|\langle f, \omega_{\ell', 0, k'} \rangle|}{(1+2^{\ell'}|2^{-\ell}k - 2^{-\ell'}k'|)^{1+\varepsilon}} \right\}\end{aligned}$$

for some $C < \infty$ and $\varepsilon > 0$. By applying Lemma 1.2.13(a) with $r = 1$, we obtain

$$|A_1(\ell, n, k)| \leq C \sum_{\ell' \leq \ell} 2^{(\ell'-\ell)(\frac{1}{2}+s+1)} \left[\mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle| \chi_{I_{\ell', k'}} \right) (x) \right]$$

for all $x \in I_{\ell, k}$. Since $\{I_{\ell, k} : k \in \mathbb{Z}\}$ is a collection of disjoint dyadic intervals, we have

$$\begin{aligned}&\left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(\ell, n, k)|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell, k}} \right\}^{1/2} \right\|_{L^p} \\&\leq C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} 2^{2\ell s} 2^\ell \left[\sum_{\ell' \leq \ell} 2^{(\ell'-\ell)(\frac{1}{2}+s+1)} \mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle| \chi_{I_{\ell', k'}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p} \\&= C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \left[\sum_{\ell' \leq \ell} 2^{\ell'-\ell} \mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle| 2^{\ell' s} 2^{\ell'/2} \chi_{I_{\ell', k'}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p} \\&\leq C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle| 2^{\ell' s} 2^{\ell'/2} \chi_{I_{\ell', k'}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p}\end{aligned}$$

where we have used Young's inequality for convolutions

$$\|\{a_\ell\} * \{b_{\ell'}\}\|_{\ell^2} \equiv \left\| \left\{ \sum_{\ell'} a_{\ell-\ell'} b_{\ell'} \right\} \right\|_{\ell^2} \leq \|\{a_\ell\}\|_{\ell^1} \|\{b_{\ell'}\}\|_{\ell^2} \quad (4.4.17)$$

with

$$a_\ell = \begin{cases} 2^{-\ell} & \text{if } \ell \geq 0 \\ 0 & \text{if } \ell < 0 \end{cases}$$

and

$$b_{\ell'} = \mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle| 2^{\ell' s} 2^{\ell'/2} \chi_{I_{\ell', k'}} \right) (x).$$

Now, using the vector-valued inequality for the Hardy–Littlewood maximal function with $q = 2$, we obtain

$$\begin{aligned} & \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_1(\ell, n, k)|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}} \right\}^{1/2} \right\|_{L^p} \\ & \leq C_p \left\| \left\{ \sum_{\ell' \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell',0,k'} \rangle|^2 2^{2\ell' s} 2^{\ell'} \chi_{I_{\ell',k'}} \right\}^{1/2} \right\|_{L^p} \\ & = C_p \left\| \widetilde{\mathcal{W}}_{\omega_0}^s f \right\|_{L^p}. \end{aligned} \quad (4.4.18)$$

For estimation of $A_2(\ell, n, k)$, we use Theorem 4.4.7(b) (\mathcal{D}^s and \mathcal{M}^s are contained in \mathcal{D}^{-1}), together with Lemma 1.2.13(b) with $r = 1$ and obtain

$$\begin{aligned} |A_2(\ell, n, k)| & \leq C \sum_{\ell' > \ell} \sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell',0,k'} \rangle| \frac{2^{\frac{1}{2}(\ell-\ell')}}{(1 + 2^\ell |2^{-\ell}k - 2^{-\ell'}k'|)^{1+\varepsilon}} \\ & \leq C \sum_{\ell' > \ell} 2^{\frac{1}{2}(\ell-\ell')} 2^{\ell'-\ell} \left[\mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell',0,k'} \rangle| \chi_{I_{\ell',k'}} \right)(x) \right] \end{aligned}$$

for some $C < \infty$ and $\varepsilon > 0$ and for all $x \in I_{\ell,k}$. Since $\{I_{\ell,k} : k \in \mathbb{Z}\}$ is a collection of disjoint dyadic intervals, we have

$$\begin{aligned} & \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(\ell, n, k)|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}} \right\}^{1/2} \right\|_{L^p} \\ & \leq C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} 2^{2\ell s} 2^\ell \left[\sum_{\ell' > \ell} 2^{-\frac{1}{2}(\ell-\ell')} \mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell',0,k'} \rangle| \chi_{I_{\ell',k'}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p} \\ & = C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \left[\sum_{\ell' > \ell} 2^{(\ell-\ell')s} \mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell',0,k'} \rangle| 2^{\ell' s} 2^{\ell'/2} \chi_{I_{\ell',k'}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p}. \end{aligned}$$

The series $\sum_{\ell' > \ell} 2^{(\ell-\ell')s}$ converges because $s \geq 1$. Therefore, by using Young's inequality for convolutions and the vector-valued inequality for the Hardy–Littlewood maximal function (Lemma 1.2.9) with $q = 2$, we obtain

$$\begin{aligned}
& \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \sum_{n=2^u}^{2^{u+1}-1} |A_2(\ell, n, k)|^2 2^{2\ell s} 2^\ell \chi_{I_{\ell,k}} \right\}^{1/2} \right\|_{L^p} \\
& \leq C \left\| \left\{ \sum_{\ell' \in \mathbb{Z}} \left[\mathcal{M} \left(\sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle|^2 2^{\ell' s} 2^{\ell'/2} \chi_{I_{\ell', k'}} \right) \right]^2 \right\}^{1/2} \right\|_{L^p} \\
& \leq C \left\| \left\{ \sum_{\ell \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} |\langle f, \omega_{\ell', 0, k'} \rangle|^2 2^{2\ell' s} 2^{\ell'} \chi_{I_{\ell', k'}} \right\}^{1/2} \right\|_{L^p} \\
& = C \left\| \widetilde{\mathcal{W}}_{\omega_0}^s f \right\|_{L^p}. \tag{4.4.19}
\end{aligned}$$

Finally, inequality (4.4.15) follows from (4.4.16), (4.4.18), and (4.4.19).

Theorem 4.4.9 *Let $s = 1, 2, 3, \dots$, and suppose that ω_n be orthonormal wavelet packets such that $\omega_n \in \mathbb{R}^s$, $n = 0, 1, 2, \dots$. Then, for $1 < p < \infty$, there exist two constants $A_{p,s}$ and $B_{p,s}$, $0 < A_{p,s} \leq B_{p,s} < \infty$, such that*

$$A_{p,s} \|f\|_{L^{p,s}(\mathbb{R})} \leq \left\| \mathcal{W}_{\omega_n}^s(f) \right\|_{L^p} \leq B_{p,s} \|f\|_{L^{p,s}}$$

for all $f \in L^{p,s}(\mathbb{R})$, where $\mathcal{W}_{\omega_n}^s(f)$ is defined by (4.4.11).

Proof Applying Theorems 4.4.5 and 4.4.7, we observe that all band-limited wavelet packets which belong to Schwartz class “S” are contained in \mathbb{R}^s .

Chapter 5

Applications in Signal Processing



5.1 Introduction

Signal processing has become an essential part of contemporary scientific and technological activity. Signal processing is used in telecommunications (telephone and television), in the transmission and analysis of satellite images, and in medical imaging (echography, tomography, and nuclear magnetic resonance), all of which involve the analysis and interpretation of complex time series. A record of stock price fluctuations is a signal, as is a record of temperature readings that permit the analysis of climatic variations and the study of global warming.

The signals we study will always be series of numbers and not series of letters, words, or phrases. These numbers come from measurements, which are typically made using some recording method. The signals ultimately appear as functions of time.

A “signal” can be defined as a function that conveys information, generally about the state or behavior of a physical system. Although signals can be represented in many ways, in all cases, the information is contained in a pattern of variations of some forms. Signals are represented mathematically as functions of one or more independent variables. The independent variable of the mathematical representation of a signal may be either continuous or discrete. “Continuous time signals” are signals that are defined at a scale of time and thus are represented by continuous variable functions. “Discrete time signals” are defined at discrete time, and thus, the independent variable takes place on only discrete values; i.e., discrete time signals are represented as sequences of numbers.

In addition to the fact that the independent variables can be either continuous or discrete, the signal amplitude (dependent variable) may be either continuous or discrete. “Digital signals” are those for which both time and amplitude are discrete. Continuous time and continuous amplitude signals are called “analog signals.”

Development of signal processing techniques and systems is of great importance. These techniques usually take the form of a transformation of a signal into another

signal that is in some sense more desirable than the original. For example, we may wish to design transformations for separating two or more signals that have been combined in some way; we may wish to enhance some component or parameter of a signal, or we may wish to estimate one or more parameters of a signal.

5.2 Applications of Wavelets in Speech Denoising

5.2.1 *Speech Signal Processing*

In almost every area of science and technology, signals must be processed to assist the extraction of information. Thus, the development of signal processing techniques and systems is of great importance. From the ancient time, when human invented spoken languages, speech has been the most direct way for human to convey information to one another. Up to the present, the communication using spoken speech is still the most dominant and common service in the communication networks. The speech signal is now extended, through technological mediation such as telephone, movies, radio, television, and the Internet.

In particular, the speedy enlargements of very large-scale integrated circuit (VLSI) and personal computers (PCs) have made the digital speech processing becoming a remarkable progress. When these devices are used under noisy conditions, the quality and intelligibility of a speech signal greatly reduced. There are two major sources of noises: The first one is the environment noise that is recorded with real speech signal by a microphone, and the second one is the noise that is caused due to the property of microphone and the recording system. If either of them becomes large compared with the observed speech signal, the quality of speech recognition and processing system becomes considerably worse. Generally, the researches of speech signal processing techniques can be divided into several areas. They are: digital transmission and storage of speech, speech synthesis system, speaker verification and identification, speech recognition system, aids to the handicapped, enhancement of speech signal quality.

Although a great deal of work on digital speech signal processing has been proposed in the literature, the accuracy and robustness (denoising) of speech signal processing still remain an open problem. The main reason is that the human speech production system is time varying, and the natural speech signal is a nonstationary random process.

Digital signal processing is a field, which came into existence in the seventeenth- and eighteenth-century mathematics. It has become an important tool in diverse fields of science and technology. The field of digital signal processing has grown a great deal in the past decade to cover and provide firm theoretical backgrounds for a large number of individual areas [209].

The term “digital signal processing” may have different meanings for different people. For example, a binary bit system can be considered a “digital signal” and the

various manipulations, or “signal processing”, performed at the bit level by digital hardware may be constructed as “digital signal processing.” Implicit in the definition of digital signal processing (DSP) is the notion of an information-bearing signal that has an analog counterpart, which is manipulated samples of this implicitly analog signal. Further, these samples are quantized, i.e., represented using finite precision, with each word representative of the value of the sample of an analog signal. These manipulations, or filters, performed on these samples are arithmetic in nature—additions and multiplications. The definition of DSP includes the processing associated with sampling, conversation between analog and digital domains, and changes in word length [248]. Digital signal processing is concerned with the representation of signals by sequences of numbers or symbols and the processing of these sequences. The purpose of such processing may be to estimate characteristic parameters of a signal or to transform a signal into a form which is in some sense more desirable.

In general, signal processing has a rich history, and its importance is evident in such diverse fields as engineering, acoustics, sonar, radar, seismology, speech communication, data communication, nuclear science, and many others. In many applications, as, for example, in speech transmission and speech recognition, we may wish to remove interference, such as noise, from the signal or to modify the signal to present it in a form which is more easily interpreted. As another example, a signal transmitted over a communication channel is generally perturbed in a variety of ways, including channel distortion, fading, and the insertion of background noise. One of the objectives at the receiver is to compensate for these disturbances. In each case, processing of the signal is required.

There are many reasons why digital signal processing of an analog signal may be preferable to processing the signal directly in the analog domain. First, a digital programmable system allows flexibility in reconfiguring the digital signal processing operations simply by changing the program. Reconfiguration of an analog system usually implies a redesign of the hardware followed by testing and verification to see that it operates properly.

Accuracy considerations also play an important role in determining the form of the signal processor. Tolerances in analog circuit components make it extremely difficult for the system designer to control the accuracy of an analog signal processing system. On the other hand, a digital system provides much better control of accuracy requirements.

Digital signals are easily stored on magnetic media (tape or disk) without deterioration or loss of signal fidelity beyond that introduced in the A/D conversion. As a consequence, the signals become transportable and can be processed off-line in a remote laboratory. The digital signal processing method also allows for the implementation of more sophisticated signal processing algorithms. It is usually very difficult to perform precise mathematical operations on signals in analog form, but these operations can be routinely implemented on a digital computer using software. The digital implementation of the signal processing system is almost always cheaper than its analog counterpart as a result of the flexibility for modifications. As a consequence of these advantages, digital signal processing has been applied in practical systems covering a broad range of disciplines.

The techniques and applications of digital signal processing are expanding at a tremendous rate. With the advent of large-scale integration and the resulting reduction in cost and size of digital components, together with increasing speed, the class of applications of digital signal processing techniques is growing.

Signal processing systems may be classified along the same lines as signals; i.e., “continuous time systems” are systems for which both input and output are continuous signals and “Discrete time systems” are those for which the input and output are discrete time signals. Similarly, analog systems are systems for which the input and output are analog signals and “digital systems” are those for which the input and output are digital signals. “Digital signal processing,” then, deals with transformations of signals that are discrete in both amplitude and time [209].

Discrete time signals may arise by sampling a continuous time signal, or they may be generated directly by some discrete time process. Whatever the origin of the discrete time signals, digital signal processing systems have many attractive features, as noted above. They can be realized with great flexibility using general-purpose digital computers, or they can be realized with digital hardware. Thus, digital representations of signals are often desirable when complicated signal processing is required.

Based on the multiresolution analysis (MRA), a signal is decomposed into an approximation and details at various scales. In other words, various information levels across successive resolutions can be extracted by decomposing the original signal using an orthonormal wavelet basis. As a conventional transform, the well-known Fourier transform and short-time Fourier transform (STFT) are used widely in mathematics and engineering. A limitation of the STFT is that, because a single window is used for all frequencies, the resolution of the analysis is the same at all locations in the time–frequency plan [282]. In speech and audio signals since human hearing system uses a frequency-dependent resolution, so these methods have poor effect. Amplitude modulation (AM)-based radio channels and frequency modulation (FM)-based radio channels may be one of the best examples. The discrete wavelet transform (DWT) can solve this problem with the rectangular tiling of the time–frequency plane.

In mathematics, research on singularities and irregular structures is very necessary because they often carry the most useful information in signals (e.g., transient, discontinuous, and nonstationary sounds). The greatest challenge is the selection of appropriate techniques which are able to study irregularities of signal structures. Until now, the Fourier transform was the main mathematical tool for analyzing singularities. However, the Fourier transform with sinusoidal waveforms extending over a fixed window length provides only a description of the global regularity of signals without well adapting to the localization of singularities in the time–frequency domain. This motivates the study of the wavelet transform which can characterize the local regularity of signals by decomposing signals into well-localized time–frequency components. As proved in [191], the detection of all the singularities of the signal is based on the local maxima property which is measured from the evolution across scales of these maxima. The detection of singularities with multiscale transforms has been studied not only in mathematics but also in signal processing

and application domains. The flexible analysis in the time–frequency plane of the DWT in comparison with the STFT [191, 282] shows its advantages for speech processing. With Heisenberg’s uncertainty principle, it is known that no transform can provide high resolution in both time and frequency domains at the same time. The useful locality property is broken in this context. Because the wavelet basis functions are short waves and generated by scaling from the mother wavelet, they are well-localized in time and scale domains. This automatic behavior of wavelet decomposition is absolutely suitable for processing of speech signals which requires high frequency resolution to analyze low-frequency components (voiced sounds, formant frequencies) and high temporal resolution to analyze high-frequency components (mostly unvoiced sounds). This elegant behavior of the DWT is employed for auditory representations of acoustic signals [298], speech coding [48, 185] using wavelet packet representation in the context of auditory modeling, speech segmentation [272], and phonetic classification [57, 214, 216]. In recent years, the wavelet shrinking approach to speech enhancement has been developed rapidly, starting with the simple hard thresholding proposed by Donoho [97]. Many improvements in wavelet thresholding to enhance speech signal have been done as semisoft thresholding with selected threshold for unvoiced regions [237], efficient hard and soft thresholding [304], smooth hard thresholding function based on μ -law [52, 216, 246]. In [177], the combination of soft and hard thresholding is applied to adapt with different properties of the speech signal. Motivated by lower speech distortion, wavelet thresholding or shrinkage methods are integrated with other techniques such as the Teager energy operator and masked adaptive threshold [26]. Critical-band wavelet decomposition is used with noise masking threshold in [176], and perceptual wavelet packet decomposition (PWD) which simulates the critical bands of the psychoacoustic model is proposed in [54, 147]. A blind adaptive filter of speech from noise is designed in the wavelet domain in [281]. Dealing with enhancement and feature extraction for robustness, several parameterization methods which are based on the DWT and wavelet packet decomposition (WPD) have been proposed in [126, 133]. More complicated shrinking functions with better characteristics than soft and hard thresholding are optimized for speech enhancement and speech recognition [172]. The usage of wavelet-based features extracted from the WPD leads to the improvement of recognition rate compared with the well-known conventional feature Mel-frequency cepstral coefficients (MFCCs) [52, 173, 234]. In addition to this approach, wavelet denoising is applied as preprocessing stage before feature extraction to balance environmental mismatches [33, 108].

A speech signal originates as subtle time variations of air pressure and becomes a curve whose complex graphical characteristics are an “adapted copy” of the voice.

Experts in signal processing are called on to describe, for a given class of signals, algorithms that lead to the construction of microprocessors and that allow certain operations and tasks to be done automatically. These tasks may be: analysis and diagnostics, coding, quantization and compression, transmission or storage, and synthesis and reconstruction.

To specify a diagnostic, it is essential to analyze, and then to erase, these natural fluctuations (which play the role of noise) in order to have access to the “artificial”

heating of the planet resulting from human activity. The diagnostic often depends on extracting a small number of significant parameters from a signal whose complexity and size are overwhelming. Thus, the analysis and the diagnostic rely naturally on data compression. If this compression is done inappropriately, it can falsify the diagnostic.

Data compression also occurs in the problem of transmission. Indeed, transmission channels have a limited capacity, and it is therefore important to reduce as much as possible the abundance of raw information so that it fits within the channel's "bit allocation."

We now consider problems posed by coding and quantization. Different coding algorithms will be presented and studied in this work: subband coding, transform coding, and coding by zero crossings. In each case, coding involves methods to transform the recorded numerical signal into another representation that is, depending on the nature of the signals studied, more convenient for some task or further processing. Quantization is associated with coding. The "exact" numerical values given by coding are replaced with nearby values that are compatible with the bit allocation dictated by the transmission capacity.

Quantization is an unavoidable step in signal and image processing. Unfortunately, it introduces systematic errors, known as "quantization noise." The coding algorithms that are used (taking into account the nature of the signals) ought to reduce the effects of quantization noise when decoding takes place. One of the advantages of quadrature mirror filters is that they "trap" this quantization noise inside well-defined frequency channels.

We have just defined a set of tasks, or operations, to be performed on signals or images. These tasks form a coherent collection. The purpose of this section is to describe certain coding algorithms that have, during the last few years, been shown to be particularly effective for analyzing signals having a fractal structure or for compression and storage. We will also describe certain "meta-algorithms" that allow one to choose the coding algorithm best suited to a given signal. To better approach this problem of choosing an adaptive algorithm, we briefly classify signals by distinguishing stationary signals, quasi-stationary signals, and transient signals.

A signal is stationary if its properties are statistically invariant over time. A well-known stationary signal is white noise, which, in its sampled form, appears as a series of independent drawings. A stationary signal can exhibit unexpected events, but we know in advance the probabilities of these events. These are the statistically predictable unknowns.

The ideal tool for studying stationary signals is the Fourier transform. In other words, stationary signals decompose canonically into linear combinations of waves (sines and cosines). In the same way, signals that are not stationary decompose into linear combinations of wavelets.

The study of nonstationary signals, where transient events appear that cannot be predicted (even statistically with knowledge of the past), necessitates techniques different from Fourier analysis. These techniques, which are specific to the nonstationarity of the signal, include wavelets of the "time–frequencies" type and wavelets of the "timescale" type. "Time–frequency" wavelets are suited, most specifically, to

the analysis of quasi-stationary signals, while “timescale” wavelets are adapted to signals having a fractal structure.

Before defining “time–frequency” wavelets and “timescale,” we indicate their common points. They belong to a more general class of algorithms that are encountered as often in mathematics and in speech processing. Mathematicians speak of “atomic decompositions,” while speech specialists speak of “decompositions in time–frequency atoms”; the specific reality is the same in both cases.

An “atomic decomposition” consists in extracting the simple constituents that make up a complicated mixture. However, contrary to what happens in chemistry, the “atoms” that are discovered in a signal will depend on the point of view adapted for the analysis. These “atoms” will be “time–frequency atoms” when we study quasi-stationary signals, but they could, in other situations, be replaced by “timescale wavelets” or “Grossman–Morlet wavelets.”

5.2.2 Noise Reduction in Speech Signal Using Thresholding in Wavelet Transform Domain

The presence of noise in speech can significantly reduce the intelligibility of speech and degrade automatic speech recognition (ASR) performance. These noises may be due to the background noise of the environment in which the speaker is speaking, or it may be introduced by the transmission media during transmission of the speech signal. It is often necessary to perform speech denoising as the presence of noise, which severely degrades the speech signal.

Reduction of noise has become an important issue in speech signal processing system, such as speech coding and speech recognition system. A number of methods for the noise reduction in speech signal have been proposed in the literature based on spectral subtraction [34, 38, 125, 161] and adaptive filtering [180]. Spectral subtraction method is one of the simplest methods employed to reduce noise in speech. This technique is based on the direct estimation of the short-term spectral magnitude [34]. Despite its ability of canceling background noise, it introduces additional artifacts called musical noise and faces difficulties in pause detection. This problem is caused due to the inaccuracies in the short-time noise spectrum estimate. Adaptive techniques to reduce noise are effective only when the reference noise is highly correlated with the corrupting noise. However, due to highly random nature of the corrupting noise, it is difficult to estimate it. Discrete Fourier transform (DFT) has also been used widely for signal analysis and speech denoising in recent years [302]. However, for nonperiodic time-varying signals DFT has several inconveniences.

In recent years, wavelet transform [284] has become a powerful tool for the multiscale representation and analysis of signals. Wavelet transform localizes the information in the time–frequency plane; in particular, they are capable of trading one type of resolution for another that makes them especially suitable for speech signal analysis. A number of researchers have applied this new technique for the

reduction of noise in speech signals [99, 123, 124, 246]. Denoising of speech signals using wavelet transform is usually based on thresholding and shrinking wavelet coefficients of noisy signals. However, the critical problems are to choose right wavelet, determining appropriate threshold value and level of decomposition. A number of papers have been reported in this field that focus on the determination of threshold value and level of decomposition [22, 98, 276].

In this section, a speech denoising method based on wavelet decomposition of speech signal is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficient of noisy speech signal and the original noise-free signal. The proposed method can effectively remove the noise from noisy speech signal degraded by additive white noise. To evaluate the denoising performance of the proposed method, mean square error (MSE) is computed at different values of signal-to-noise ratio (SNR). The simulation is done using MATLAB 7.0 version.

Wavelet transform is successfully applied to nonstationary signal processing and provides an alternative to the short-time Fourier transform (STFT). STFT uses a single window for signal analysis, whereas the wavelet transform uses short windows at high frequencies and long windows at low frequencies. This results in a high-frequency resolution in low frequency band and low-frequency resolution in high frequency band. This makes wavelet transform a powerful tool for modeling nonstationary signal like speech that exhibits slow temporal variation at low frequency and abrupt temporal changes at high frequency.

There are number of wavelets proposed in the literature. However, in this section some of the most popular wavelet functions are explored for the reduction of noise in speech signal, such as Daubechies, Symlets, and Coiflets wavelet functions [284].

The Proposed Method

In the present section, soft thresholding is used for the shrinkage of wavelet coefficients. The proposed method is implemented using the following steps.

1. Consider a clean speech signal $x(n)$.
2. Generate a random white Gaussian noise $w(n)$, and add it to the original signal $x(n)$; mathematically, it can be written as

$$y(n) = x(n) + w(n).$$

3. Compute the discrete wavelet transform of the noisy speech signal $y(n)$.
4. Choose a threshold value for thresholding. The optimal threshold is chosen by taking the minimum error between the detailed coefficients of noisy signal and those for original signal. A soft thresholding is used to shrink the wavelet detailed coefficient of the noisy signal.
5. The original speech signal is reconstructed by taking the inverse discrete wavelet transform.
6. To evaluate the performance of the proposed method, mean square error (MSE) between original signal and estimated signal is computed, which is given by

Fig. 5.1 Original speech signal

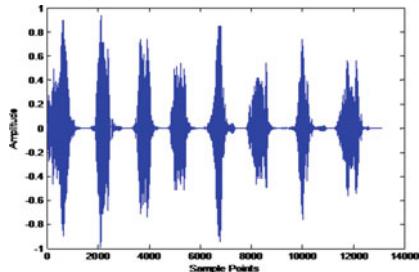
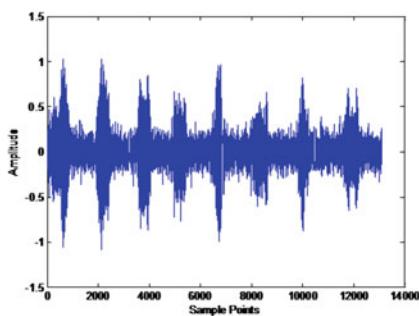


Fig. 5.2 Speech signal corrupted with noise



Haykin [138]

$$MSE = \frac{1}{N} \left(\sum_{n=0}^{N-1} (x(n) - \tilde{x}(n))^2 \right)$$

where $x(n)$ is the original speech signal and $\tilde{x}(n)$ is estimated speech signal obtained by the proposed method and N is number of sample points in the signal.

Simulation and Results

For the simulation of the proposed method, a bird's sound signal has been taken as original speech signal, shown in Fig. 5.1. White Gaussian noise (WGN) is used to model the background noise. This WGN is added to the original speech signal to introduce distortions. This noisy speech signal, shown in Fig. 5.2, is used as the test signal for the simulation of proposed method. The estimated signal from the noisy speech signal using proposed method is shown in Fig. 5.3. From Figs. 5.2 and 5.3, it is clear that the noise has been greatly reduced.

To study the performance of the proposed method, three different types of wavelet transforms are used for the estimation of the original signal from the noisy signal. These are Daubechies, Symlets, and Coiflets. The level of decomposition is arbitrarily chosen as five. To evaluate the performance of the proposed method, MSE is computed for various values of signal-to-noise ratio (SNR). The results obtained, from the simulation, are given in Table 5.1. For the graphical representation, a plot

Fig. 5.3 Estimated speech signal

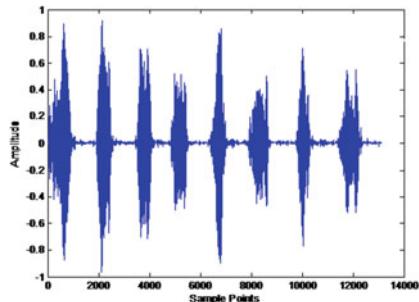


Table 5.1 Computation of MSE for SNR

S.No.	SNR(in dB)	Mean square error (MSE)		
		Daubechies(db2)	Coiflets (coif2)	Symlets (sym2)
1	0	0.0689	0.0715	0.0689
2	2	0.0572	0.0583	0.0572
3	5	0.0473	0.0479	0.0473
4	10	0.0404	0.0406	0.0404
5	15	0.0162	0.0151	0.0204
6	20	0.0088	0.0073	0.0154
7	25	0.0032	0.0026	0.0107
8	30	0.0011	0.0010	0.0011

mean square error verses SNR is shown in Fig. 5.4. From the plot, it is obvious that as the SNR increases, the MSE decreases; i.e., at high SNR, denoising of the speech signal is better than at low SNR.

For the comparison purpose, the results obtained from the proposed method are compared with results obtained by using Donoho thresholding [97]. Here, the wavelet transform used in both cases is Coiflets (coif2). MSE obtained at different SNRs in

Fig. 5.4 Performance of proposed method

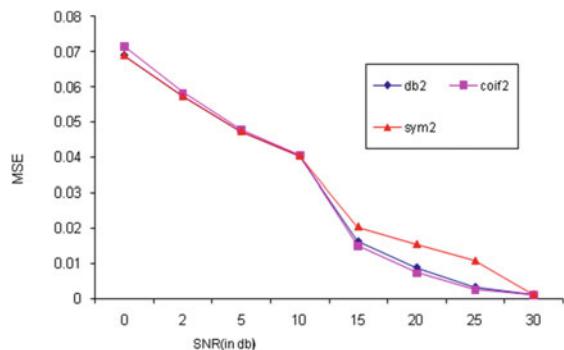
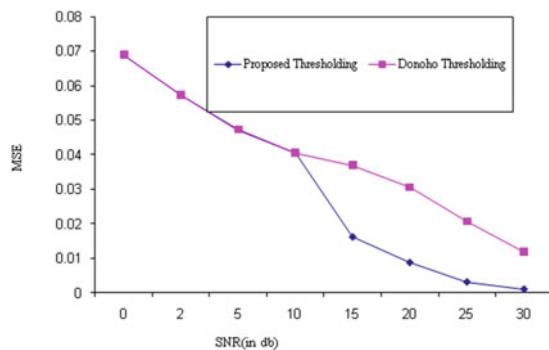


Table 5.2 Computation of MSE for SNR

S.No.	SNR(in dB)	Mean square error (MSE)	
		Donoho threshold	Proposed method
1	0	0.0709	0.0715
2	2	0.0585	0.0583
3	5	0.0475	0.0479
4	10	0.0405	0.0406
5	15	0.0366	0.0151
6	20	0.0301	0.0073
7	25	0.0199	0.0026
8	30	0.0108	0.0010

Fig. 5.5 Comparison of proposed method

both cases is given in Table 5.2, and a comparative plot is shown in Fig. 5.5. From the plot, it is observed that the proposed method gives similar result at lower SNR. However, it gives better result at higher SNR.

Conclusions

In the present work, a speech denoising method based on wavelet decomposition of speech signal is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy speech signal and the original noise-free signal. The three most popular wavelet transforms Daubechies (db2), Symlets (sym2), and Coiflets (coif2) are used. To evaluate the performance of the proposed method, mean square error is computed for different values of SNR. The result obtained is compared with Donoho's thresholding method, and it is found that the proposed method gives better result at high SNR.

5.2.3 Estimation of Speech Signal in the Presence of White Noise Using Wavelet Transform

In this section, a speech denoising method based on wavelet decomposition of speech signal is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy speech signal and the original noise-free signal. The proposed method can effectively remove the noise from noisy speech signal degraded by additive white noise. To evaluate the denoising performance of the proposed method, mean square error (MSE) is computed at different values of signal-to-noise ratio (SNR). The simulation is done using MATLAB 7.0 version.

The Proposed Method

Here, soft thresholding is used for the shrinkage of wavelet coefficients. The proposed method is implemented using the following steps:

1. Consider a clean speech signal $x(n)$.
 2. Generate a random white Gaussian noise $w(n)$, and add it to the clean speech signal $x(n)$, mathematically, it can be written as
- $$y(n) = x(n) + w(n).$$
3. Compute the discrete wavelet transform of the noisy speech signal $y(n)$.
 4. Choose a threshold value for thresholding. The optimal threshold value is chosen by taking the minimum error between the detailed coefficients of noisy signal and those for original signal. A soft thresholding is used to shrink the wavelet detailed coefficients of the noisy signal.
 5. The original speech signal is reconstructed by taking the inverse discrete wavelet transform.
 6. To further reduce the noise, it is passed through the Wiener filter.
 7. To evaluate the performance of the proposed method, mean square error (MSE) [138] between original signal and estimated signal is computed, which is given by

$$MSE = \frac{1}{N} \left(\sum_{n=0}^{N-1} (x(n) - \tilde{x}(n))^2 \right)$$

where $x(n)$ is the original speech signal and $\tilde{x}(n)$ is estimated speech signal obtained by the proposed method and N is number of samples in the signal.

Wiener Filter

Wiener filters compute linear estimators of signals embedded in additive noises while maximizing the average mean square error. The performances of these estimators are well suited for Gaussian as well as non-Gaussian processes.

Let $Y[n]$ be the signal which is obtained after step 5, with some noise $W[n]$ (zero mean white Gaussian noise of variance σ^2) of length N . Now, the noisy measurements are

$$Z[n] = Y[n] + W[n],$$

Then,

$$E\{W[n]W[k]\} = \sigma^2 \delta[n - k].$$

The signal and noise values $Y[n]$ and $W[k]$ are independent for any $0 \leq k, n < N$; we suppose that $E\{Y[n]\} = 0$. If not, then $E\{Y[n]\}$ is subtracted from $Z[n]$ to obtain a zero mean signal.

Let \tilde{Y} be an estimator of Y from the data Z . Now, the main aim is to minimize mean square error of the estimated signal or maximize signal-to-noise ratio of the estimated signal, which is measured in decibels by

$$\begin{aligned} SNR_{dB} &= 10 \log_{10} \left(\frac{E\{\|Y\|^2\}}{E\{\|Y - \tilde{Y}\|^2\}} \right) \\ &= 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} E\{\|Y[n]\|^2\}}{\sum_{n=0}^{N-1} E\{\|Y[n] - \tilde{Y}[n]\|^2\}} \right) \end{aligned}$$

For a fixed n , the estimator $\tilde{Y}[n]$ that minimizes the mean square error

$$E\{\|Y[n] - \tilde{Y}[n]\|^2\}$$

is the conditional expectation,

$$\tilde{Y}[n] = E\{Y[n]|Z[0], Z[1], \dots, Z[N-1]\}$$

The Wiener filter simplifies the estimation by computing the best estimator $\tilde{Y}[n]$ that is a linear combination of $\{Z[k]\}_{0 \leq k < N}$. The following theorem proves that the resulting error is uncorrelated with the data.

Theorem A linear estimator \tilde{X} of a random variable X :

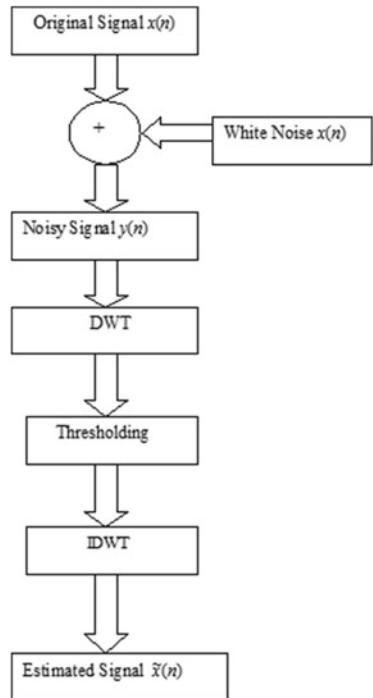
$$\tilde{X} = \sum_{k=0}^{N-1} h[k]Z[k]$$

minimizes $E\{(X - \tilde{X})^2\}$ if and only if

$$E\{(X - \tilde{X})Z[k]\} = 0 \text{ for } 0 \leq k < N \quad (5.2.1)$$

Proof Let us minimize

Fig. 5.6 Noise removal algorithm



$$Er(h) = E \left\{ \left(X - \sum_{k=0}^{N-1} h[k]Z[k] \right)^2 \right\} \quad (5.2.2)$$

at the minimum

$$\frac{\partial Er(h)}{\partial h[n]} = -2E \left(X - \sum_{k=0}^{N-1} h[k]Z[k] \right) Z[n] \quad (5.2.3)$$

which verifies Eq. (5.2.1), since

$$\frac{\partial^2 Er(h)}{\partial h[n]^2} = 2E \{ |Z[n]|^2 \} \geq 0.$$

This gives the minimum value of Eq. (5.2.2) (Fig. 5.6).

Simulation and Results

For the simulation of the proposed method, a male speech signal has been taken as original signal, shown in Fig. 5.7. The sampling frequency is 16 kHz, and 11025 samples of the signal are used. White Gaussian noise (WGN) is used to model the background noise. This WGN is added to the original speech signal to introduce

Fig. 5.7 Original speech signal

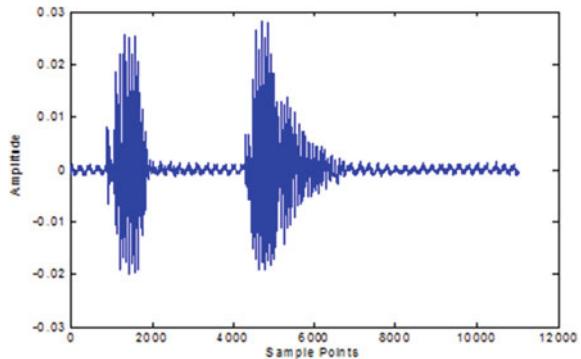
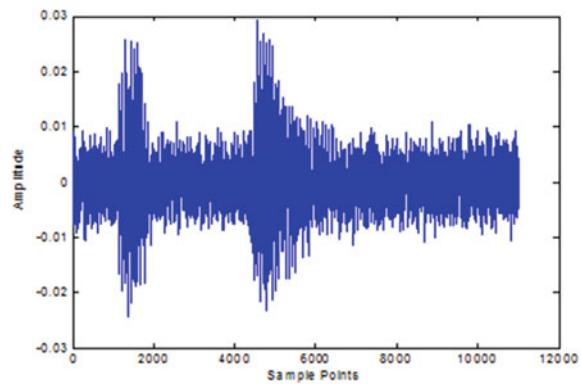


Fig. 5.8 Speech signal corrupted with noise



distortions. This noisy speech signal, shown in Fig. 5.8, is used as the test signal for the simulation of proposed method. The estimated signal from the noisy speech signal using proposed method is shown in Fig. 5.9. From the figure, it is clear that the noise has been greatly reduced. Daubechies (db4) wavelet is used to decompose the signal up to four levels.

To evaluate the performance of the proposed method, MSE is computed for various values of signal-to-noise ratio (SNR). The results obtained, from the simulation, are given in Table 5.3.

For the graphical representation, a plot of mean square error verses SNR is shown in Fig. 5.10. From the plot, it is observed that MSE of estimated signal has drastically reduced as compared to the MSE of noisy signal.

For the comparison purpose, the result obtained from the proposed method is compared with the result obtained by using Donoho's thresholding [97]. Here, the wavelet transform used in both cases is db4. A comparative plot is shown in Fig. 5.11. From the plot, it is observed that the proposed method gives better result as compared to Donoho's thresholding.

Fig. 5.9 Estimated speech signal

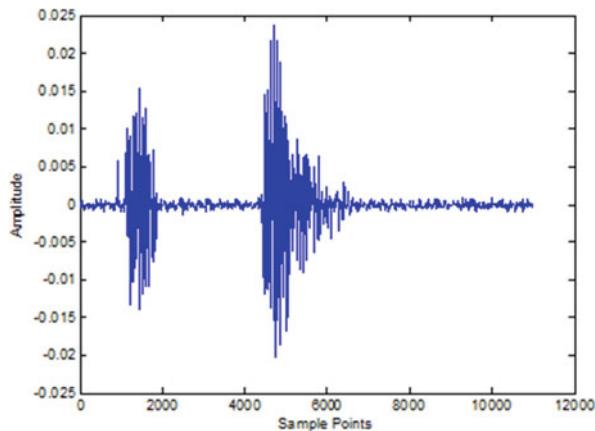


Table 5.3 Computation of MSE for SNR

S.No.	SNR(in dB)	Mean square error (MSE)		
		Noisy signal	Donoho's method	Proposed method
1	0	1.0009	0.4934	0.3280
2	5	0.5678	0.2780	0.1877
3	10	0.3196	0.1555	0.1033
4	15	0.1768	0.0909	0.0589
5	20	0.1004	0.0498	0.0331
6	25	0.0565	0.0289	0.0186
7	30	0.0317	0.0159	0.0112
8	35	0.0179	0.0095	0.0064
9	40	0.0100	0.0056	0.0042

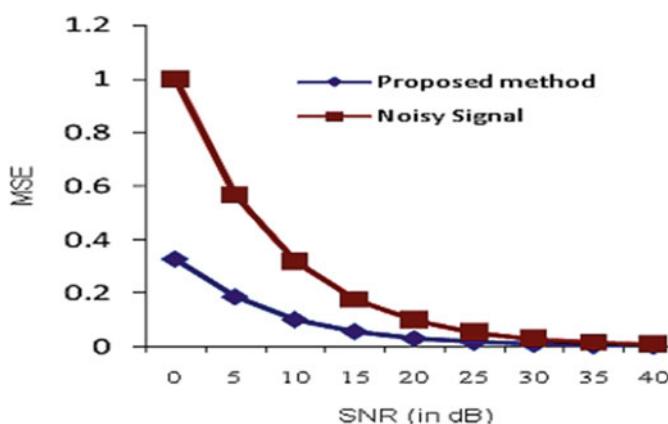


Fig. 5.10 Performance of proposed method

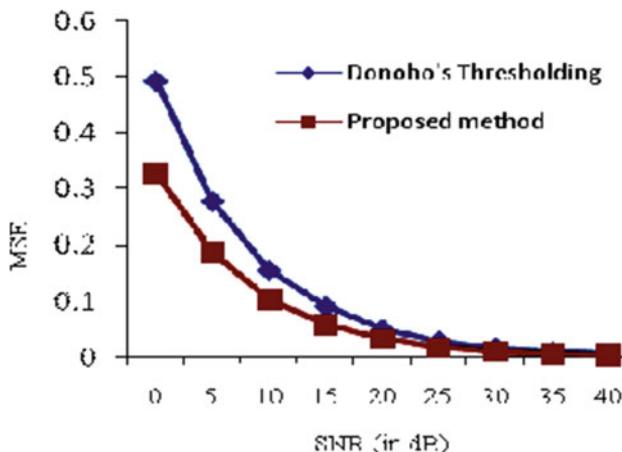


Fig. 5.11 Comparison of proposed method

Conclusions

In the present work, a speech denoising method based on wavelet decomposition of speech signal and Wiener filter as post filtering is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy speech signal and the original noise-free signal. A very popular wavelet transform proposed by Daubechies is used. To evaluate the performance of the proposed method, mean square error is computed for different values of SNR. The result obtained is compared with Donoho's thresholding method, and it is found that the proposed method gives better result.

5.3 Applications of Wavelet Packets in Speech Denoising

5.3.1 Denoising of Speech Signal by Using Wavelet Packet Transform

The purpose of the speech is to communicate the information. According to information theory, speech may be represented in terms of its message contents. In speech communication systems, the speech signal is transmitted, stored, and processed in many ways. In general, there are two major concerns in any speech communication systems [203].

1. Representation of speech signal in the form that is convenient for transmission and storage.
2. Preservation of message content in the speech signal.

In recent years, with the rapid development of information technology, many peripherals using speech interfaces, such as mobile phone and speech recognition software, have been developed. However, when these devices are used under noisy conditions, the quality and intelligibility of a speech signal are greatly reduced. There are two major sources of noises: The first one is the environment noise that is recorded with real speech signal by a microphone, and the second one is the noise that is caused due to the property of microphone and the recording system. If either of them becomes large compared with the observed speech signal, the quality of speech recognition and processing system becomes considerably worse. To overcome this important problem, some noise reduction techniques and speech enhancement methods have been developed, which is normally known as speech denoising.

Many approaches [23, 35, 38, 95, 98, 206] have been developed to address the problem of denoising speech signals that are degraded by noise. The fractional Fourier transform (FRFT) method has been developed for analyzing and processing of nonstationary signals [206]. The Mel-frequency cepstral coefficients (MFCCs) [95], derived on the basis of short-time Fourier transform (STFT) and power spectrum estimation, have been introduced for speech recognition system. But the weakness of MFCC speech parameterization is that MFCCs do not lead always to good representation at low signal to noise ratios. The other problems of the STFT-based approach are the fixed time resolution as well as its inappropriateness to analyze highly nonstationary parts of the speech signals.

Adaptive algorithms, which are capable of tracking a nonstationary input, always exhibit two contradictory features: (i) the convergence speed and (ii) the steady-state fluctuations [35]. It is often found that RLS tracks a nonstationary environment better than LMS because it has a greater convergence speed. But this assumption is not sufficiently justified because steady-state fluctuation is not considered. That is, tracking is not a transient problem but a steady-state problem. It has been proposed that when adaptive filtering is combined with wavelet thresholding [23], the rate of convergence of the RLS algorithm is faster than that of LMS. But this approach is very expensive due to high computational complexity. Also, this approach does not explain the correct selection of wavelets and its family.

Wavelet transform has become a promising tool for the multiscale representation and analysis of signals. A number of researchers are applying this new tool for the reduction of noise in speech signals [97, 123, 124, 246]. Denoising of speech signals using wavelet transform is usually based on thresholding and shrinking wavelet coefficients of noisy signals. However, the critical problems are to choose correct wavelet, determining appropriate threshold value and level of decomposition. In this method, a discrete time signal is split up by high-pass and low-pass filters, which is done recursively for low-pass coefficients only. The extension of wavelet packets is straightforward. The high-pass coefficients are split up as well. Actually, there are numerous denoising techniques used in speech processing. Most of them include hypotheses on the original signal, SNRs as well as distortions [289]. Ultimately, the purpose of denoising methods in speech signals can be used to improve the quality of speech processing devices like mobile telephone, digital hearing aids used in our daily life to make them more robust in noisy environment.

In this section, a novel wavelet packet denoising method has been proposed for speech enhancement. It can effectively enhance the noisy speech that is degraded by additive white Gaussian noise. To evaluate the denoising performance, the method developed in this section is compared with the method proposed in [23]. The results obtained show improved performances over existing method.

Wavelet Packet Transform

Wavelet packet transforms are particular linear combinations or superpositions of wavelets. They form bases which retain many of the orthogonality, smoothness, and localization properties of their parent wavelets. The coefficients in linear combinations are computed by factored or recursive algorithm, with the result that the expansions in wavelet packet bases have low computational complexity. The advantages of wavelet packets are the following:

1. Daubechies orthogonal wavelets are a particular case of wavelet packets.
2. Wavelet packets are organized naturally into collections, and each collection is an orthonormal basis for $L^2(\mathbb{R})$.
3. Thus, one can compare the advantages and disadvantages of the various possible decompositions of a given signal in these orthonormal bases and select the optimal collection of wavelet packets for representing the given signal.
4. Wavelet packets are described by a very simple algorithm $2^{j/2}\omega_n(2^j x - k)$, where $j, k \in \mathbb{Z}$, $n = 0, 1, 2, \dots$ and where the supports of the ω_n 's are in the same fixed interval $[0, L]$.

Now, we say that powerful tools for denoising noisy signals are wavelet packets, which are a redundant time-frequency representation.

The functions ω_n are obtained roughly speaking by superposition of $\frac{1}{2}$ -scaled translated versions of functions of lower index. But supports of all ω_n are in $[0, 2N - 1]$. Therefore, ω_n oscillates approximately n times and then n can be interpreted as a frequency parameter.

Starting from ω_n , the three-index family of wavelet packet atoms is considered which is obtained by dyadic dilations and translations of ω_n :

$$\omega_{j,n,k}(x) = 2^{-\frac{j}{2}}\omega_n(2^{-j}x - k).$$

Wavelet Packet Thresholding

Removal of the noise components by thresholding coefficients is based on the observation that in speech signal, energy is mostly concentrated in a small number of wavelet dimensions. The coefficients of these dimensions are relatively large compared to noise signal (that has energy spread over a large number of coefficients). Hence by setting smaller coefficients to zero, one can nearly optimally eliminate noise while preserving the important information of the original signal. Thus, motivation of thresholding is due to noise characteristics, which tend to be characterized by smaller coefficients across time and scale, while signal energy is concentrated in larger coefficients.

The DWT can be computed efficiently by a conjugate quadrature filter bank [35]. A discrete time signal is split up by high-pass and low-pass filters, which is done recursively for low-pass coefficients. The extension to wavelet packets is straightforward. The high-pass coefficients are split up as well. Fourth order of Daubechies wavelet (db4) has been applied in this section, because it is more efficient due to its discrete nature. The coefficients can be organized in a binary tree and are addressed by

$$C_{d,b}(n) = WP(d, b, n) \text{ with } 0 \leq d \leq D, 0 \leq b \leq 2^d - 1, 0 \leq n \leq \frac{N}{2^d} - 1$$

where d denotes the depth in the tree, b is the node number in this depth, n is the coefficient index in the specific node, D is the maximum depth, and N is the signal length. Since a wavelet packet is a redundant decomposition, a best basis can be chosen to represent the signal in few but large coefficients. Such a denoising technique in wavelet packet domain is called “wavelet packet thresholding.” The wavelet packet coefficients of the noisy signals are thresholded by a nonlinear thresholding function. In the case of white noise, the threshold can be constant for all coefficients because the energy of white noise is distributed equally across the coefficients of an orthonormal basis. We have already described that wavelet packets approach is generalized by the wavelets methodology. In the binary wavelet packet tree decomposition represented by Fig. 5.12, the detailed spaces are subdivided in the same way as the approximation ones according to:

$$W_{j+k}^l = W_{j+k+1}^{2l} + W_{j+k+1}^{2l+1}.$$

The filter bank analysis algorithm can be used to compute the wavelet packet coefficients, which consist of (Fig. 5.13)

$$d_{j+k+1}^{2l}(n) = [h_0(n) * d_{j+k}(n)]_{\downarrow 2}$$

$$d_{j+k+1}^{2l+1}(n) = [h_1(n) * d_{j+k}^{(l)}(n)]_{\downarrow 2}$$

$$k = 0, 1, 2, 3, \dots; j = 0, 1, 2, 3, \dots$$

The corresponding filter bank synthesis algorithm can be used to reconstruct the analyzed signal, which permits a finer analysis of the signal than the dyadic one.

Proposed Methodology

Let the contaminated signal f be obtained as:

$$f = s + n \tag{5.3.1}$$

where s is the original signal and n is the noise with the following two conditions:

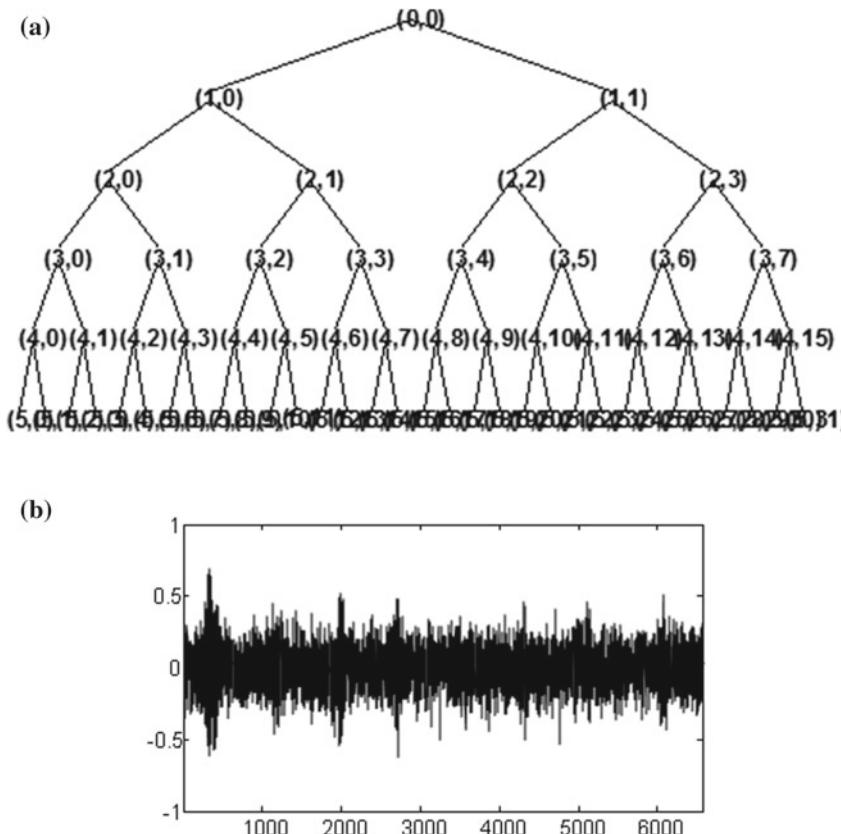


Fig. 5.12 **a** Wavelet packet tree decomposition. **b** Data for Node (1) or $(1, 1)$

1. The energy of s is captured, to a high percentage, by transform values whose magnitudes are greater than a threshold $T_s > 0$.
2. The noise signal's transform values all have magnitudes which lie below a noise threshold $T_n < T_s$.

Then, the noise in f has been removed by wavelet packet thresholding. All the values of its transform whose magnitudes lie below the noise threshold T_n are set to zero and an inverse transform, providing a good approximation of f [267].

Implementation

The proposed method can be implemented by the following steps:

1. A bird speech is recorded.
2. This sound is taken as original signal (s).
3. This signal s is then contaminated with white Gaussian noise (n) to get noisy signal as f .

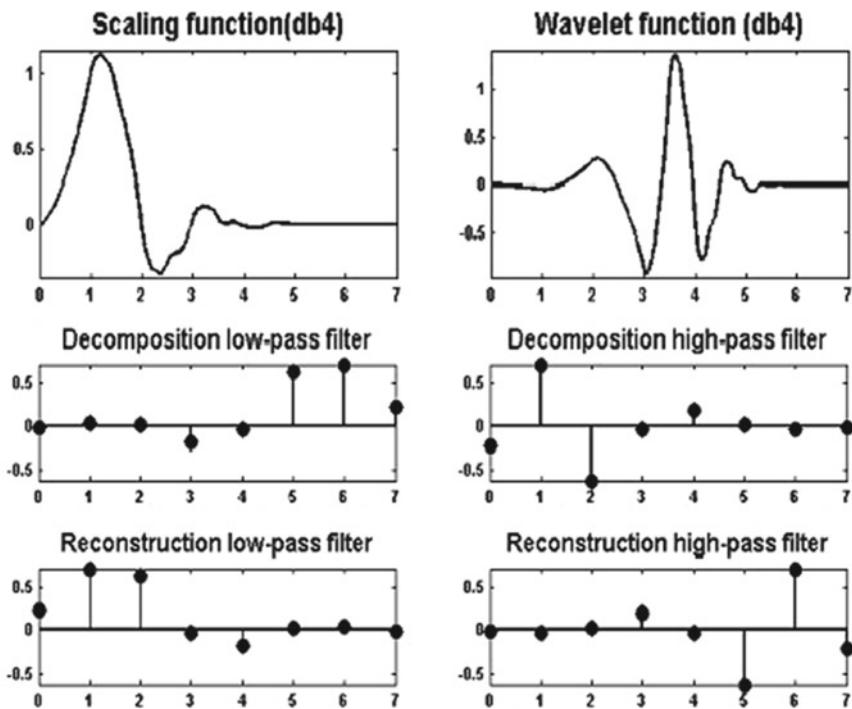


Fig. 5.13 Daubechies wavelet (db4) and its filters

4. As we have already described that wavelet packet is generalization of wavelets, so we have used fourth-order Daubechies wavelet as mother wavelet, i.e., db4.
5. The contaminated signal is treated with wavelet packet transform with proper threshold value. To choose optimal threshold value, we have considered a basic assumption. Since noise in the signal is additive white Gaussian noise, it has a variance with invariant of time. Also, most of the noise variance lies on detailed coefficients of first level of decomposition, as shown in Fig. 5.12b. So, the fixed global threshold is employed when noisy signal is analyzed in each scale.

Threshold value is estimated with the coefficients as

$$thr = \sigma \sqrt{2 \log(N)/N}$$

where the variance σ can be estimated using median estimator as

$$\sigma = \frac{\text{median}(c_i)}{0.6745}$$

where c_i are the high-frequency wavelet coefficients. The factor 0.6745 rescales the numerator so that σ is a suitable estimator for the standard deviation for

Table 5.4 Computation of MSE for SNR

SNR (in dB)	Mean square error (MSE)	
	Adaptive filter + WT	Wavelet packet transform
-5	0.6919	0.1456
0	0.3202	0.0655
5	0.0908	0.0468
10	0.0498	0.0289
15	0.0397	0.0167
20	0.0263	0.0092
25	0.0135	0.0050
30	0.0058	0.0027
35	0.0029	0.0015

Gaussian white noise. A soft thresholding is used to shrink the detailed coefficients of the noisy signal.

6. Results obtained by this method are given in Table 5.4.
7. The estimated signal \hat{f} is reconstructed by applying inverse wavelet packet transform by keeping all approximated coefficients and all thresholded detailed coefficients.
8. To evaluate the performance of the proposed method, the MSE values are found using the formula

$$MSE = \frac{1}{N} \sum (\hat{f} - f)^2$$

where f is the original signal and \hat{f} the estimated signal.

Simulation and Results

The proposed scheme is simulated using MATLAB 7.0. In the first, white Gaussian noise is added to original signal to produce different SNR levels, i.e., -5, 0, +5, +10, +15, +20, +25, +30, +35 dB. These contaminated signals are then filtered with the methods described in [23], and we propose the method. The MSE values obtained by these methods are reported in Table 5.4 (Figs. 5.14 and 5.15).

Conclusions

In the present work, a speech denoising method based on wavelet packets decomposition is proposed. To choose optimal threshold value, a basic assumption is considered. Since noise in the signal is additive white Gaussian noise, it has a constant power spectral density (PSD) for all frequencies. Hence, the fixed global threshold is employed when noisy signal is analyzed in each scale. A very popular wavelet proposed by Daubechies (db4) is used as mother wavelet. To evaluate the performance of the proposed method, mean square error is computed for different values of SNR. The result obtained is compared with J. P. Areanas's method [23], and it is found that the proposed method gives better result.

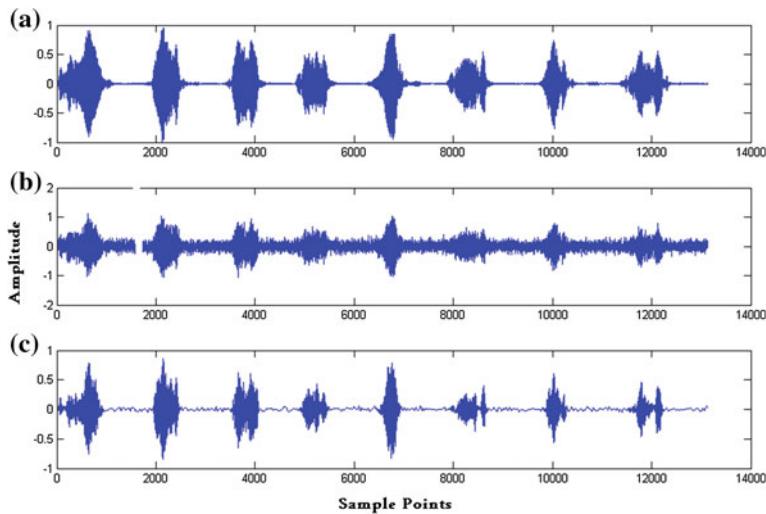
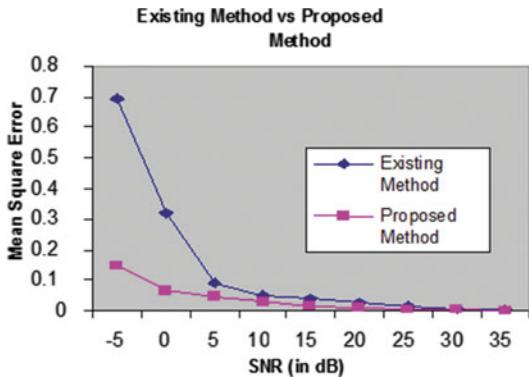


Fig. 5.14 **a** Original speech signal. **b** Noisy speech signal. **c** Estimated speech signal

Fig. 5.15 Comparative plot between existing method and proposed method



5.3.2 Selection of Optimal Decomposition Level Based on Entropy for Speech Denoising Using Wavelet Packets

Among many denoising techniques, the spectral subtraction noise cancelation introduced by Boll [38] is based on an estimation of the short-time spectrum magnitude of the original signal, taking into account the human auditory perception and the phase information. Despite its ability of canceling background noise, it introduces additional artifacts called musical noise and faces difficulties in pause detection. This problem is caused due to the inaccuracies in the short-time noise spectrum estimate. Fourier and fractional Fourier methods are other methods to suppress noise

[95]. Recently, wavelet transform methods have become more popular, which provide a powerful tool for nonlinear filtering of signals contaminated by noise. Mallat and Hwang [191] have shown that effective noise suppression may be obtained by transforming the noisy signal into the wavelet domain and preserving only the local maxima of the transformation. Alternatively, the estimated signal is achieved by thresholding the wavelet transform of the noisy signal. The wavelet threshold denoising method is based on the principle of multiresolution analysis [87]. The discrete detailed coefficients and the approximation coefficients can be obtained by multilevel wavelet decomposition, which are generally known as high-pass and low-pass filters, which are done recursively for low-pass coefficients. So, we are losing data from high-pass coefficients at each level of decomposition for further analysis and reconstruction of the original signal. The extension to wavelet packets is straight forward. The high-pass coefficients are split up as well to get more signal coefficients for further analysis and reconstruction.

Denoising of speech signals using wavelet transform is usually based on thresholding and shrinking wavelet coefficients of noisy signals. However, the critical problems are to choose right wavelet, determining appropriate threshold value and level of decomposition. A number of papers have been reported in this field that focus mainly on the determination of optimum threshold value. In [97], Donoho proposed a thresholding method in wavelet transform domain and has shown to have near-optimal properties for numbers of signals which are corrupted by white Gaussian noise. Several methods have been reported in this area, which mainly focus on the computation of threshold [98, 158]. The SureShrink algorithm [300] uses Stein's method of unbiased risk estimation to obtain the proper value of threshold. Lu and Wang [176] proposed employing the masking properties in the adjustment of the threshold value. In this reported work, generally a fixed decomposition level of wavelet transformation is considered before the processing. However, if the wavelet decomposition level is too high, the quality of the denoised signal may be reduced because the data are too much reduced. Hence, it is important to select an optimum level of decomposition based on the nature of signal being processed.

To overcome this problem, a Shannon entropy criterion has been reported in this chapter. Entropy is a quantitative measure of how uncertain the outcome of a random experiment is. Its definition and interpretation were introduced by C. E. Shannon [242]. Entropy has been used in speech signals as a segmentation criterion in noisy environments [247] and in deciding the desirable frame rate in the analysis of speech signals. Since entropy is a metric of uncertainty for random variables, the entropy of speech segments is obviously different from that of the noise signals because of the inherent characteristics of speech spectrum. Thus, the entropy will be maximal if the signal is only noise and minimal if it is a noiseless speech signal. The assumption is that the signal spectrum is more organized in speech segments than in noise segments.

Also for the noisy signals, Grossman proved that the variances and amplitude of the details of the white noise at the various levels decrease continuously as the level increases. Using this fact, present chapter deals with joint application of DWT and wavelet packet transform. Shannon entropy [79] plays an important role to determine optimal decomposition level in DWT domain. Actually, there are numerous

denoising techniques used in speech processing. Most of them include hypotheses on the original signal, SNRs as well as distortions [289]. Ultimately, the purpose of denoising methods in speech signals can be used to improve the quality of speech processing devices like mobile telephone, digital hearing aids used in our daily life to make them more robust in noisy environment.

When the signal is decomposed into approximation coefficients and detailed coefficients, the coefficients which are below the threshold value are considered as noise and discarded. After certain level of decompositions, discarded coefficients may constitute some portion of the signal and this results in the loss of information. To arrive at the optimal level of decomposition, the entropy of the signal is measured at each node. Most of the information contented on the approximation coefficients, while the wavelet transformation takes place. Thus, the decomposition level at which the entropy of the approximation coefficients becomes less than that of the detailed coefficients is considered as optimal level of decomposition. Proposed method selects the optimal decomposition level, and to choose optimal threshold value wavelet packet, thresholding method has been used. In both cases, noisy speech signal is considered under different signal-to-noise ratios (SNRs). Experimental results show the advantage of the proposed method.

Wavelet transforms are relatively recent developments that have fascinated the scientific, engineering, and mathematics community with their versatile applicability [174, 208, 284]. Wavelet is generally called a time–frequency localization that splits data into different frequency components, and then each component can be studied with a resolution matched to its scale. The main properties of the wavelets include the ability to concentrate the energy of a smooth signal in a few wavelet coefficients, while at the same time the transformation of white noise retains its attributes. That is why the concept of entropy has been used. One can obtain both redundant and nonredundant representation using appropriate choices of wavelets and discrimination schemes [35].

In the orthogonal decomposition procedure, the generic step splits the approximation coefficients into two parts. After splitting, we obtain a vector of approximation coefficients and a vector of detailed coefficients, both at a coarser scale. The information lost between two successive approximations is captured in the detailed coefficients. Then, the next step consists of splitting the new approximation coefficient vector, and successive details are never reanalyzed. In the corresponding wavelet packet situation, each detailed coefficient vector is also decomposed into two parts using the same approach as in approximation vector splitting. This offers the richest analysis. The idea of this decomposition is to start from a scale-oriented decomposition and then to analyze the obtained signals on frequency subbands. This property of wavelet packet transform inspired us to use it for denoising noisy speech signals.

Fundamentals of Entropy in Information Theory

The entropy H is the expected information content in a sequence, so it is the average of all the information contents weighted by their probabilities to occur as

$$H(x) = E[I(p)] = \sum_{i=1}^n p_i I(p) = -\sum_{i=1}^n p_i \ln(p_i) \quad (5.3.2)$$

or its continuous version called “differential entropy” represented by

$$H(x) = - \int_{-\infty}^{\infty} p(x) \log[p(x)] dx. \quad (5.3.3)$$

The entropy of a signal is a measure of how unpredictable it is; if the signal is a constant k , then its probability density function (PDF) is a unitary impulse located at k , that is $p_i = \delta(k)$; its entropy or unpredictability is zero as given by

$$H_{\min} = - \sum_i \delta(k) \ln[\delta(k)] = -\ln(1) = 0 \quad (5.3.4)$$

on the opposite case; if the signal has a uniform distribution, then the entropy would be maximum; that is, if $p_i = \frac{1}{n}$ for n possible values, then its entropy would be $\log(n)$ as given by

$$H_{\max} = - \sum_{i=1}^n \frac{1}{n} \ln\left(\frac{1}{n}\right) = -\ln\left(\frac{1}{n}\right) = \ln(n). \quad (5.3.5)$$

For two-dimensional data, Shannon's entropy is computed by

$$H = - \sum_{i=1}^n \sum_{j=1}^n p_{i,j} \ln(p_{i,j}). \quad (5.3.6)$$

For the uniform distribution, the $2D$ entropy is $2\ln(n)$, i.e.,

$$H_{\max} = - \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n^2} \ln\left(\frac{1}{n^2}\right) = -\ln\left(\frac{1}{n^2}\right) = 2\ln(n). \quad (5.3.7)$$

Computing the entropy of a signal requires some estimation of the PDF

$$p_1, p_2, \dots, p_i, \dots, p_n$$

we can use parametric methods, nonparametric methods.

In parametric methods, first a kind of distribution is chosen and for that its parameters are determined [31]. If, for example, the speech signal follows a Gaussian distribution with mean zero and variance σ^2 , then the PDF would be described by

$$p(x) = \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma}. \quad (5.3.8)$$

Replacing $p(x)$ into Eq. (5.3.3), we get this known formula for determining the entropy of a random variable with a Gaussian distribution

$$\begin{aligned}
 H &= - \int_{-\infty}^{\infty} \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \ln \left[\frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \right] dx \\
 &= \frac{\ln(\sqrt{2\pi}\sigma)}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx + \frac{1}{\sqrt{2\pi}2\sigma^3} \int_{-\infty}^{\infty} x^2 e^{-x^2/2\sigma^2} dx \\
 &= \frac{\ln(\sqrt{2\pi}\sigma)}{\sqrt{2\pi}\sigma} \sqrt{2\sigma^2\pi} + \frac{4\sqrt{\pi}(\sqrt{2}\sigma/2)^3}{\sqrt{2\pi}2\sigma^3} \\
 &= \frac{1}{2} \ln(2\pi e) + \frac{1}{2} \ln(\sigma^2).
 \end{aligned} \tag{5.3.9}$$

Similarly, for the n -dimensional case it is not difficult to show that the entropy of a random variable with distribution $N(0, R)$ is computed, where R is the covariance matrix of size $n \times n$ [202]. We have

$$H = \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \ln[\det(R)]. \tag{5.3.10}$$

Entropy is a common concept in many fields, mainly in signal processing and information theory. It is a quantitative measure of how uncertain the outcome of a random experiment is. The “measure of organization” of a discrete magnitude spectrum of a signal can be described similarly as the entropy of an information source by Shannon [242]. In thermodynamics, entropy is a measure of quantifying the imbalance degree of thermostat. In mathematics, entropy is used to measure the uncertainty of problems. While in information science, entropy is the average uncertainty of information source. In other words, entropy is a measure of irregularity of states such as imbalance, uncertainty. A method for measuring the entropy appears as an ideal tool for quantifying the ordering of nonstationary signals.

In this section, the nonnormalized Shannon entropy is used since the speech signal has imbalance, nonstationary, different frequency component, and the different energy distribution. Shannon entropy is used to extract the features from different power quality disturbance signals as a measurement of these irregularities. Let the disturbance signal be $U = \{U_j : j = 1, 2, 3, \dots, N\}$. The Shannon (nonnormalized) entropy of the n th point is represented as

$$SE_j = \sum_{k=1}^N E_{jk} \log E_{jk}$$

where E_{jk} is the wavelet energy spectrum at scale j and instant k and is defined as follows. The wavelet energy spectrum for detail and approximation coefficients in each decomposition level is obtained from the wavelet multiresolution analysis of

different disturbances and can be calculated as follows:

$$ED_{jk} = |D_j(k)|^2, \quad EA_{jk} = |A_j(k)|^2$$

where $j = 1, 2, 3, \dots, n$, n represents the total number of decomposition levels. ED_{jk} and EA_{jk} are the energy spectrums for detail and approximation coefficients, respectively. $D_j(k)$ are the detailed coefficients, and $A_j(k)$ are the approximation coefficients.

The entropy will be maximum if the signal is only noise and minimum if it is a noiseless speech signal. The assumption is that the signal spectrum is more organized in speech segments than in noise segments [59].

Proposed Methodology

Let noisy signal $Z[n]$ be represented by the following equation

$$Z[n] = W[n] + S[n]$$

where S is the original signal and W is the white Gaussian noise. We apply the proposed method as follows:

1. Compute the discrete wavelet transform of the signal Z down to the level based on entropy criterion.
2. Compute Shannon entropy for approximation coefficients and detailed coefficients, i.e., Ea_j and Ed_j , where $1 \leq j \leq n$, $\forall n \in \mathbb{N}$.
3. If we reach the stage where $Ea_j < Ed_j$ successively, then we stop decomposition.
4. For determination of threshold value, wavelet packet thresholding method has been used.

So, applying proposed technique simultaneously minimizes noise as well as chooses optimal decomposition level.

Implementation

The proposed method can be implemented by the following steps:

1. A whistle sound has been recorded for three seconds to generate a speech signal.
2. This sound has been considered as original signal (S).
3. This signal S is then contaminated with white Gaussian noise (W) to get noisy signal as Z having different SNRs.
4. A fourth-order Daubechies wavelet (db4) is used as mother wavelet in both cases, i.e., for decomposition and for choosing threshold value.
5. For decomposition level, we have used Shannon entropy under consideration of DWT.
6. When $Ea_j < Ed_j$ successively, then we stop decomposition, where Ea_j represents entropy value for approximation coefficients, Ed_j represents entropy value for detailed coefficients, and j is the level of decomposition.

7. The contaminated signal is then decomposed to the level obtained from step (6) with wavelet packet transform. To choose optimal threshold value, we have considered a basic assumption. Since noise in the signal is additive white Gaussian noise, it has a variance with invariant of time. Also, most of the noise variance lies on detailed coefficients of first level of decomposition, as shown in Fig. 5.16. So, the fixed global threshold is employed when noisy signal is analyzed in each scale.
8. Threshold value is estimated with wavelet packet coefficients selection rule, i.e.,

$$thr = \sigma \sqrt{2(\log N)/N}$$

using a penalization method proposed by Lucien Birge and Pascal Massart [36]. Assuming α (must be real number and greater than one) as a tuning parameter, one can minimize the penalized criteria given by

$$\text{crit}(t) = - \sum [\{c(k)\}^2, k \leq t] + 2\sigma^2 t(\alpha + \log(n/t))$$

where $c(k)$ are the wavelet packet coefficients sorted in decreasing order of their absolute value and n is the number of coefficients. In the present case, α has been taken as two.

9. To estimate signal, inverse wavelet packet transform has been used by keeping all approximation coefficients and all thresholded detailed coefficients.
10. To check the performance of the proposed method, the SNR of the estimated signal is found using the formula

$$SNR = 10 \log_{10} \left(\frac{\sum |S|^2}{\sum |\hat{S} - S|^2} \right).$$

11. Here, \hat{S} is the estimated signal, reported in Fig. 5.3.

Simulation and Results

The proposed scheme is simulated using MATLAB 7.0. Initially, white Gaussian noise is added to original signal to produce different SNRs, i.e., $-22, -17, -12, -6, -2, 8$ dB. These contaminated signals are then decomposed with DWT, and note Shannon entropy at each level, which is given in tables (5.5–5.10), presented below. For removing noise from noisy signal, a wavelet packet method has been used. The SNR values obtained by these methods are reported in the same tables. To choose optimal threshold value, a basic assumption is considered. Since noise in the signal is additive white Gaussian noise, it has a variance with invariant of time. Also, most of the noise variance lies on detailed coefficients of first level of decomposition. So, the fixed global threshold using median estimator is employed when noisy signal is analyzed in each scale (Fig. 5.17), Tables 5.5, 5.6, 5.7, 5.8, 5.9, and 5.10).

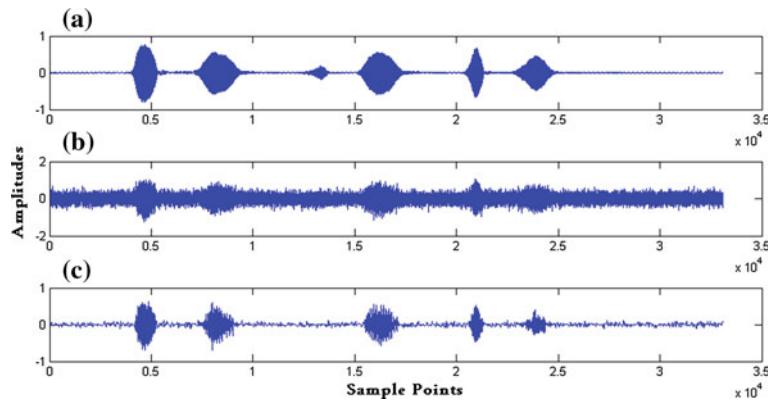


Fig. 5.16 **a** Original speech signal, **b** Noisy speech signal. **c** Estimated speech signal

Fig. 5.17 **a** Wavelet packet tree decomposition. **b** Data for node: (1, 1)

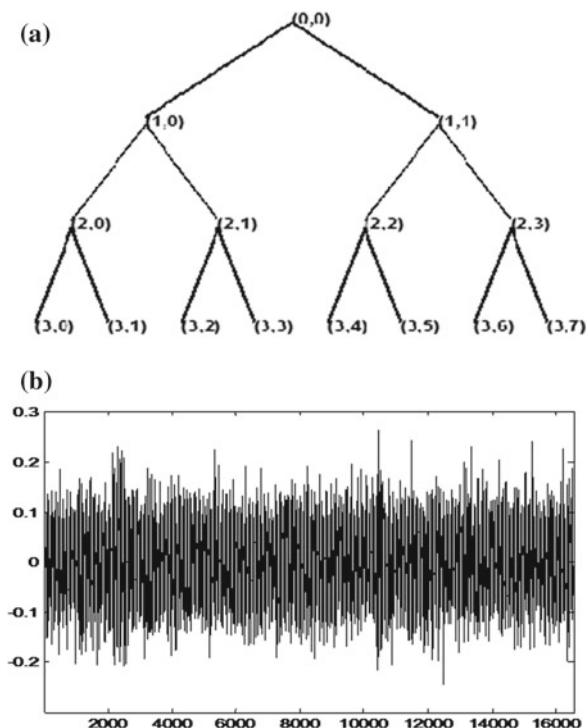


Table 5.5 Mother wavelet (db4) for SNR = -22 dB

Level of decompositions	Shannon entropy		SNR (Estimated signal)
	Approx. coeffs.	Detailed coeffs.	
1	-9.8807e + 004	-9.8153e + 004	-21.9121
2	-4.9070e + 004	-5.0158e + 004	-21.9040
3	-2.3673e + 004	-2.5212e + 004	-21.9034
4	-1.1793e + 004	-1.2094e + 004	-13.0115
5	-5.9967e + 003	-6.1832e + 003	-13.2646
6	-3.6172e+003	-2.8378e+003	-15.9814
7	-2.5494e + 003	-1.5475e + 003	-18.9809
8	-3.0377e + 003	-731.5272	-21.9360
9	-4.6318e + 003	-914.4873	-23.0635

Table 5.6 Mother wavelet (db4) for SNR = -17 dB

Level of decompositions	Shannon entropy		SNR (Estimated signal)
	Approx. coeffs.	Detailed coeffs.	
1	-1.3462e + 004	-1.2238e + 004	-16.8915
2	-6.8562e + 003	-6.6328e + 003	-11.0744
3	-2.9958e + 003	-3.7890e + 003	-16.9188
4	-1.4406e + 003	-1.4948e + 003	-11.0337
5	-733.6843	-718.4768	-11.1697
6	-321.7882	-458.4274	-8.4689
7	-214.9453	-104.3466	-3.9164
8	-183.4416	-69.9529	-6.0590
9	-97.4970	-66.5513	-5.8274
10	-65.1301	-55.4260	-5.9882
11	-126.9126	-3.4198	-6.8508

Conclusions

In the present work, a novel wavelet packet denoising method based on optimal decomposition and global threshold value has been proposed for speech denoising. Proposed method selects the optimal decomposition level using the method of Shannon entropy, and to choose optimal threshold value wavelet packet, thresholding method has been used. In both cases, noisy speech signal is considered under different signal-to-noise ratios (SNRs). To choose optimal threshold value, we have considered a basic assumption. Since noise in the signal is additive white Gaussian noise, it has a constant power spectral density for all frequencies. A very popular

Table 5.7 Mother wavelet (db4) for SNR = -12 dB

Level of decompositions	Shannon entropy		SNR (Estimated signal)
	Approx. coeffs.	Detailed coeffs.	
1	1.6393e + 003	2.2223e + 003	-8.8625
2	735.3587	895.2607	-6.2985
3	502.4374	131.0649	-4.7110
4	271.3682	266.1975	-2.7200
5	138.8998	128.7698	-2.8682
6	70.4800	72.0480	-0.2273
7	32.8967	29.4306	2.1259
8	-0.9650	19.1912	1.9140
9	-64.8198	13.9416	1.5536
10	-187.7942	5.8582	1.2208
11	-474.3511	3.8325	1.7181

Table 5.8 Mother wavelet (db4) for SNR = -6dB

Level of decompositions	Shannon entropy		SNR (Estimated signal)
	Approx. coeffs.	Detailed coeffs.	
1	2.6406e + 003	2.5830e + 003	-6.8882
2	1.2846e + 003	1.3255e + 003	-68594
3	648.0779	489.9936	-6.9298
4	327.5391	317.9025	-6.8706
5	167.6813	158.2608	1.5040
6	91.4766	88.0061	2.8143
7	48.4708	43.4288	3.6833
8	20.5510	20.8542	4.7621
9	-47.6644	10.3474	4.8475
10	-121.7673	-0.2152	4.4251
11	-319.0546	2.1629	3.5179

wavelet proposed by Daubechies (db4) is used as mother wavelet. To check the performance of the proposed method, signal-to-noise ratios is computed for denoised speech signal.

Table 5.9 Mother wavelet (db4) for SNR = -2 dB

Level of decompositions	Shannon entropy		SNR (Estimated signal)
	Approx. coeffs.	Detailed coeffs.	
1	1.8403e+003	1.4399e+003	-1.8946
2	906.0067	883.1308	-1.9179
3	360.0899	343.3386	19052
4	180.4945	182.3653	3.6039
5	95.2886	86.9895	4.2153
6	50.4727	45.2751	5.9480
7	25.7655	25.7755	6.7151
8	11.2449	12.1686	7.5685
9	-6.6604	6.6355	7.6364
10	-33.3566	4.4377	7.1289
11	-106.5147	3.5094	5.8657

Table 5.10 Mother wavelet (db4) for SNR = 8 dB

Level of decompositions	Shannon entropy		SNR (Estimated signal)
	Approx. coeffs.	Detailed coeffs.	
1	977.8638	280.2230	10.8059
2	473.5591	432.5560	9.5386
3	70.7909	135.8903	10.0008
4	39.2111	31.9754	10.5984
5	22.8755	17.0418	12.1041
6	14.7612	8.7214	13.0400
7	8.4187	7.7579	13.3909
8	6.0293	2.8484	14.1052
9	2.4525	2.6736	14.1389
10	-4.2293	2.1485	13.7834
11	-24.7866	1.5269	12.4053

5.3.3 Thresholding Method for Denoising Speech Signal Using Wavelet Packet Transform

Signal plays an important role in our daily life. It can be generated by a single source or by multiple sources. In the former case it is scalar signal and in the latter case, it is vector (or multichannel) signal. Speech is a one-dimensional signal in which independent variable is time. If the independent variable is continuous, the signal is

called continuous time signal. If time is represented in discrete form, the signal is called discrete time signal. A continuous time signal is defined at every instant of time, while discrete time signal is defined at discrete instants of time.

In speech communication system, noise is a summation of unwanted or disturbing energy from natural sources and sometimes they are from man-made sources. If the noise becomes large compared with the observed speech signal, the quality of speech recognition and processing system become less efficient. In general, we can say that noise is the main disturbing factor in the field of digital signal processing. There are several kinds of noises among which white noise and colored noise are more popular. White noise has constant power spectral density across the entire frequency spectrum. When the power spectral density of the noise is not uniform across the entire frequency spectrum, then it is known as colored noise.

Using noise reduction techniques for noise reduction in speech signal, one can try to increase the quality of the speech processing devices like digital hearing aids, mobile phones, and human-machine communication systems in our daily life. Also, under noisy conditions these systems can become more robust in nature after applying different noise reduction techniques. So, one can conclude that the reduction of noise in speech signal improves the quality and intelligibility of digital hearing aids which reduces the listener's fatigue.

In the literature, a number of noise reduction techniques have been proposed by the researchers [23, 35, 38, 95]. Boll [38] proposed spectral subtraction method for noise reduction in speech signal which is based on an estimation of the short-time spectrum magnitude of the original signal, taking into account the human auditory perception and the phase information. But, due to the inaccuracies in the short-time noise spectrum, it introduces additional artifacts which create other problems.

From last few years, wavelet transform has become one of the popular and powerful tools to denoise noisy speech signals. The denoising technique used in wavelet analysis is based on the assumption that the amplitude rather than the spectrum of the signal is different from the noise. The localizing property of the wavelet adds thresholding and shrinking the wavelet coefficients which help us to filter noise from the signal. For denoising noisy speech signal using wavelet transform, the vital problem is to choose good wavelet, determine optimal level of decomposition, and to find appropriate threshold value. In wavelet thresholding techniques, noise reduction is made by thresholding the wavelet coefficients. The estimated threshold should define a boundary between the wavelet coefficients of noise and speech signal. Several techniques have been reported in this field, which are mainly focused on the computation of the threshold value [97, 98, 158].

Wavelet packet transform is an extension of the wavelet transform. In wavelet transformation, signal decomposes into approximation coefficients and detailed coefficients, in which further decomposition takes place only at approximation coefficients, whereas in wavelet packet transformation detailed coefficients are decomposed as well which give more wavelet coefficients for further analysis.

The chosen global threshold value at initial decomposition level does not well suited always because at each decomposition level we are ignoring the detailed coefficients obtained by using wavelet transform. Also, if we choose global thresholding,

then getting more number of wavelet coefficients seems to be meaningless. In universal thresholding, when the high-level wavelet decomposition is applied, the quality of the denoised signal may be low because the data are too much reduced. Therefore, in the present section in order to choose appropriate threshold value we have applied median threshold estimator in wavelet packet domain. To find the optimal threshold value, we have considered the threshold value of the detailed wavelet coefficients at each decomposition level and then taking the average of these values, which retain maximum possible signal coefficients and remove much more noisy coefficients. Experimental results show the performance of the proposed method.

Methodology

Consider noisy speech signal $y(t)$ given by

$$y(t) = x(t) + n(t)$$

where $x(t)$ is the original speech signal and $n(t)$ is white Gaussian noise. Taking wavelet transform of the above equation, we get

$$Y_{j,k} = X_{j,k} + N_{j,k}$$

where $Y_{j,k}$ is the k th wavelet coefficient with the dyadic scaling factor j .

Now, a thresholding technique has been used to minimize noise coefficients in noisy signal, which consists of two types, i.e., hard thresholding and soft thresholding. In hard thresholding technique, the following characteristics exist, i.e.,

$$Y_{j,k}^H = \begin{cases} Y_{j,k}, & |Y_{j,k}| > \delta \\ 0, & \text{otherwise} \end{cases}$$

where δ is threshold value. Soft thresholding is an extension of hard thresholding, which is given by

$$Y_{j,k}^S = \text{sign}(Y_{j,k}) \max\{|Y_{j,k}| - \delta, 0\}.$$

Several techniques for estimating the value of threshold have been proposed. Donoho and Johnstone [98] proposed a thresholding technique in which threshold value has been calculated as:

$$\delta_1 = \sigma \sqrt{2 \ln N} \quad (5.3.11)$$

where σ is the standard deviation and N the total number of sample points of the noisy signal. Standard deviation of the noisy signal is given by

$$\sigma_j = \text{median}\{|W_{j,k}|\}/0.6745, \quad 0 \leq k \leq 2^{j-1} - 1 \quad (5.3.12)$$

where $W_{j,k}$ is the wavelet coefficient at the j th scale and 0.6745 is a tuning parameter.

Rigrsure threshold is a soft threshold evaluator of unbiased risk. Suppose $W = (\omega_1, \omega_2, \dots, \omega_n)$ is a vector consisting of the square of wavelet decomposition coefficients from small to large. Choose the minimum value r_b from the risk vector $R = \{r_i\}_{i=1,2,\dots,n} = [n - 2i + (n - 1)\omega_1 + \sum_{k=1}^l \omega_k]/n$ as the risk value. Then, rigrsure threshold is given by

$$\delta_2 = \sigma \sqrt{r_b} \quad (5.3.13)$$

where σ is the standard deviation of the noisy signal.

If $s = \sum_{i=1}^n \omega_i^2$ is the square sum of the wavelet coefficients, $a = (s - n)/n$ and $b = (\log_2 n)^{3/2} \sqrt{n}$, where n is the length of the wavelet coefficients, then heursure threshold is given by

$$\delta = \begin{cases} \delta_1, & a < b \\ \min\{\delta_1, \delta_2\}, & a \geq b \end{cases}$$

where δ_1 and δ_2 are defined in Eqs. (5.3.11) and (5.3.13), respectively.

Minimax threshold is based on the minimum mean square error (MMSE) extreme value and is determined by

$$\delta = \begin{cases} \sigma(0.3936 + 0.1829 \log_2 n), & n > 32 \\ 0, & n < 32 \end{cases}$$

Proposed Rule

In the existing technique [98], it is assumed that most of the noise coefficients lie on the first level of decomposition. So, they have used universal thresholding. But in many practical situations, it is found that if we apply thresholding at first decomposition level, then many useful information can be suppressed at further decomposition level by taking threshold value at this level. In this section, an extension of this methodology has been proposed. In the proposed technique, for threshold value, median estimator has been applied at each level and then average value of these obtained values has been considered as optimal threshold value to restore more informative wavelet coefficients by fixing the level of decomposition, i.e.,

$$\delta = \sum_{j=1}^5 \{\sigma_j\}/5 \quad (5.3.14)$$

where σ_j is defined by Eq. (5.3.12).

Thus, the proposed technique is more efficient because of considering the importance of wavelet coefficients lying at each level of decomposition, which is probably not found in the previous existing technique.

The complete algorithm is summarized in steps as below:

Step1: Generate a whistle sound, and consider it as a clean signal $x(n)$.

Step2: Generate a random white Gaussian noise $w(n)$.

Step3: Add these two signals to get a noisy signal $y(n)$, i.e.,

$$y(n) = x(n) + w(n).$$

Step4: Compute the discrete wavelet packet transform of the noisy speech signal $y(n)$.

Step5: Choose optimal threshold value by considering Eq. (5.3.14).

Step6: Compute the inverse discrete wavelet packet transform to get estimated signal.

Step7: Compute the signal-to-noise ratio (SNR), i.e.,

$$SNR = 10 \log_{10} \left\{ \frac{\left(\sum_{n=0}^{N-1} x(n) \right)^2}{\left(\sum_{n=0}^{N-1} (y(n) - x(n))^2 \right)} \right\}$$

to evaluate the performance of the proposed method, where $x(n)$ is the original speech signal, $y(n)$ is noisy speech signal, and N is number of sample points.

Simulation and Experimental Results

For the simulation of the proposed method, a whistle sound has been recorded and considered as original speech signal, shown in Fig. 5.18. White Gaussian noise (WGN) is used to model the background noise. This WGN is added to the original speech signal to introduce distortions. This noisy speech signal, shown in Fig. 5.19, is used as the test signal for the simulation of proposed method. The estimated signal from the noisy speech signal using proposed method is shown in Fig. 5.20.

To study the performance of the proposed method, db5 has been used as a mother wavelet. The level of decomposition is arbitrarily chosen as five. To evaluate the performance of the proposed method, SNR is computed for various distortions. The results obtained from the simulation are given in Table 5.11.

The performance comparison between existing method and proposed method is shown in Fig. 5.21. The performance of the proposed method is shown in the flowchart given below.

Conclusions

In the present work, a speech denoising method based on wavelet packet transform of speech signal is proposed. In the proposed technique, for fixed threshold value, median estimator has been applied at each level of decomposition and then average value of these obtained values has been considered as optimal threshold value to restore more informative wavelet coefficients by fixing the level of decomposition. A very popular wavelet transform proposed by Daubechies, i.e., db5, is used. To evaluate the performance of the proposed method, SNR is computed for noisy signal as well as estimated signal. The result obtained is compared with universal thresholding method, and it is found that the proposed method gives better result.

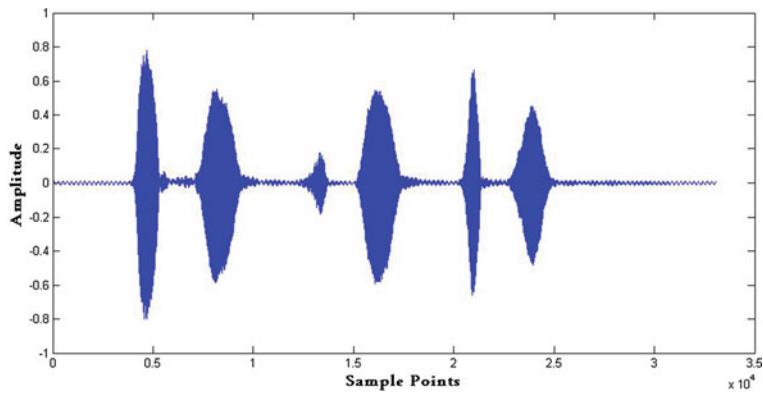
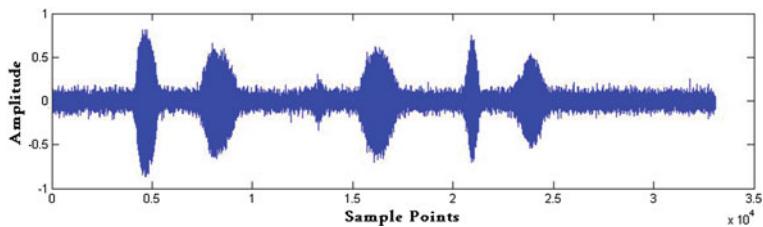
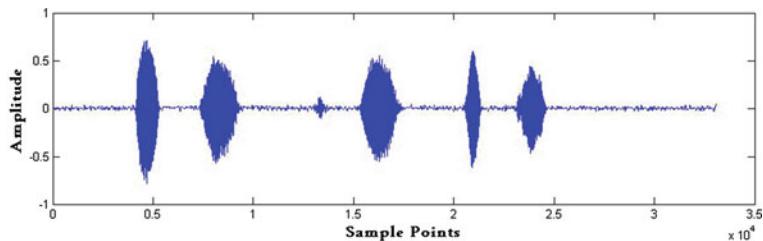
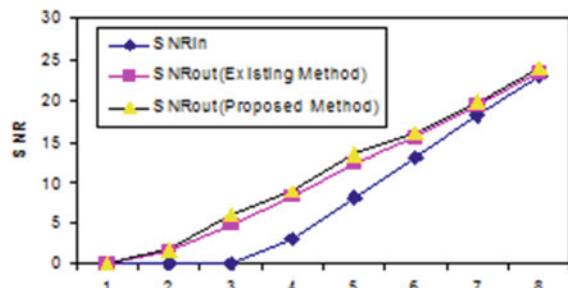
**Fig. 5.18** Original signal**Fig. 5.19** Noisy signal**Fig. 5.20** Estimated signal**Fig. 5.21** Performance comparison between existing method and proposed method

Table 5.11 Computation of SNR

S.No.	Universal threshold		Proposed threshold	
	SNRin	SNRout	SNRin	SNRout
1	-11.8882	-1.4753	-11.8882	-1.3578
2	-6.8262	1.5409	-6.8262	1.6844
3	-1.9024	4.8946	-1.9024	5.9988
4	3.0893	8.2683	3.0893	8.9391
5	8.0653	12.3668	8.0653	13.4596
6	13.0450	15.4930	13.0450	16.0044
7	18.1427	19.4280	18.1427	19.9567
8	23.0570	23.4085	23.0570	23.9362

5.4 Applications of Wavelets in Biomedical Signals

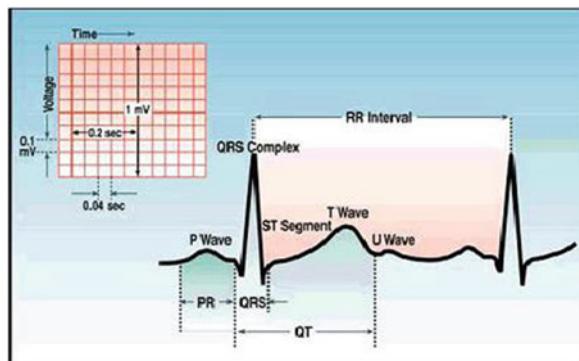
5.4.1 ECG Signal Characteristics

An electrocardiogram (ECG) signal is an electrical impulse recording of the heart and is used in the investigation of heart disease. These impulses are recorded as waves called P-QRS-T deflections. Each cardiac cell is surrounded by and filled with a solution that contains, in part, sodium (Na^+), potassium (K^+), and calcium (Ca^{++}). In its resting condition, the interior of the cell membrane is considered negatively charged, with respect to the outside. When an electrical impulse is initiated in the heart, the inside of a cardiac cell rapidly becomes positive in relation to the outside of the cell. The electrical impulse causes this excited state, and this change of polarity is called depolarization. Immediately after depolarization, the stimulated cardiac cell returns to its resting state, which is called repolarization. The resting state is maintained until the arrival of the next wave of depolarization. This change in cell potential from negative to positive and back to negative is called an action potential. That action potential initiates a cardiac muscle contraction. The ECG signal is a measurement of the effect of this depolarization and repolarization for the entire heart on the skin surface and is also an indirect indicator of heart muscle contraction, because the depolarization of the heart leads to the contraction of the heart muscles. Although the phases of the ECG signal are due to action potentials traveling through the heart muscle, the ECG signal is not simply a recording of an action potential. During the heartbeat, cells fire action potentials at different times, and the ECG signal reflects patterns of that electrical activity. Figure 5.22 shows ECG waves and intervals as well as standard time and voltage measures.

Arrhythmia

The rhythm of the heart is normally generated and regulated by pacemaker cells within the sinoatrial (SA) node, which is located within the wall of the right atrium. SA nodal 6 pacemaker activity normally governs the rhythm of the atria and ventricles. Normal rhythm is very regular, with minimal cyclical fluctuation. Furthermore,

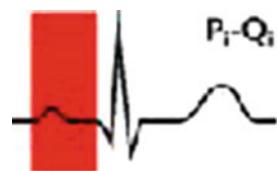
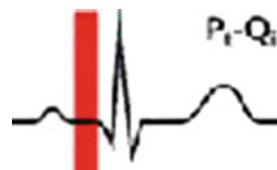
Fig. 5.22 Components of an ECG signal



atrial contraction is always followed by ventricular contraction in the normal heart. When this rhythm becomes irregular, too fast (tachycardia) or too slow (bradycardia), or the frequency of the atrial and ventricular beats are different, this is called an arrhythmia. The term “dysrhythmia” is sometimes used and has a similar meaning. About 14 million people in the USA have arrhythmias (5% of the population) [275]. The most common disorders are atrial fibrillation and flutter. The incidence is highly related to age and the presence of underlying heart disease; the incidence approaches 30% following open heart surgery. Patients may describe an arrhythmia as a palpitation or fluttering sensation in the chest. For some types of arrhythmias, a skipped beat might be sensed because the subsequent beat produces a more forceful contraction and a thumping sensation in the chest. A “racing” heart is another description. Proper diagnosis of arrhythmias requires an electrocardiogram, which is used to evaluate the electrical activity of the heart. Depending on the severity of the arrhythmia, patients may experience dyspnea (shortness of breath), syncope (fainting), fatigue, heart failure symptoms, chest pain, or cardiac arrest.

A frequent cause of arrhythmia is coronary artery disease because this condition results in myocardial ischemia or infarction. When cardiac cells lack oxygen, they become depolarized, which leads to altered impulse formation and/or altered impulse conduction. The former concerns changes in rhythm that are caused by changes in the automaticity of pacemaker cells or by abnormal generation of action potentials at sites other than the SA node (termed ectopic foci). Altered impulse conduction is usually associated with complete or partial block of electrical conduction within the heart. Altered impulse conduction commonly results in reentry, which can lead to tachyarrhythmias. Changes in cardiac structure that accompany heart failure (e.g., dilated or hypertrophied cardiac chambers) can also precipitate arrhythmias. Finally, many different types of drugs (including antiarrhythmic drugs) as well as electrolyte disturbances (primarily K⁺ and Ca⁺⁺) can precipitate arrhythmias.

Arrhythmias can be either benign or more serious in nature depending on the hemodynamic consequence of the arrhythmia and the possibility of evolving into a lethal arrhythmia. Occasional premature ventricular complexes (PVCs), while annoying to a patient, are generally considered benign because they have little hemody-

Fig. 5.23 PR interval**Fig. 5.24** PR segment

namic effect. Consequently, PVCs if not too frequent are generally not treated. In contrast, ventricular tachycardia is a serious condition that can lead to heart failure or, worse, to ventricular fibrillation and death.

When arrhythmias require treatment, they are treated with drugs that suppress the arrhythmia. These drugs are called antiarrhythmic drugs. There are many different types of antiarrhythmic drugs and many different mechanisms of action. Most of the drugs affect ion channels that are involved in the movement of sodium, calcium, and potassium ions in and out of the cell. These drugs include mechanistic classes such as sodium channel blockers, calcium channel blockers, and potassium channel blockers. By altering the movement of these important ions, the electrical activity of the cardiac cells (both pacemaker and nonpacemaker cells) is altered, hopefully in a manner that suppresses arrhythmias. Other drugs affect autonomic influences on the heart, which may be stimulating or aggravating arrhythmias.

Waves and Intervals of the ECG Signal

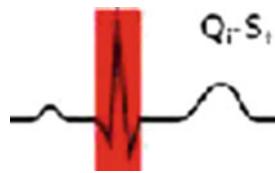
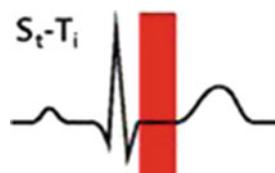
(a) P Wave

The PR interval begins with the onset of the P wave (P_i) and ends at the onset of the Q wave (Q_i). It represents the duration of the conduction through the atria to the ventricles. Normal measurement for PR interval is 120–200 ms. It is shown in Fig. 5.23.

The PR segment begins with the endpoint of the P wave (P_t) and ends at the onset of the Q wave (Q_i). It represents the duration of the conduction from the atrioventricular node, down the bundle of its end through the bundle branches to the muscle. It is shown in Fig. 5.24.

(b) QRS Complex

The QRS complex which is shown in Fig. 5.25 is a structure on the ECG signal that corresponds to the depolarization of the ventricles. Because the ventricles contain more muscle mass than the atria, the QRS complex is larger than the P wave. In addition, because the His/Purkinje system coordinates the depolarization of the ventricles, the QRS complex tends to look “spiked” rather than rounded due to the increase in conduction velocity. A normal QRS complex is 0.06–0.10 s (60–100 ms) in duration.

Fig. 5.25 QRS complex**Fig. 5.26** ST segment

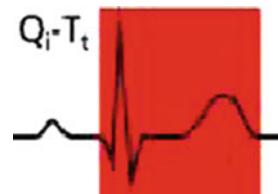
Not every QRS complex contains a Q wave, R wave, and S wave. By convention, any combination of these waves can be referred to as a QRS complex. However, correct interpretation of difficult ECG signal requires exact labeling of the various waves. The duration, amplitude, and morphology of the QRS complex are useful in diagnosing cardiac arrhythmias, conduction abnormalities, ventricular hypertrophy, myocardial infarction, electrolyte derangements, and other disease states. Q waves can be normal (physiological) or pathological. Normal Q waves, when present, represent depolarization of the interventricular septum. For this reason, they are referred to as septal Q waves and can be appreciated in the lateral leads I, a VL, V5, and V6. Q waves greater than 1/3 the height of the R wave, greater than 0.04 s (40 ms) in duration, or in the right precordial leads are considered to be abnormal and may represent myocardial infarction.

(c) ST Segment

The ST segment which is shown in Fig. 5.26 connects the QRS complex and the T wave and has duration of 0.08–0.12 s (80–120 ms). It starts at the J-point (junction between the QRS complex and ST segment) and ends at the beginning of the T wave. However, since it is usually difficult to determine exactly where the ST segment ends and the T wave begins, the relationship between the ST segment and T wave should be examined together. The typical ST segment duration is usually around 0.08 s (80 ms). It should be essentially level with the PR and TP segments. The normal ST segment has a slight upward concavity. Flat, down sloping, or depressed ST segments may indicate coronary ischemia. ST segment elevation may indicate myocardial infarction. An elevation of greater than 1 mm and longer than 80 ms follows the J-point. This measure has a false positive rate of 15–20% (which is slightly higher in women than men) and a false negative rate of 20–30%.

(d) T Wave

The T wave represents the repolarization (or recovery) of the ventricles. The interval from the beginning of the QRS complex to the apex of the T wave is referred to as the absolute refractory period. The last half of the T wave is referred to as the relative refractory period (or vulnerable period). In most leads, the T wave is positive. However, a negative T wave is normal in lead a VR. Lead V1 may have a positive,

Fig. 5.27 QT interval

negative, or biphasic T wave. In addition, it is not uncommon to have an isolated negative T wave in lead III, a VL, or a VF. Inverted (or negative) T waves can be a sign of coronary ischemia, Wellens' syndrome, left ventricular hypertrophy. Tall or “tented” symmetrical T waves may indicate hyperkalemia. Flat T waves may indicate coronary ischemia or hypokalemia. The earliest electrocardiographic finding of acute myocardial infarction is sometimes the hyperacute T wave, which can be distinguished from hyperkalemia by the broad base and slight asymmetry. When a conduction abnormality (e.g., bundle branch block, paced rhythm) is present, the T wave should be deflected opposite the terminal deflection of the QRS complex. This is known as appropriate T wave discordance.

(e) QT Interval

The QT interval which is shown in Fig. 5.27 is measured from the beginning of the QRS complex to the end of the T wave. A normal QT interval is usually about 0.40 s. The QT interval as well as the corrected QT interval is important in the diagnosis of long QT syndrome and short QT syndrome. The QT interval varies based on the heart rate, and various correction factors have been developed to correct the QT interval for the heart rate. The most commonly used method for correcting the QT interval for rate is the one formulated by Bazett and published in 1920.

Bazett's formula is

$$QT_c = \frac{QT}{\sqrt{RR}}$$

where QT_c is the QT interval corrected for rate and RR is the interval from the onset of one QRS complex to the onset of the next QRS complex, measured in seconds. However, this formula tends to be inaccurate, overcorrects at high heart rates, and under-corrects at low heart rates.

(f) U Wave

The U wave is not always seen. It is typically small and, by definition, follows the T wave. U waves are thought to represent repolarization of the papillary muscles or Purkinje fibers. Prominent U waves are most often seen in hypokalemia, but may be present in hypercalcemia, thyrotoxicosis, or exposure to digitalis, epinephrine, and Class 1A and 3 antiarrhythmics, as well as in congenital long QT syndrome and in the setting of intracranial hemorrhage. An inverted U wave may represent myocardial ischemia or left ventricular volume overload.

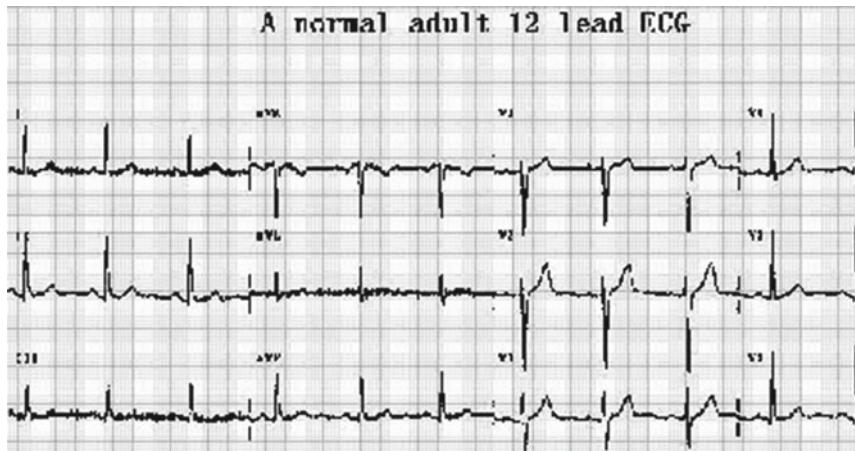


Fig. 5.28 Normal ECG signal

ECG Signal Monitoring Method

Electrodes are placed on designated areas of the patient's body, and these various combinations of the electrodes are used for analysis of the heart condition. Each separate view of the heart is called an ECG signal lead. The two ECG signal monitoring methods are standard 12-lead ECG signal monitoring [42] and continuous ECG signal monitoring or Holter monitoring. Twelve-lead ECG signal consists of three standard leads, designated as lead I, II, III and three augmented leads, designated as lead a VR, a VL, and a VF, that view the heart in the frontal plane, and six precordial or chest leads, designated V1 through V2, that view the heart in the horizontal plane. Both the standard leads and the augmented leads are limb leads. The standard leads are called bipolar because they are composed of two electrodes, one that is negative and other is positive, and the ECG signal records the difference in electrical potential between them. The standard 12-lead ECG signal records 12 different views of the same electrical activity on the ECG graph paper. Holter monitoring provides a continuous recording of heart rhythm during normal activity, and the monitor is usually worn for 24 h. In Holter monitoring, electrodes (small conducting patches) are placed on the chest and attached to a small recording monitor that can be carried or in a small pouch worn around the neck.

Normal ECG signal

In the normal ECG signal which is shown in Fig. 5.28, the PR interval should not exceed 0.20 s. The QRS duration should not exceed 0.10 s. The P wave duration should not exceed 0.10 s. The T wave should be at least 0.20 s wide. A heartbeat rate between 60 and 100 is considered "normal", so the RR interval should be between 0.6 and 1 s [102].

Artifacts

Electrocardiograph (ECG) signal artifacts are disturbances on ECG signal which is a measurement of cardiac potentials on the body surface. As a result of artifacts, normal components of the ECG signal can be distorted.

The word artifact is similar to artificial in the sense that it is often used to indicate something that is not natural (i.e., man-made). In electrocardiography, an ECG signal artifact is used to indicate something that is not “heart-made”. These include (but are not limited to) electrical interference by outside sources, electrical noise from elsewhere in the body, poor contact, and machine malfunction. Artifacts are extremely common, and knowledge of them is necessary to prevent misinterpretation of a heart’s rhythm.

Causes

Artifacts can be generated by patient’s motion or other electrical devices attached to or implanted (e.g., deep brain stimulator) in the body. Tremors and shivering are good examples of motion-induced artifacts. Simple movements such as brushing and combing the hair can cause ECG signal disturbances during ambulatory ECG signal monitoring. External sources of ECG signal artifacts mainly include power line electrical disturbances and radio frequency-based commercial (e.g., mobile phones) or medical devices. In operation theaters and intensive care unit, various equipments can affect ECG signal measurement system (e.g., electrodes, leads, amplifier, and filters). Examples of equipment which can cause ECG signal artifacts are electrocautery, transcutaneous nerve stimulator, hemofiltration machines. [139].

Noise Originating From Sources Outside the Patient

(a) Electrostatic Sources

Patient acts as one plate of a capacitor seen when a charged body is brought up close to an uncharged one, and an equal opposite charge develops on the uncharged body; e.g., if an unearthed body is close to any cable or lamp element that is connected to mains, he will develop a surface charge of equal and opposite potential even though no current is flowing between the two bodies. As the mains potential has a frequency of 50 Hz, the induced potential will also have this frequency. Other sources of electrostatic charge include the operating table, other persons, and electronic equipment.

(b) Electromagnetic Induction

An interference that occurs in the vicinity of wires carrying alternating current results in 50 Hz interference. Due to the generation of a magnetic field by the flow of a current, all conductors carrying mains current are surrounded by electromagnetic fields. The 50 Hz interference is a difference in potential, relative to the ground, that is impressed upon any subject in proximity to the wire carrying alternating current; the subject takes on a potential that is neither that of ground nor of the power line, but somewhere in between. Since the utility current is fluctuating, the voltage of the subject is also fluctuating. Effect is minimized by the fact that the electromagnetic

field generated by the live wire is to a greater degree canceled out by the neutral cable flowing adjacent to the live cable but flowing in the opposite direction. However, if leakage of current occurs, the two currents are no longer equal self-canceling, thereby generating an e.m.f. The effect is multiplied if the wires are coiled. Leads connecting patient electrodes to sensitive amplifiers are most frequently affected.

(c) Radio Frequency Interference [$> 100 \text{ kHz}$]

Radio frequency interference may enter via mains distribution system mixed up with 50 Hz current; sources include diathermy and electric motors. Radio propagation whereby activated diathermy probe held in air acts as radio transmitter aerial, while the patient ECG signal lead acts as a receiving aerial.

Massachusetts Institute of Technology/Beth Israel Hospital (MIT/BIH) Data base

This is a rich database of several hundred ECG signal recordings, extending over 200 h [204]. Each recording contains one to three signals and ranges from 20 s to 24 h in duration. Most of the signals have been annotated on beat-to-beat basis. In August 1989, a CD-ROM was produced containing the original MIT/BIH Arrhythmia Database (developed between 1975 and 1979, and first released in 1980), as well as a large number of supplementary recordings assembled for various research projects between 1981 and 1989. The CD-ROM contains approximately 600 megabytes of digitized ECG signal recordings, most with beat-by-beat annotations, having a total duration in excess of 200 h. In September 1991, these data were made available on Internet (www.physionet.org) on PhysioNet, which is a Web-based resource supplying well-characterized physiologic signals and related open-source software to the biomedical research community. From September 2000, the data archive named PhysioBank, containing roughly 35 gigabytes of recorded signals and annotations, was made available via PhysioNet [204].

5.4.2 Performance Comparison of Wavelet Threshold Estimators for ECG Signal Denoising

Heart disease is one of the leading causes of death in the world. The majority of this is due to coronary heart disease (CHD). It is estimated that the economic cost of CHD alone is about \$142.5 billion in the USA [275]. The electrocardiogram (ECG) signal is a recording of the heart's electrical activity and provides valuable clinical information about the performance of the heart. ECG signal is a type of electrical signal generated as myocardial tissues making up the heart constrict and relaxes under the regulation of the heart's impulse conduction system. Components of an ECG signal are shown in Fig. 5.29. The ECG signal contains pulses with different frequencies and amplitude. It has a high-frequency Q, R, S waves that forms QRS complex and low-frequency P and T waves. The time-varying behavior of these pulses makes ECG as highly nonstationary signal. One of the most serious problems in the recording of ECG signal is the parasitic interference by other high-frequency signals during recording process

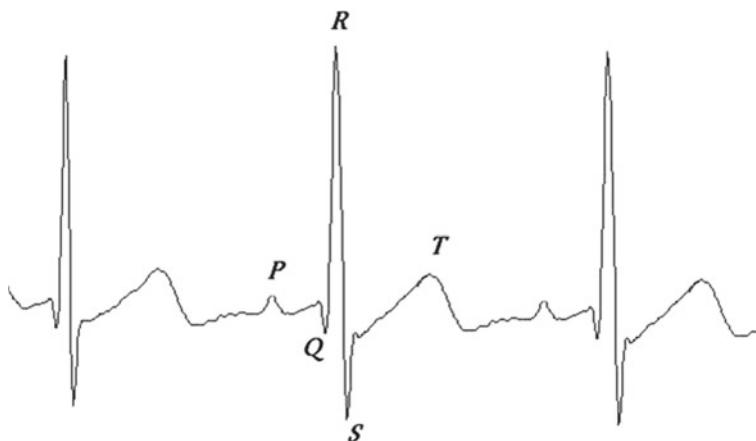


Fig. 5.29 Components of an ECG signal

in biomedical field. Most common artifacts are power line interference, EMG signal, motion artifacts, and base line interference [210]. In order to support clinical analysis, the ECG signal must be free from all noises and artifacts. Since ECG signal is very weak compared to noise, digital signal processing methods are used for removal of noises and artifacts. In order to remove noises of ECG signal, a number of techniques are available like digital filtering, adaptive methods, and wavelet transform methods [58, 162, 199]. However, digital filtering and adaptive methods are suitable only for stationary signals, whereas ECG signal is one of the biosignals that are considered as nonstationary signal. Recent works in the signal processing show that wavelet transform is the best suited for nonstationary signal analysis [97, 98, 118]. Denoising of signals using wavelet transform is based on thresholding of wavelet coefficients of noisy signals. Thresholding is used to smooth out or to remove some coefficients of wavelet transform subsignals of the measured signal. This reduces the noise content of the signal under the nonstationary environment. There are two common ways to threshold the resulting wavelet coefficients [97].

In the first case, the coefficients values are set to zero whose absolute value is below a threshold value. This is commonly known as hard thresholding. The second one, known as soft thresholding, goes one step further and reduces the magnitude of the remaining coefficients by the threshold value.

Hard thresholding maintains the scale of the signal but introduces ringing and artifacts after reconstruction due to discontinuity in the wavelet coefficients. Soft thresholding eliminates this problem resulting in smoother signal [97]. However, the critical problem in wavelet thresholding is to choose appropriate threshold value. A number of methods have been reported in the literature that focused on the computation of appropriate threshold value [97–101]. Universal threshold proposed by Donoho [97] is one of the most popular and widely used in wavelet thresholding. Minimax uses precomputed thresholds to minimize a constant term in upper bound

for the minimax risk of estimating a function using a threshold estimator [101]. SureShrink threshold is based on Stein's unbiased risk estimation (SURE) [100]. Hybrid thresholding technique based on the principle of SURE and universal thresholding is proposed by Donoho and Johnstone [98]. In this method, if the coefficients are not very sparse, then use SureShrink, if not then use universal threshold, thus becoming a hybrid threshold.

In this section, soft thresholding as well as hard thresholding have been used for the denoising of the ECG signal. Some of the most promising methods for the estimation of threshold value are investigated, and their results are compared.

The Proposed Method

While registering the ECG signal, it may get contaminated by random noises uncorrelated with the ECG signal. These noises can be approximated by white Gaussian noise.

The proposed method is implemented using the following steps.

1. White Gaussian noise with zero mean and constant variation is generated and added to the noise-free ECG signal. Mathematically, this may be written as

$$y(n) = x(n) + w(n)$$

where $x(n)$ is the noise-free ECG signal, $w(n)$ is the white Gaussian noise, and $y(n)$ is the noisy ECG signal.

2. Using an appropriate mother wavelet, the noisy ECG signal is decomposed to obtain approximate and detailed coefficients.
3. Choose a threshold value for thresholding. Selection of threshold value plays an important role in denoising of ECG signal. A number of methods for the threshold estimation have been proposed. In this section, we have evaluated the performance of the following threshold estimators on the denoising of ECG signal.

(a) Universal Thresholding

This is proposed by Donoho. The threshold value σ is given by

$$\sigma = \sqrt{2 * \log(n)}$$

where n is the number of samples in the signal.

(b) SureShrink Threshold Rule

SureShrink threshold is based on Stein's unbiased risk estimation (SURE). This minimizes the risk of the threshold value

$$\sigma = \sqrt{2 * \log(n * \log_2(n))}.$$

(c) Minimax Threshold Rule

Minimax uses precompounded thresholds to minimize a constant term in upper bound for the minimax risk of estimating a function using a threshold estimator.

The minimax threshold does not give an estimate with a good visual appearance, but it has the advantage of good predictive performance.

(d) Heursure Threshold Rule

This is a heuristic variant of SURE. It is obtained by combining universal threshold and SureShrink threshold rules.

4. After estimating the threshold values, apply thresholding to shrink the wavelet detailed coefficients of the noisy signal. Normally, there are two types of thresholding methods, hard thresholding and soft thresholding. In hard thresholding, all coefficients below the threshold value are set to zero. But in soft thresholding, in addition to that the remaining coefficients are also reduced linearly.
5. After thresholding, compute the inverse discrete wavelet transform to estimate the original ECG signal.
6. To evaluate the performance of the proposed method, mean square error (MSE) [138] between original signal and estimated signal is computed, which is given by

$$MSE = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \tilde{x}(n))^2$$

where $x(n)$ is the original ECG signal and $\tilde{x}(n)$ is estimated ECG signal.

Simulation and Results

For the simulation of the proposed method, an ECG signal has been taken as original signal, shown in Fig. 5.30a. The sampling frequency is 16 kHz, and 4000 samples of the signal are used. White Gaussian noise is added to the original ECG signal to introduce distortions. This noisy ECG signal shown in Fig. 5.30b is used as the test signal for the simulation of the proposed method. Daubechies (db3) wavelet is used to decompose the signal up to four levels. The simulation is done using MATLAB 7.0. The estimated signal from the noisy ECG signal using proposed method is shown in Fig. 5.30c. From the figure, it is clear that the noise has been greatly reduced.

Denoising of ECG signal is performed using all the four threshold value estimators, to compare their performance. Both hard and soft thresholding are used. To evaluate the performance of the proposed method, MSE is computed for various values of signal-to-noise ratio (SNR). The results obtained, from the simulation, are given in Tables 5.12 and 5.13. For the graphical representation, a plot of mean square error versus SNR is shown in Figs. 5.31 and 5.32. From the plots, it is observed that MSE is minimum in case of universal threshold for both hard and soft wavelet thresholdings.

Conclusions

Denoising of ECG signal is essential for the proper analyses of patient's heart condition. Wavelet thresholding is considered as one of the most preferred techniques for the signal denoising. In the wavelet thresholding, optimum threshold estimation is important for the proper denoising of the signal. In the present work, a comparative

Fig. 5.30 **a** Original ECG signal. **b** Noisy ECG signal. **c** Estimated ECG signal

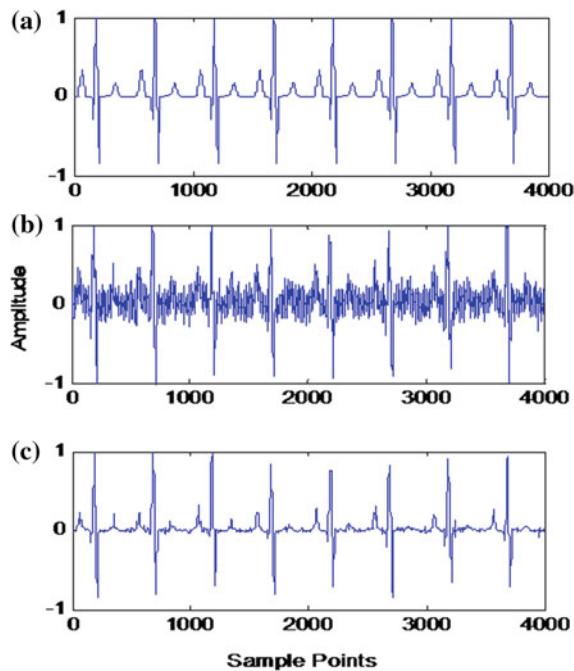


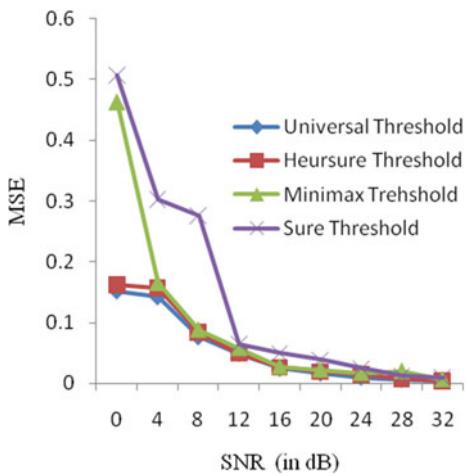
Table 5.12 Soft thresholding method

S.No.	SNR (in dB)	Mean square error (MSE)			
		Universal threshold	Heursure threshold	Minimax threshold	SURE threshold
1	0	0.1512	0.1623	0.4634	0.5066
2	4	0.1416	0.1567	0.1653	0.3028
3	8	0.0762	0.0844	0.0882	0.2762
4	12	0.0488	0.0494	0.0579	0.0647
5	16	0.0248	0.0267	0.0270	0.0499
6	20	0.0170	0.0194	0.0224	0.0390
7	24	0.0101	0.0153	0.0170	0.0245
8	28	0.0065	0.0073	0.0199	0.0123
9	32	0.0041	0.0041	0.0065	0.0094

study of various threshold estimators has been made for the ECG signal denoising. Mean square error is computed for different values of SNR of noisy ECG signal. The simulation is done using MATLAB 7.0. The simulation result shows that universal threshold, proposed by Donoho, gives best MSE performance at all SNR.

Table 5.13 Hard thresholding method

S.No.	SNR(in dB)	Mean square error (MSE)			
		Universal threshold	Heursure threshold	Minimax threshold	SURE threshold
1	0	0.3842	0.4597	0.7800	1.1972
2	4	0.1807	0.1796	0.3823	0.5314
3	8	0.0820	0.0833	0.1982	0.2685
4	12	0.0452	0.0542	0.0743	0.0742
5	16	0.0212	0.0328	0.0287	0.0278
6	20	0.0107	0.0142	0.0140	0.0172
7	24	0.0054	0.0079	0.0066	0.0072
8	28	0.0032	0.0044	0.0041	0.0043
9	32	0.0013	0.0023	0.0020	0.0025

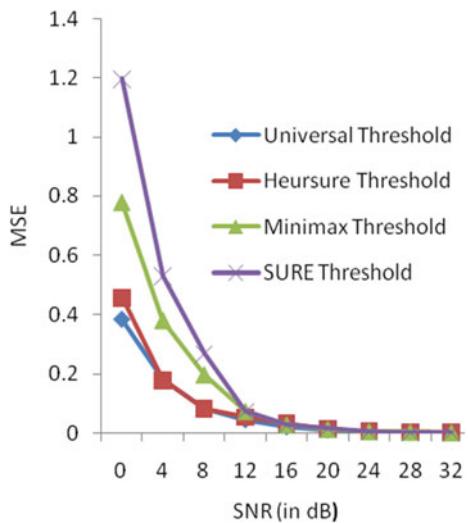
Fig. 5.31 Performance comparison threshold estimators using soft thresholding

5.4.3 Denoising of ECG Signal Using Wavelet Transform

ECG signal is constructed by measuring electrical potentials between various points of the body using a galvanometer. Understanding the various waves and normal vectors of depolarization and repolarization is very important to obtain useful diagnostic information. ECG signals have a wide array of applications throughout the medical field in determining whether the heart is functioning properly or suffering from any abnormalities.

ECG signal analysis is the gold standard for the evaluation of cardiac arrhythmias. It guides therapy and risk stratification for patients with suspected acute myocardial infarction. The baseline voltage of the electrocardiogram is known as the isoelectric line. A typical ECG signal tracing of a normal heartbeat (or cardiac cycle) consists

Fig. 5.32 Performance comparison threshold estimators using hard thresholding



of a P wave, a QRS complex, and a T wave. A small U wave is normally visible in 50–75% of ECG signal [218]. Figure 5.33 shows an example of a normal ECG signal trace, which consists of a P wave, a QRS complex, and a T wave. The small U wave may also be sometimes visible, but is neglected in this work for its inconsistency. The P wave is the electrical signature of the current that causes atrial contraction; the QRS complex corresponds to the current that causes contraction of the left and right ventricles; the T wave represents the repolarization of the ventricles; the U wave,

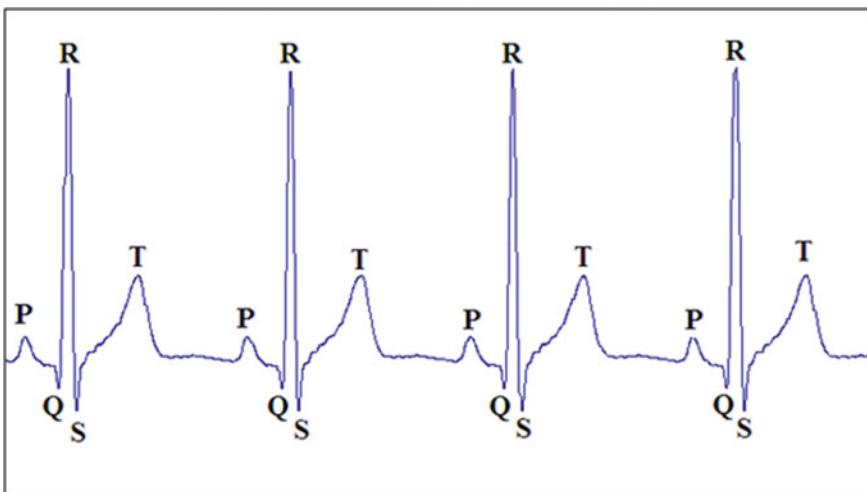


Fig. 5.33 Four cycles of normal ECG signal

although not always visible, is considered to be a representation of the papillary muscles or Purkinje fibers. The presence or lack of presence of these waves as well as the QT interval and PR interval is meaningful parameters in the screening and diagnosis of cardiovascular diseases.

The objective to analyze accurately an ECG signal is especially important in the applications where the feature extraction of the ECG signal is to locate the interested characteristic points that can be used to detect possible cardiovascular abnormalities. The topic is further complicated, since most of the time the desired ECG signals are either corrupted or embedded in noises. The answer to all of these problems is wavelet analysis. Wavelet theory provides a unified framework for a number of techniques, which had been developed independently for various signal processing applications.

Many approaches have been proposed for denoising ECG signal. Some of them, however, use either linear or nonlinear filter. The median filter fails to remove Gaussian noise since the larger the window size used, the more the characteristic of the ECG signal is removed [21]. On the other hand, linear filtering removes both the noise and high-frequency components of the ECG signal and the ECG signal then becomes blurred. Other algorithms use wavelet transforms. There are many applications of wavelets to biomedical signal processing which include some tasks such as noise reduction, enhancement, and detection [146, 231]. One application of the wavelet transform (WT) is the detection of the QRS complex in ECG signal. The detection properties of the WT have also been used extensively for biomedical signal processing tasks. Different QRS detection algorithms naturally need noise-free ECG signal to reduce the effect of artifacts. However, the original ECG signal should be distorted as little as possible by the filtering process. If noise has the same frequency band as the original signal, conventional filtering processes run into serious difficulty. Because of these reasons, noise removal from ECG signal is an interesting area, and moreover, one in which the application of wavelets has not been intensively studied [115, 134, 178, 277].

Thresholding is used in wavelet domain to remove some coefficient of wavelet transform subsignals of the measured signal. Many methods exist for choosing the value of threshold. Donoho and Johnstone proposed several thresholds, which work excellent in denoising the signal such as universal threshold, SureShrink, and minimax. It has been proved that Donoho's method for noise reduction works well for a wide class of one-dimensional and two-dimensional signals.

The major interest of this section is to reduce noise in noisy ECG signal. The determination of the wavelet transform and the choice of thresholding parameters are considered. We use mean square error (MSE) to check minimization of noise from noisy ECG signal using some of existing wavelet families. In this section, we find the different values of signal-to-noise ratio and mean square error using wavelet transform with appropriate threshold. To evaluate the performance, we compare our method with universal threshold method [199]. Experimental results show the advantage of proposed method.

Mathematical Representation of Wavelet

Wavelet is a powerful tool for the representation and analysis of ECG signal. They have been implemented for the analysis of physiological waveforms like ECG, phonocardiogram. [18, 192, 233]. This is because wavelet has finite duration as compared to Fourier methods based on sinusoids of infinite duration. Wavelet transform involves the decomposition of signal into various components. They provide both time and frequency view. Unlike Fourier transform, they are very efficient for nonstationary signals like ECG.

The Fourier transform is a widely used tool for many scientific purposes, but it is well suited for stationary signals. Gabor introduced a local Fourier analysis. He used the concept of a sliding window. This method, however, gives results when the coherence time is independent of frequency. Morlet introduced wavelet transform to have a coherence time proportional to the period. In wavelet transform, a fully scalable modulated window is used which solves the signal-cutting problem. The window is shifted along the signal. Spectrum is calculated for every position. This process is repeated by varying the length of the window. So, we have a collection of representations, hence the name multiresolution analysis.

Definition A function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is called a mother wavelet of order m if the following five properties are satisfied:

1. If $m > 1$, then ψ is $(m - 1)$ -times differentiable.
2. Let $\psi \in L^\infty(\mathbb{R})$. If $m > 1$, for each $j \in \{1, 2, \dots, m - 1\}$, $\psi^{(j)} \in L^\infty(\mathbb{R})$.
3. ψ and all its derivatives up to order $(m - 1)$ decay rapidly: For each $r > 0$, there is $\lambda > 0$ such that $|\psi^{(j)}(t)| < \frac{1}{t^r}$, $j \in \{1, 2, \dots, m - 1\}$ for each $|t| > \gamma$.
4. For each $j \in \{0, 1, \dots, m\}$, we have $\int t^j \psi(t) dt = 0$. This property of mother wavelet also called property of vanishing moment is very useful as it leads to economical representations of functions under study.
5. The set $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$, where $\psi_{j,k}$ are derived from the mother wavelet by relationship $\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$. Thus, expression for wavelet coefficients is given as

$$f_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt.$$

The Proposed Method

(a) Optimal Wavelet Selection

In wavelet domain, the selection of suitable mother wavelet is necessary for the ECG signal processing. For the ECG signal under investigation, an optimal wavelet will lead maximization of coefficient values in wavelet domain. This will produce highest local maxima of the ECG signal in wavelet domain. The possibility of best characterization of frequency content of the ECG signal is possible with optimally selected wavelet filter bank. In this section, Daubechies wavelet (db3) has been used as a mother wavelet.

(b) Wavelet Thresholding

The recovery of signal from its noisy version in wavelet domain is carried out with an assumption that smooth functions have economical wavelet representations and thereby most of the coefficients are set to zero without introducing larger error. Due to its orthogonal properties, a DWT transforms white noise to white noise with same variance. Thus, wavelet thresholding leads setting of small wavelet coefficients to zero and retaining or shrinking the coefficients corresponding to desired signal. Classical thresholding assures that the wavelet transform of smooth functions has economical representations, so that most of the coefficients are nearly zero, and white noise is transformed to white noise. Therefore, it is reasonable to assume that small coefficients are due to noise and can be set to zero, while the signal is stored in a few large coefficients, which should be retained. In fact, preprocessing steps are necessary for removing noise from the ECG signal before extracting the morphological parameters. There are two common ways to threshold the resulting wavelet coefficients. In the first case, the coefficient values are set to zero whose absolute value is below a threshold value. This is commonly known as hard thresholding. The second one, known as soft thresholding, goes one step further and reduces the magnitude of the remaining coefficients by the threshold value. Hard thresholding maintains the scale of the signal but introduces ringing and artifacts after reconstruction due to discontinuity in the wavelet coefficients. Soft thresholding eliminates this problem resulting in smoother signal.

In many research works, a common problem consists of recovering a true signal from incomplete, indirect, or corrupted signal. The development of fast computers has allowed the practical implementation of wavelets to solve this problem, through an algorithm called wavelet thresholding methods.

Assuming white Gaussian noise $n(t)$ is modeled with the ECG signal $y(t)$, which is represented by

$$z(t) = y(t) + n(t)$$

After performing the wavelet transform, we find

$$Z_{j,k} = Y_{j,k} + N_{j,k}$$

where $Z_{j,k}$ is the k th wavelet coefficient with the scale j . Mathematically, hard threshold is given by

$$\hat{Z}_{j,k}^{hard} = \begin{cases} Z_{j,k} & ; |Z_{j,k}| > \lambda \\ 0 & ; \text{otherwise} \end{cases}$$

where λ is threshold value.

Soft thresholding which is an extension of hard thresholding is given by

$$\hat{Z}_{j,k}^{soft} = sign(Z_{j,k}) \max\{|Z_{j,k}| - \lambda, 0\}.$$

The proposed method is implemented using the following steps.

1. Consider a clean ECG signal $x(n)$.
2. Generate a random white Gaussian noise $w(n)$, and add it to the original signal $x(n)$. Mathematically, it can be written as

$$y(n) = x(n) + w(n).$$

3. Compute the discrete wavelet transform of the noisy ECG signal $y(n)$.
4. A soft thresholding is used to shrink the wavelet detailed coefficient of the noisy signal.
5. The original ECG signal is reconstructed by taking the inverse discrete wavelet transform.
6. To evaluate the performance of the proposed method, mean square error (MSE) between original signal and estimated signal is computed, which is given by Haykin in [138]:

$$MSE = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \tilde{x}(n))^2$$

where $x(n)$ is the original ECG signal and $\tilde{x}(n)$ is estimated ECG signal obtained by the proposed method and N is the number of samples in the signal.

Simulation and Results

For the simulation of the proposed method, an ECG signal has been taken as original signal, shown in Fig. 5.34a. The sampling frequency is 16 kHz, and 10000 samples of the signal are used. White Gaussian noise (WGN) is used to model the background noise. This WGN is added to the original ECG signal to introduce distortions. This noisy ECG signal, shown in Fig. 5.34b, is used as the test signal for the simulation of proposed method. The estimated signal from the noisy ECG signal using proposed method is shown in Fig. 5.34c. From Fig. 5.34c, it is clear that the noise has been greatly reduced. Daubechies (db3) wavelet is used to decompose the signal up to four levels.

To evaluate the performance of the proposed method, MSE is computed for various values of signal-to-noise ratio (SNR). The results obtained, from the simulation, are given in Table 5.14. For the graphical representation, a plot of mean square error verses SNR is shown in Fig. 5.35. From the plot, it is observed that MSE of estimated signal has greatly reduced as compared to the MSE of existing method.

Conclusions

In the present work, an ECG signal denoising method based on wavelet transform is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy ECG signal and the original noise-free ECG signal. A very popular wavelet transform proposed by Daubechies (db3) is

Fig. 5.34 **a** Original ECG signal. **b** Noisy ECG signal. **c** Estimated ECG signal

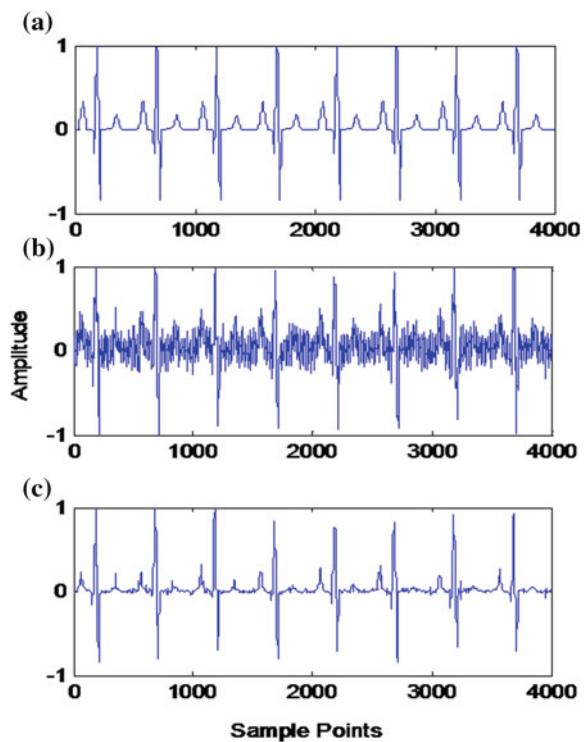
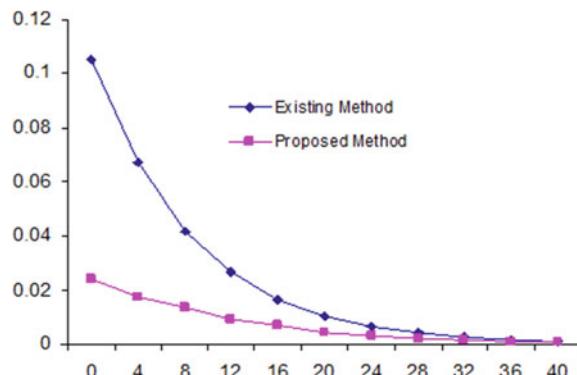


Table 5.14 Computation of MSE for SNR

S.No.	SNR (in dB)	Mean square error (MSE)	
		Existing method	Proposed method
1	0	0.1052	0.0243
2	4	0.0669	0.0175
3	8	0.0417	0.0139
4	12	0.0268	0.0095
5	16	0.0167	0.0069
6	20	0.0106	0.0045
7	24	0.0067	0.0033
8	28	0.0042	0.0022
9	32	0.0027	0.0019
10	36	0.0017	0.0013
11	40	0.00117	0.0010

Fig. 5.35 Performance of proposed method



used. To evaluate the performance of the proposed method, mean square error is computed for different values of SNR. The result obtained is compared with method used in [199], and it is found that the proposed method gives better result.

5.5 Applications of Wavelet Packets in ECG Signal Denoising

The analysis of the ECG signal is widely used for diagnosing many cardiac diseases, heart diseases, and arrhythmias. Most of the clinically useful information carried by the ECG signal is found in the morphology of the QRS complex, ST segment, and in the T wave, measured as amplitude, deformations, and duration (represented by Fig. 5.36). The ECG signal is characterized by a cyclic occurrence of patterns with different frequency contents (QRS complex, P and T waves). The QRS complex is the most important characteristic waveform of the ECG signal, reflecting the left ventricular depolarization, as expression of the electrical activity within the heart during the ventricular activation. Its shape, duration (0.6–1 s.), and time of occurrence provide valuable information about the current state of the heart. Cardiologists can use features of these signals to obtain important data about the clinical condition of their patients. These features are reflected by the morphology and duration of the individual waves of the ECG (P, QRS complex, and T waves). From the clinical (e.g., cardiologists and other specialties) point of view, the ECG signal analysis is a milestone in the assessment of a patient, and we consider the clinical criteria for determining the starting points and endpoints of the P wave, QRS complex, and T wave as essential. But in the mean time, we consider also that in the body surface ECG lays more information. Also, a lot of artifacts take place while recording ECG signal from the patients.

These artifacts are the main factors that cause problems in ECG signal processing. The removal of such kind of artifacts in ECG signal is essential to further diagnostic analysis for patient having a heart disease. The aim to remove these artifacts in ECG

signal is to separate the valid signal components from the undesired artifacts, so as to present an ECG signal that facilitates easy and accurate interpretation.

Commonly encountered artifacts include baseline drift, power line interference, physiological signals generated by other organs of the body or induced by muscular contractions related to breathing, and high-frequency random noise [67]. Several techniques have been proposed to address the problem of denoising ECG signals that are degraded by the noise. Adaptive filtering, Fourier transform, cubic spline, linear time-varying filtering, and wavelet transform are some of the methods, which are found in the literature [8, 37, 123, 135, 163, 197, 246, 258].

Among several artifacts, baseline drift is considered as an artifact which produces inaccurate data when measuring ECG signal parameters. To resolve this problem, the baseline estimation method which is based on cubic spline of the Maclaurin series is proposed [197]. In this method, a third order approximation is used because higher than forth order derivative are neglected, and then subtracted from original ECG signal. This method is nonlinear, and its performance is based on estimation of the signal in PR intervals only. Therefore, the main disadvantage is that estimating reference points does not belong to the baseline.

A linear time-varying filtering method is also undertaken, which suppresses the baseline drift in the ECG signal [258]. In this case, the beat average is subtracted from the signal and then decimated. Low-pass filtering is applied to estimate the baseline drift, interpolated, and then subtracted from the original signal. This is again a nonlinear approach; therefore, it is complex and highly dependent on the beat rate calculations and becomes less accurate for low heart rate. However, these techniques are not effective and reliable for analysis of ECG signals [124].

Morlet [134] introduced wavelet transform to have a time proportional to the period, which is generally based on decomposition of signal into various components. In this case, we get information of the signal in both time and frequency view. Unlike Fourier transform, they are very efficient for nonstationary signals like ECG [144]. Wavelet transform has become an important tool for the multiscale representation and analysis of signals. A number of researchers are applying this new tool for the reduction of noise in ECG signals [145, 201]. The signal denoising using wavelet transform is based on the threshold and shrinkage function selection. In this method, a discrete time signal is split up by high-pass and low-pass filters, which is done recursively for low-pass coefficients only.

Therefore, the wavelet transform does not give us detailed signal analysis. The wavelet packet method is a generalization of wavelet decomposition that offers a detailed signal analysis. Wavelet packets give a more complex and flexible analysis, because in wavelet packet analysis, the detailed coefficients as well as the approximation coefficients are split simultaneously. So, even few levels of wavelet packet decomposition give a lot of bases from which we can choose the best basis.

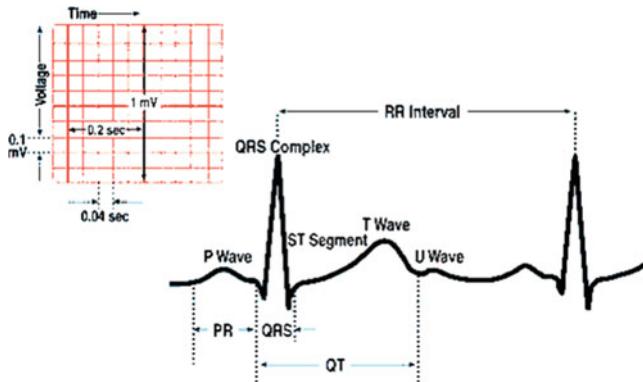


Fig. 5.36 Cardiac cycle of ECG signal

5.5.1 Wavelet Packet Thresholding Method for Denoising ECG Signal

In this section, a wavelet packet thresholding method has been proposed for denoising ECG signal. To choose optimal threshold value, we have considered a basic assumption. Since noise in the signal is additive white Gaussian noise, it has a variance with invariant of time. Hence, the fixed global threshold is employed when noisy signal is analyzed in each scale. A very popular wavelet proposed by Daubechies (db3) is used as mother wavelet. It can be used effectively in noisy ECG signals and other signals. To evaluate the denoising performance, the proposed method based on wavelet packet transform is compared with the method used in [228].

Proposed Methodology

The proposed method can be implemented by the following steps:

1. White Gaussian noise with zero mean and constant variation is generated and added to the noise-free ECG signal. Mathematically, this may be written as

$$f(x) = s(x) + n(x)$$

where $s(x)$ is the noise-free ECG signal, $n(x)$ is the white Gaussian noise, and $f(x)$ is the noisy ECG signal.

2. Using third-order Daubechies wavelet (db3) as mother wavelet, the noisy ECG signal is decomposed to obtain approximate and detailed coefficients.
3. The contaminated signal is decomposed with wavelet packet transform with proper threshold value. To choose optimal threshold value, we have considered a basic assumption. Since noise in the signal is additive white Gaussian noise, it has a variance with invariant of time. Also, most of the noise variance lies on detailed coefficients of first level of decomposition. So, the fixed global threshold

is employed when noisy signal is analyzed in each scale.

Threshold value is estimated with the coefficients as

$$thr = \sigma \sqrt{2 * \log(N)/N}$$

where the variance σ can be estimated using median absolute deviation (MAD) as

$$\sigma = \frac{\text{MAD}(c_i)}{0.6745}$$

where c_i are the high-frequency wavelet coefficients. The factor 0.6745 rescales the numerator so that σ is a suitable estimator for the standard deviation for Gaussian white noise. A soft thresholding is used to shrink the detailed coefficients of the noisy signal.

4. Results obtained by this method are given in Table 5.15.
5. The estimated signal $\hat{f}(x)$ is reconstructed by applying inverse wavelet packet transform by keeping all approximated coefficients and all threshold detailed coefficients.
6. To evaluate the performance of the proposed method, the MSE values are found using the formula

$$MSE = \frac{1}{N} \sum \left[\hat{f}(x) - f(x) \right]^2$$

where $f(x)$ is the original signal, $\hat{f}(x)$ is the estimated signal, and N represents the number of sample points

Simulation and Results

In order to illustrate the effectiveness of proposed algorithm, ECG signal has been taken from the data records MIT/BIH [296]. This ECG signal which is taken from data record is distorted by white Gaussian noise which is shown in Fig. 5.37b. The algorithm based on wavelet packet transform is successfully removed to this noise. The estimated signal is shown in Fig. 5.37c, which indicates that wavelet packet transform is an improved method when compared with wavelet transform method [228].

The wavelet transform and wavelet packet transform are used for denoising of ECG signal. To evaluate the performance of this method, mean square error has been calculated at different values of signal-to-noise ratio (SNR). The results obtained, from the simulation, are given in Table 5.15. For the graphical representation, a plot of mean square error versus SNR is shown in Fig. 5.38. From the plot, it is observed that mean square error (MSE) is minimum in the proposed method (wavelet packet transform).

Fig. 5.37 **a** Original ECG signal. **b** Noisy ECG signal. **c** Estimated ECG signal

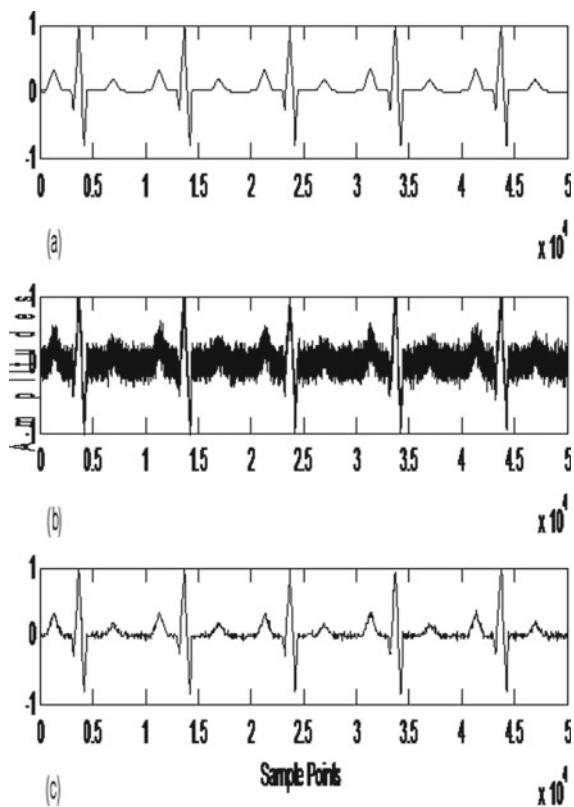


Fig. 5.38 Wavelet transform versus wavelet packet transform

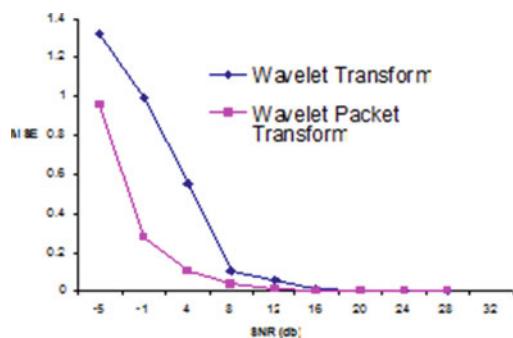


Table 5.15 Computation of MSE for SNR

SNR (in dB)	Mean square error (MSE)	
	Universal	Heursure
-5	3.6919	2.2334
-1	1.3202	0.9598
4	0.9908	0.2803
8	0.5498	0.1039
12	0.0997	0.0355
16	0.0563	0.0123
20	0.0135	0.0052
24	0.0068	0.0025
28	0.0049	0.0012
32	0.0030	0.0010

Conclusions

In the present work, an ECG signal denoising method based on wavelet packet decomposition is proposed. To choose optimal threshold value, we have considered a basic assumption. Since noise in the signal is additive white Gaussian noise, it has a variance with invariant of time. Hence, the fixed global threshold is employed when noisy signal is analyzed in each scale. A very popular wavelet proposed by Daubechies (db3) is used as mother wavelet. To evaluate the performance of the proposed method, mean square error is computed for different values of SNR. The result obtained is compared with [55], and it is found that the proposed method (wavelet packet transform) gives better result.

5.5.2 Baseline Drift Correction in ECG Signal Using Wavelet Packets Transform

In ECG signals, baseline drift is the problem that can influence the accurate diagnosis of heart diseases, such as ischemia and arrhythmia. Muscle contraction and electrode impedance changes due to movement of the body are the important sources of baseline drift in most types of ECG signal recordings. Detection of ischemia can be achieved by analyzing the ST segment of the ECG signal, and in some cases, the analysis may be influenced by baseline drift as noise [139]. Additionally, the determination of an accurate ECG baseline is generally needed for the localization of ventricular arrhythmias with body surface potential mapping [30]. The comparison of isopotential maps from different beats requires reliable determination of the baseline to achieve reproducible and consistent results. The importance of baseline correction

for the reconstruction of activation time imaging from electrocardiographic mapping is indicated in [299].

Many methods of removing the artifacts in ECG signals were proposed in last two decades [83, 236]. Cubic spline method has been used for the estimation of baseline drift in ECG signal. This is a non linear method, and performance of this method depends on the knots determination accuracy. The main disadvantage of this method is estimating undesirable signal distortion due to the overlapping of signal and disturbance. The interferences are estimated, but the useful components of ECG signal are removed.

In [24], a linear filtering approach has been used. In this approach, the high-pass filter with 0.5 Hz cutoff frequency can be used to remove the interference of baseline drift which can filter out signal components with frequency below 0.5 Hz, while frequency above 0.5 Hz is preserved. Nonlinear phase filters are usually avoided because they can introduce significant distortions to the ECG signal and, consequently, increase the chance of heart disease misdiagnosis.

Adaptive filtering has been also used to estimate baseline drift. This filtering makes the assumption that extremities of the features of the signal are known. The adaptive techniques first apply in combination with a least mean square-driven adaptive impulse correlated filter and a two-stage cascade filter for the correction of baseline drift. This technique requires detection of QRS complex and the transfer function of the cascade filter [200]. Another drawback in these filters is slow tracking properties.

Discrete time-frequency transform has also been used to estimate base line drift. In [211], short-time Fourier transform (STFT) has been used to estimate the presence of baseline drift in ECG signals, which can then be removed using a time-varying filter. Two problems may be identified by using this approach. First, the STFT uses a window of constant length which needs to be properly chosen and which fixes the resolution both in time. Second, the redundant STFT is used only for estimation of the baseline drift.

Our approach based on the wavelet packet transform corrects the baseline drift successfully and computationally less intensive than the approach based on the STFT described in [211]. Additionally, because the wavelet packet transform decomposes the signal at different scales, the size of the window can be easily adopted to signal by changing the number of levels in the decomposition. This approach of wavelet packet transform makes more convenient to remove baseline drift than the STFT approach. This approach to remove baseline drift based on the wavelet packet transform assumes that the optimal correction must satisfy the two requirements. First, it must remove the low frequency elements that are not related to the cardiac electrical activity; second, it must preserve the shape and amplitude of the PQRST complexes.

ECG Signal Characteristics

The electrocardiogram (ECG) signal describes the electrical activity of the heart. It is obtained by placing electrodes on the chest, arms, and legs. With every heartbeat, an impulse travels through the heart, which determines its rhythm and rate and causes the heart muscle to contract and pump blood. The voltage variations measured by the electrodes are caused by the action potentials of the excitable cardiac cells, as

they make the cells contract. The ECG signal is characterized by a series of waves whose morphology and timing provide information used for diagnosing diseases reflected by disturbances of the electrical activity of the heart. The time pattern that characterizes the occurrence of successive heartbeats is also very important.

The first ECG signal-recording device was developed by the Dutch physiologist Willem Einthoven, using a string galvanometer which was sensitive enough to record electrical potentials on the body surface. He also defined sites for electrode placement on the arms and legs which remain in use today. Since then, ECG signal recording has developed incredibly and become an indispensable tool in many different contexts. The ECG signal record is used today in a wide variety of clinical applications. Its importance has been strengthened thanks to the discoveries of subtle variability patterns which are present in rhythm or wave morphology.

The electrodes used for ECG signal recording are positioned so that the spatiotemporal variations of the cardiac electrical field are sufficiently well reflected. The difference in voltage between a pair of electrodes is referred to as a lead. The ECG signal is typically recorded with a multiple-lead configuration. The electrode wires are connected to a differential amplifier specially designed for bioelectrical signals. The ECG signal ranges from a few microvolts to about 1 V in magnitude. Whereas the characteristic waves of an ECG signal have a maximal magnitude of only few millivolts, baseline drift in the ECG signal due to variations in electrode-skin impedance may reach 1 V.

The characteristic waves of an ECG signal are shown in Fig. 5.39. Atrial depolarization is reflected by the P wave and ventricular depolarization is reflected by the QRS complex, whereas the T wave reflects ventricular repolarization. The amplitude of a wave is measured with reference to the ECG signal baseline level, commonly defined by the isoelectric line which immediately precedes the QRS complex.

One of the main reasons for computer-based ECG signal analysis is the capability to improve poor signal quality using signal processing algorithms. There are several common types of noise and artifacts in the ECG signal. The baseline drift is a low-frequency activity in the ECG signal which may interfere with the signal analysis, making the clinical interpretation inaccurate. When baseline drift takes place, ECG signal measurements related to the isoelectric line cannot be computed since it is not well defined. Baseline drift is often exercise-induced and may have its origin in a variety of sources, including perspiration, respiration, body movements, and poor electrode contact. The spectral content of the baseline drift is usually in the range between 0.05 and 1 Hz, but during strenuous exercise, it may contain higher frequencies [137].

The electromyography noise is caused by the electrical activity of skeletal muscles during periods of contraction, commonly found in ECGs recorded during ambulatory monitoring exercise. Different muscles are active in producing the noise which corrupts the ECG signal. This kind of noise can either be intermittent in nature, due to a sudden body movement, or have more stationary noise properties. The frequency components of this noise considerably overlap those of the QRS complex wave.

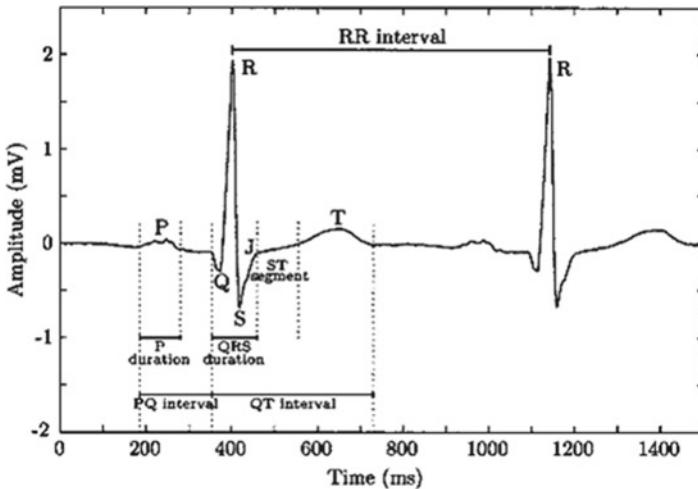


Fig. 5.39 ECG signal

Methodology

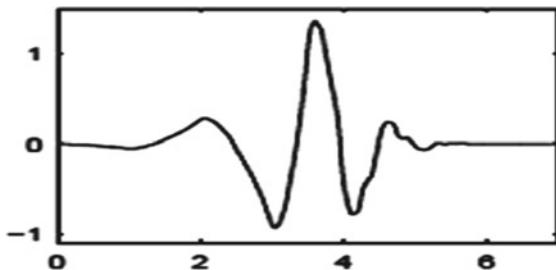
Most of the ECG signals are accompanied with noise which may be modeled as Gaussian white noise. One of the most popular noises known as baseline drift is generally found in ST segment of ECG signal. Due to the presence of such type of noises, the received signal may be of very little use. Hence, the process of information extraction from the ECG signal due to baseline drift is difficult task and gives incorrect information. Although many methods have been found in the literature [301], these methods contribute some additional artifacts while removing baseline drift present in ECG signals. In this section, a wavelet packet transformation has been employed to recover a signal from the signal with baseline drift. The proposed algorithm involves the following processes to remove the baseline drift in an ECG signal:

1. Analysis
2. Decomposition
3. Thresholding
4. Reconstruction
5. Percent Root Mean Square Difference (PRD).

Analysis

One of the measured steps is to choose a mother wavelet while applying wavelet packet transformation. The wavelet chosen should be similar to the signal that has to be filtered to give the best possible results. This similarity can be decided on the basis of the cross-correlation between the two functions. In this chapter, one of the Daubechies families (Fig. 5.40) of wavelets, i.e., db4, has been taken, because of their high number of vanishing moments making them capable of representing complex

Fig. 5.40 Daubechies wavelet (db4)



high-degree polynomials [268]. The result of our simulations shows that db4 wavelet provides sufficiently good output.

Decomposition

In wavelet packet decomposition, detailed coefficients and approximation coefficients are split up to best level. Finding the best level of decomposition is again another fundamental characteristic of denoising an ECG signal with baseline drift as noise. Here a trial and error method has been considered for proposed methodology. We decompose the signal to the level 6, 7, 8, and 9 by db4 as mother wavelet in wavelet packet domain. A Percent Root Mean Square Difference (PRD) has been used for analysis of these levels.

Thresholding

Thresholding is used in wavelet domain to smooth out or to remove some coefficients of wavelet transform subsignals of the measured signal. This reduces the noise content of the signal under the nonstationary environment. There are two common ways to threshold the resulting wavelet coefficients. In the first case, the coefficient values are set to zero whose absolute value is below a threshold value. This is commonly known as hard thresholding. The second one, known as soft thresholding, goes one step further and reduces the magnitude of the remaining coefficients by the threshold value. Hard thresholding maintains the scale of the signal but introduces ringing and artifacts after reconstruction due to discontinuity in the wavelet coefficients. Soft thresholding eliminates this problem resulting in smoother signal. In this section, we have evaluated the performance of ECG signal using wavelet packet threshold. A discrete time signal is split up by high-pass and low-pass filters, which is done recursively for low-pass coefficients in wavelet transform. In wavelet packet transform, high-pass coefficients are split as well so that the best basis can be chosen to represent the signal in few but large coefficients. Such a denoising technique in wavelet packet domain is called “wavelet packet thresholding”. The wavelet packet coefficients of noisy signals are thresholded by a nonlinear thresholding function, which is obtained by taking the difference of maximum value of ECG signal coefficients and mean value of ECG signal coefficients, and then divide this value by 2 [82].

Reconstruction

Reconstruction procedure converts wavelet coefficients into the original sampled data. Inverse wavelet packet transform is used after applying the wavelet packet transform and threshold procedures [55, 116, 155]. In inverse process, we have taken all approximation coefficients and the thresholded detailed coefficients up to level 8 to obtain the output, which gives us the baseline-free ECG signal.

Percent Root Mean Square Difference (PRD)

The performance of the proposed method has been computed using Percent Root Mean Square Difference (PRD) method [82].

$$PRD = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}} \times 100$$

where x_i is the original ECG signal, \hat{x}_i is the estimated ECG signal, and \bar{x}_i is the mean of original ECG signal.

Simulation and Results

For the simulation of proposed method, an ECG signal has been taken from the MIT/BIH ECG database [296]. All segments are of 360Hz sampling rate and also present quite important baseline deviation. In addition, the segments include both normal and abnormal heartbeats. In this section, an ECG signal is taken from data

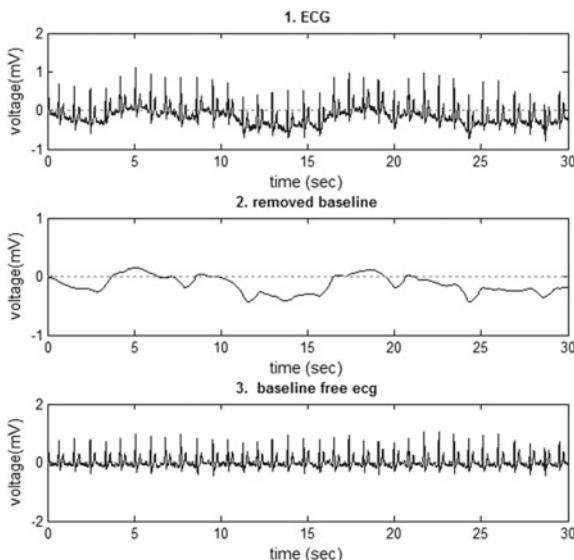


Fig. 5.41 ECG MIT/BIH record with B0 baseline

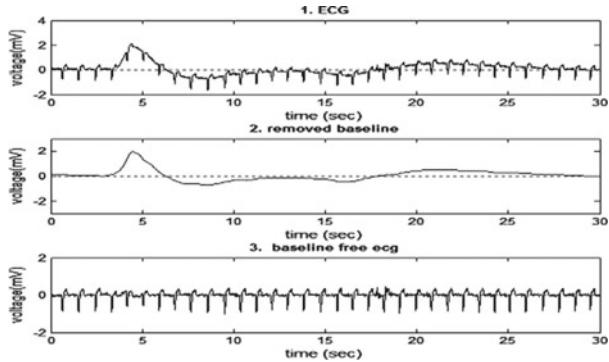


Fig. 5.42 ECG MIT/BIH record with B1 baseline

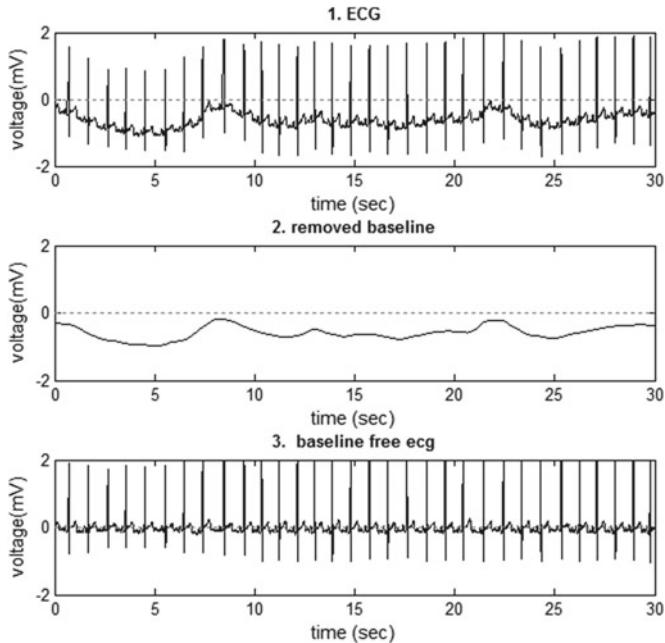


Fig. 5.43 ECG MIT/BIH record with B2 baseline

record and add three different artificially baseline; i.e., B0 is a 0.1 Hz sin wave, B1 is a 0.05 Hz triangular wave, and B2 is a 0.05 Hz sin wave to generate three distorted ECG signals. The simulation is done using MATLAB 7.0. After applying wavelet packet transform, we obtained correct baseline drift. Figure 5.41 shows ECG signal with baseline drift B0 and baseline-free ECG signal, Fig. 5.42 shows ECG signal with baseline drift B1 and baseline-free ECG signal, and Fig. 5.43 shows ECG signal with baseline drift B2 and baseline-free ECG signal. To evaluate the performance of the

Table 5.16 PRD value (%) for different decomposition levels 6, 7, 8, and 9

Level of decomposition	Percent Root Mean Square Difference (PRD)		
	Baseline B0	Baseline B1	Baseline B2
6	1.3243	1.4321	1.4129
7	1.6831	1.6820	1.5831
8	0.2708	0.2753	0.3832
9	1.5386	1.8613	1.4192

proposed method, PRD is calculated. The level of decomposition plays a crucial role in the removal of baseline drift. The PRD for the different levels of decomposition is also evaluated. The result obtained is given in Table 5.16. From this table, it is observed that the best level of decomposition is level 8.

Conclusions

Estimation of baseline drift is an important task for the proper analysis of patient's heart conditions. Wavelet packet is considered one of the preferred techniques for estimation of baseline drift in ECG signal. In the wavelet packet transform, proper estimation is important for preserving ECG signal characteristics. In the present work, wavelet packet transform has been used for the correct estimation of baseline drift in ECG signal. Percent root mean square difference is computed for different levels. The simulation is done using MATLAB 7.0. The simulation result shows that level 8 is best for correct estimation of baseline drift in ECG signal.

Chapter 6

Applications in Image Processing



6.1 Introduction

Signal processing problems are not confined to one-dimensional signals. Many image processing applications require the use of two-dimensional signal processing techniques. This is the case, for example, in x-ray enhancement, the enhancement and analysis of aerial photographs for detection of forest fires or crop damage, the analysis of satellite weather photos, and the enhancement of television transmissions from lunar and deep space probes. Seismic data analysis as required in oil exploration, earthquake measurements, and nuclear test monitoring also utilizes multidimensional signal processing techniques [224].

It is equally important to consider two-dimensional signals, which is to say, images. Here again, image processing is done on the numerical representation of the image. For a black and white image, the numerical representation is created by replacing the x and y coordinates of an image point with those of the closest point on a sufficiently fine grid. The value $f(x, y)$ of the “grayscale” is then replaced with an average coefficient, which is then assigned to the corresponding grid point.

The image thus becomes a large, typically square, matrix. Image processing is done on this matrix. These matrices are enormous, and as soon as one deals with a sequence of images, the volume of numerical data that must be processed becomes immense. Is it possible to reduce this volume by considering the “hidden laws” or correlations that exist among the different pieces of numerical information representing the image? This question leads us naturally to define the goals of the scientific discipline called “signal processing.”

Image encoding techniques are broadly classified into two categories, namely first-generation image encoding techniques and second-generation image coding techniques. The most important first-generation techniques use an orthonormal basis for representing an image, while second-generation techniques exploit image characteristics and the psychophysics of human visual perception.

An image is a compactly supported, real-valued function in $L^2(\mathbb{R}^2)$. In practice, it is discretely sampled and represented by a matrix $c = (c(i,j))_{(RC)}$ with R rows and C columns, with integer entries ranging from 0 to $K - 1$, for some positive integer K . The discrete spatial positions (i, j) are called pixels. The entries $c(i, j)$ are called the pixel values or grayscale levels and represent an amplitude discretization of the image intensity (gray-level value) at the corresponding pixels.

In general, an image can contain complex features at several different scales and hence representing it at an adequate resolution often requires a very large matrix. Faithful representation, storage, and transmission of images are thus costly exercises in terms of computation, memory, and time.

A feature commonly found in images is the high correlation between the values of neighboring pixels, often resulting in high redundancy of information content in the images. The representation of images verbatim as discrete pixels is thus, frequently, inefficient. The principal task of an efficient coding scheme for images is, therefore, the elimination of this redundancy, leading to significant data compression. Typically, these schemes fall into two broad categories, namely the casual and the noncasual predictive schemes.

In casual predictive coding schemes, pixels are encoded row-wise sequentially (raster format). Before encoding each pixel value, it is “predicted” from already encoded pixel values. This predicted value represents redundant information and, therefore, is subtracted from the actual value of the pixel, and only the difference (prediction error) is encoded. These encoding schemes are referred to as casual since they employ only previously encoded pixels for prediction. A pixel value is decoded by computing its predicted value from previously decoded pixel values and adding it to the encoded prediction error.

Noncasual (neighborhood-based) predictive schemes are more natural and result in more accurate prediction, leading to higher compression. It is, however, harder to implement such schemes as they do not permit sequential coding but involve application of image transforms or the solution of large systems of simultaneous equations, resulting in high computational overheads.

Multiscale Edge Representation of Images

Edges in images are characterized by sharp variations in intensity values. However, these variations can occur at several scales, ranging from edges of large objects (at low resolutions), and contours of smaller objects (at higher resolutions) to texture (at even higher resolutions). The distinction between edge information and texture information in an image is largely contextual in the sense that what might appear as edges at a particular resolution may appear as texture at a lower resolution. In any given image, this distinction is largely based on the psychophysics of human vision. Nevertheless, edges and texture are the two most important characteristics of images, from the point of human visual perception. Edges are more important than texture for image understanding and object recognition, and distortion or degradation of edge information markedly affects the quality of the image and the recognizability of various features in it. Texture carries contextual information about lighting, surface features, depth, and other perceptual cues of objects in an image, and while being less

structured and harder to characterize than edge information, it affects the perception and quality of an image to a significant degree. However, distortions introduced in texture are not as perceptible to the human eye as edge distortions, and this aspect is taken advantage of to obtain high compressions by encoding texture information less accurately, while at the same time introducing distortions that are not very apparent to the human eye. Contours of objects, while being crucial to characterizing the objects, can be encoded more economically than the objects themselves as they are spatially much sparser than the image, and since the range of intensities is considerably reduced, far fewer bits are required to encode the contours. The question then is, can one faithfully represent objects images by means of edge and texture information alone? In other words, how close is the image recovered from its edge-texture representation to the original image? Another issue is that of distinguishing “important” edge corresponding to contours of objects from those corresponding to noise or texture.

As mentioned before, an edge in an image is characterized by a change in image intensity. To examine the nature of this change, we may for the sake of simplicity (and without loss of generality) consider the case of one-dimensional signals. A transition in the signal corresponding to an (ideal) edge may be modeled by an appropriately scaled Heaviside function, whose discontinuity is located at the place where the transition in the signal occurs. However, in reality, changes in intensity do not occur abruptly at a single point, as modeled by a Heaviside function, but rather over a transition band or an interval. The width of this band may be very narrow or very wide, depending on the scale at which the intensity takes place.

The image intensity changes occur over a wide range of scales, and a gradual change may go undetected at a fine scale, while sharp transitions may be hard to localize, or appear as noise or texture, at a coarse scale. Thus, it does not make much sense to talk of intensity changes without reference to the scale at which these changes are taking place. In order to analyze an image at different scales, it is necessary to smooth the image with filters of appropriate time-scale characteristics. The effect of smoothing an image with a low-pass filter is that of taking local averages of the image intensities. This results in a low-pass filtered image in which the range of scales over which intensity changes take place is decreased (e.g., halved). The smoothed (and downsampled) image may now be examined for intensity changes occurring in its range of scales. This process is repeated over and over again to obtain a multiscale analysis of intensity variations of the image. Since the purpose of filtering the image is to reduce the range of scales over which its intensity variations occur, the filter function must be smooth and well localized in the Fourier (scale/frequency) domain. While the averaging is done to reduce the range of scales over which the variations are observed, it is also important to keep these averages local, since transitions in intensity values at each scale correspond to phenomena that are spatially localized at that scale. This requirement imposes that the corresponding filter function be well localized and smooth in the spatial domain as well. Thus, the filter function is so chosen that both it and its Fourier transform are smooth and well-localized functions.

At any given resolution, the edges in the smoothed image may be locally modeled by smoothed and appropriately scaled versions of the Heaviside function. Edge points thus correspond to inflection points of the smoothed image. These inflection points in turn correspond to local maxima of the absolute value of the first derivative or, equivalently, to the zero crossings of the second derivative of the smoothed image (intensity function). In the two-dimensional (actual) case, the first and second derivatives are replaced by the first and second directional derivatives. When the smoothing function is a Gaussian, the method of detecting edge points by means of the local maxima of the modulus of the first directional derivative corresponds to a Canny edge detector, while the employment of zero crossings of the second directional derivative corresponds to a Marr–Hildreth edge detector.

The second derivative, however, also vanishes at inflection points of the image that correspond to gradual variations, i.e., points at which the rate of change of intensity is very slow. However, these points do not correspond to edges, and therefore, it is hard to distinguish these inflection points from the ones that correspond to rapid intensity changes using zero crossings of the second derivative alone. On the other hand, the “slow” inflection points correspond to local minima of the modulus of the first directional derivative. This means that the modulus of the first derivative of the intensity map is capable of distinguishing between the fast and the slow inflection points (corresponding to maxima and minima, respectively, of the modulus of the first derivative). Thus, one can select the inflection points corresponding to edges using the modulus of the first derivative more easily than from the zero crossings of the second derivative. Also, the zero crossings of the second derivative indicate only the position of the intensity changes and not their magnitudes. It is, therefore, not possible to distinguish small variations from prominent transitions using the zero crossings. The values of the maxima of the modulus of the first derivative, however, are a measure of the intensity changes. Thus, the modulus maxima method of edge detection (Canny edge detection) is more advantageous than the zero crossing approach (Marr–Hildreth edge detection) for purposes of edge detection.

Edge information of an image can be obtained at various scales by extracting edges from smoothed versions of the image at various resolutions. This process is called multiscale edge detection and is very useful for object/pattern recognition and image compression.

Multiscale Edge-Based Image Coding

In this encoding technique, the most significant edge (maxima) curves are first selected and then efficiently encoded. The significance of an edge curve is, in general, a contextual valuation and depends to a considerable extent on the importance of the semantic information provided by the curve from the point of pattern or object recognition. In general, it is hard to evaluate a curve’s significance, as it is image dependent. In this technique, however, the significance of an edge curve is based on whether its length is above a certain threshold value, the rationale behind this criteria being that the contours of the salient objects in an image give rise to long edge curves. The efficient coding capitalizes on the positional and structural similarity of the important edge curves across scales. Since most of the frequency (detail)

information is concentrated at the finer scales, the edge encoding is restricted to these scales. Thus, only the scales 2^j , $1 \leq j \leq j_0$ for some appropriately chosen j_0 (say $j_0 = 3$) are considered for edge extraction, while the larger scale information is stored in the low-frequency image $S_{2^{j_0}}^d I$. The edge information at very fine scales contains a large number of “insignificant” edges corresponding to texture and noise and is computationally expensive at the edge selection stage. Hence, an appropriate intermediate scale 2^j , $1 \leq j \leq j_0$ (say $j = 2$) is chosen from which to select the important edge curves. First, the edge curves whose lengths are below a threshold value are eliminated. Next, from the rest of the curves, only those that correspond to rapid variations in image intensity are retained. This operation is performed by computing the average value of the modulus $M_{2^j} I$ and thresholding it below. The positional information (maxima locations) of these selected edge curves is recorded at scale 2^j and taken to be the same at all other scales 2^j , $1 \leq j \leq j_0$. These edge curves are encoded predictively by encoding the first pixel of each curve and encoding only the difference in position of each successive pixel from that of the previous pixel. This method of encoding is shown in [50] on an average to take 10 bits for the first pixel and 1.3 bits for each successive pixel. The cost of this encoding is further halved by performing it after downsampling the edge image by 2 and linearly interpolating the edge point locations after upsampling the encoded edge image. Since at any scale the gradient at the edge points is orthogonal to the tangent to the curve to which the edge point belongs, the angles $\theta_{2^j} I$ at the edge points can be directly approximated from the tangents to the encoded curves at all scales 2^j , $1 \leq j \leq j_0$, and, hence, are not recorded. Thus, the position and angle information of the edge points at the scales 2^j , $1 \leq j \leq j_0$ are the same. The values of the modulus $M_{2^j} I$ are encoded at each scale 2^j , $1 \leq j \leq j_0$ along the edge curves in a predictive fashion employing a coarse quantization of the predicted values. The energy of the low-frequency image $S_{2^{j_0}}^d I$ is largely concentrated in the frequency domain over a region whose size is 2^{-2j_0} times smaller than the bandwidth of the image I . Hence, it is downsampled by a factor of 2^{j_0} in each direction to obtain a low-frequency image that is 2^{-2j_0} times smaller than the original image. This method of encoding results in high compression ratios that vary with the number of edge points remaining after edge selection.

Texture-Based Image Encoding

The difference between the original image and the image reconstructed from the multiscale edge-based encoding yields an error image. Since the edge-based encoding disregarded texture information, the error image contains the texture information of the image along with certain object features whose edges were omitted. Due to the high compression achieved in the edge-based encoding, the error image tends to contain more information than the edge-encoded image. The texture information in the form of the error image is encoded using an orthonormal wavelet pyramidal technique. As most of the edge features are encoded, one can achieve high compression ratios in encoding the error image without creating rippling effects (Gibbs phenomena) at points of rapid intensity variations. To economize on the number of wavelet coefficients to be encoded, each of the three high-pass images at each scale

is partitioned into smaller square blocks, and in each block the total energy of the wavelet coefficients in that block is computed (by evaluating the sum of the squares of these coefficients); if the energy is below a threshold value, all the coefficients in that block are reset to zero. The justification for this operation is that there is no texture in this block and any nonzero coefficient is either due to noise or due to some insignificant fluctuation in image intensity. The rest of the coefficients values are quantized based on the frequency of occurrence. This encoding preserves most of the image's texture.

The reconstructed edge-based encoding and the reconstructed texture-based encoding are added to obtain the approximate image with edge and texture information. The compression factors obtained by this double-layered encoding technique are impressive, ranging from 40 for a 256×256 pixel image to 100 for a 512×512 image.

Image Compression Algorithms

Some of the main issues confronted in the design of image compression algorithms are:

- (i) choice of wavelet basis;
- (ii) choice of metric to measure compression error;
- (iii) compression strategy: linear versus nonlinear;
- (iv) smoothness of images;
- (v) encoding of compressed wavelet coefficients;
- (vi) computational efficiency.

Choice of Compression Metric

The criteria used to evaluate compressed images should depend on the intended application. For example, producing compressed images for video are quite a different matter from producing them for medical imagery. In the former, the human visual system would be the ultimate judge of quality, while retention of diagnostic fidelity is the main issue in the latter. In some applications, image compression is done as a preprocessor to other image processing tasks such as feature classification or object recognition. Compression can then have the effect of reducing the complexity of the postprocessing task and enhancing the features important for the postprocessing application. For example, some compression strategies can be viewed as removing noise from images.

To carry out a sensible mathematical analysis of compression, it is desirable to quantify in mathematical terms the compression criteria. One way of doing this is to attempt to find a metric defined on images to measure the error between the original and the compressed image. The metric chosen should model the goals of the compression algorithm.

Most metrics used for measuring compression error arise by considering images as functions defined on a square or rectangle Ω . Even there is some ambiguity in how to interpret the image as a function. One possibility is to consider the image as the piecewise constant function given by its pixel representation. Another is to

consider the image as the function which is its wavelet representation. Still another possibility is to think of the image as a function defined on a continuum and its pixel values as samples obtained by averaging over small squares (the function F introduced earlier). Fortunately, the decision of how to view the image as a function has no essential effect on the design and analysis of compression algorithms

$$I \sim \sum_{j \in \mathbb{Z}^2} c_j \varphi_{j,m}(x, y) \quad (6.1.1)$$

$$I \sim \sum_{j \in \mathbb{Z}^2} c_j \varphi_{j,m}(x, y) = \sum_{j \in \mathbb{Z}^2} d_{j,0} \varphi_{j,0} + \sum_{j=0}^{m-1} \sum_{j \in \mathbb{Z}^2} \sum_{\psi \in \Psi} c_{j,k} \psi_{j,k} \quad (6.1.2)$$

For the purposes of the following discussion of compression algorithms, it will be convenient to think of an image τ as its wavelet representation, i.e., as the function appearing on the right side of (6.1.1) (or equivalently (6.1.2)).

The most frequently used compression metric is the L^2 -norm. The reason for this seems to be that it is easiest to describe optimal compression strategies with this choice since we can compute compression error exactly in terms of pixel values or wavelet coefficients (in the orthogonal wavelet case). On the other hand, there is much evidence to indicate that this may not always be the best choice given the intended application of compression.

For example, to produce visually pleasing compressed images, it seems more reasonable to try to model the human visual system. Experiments of this type in the psychological community suggest that the L^1 -norm models better some aspect of the human visual system (see for example the discussion in [90]).

While other possibilities could be considered, we shall assume that our compression metric comes from one of the L^p -(quasi) norms, $0 < p \leq \infty$.

Compression Strategies: Linear and Nonlinear

We suppose that we have chosen a compression metric that models our compression criteria; we assume that this is one of the L^p -metrics $0 < p \leq \infty$. We also assume that we have chosen the wavelet basis we shall use to represent our image. We obtain compression by retaining some of these terms in the wavelet decomposition (6.1.2) of the image and deleting others. But what exactly should be our strategy for the retention of terms.

For notational convenience, we shall assume that all terms corresponding to $\varphi_{j,0}$ will be retained in the compressed image; in any case, there are very few of them.

We want to draw the distinction between two classes of compression algorithms: linear and nonlinear. In linear algorithms, one choose a dyadic level $0 \leq K \leq m$ and retains in the compressed image \tilde{I} all terms in (6.1.2) corresponding to $k < K$. We say that this is a linear algorithm because we can view the compressed image as an approximation to I from the linear space $S^K = S^K(\varphi)$.

If the L^2 -projector was used in the construction of the wavelet ψ , then the linear compression strategy is optimal with respect to compression in the L^2 -metric in the

sense that once we have decided to approximate the image I by functions from the linear space S^K then our strategy for choosing the approximation is optimal. Indeed, the portion of the wavelet decomposition (6.1.2) corresponding to the values of $k < K$ gives the L^2 projection of I on the space S^K and is therefore the best L^2 approximation from that space. It is also important to note that in linear methods the compression strategy (the terms that we retain) does not depend on the image.

In nonlinear compression methods, we allow the terms that we retain in the compressed image \tilde{I} to depend on the image I . Thus, typically some but not necessarily all terms are chosen from each dyadic level. If n is the total number of terms to be retained, this can be thought of as approximation from the nonlinear manifold consisting of all functions of g which are a linear combination of at most n of the functions appearing on the right side of (6.1.2).

There is one case where it is particularly simple to describe the optimal nonlinear compression strategy, namely if φ and ψ have orthogonal shifts and the compression is done in the metric L^2 . If we fix the number n of terms we wish to retain in the compressed image (which obviously is related to the rate of compression), then the optimal strategy is to retain the n terms in (6.1.2) for which

$$\|C_{j,k}\psi_{j,k}\|_{L^2} = |2^{-k}C_{j,k}|\|\psi\|_{L^2}$$

is largest (with any criteria used to break ties).

In numerical implementation, to avoid the time expensive sorting, the following modification of this strategy is implemented. We fix a positive number $\varepsilon > 0$ which will measure the quality of the compressed image. We retain in the compressed image \tilde{I} all terms of (6.1.2) for which

$$|2^{-k}C_{j,k}| > \varepsilon \quad (6.1.3)$$

The selection (6.1.3) is called thresholding. The smaller value of ε , the more terms are retained in the compressed image and the smaller the compression error.

A disadvantage of thresholding is that one does not know in advance the relationship between the choice of ε and the compression ratio. One possible way to circumvent this difficulty is to adaptively choose ε . One begins with a large value of ε so that no coefficients satisfy the threshold criteria (6.1.3). Given in hand any value of ε , we can replace ε by $\varepsilon/2$ and add to our compressed image the additional terms that satisfy (6.1.3). We proceed until we have the desired quality in the compressed image or the desired compression ratio. The buildup of the compressed image in this way is progressive since we only add to the previously compressed image and do not have to redo any previous calculations. A discussion of progressive transmission for nonlinear-based wavelet compression and their encoding can be found in Shapiro [243].

There is no another important variant in the nonlinear strategy. Rather than using the entire coefficient $C_{j,k}$ in the terms that are retained, we replace it by an approximation $\tilde{C}_{j,k}$ which satisfies

$$2^{-k} |C_{j,k} - \tilde{C}_{j,k}| \leq \varepsilon. \quad (6.1.4)$$

This is called quantization and is implemented numerically by retaining only the leading bits of the coefficients $C_{j,k}$. The effect of the L^2 -quantization strategy (6.1.4) is to retain one less bit in the coefficients as we move from one dyadic level to a finer dyadic level.

If in place of the L^2 -metric, $p \neq 2$, or if we use nonorthogonal wavelets, then it is not so clear what should be the strategy for thresholding or quantization. The following strategy was derived in [92] and shown to be optimal in a sense that will be described in the next subsection. If we consider the compression problem as retaining the best n terms of the wavelet decomposition of the image so as to minimize L^p compression error then the strategy put forth in [92] is to choose the n terms for which

$$\|C_{j,k} \psi_{j,k}\|_p = |2^{-\frac{k}{p}} C_{j,k}| \|\psi\|_{L^2} \quad (6.1.5)$$

are largest. Again it leads to corresponding thresholding and quantization strategies obtained from (6.1.3) and (6.1.4) by replacing 2^{-k} by $2^{-k/p}$. For example, the quantization strategy when using the L^1 error metric is to choose two less bits as we move from one dyadic level to the next finer dyadic level.

Smoothness of Images

A common way of making comparisons between compression algorithms is to test their performance on certain standardized images. This obviously has its limitations. Another possibility for comparing performance of compression algorithms is to try to classify images according to their compressibility. Compression is a form of approximation and we know from such questions in approximation theory that the rate of approximation of a function is connected with its smoothness. This leads us to consider the classification of images by their membership in smoothness spaces.

While there are many smoothness spaces, we shall limit our discussion to the Besov spaces. They are robust enough to measure all orders of smoothness in all L^p -spaces. As we shall see, there is room for discussion as to whether these are appropriate spaces for measuring the smoothness of images.

For any $0 < s \leq \infty$ and $\alpha > 0$, we let $B_q^\alpha(L^s)$ to denote the Besov space of smoothness of order α in L^s with secondary parameter q . We do not give the definition of these spaces here (for definition see [92]). One can heuristically view the space $B_q^\alpha(L^s)$ as consisting of the functions f from L^s whose α th derivative is in L^s (the parameter q gives a more subtle delineation of smoothness). However, note that we allow α to be noninteger and s to be less than one.

We can classify the smoothness of an image by its membership in the Besov spaces. Since Besov space norms are applied to functions, we need to view the image as a function. One possibility is to consider the image as the underlying function F (introduced earlier) whose cell averages gave the pixel values. It is also reasonable to consider either the pixel representation (piecewise constant function) or the wavelet representation as being the function representing the image. We adopt the latter

viewpoint in this section. One could argue that the wavelet function representation is always smooth (because it consists of a finite sum of smooth functions $\psi_{j,k}$) and that the image would be in all Besov spaces within the smoothness of the wavelet. But correctly viewed, the wavelet representation is just a partial sum of some larger series and this larger series has some intrinsic smoothness. For the purpose of the following discussion, one should adopt this latter viewpoint.

The value of measuring smoothness of images is that it gives us a way of evaluating the performance of compression algorithms. For example, if the dual functionals for the wavelet basis have r vanishing moments, then standard results in approximation theory [91] say that the linear method of compression will approximate a bivariate function f to an order $O(n^{-\alpha/2})$ (with n the number of terms retained and $\alpha < \tau$) in the L^p -metric if and only if f is in the Besov space $B_\infty^\alpha(L^p)$. Hence, we can expect this performance by the compression algorithm only if the image has this smoothness.

A similar result holds for the nonlinear approximation which gives rise to the nonlinear compression algorithm [92]. In this case, there is a characterization of the bivariate functions f which have approximation order like $O(n^{-\alpha/2})$ for $\alpha < \tau$. The function f should be in the Besov space $B_\tau^\alpha(L^\tau)$ with $\tau = \left(\frac{\alpha}{2} + \frac{1}{p}\right)^{-1}$.

It is interesting to compare the linear and the nonlinear algorithms from this vantage point. To obtain an L^p -error $O(n^{-\alpha/2})$ (with n the number of terms retained), both require smoothness of order α but the linear algorithm measures this smoothness in the space L^p (the same space as the compression error is measured) while the nonlinear algorithm measures the smoothness in L^τ with $\tau = \left(\frac{\alpha}{2} + \frac{1}{p}\right)^{-1}$. Note that $\tau < p$ and it is more likely for a function to be in the Besov space for nonlinear approximation than in the one for linear approximation. Thus, the nonlinear algorithm will compress more images to order $O(n^{-\alpha/2})$, i.e., require less smoothness for the image to be compressed with this accuracy.

For example, consider compression in the L^2 error metric. For compression error of order $O(n^{-\alpha/2})$, the image should have smoothness of order α as measured in L^2 . Since images have edges and other discontinuities they cannot be in these Besov space for large values of α . For example, the characteristic function of a square is in the Besov space $B_\infty^\alpha(L^p)$ only if $\alpha \leq \frac{1}{2}$, and hence the linear compression algorithm will not give compression error $O(n^{-\alpha/2})$ for any $\alpha > 1/2$. However, in the case of the nonlinear algorithm, the smoothness is measured in L^τ . For example, this same characteristic function of a square is in the space $B_\tau^\alpha(L^\tau)$, $\tau = \left(\frac{\alpha}{2} + \frac{1}{2}\right)^{-1}$ provided $\alpha < \frac{1}{\tau}$, or equivalently if $\alpha < 1$. In other words, images can have higher smoothness orders α when measured in the L^τ spaces and therefore can be compressed more efficiently by the nonlinear algorithms. A similar discussion applies to approximation in the metric L^p .

Choosing the coefficients for which (6.1.5) is largest provides a near optimal strategy for nonlinear approximation in the following sense. For functions in the unit ball of the Besov spaces $B_\tau^\alpha(L^\tau)$, this strategy provides an approximation error $O(n^{-\alpha/2})$. No other compression strategy based on approximation by a sum $\sum_{j,k,\psi \in \Lambda} b_{j,k} \psi_{j,k}$,

with $|\Lambda| \leq n$, can improve on this error. There are however the constants in the $O(n^{-\alpha/2})$ term about which we are not able to say anything. These constants are of importance in numerical comparisons and it may be that some other strategy could improve on (6.1.5) in these constants.

There is another sense in which the nonlinear strategy based on (6.1.5) is optimal. Given any stable nonlinear approximation scheme, not necessarily based on wavelets, it cannot improve on the error estimate $O(n^{-\alpha/2})$ of (3) (see [94]). Therefore, if we agree in advance that we want to compress all images from unit ball of the Besov space $B_\tau^\alpha(L^r)$ then no stable compression strategy can do better whether it is based on wavelets or some other ideas such as fractals or Fourier expansions.

The above strategy also gives a method for numerically computing the smoothness of an image. One computes the errors from the nonlinear approximation strategy (6.1.5), and graphs them on a log–log scale (error versus number of coefficients). This graph will look linear for a certain range of n (typically for $n \leq 30,000$ in the case of 512×512 images). The slope of this line gives the largest α for which the image is in $B_\tau^\alpha(L^r)$. We refer the reader to [90] for a more detailed discussion of the numerical determination of smoothness of images.

Further discussion of linear and nonlinear compression algorithms can be found in [90]. They show the overall superiority of the nonlinear algorithms.

Choice of Wavelet: Fixed Basis

There is the interesting question of whether any particular wavelet is preferable over another in compression algorithms. We have analyzed several wavelets for this purpose. In order to isolate the question of choice of wavelet basis, we have made two restrictions in the analysis that follows. First, in order to not have the compression strategy play a role in this analysis, we have chosen to measure the compression error in the L^2 -metric. In this case, when using orthogonal wavelets, we know the optimal compression strategy for thresholding or quantization (we use only thresholding in the analysis that follows). Secondly, our analysis considers compression error (in the L^2 -metric) versus the number of wavelet coefficients. A more accurate analysis of the performance of a compression algorithm would study the compression error versus the number of bits needed to encode the compressed image. However, the latter would depend on the encoding of the wavelet coefficients (discussed in the last section). The best encoding could very well depend on the wavelet that is utilized in the compression algorithm. To eliminate this dependence on encoding we have simply counted coefficients.

It seems that a more important issue is the playoff between the support of the coefficient functionals and the number of its vanishing moments. If the coefficient functional has large support then the dominant discontinuities in the image effect many wavelet coefficients. In low compression, the large coefficients from the dominant discontinuities are small in number compared to the number of terms retained and the deciding factor in the compression rate are the coefficients corresponding to smoother parts of the image. These coefficients are smaller because the coefficient functionals have many vanishing moments. For high compression, however, the coefficients corresponding to the dominant discontinuities are the deciding factor.

Overall, the biorthogonal spline-based wavelets perform better than the Daubechies wavelets. An explanation put forward for this by others is the symmetry in the biorthogonal wavelets. We should also reiterate the advantage of certain of the biorthogonal wavelets in that the computation of the wavelet decomposition can be done in integer arithmetic with shifts.

Hyperbolic Bases

The use of hyperbolic bases in image compression brings forward new questions as to how compression algorithms should be implemented with these wavelets. As with the standard wavelet basis, one can divide the compression question into two parts corresponding to linear methods (approximation from linear spaces) and non-linear methods. There is a theory [93] for linear methods of compression based on hyperbolic wavelets that are similar in spirit to that given in compression strategies: linear and nonlinear. We do not want to discuss in detail these results but mention only a couple highlights. This theory introduces new moduli of smoothness based on certain mixed differences and defines new smoothness spaces in a similar way to the Besov spaces but using the new moduli of smoothness in place of the usual moduli of smoothness. It is shown then that these new smoothness spaces replace the roles of the Besov spaces when characterizing rates of approximation by the linear spaces H^n .

Unfortunately, there is not yet a corresponding theory for the nonlinear compression problem using hyperbolic wavelets except for special case when the error is measured in L^2 . One of the reasons for this is that there is no characterization of the approximation spaces for nonlinear approximation using hyperbolic wavelets. This however does not prevent the experimentation with hyperbolic wavelet in image compression. The following is an initial report on some experience with hyperbolic wavelets.

We first discuss the representation of digitized images by hyperbolic wavelet bases. In order to simplify the following discussion by not having to deal with extensions near the boundary of the unit square (as discussed in wavelet representation of grayscale images), we shall limit our discussion to the case of the hyperbolic Haar basis. Using ideas similar to that presented in wavelet representation of grayscale images, we can derive expansions of images using other hyperbolic basis.

For this discussion, we let $\varphi = \chi_{[0,1]}$ and let ψ be the univariate Haar function. We start with the representation (6.1.1) of the image I as

$$I \sim \sum_{j,k \in \mathbb{Z}^2} C_j \varphi_{j,m}(x, y) \quad (6.1.6)$$

with $\varphi(x, y) = \varphi(x)\varphi(y)$ and C_j defined to be the given pixel values p_j in case the support of $\varphi_{j,m}$ is contained in Ω and zero otherwise. The univariate functions $\varphi_{s,m}$, when $0 \leq s < 2^m$ are a linear combination of $\varphi_{0,0}$ and the univariate Haar functions ψ_{j_1,k_1} , $0 \leq j_1 < 2^{k_1}$. Making this change of basis gives the hyperbolic Haar representation

$$I \sim \sum_{j,k \in \mathbb{Z}^2} C_{j,k} \eta_{j,k}(x, y) \quad (6.1.7)$$

with $j = (j_1, j_2)$ and $k = (k_1, k_2)$ and

$$\eta_{j,k}(x, y) = \psi_{j_1, k_1}(x) \psi_{j_2, k_2}(y)$$

with the convention that $\psi_{0,0}$ is replaced by $\varphi_{0,0}$ and $C_{j,k} = 0$ unless the support of $\eta_{j,k}$ is contained in Ω . As is discussed in the following section, the coefficients in the representation (6.1.7) can be found from the univariate fast wavelet transform.

Given the hyperbolic wavelet representation (6.1.7) (or its analog in the non-Haar case) of an image, we can implement compression strategies to threshold or quantize the coefficients $C_{j,k}$. We have implemented the analog of the nonlinear L^2 -compression strategy (6.1.3) for various hyperbolic wavelets. We again display L^2 -compression error versus the number of coefficients using only thresholding. The results are representative of our usual experience that for some images (such as San Francisco), the hyperbolic wavelet bases perform slightly better than the usual wavelet basis while on other the opposite is true. It seems therefore that the choice whether to use the standard wavelet basis or the hyperbolic basis in compression depends on the image to be compressed.

6.2 Applications of Wavelets in Image Denoising

The growth of media communication industry and demand of high quality of visual information in modern age have an interest to researchers to develop various image denoising techniques. In recent years, there has been a plethora of work on using wavelet thresholding techniques [99] for removing noise in both the signal processing and statistics community due to its effectiveness and simplicity. In its most basic form, technique denoises in the orthogonal wavelet domain, where each coefficient is thresholded by comparing against a threshold; if the coefficient is smaller than the threshold, it is set to zero, otherwise it is kept or modified. The intuition is that because the wavelet transform is good at energy compaction, small coefficients are more likely due to noise, and large coefficients due to important signal features. The threshold thus acts as an oracle deciding whether or not to keep the coefficients.

An image is often corrupted by noise during its acquisition or transition. The objective is to remove the noise without affecting the important features of the image. The most commonly used procedure to remove the noise is wavelet shrinkage by nonlinear method, named WaveShrink, proposed by Donoho and Johnstone [99, 100] and Donoho [97]. WaveShrink is based on the fact that for many of real-life signals, a limited number of wavelet coefficients in the lower subbands are well sufficient to reconstruct the original signal. Usually, the numerical values of these coefficients are relatively large compared to noise coefficients. In literature, a number of techniques for adaptive selection of threshold values [52, 56, 266] are reported.

In this section, an image denoising technique based on wavelet decomposition of image signal is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy image and denoised image. The present technique can effectively remove the noise. To evaluate the denoising performance of the present technique, mean square error (MSE) is computed for different noise levels. The simulation is done using MATLAB 7.0 version.

Extension of the Orthogonal Wavelet Representation to Images

Wavelets are families of functions obtained by taking dilations and translations of a particular function with sufficient decay in both time and frequency domain. In 2D wavelet transform, the 2D scaling function is given by

$$\varphi(x_1, x_2) = \varphi(x_1)\varphi(x_2)$$

and separable, “directionally sensitive” wavelets are

$$\psi^h(x_1, x_2) = \psi(x_1)\varphi(x_2)$$

$$\psi^v(x_1, x_2) = \varphi(x_1)\psi(x_2)$$

$$\psi^d(x_1, x_2) = \psi(x_1)\psi(x_2)$$

where $\psi(\cdot)$ represents translated and scaled versions of a mother wavelet function and $\varphi(\cdot)$ be the translated versions of a low-pass scaling function. Let h , v , and d be the horizontal, vertical, and diagonal directions, respectively. These wavelets measure functional variations–intensity variations for images along different directions, ψ^h measures variations along columns, ψ^v corresponds to variations along rows and ψ^d corresponds to variations along diagonals. The directional sensitivity is a natural consequence of the separability.

Given a separable 2D scaling and wavelet functions, extension of the 1D discrete wavelet transform to 2D is straight forward. The scaled and translated basis functions are given by

$$\varphi_{j,k}(X) = 2^{j/2}\varphi(2^jx_1 - k_1, 2^jx_2 - k_2)$$

$$\psi_{j,k}^i(X) = 2^{j/2}\psi^i(2^jx_1 - k_1, 2^jx_2 - k_2), \quad i = \{h, v, d\}$$

where $X = (x_1, x_2) \in \mathbb{R}^2$, $K = (k_1, k_2) \in \mathbb{Z}^2$.

The discrete wavelet transform of image $f(x_1, x_2)$ of, $M \times N$, $\forall M, N \in \mathbb{N}$ is

$$W_\varphi(j_0, k_1, k_2) = \frac{1}{\sqrt{MN}} \sum_{x_1=0}^{M-1} \sum_{x_2=0}^{N-1} f(x_1, x_2) \varphi_{j_0, k_1, k_2}(x_1, x_2)$$

$$W_{\psi}^i(j, k_1, k_2) = \frac{1}{\sqrt{MN}} \sum_{x_1=0}^{M-1} \sum_{x_2=0}^{N-1} f(x_1, x_2) \psi_{j, k_1, k_2}^i(x_1, x_2), \quad i = \{h, v, d\}$$

where j_0 is an arbitrary starting scale and the $W_{\varphi}(j_0, k_1, k_2)$ coefficients define an approximation of $f(x_1, x_2)$ at scale j_0 . The $W_{\psi}^i(j, k_1, k_2)$ coefficients add horizontal, vertical, and diagonal details for scale $j \geq j_0$. We normally consider $j_0 = 0$, $M = N = 2^j$, $j = 0, 1, 2, \dots, J - 1$, and $k_1 = k_2 = 0, 1, 2, \dots, 2^j - 1$. Using W_{φ} and W_{ψ}^i , $f(x_1, x_2)$ is expressed using IDWT as follows:

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{\sqrt{MN}} \sum_{k_1} \sum_{k_2} W_{\varphi}(j_0, k_1, k_2) \varphi_{j_0, k_1, k_2}(x_1, x_2) \\ &\quad + \frac{1}{\sqrt{MN}} \sum_{i=h, v, d} \sum_{j=j_0}^{\infty} \sum_{k_1} \sum_{k_2} W_{\psi}^i(j, k_1, k_2) \psi_{j, k_1, k_2}^i(x_1, x_2). \end{aligned}$$

Proposed Technique

In the present section, soft thresholding is used for the shrinkage of wavelet coefficients. The proposed method is implemented using following steps:

Step 1: Load a clean original image $f(x_1, x_2)$.

Step 2: Generate a random white noise $w(x_1, x_2)$ add it to the original image signal.

Mathematically it can be written as

$$\tilde{f}(x_1, x_2) = f(x_1, x_2) + w(x_1, x_2).$$

Step 3: Compute the discrete wavelet transform of the noisy image $\tilde{f}(x_1, x_2)$.

Step 4: Choose a threshold value for thresholding. The optimal threshold value is chosen by taking the minimum error between detailed coefficients of noisy image and those for original image. A soft thresholding is used for shrinkage of the wavelet detailed coefficients of the noisy signal.

Step 5: The clean image is reconstructed by taking the inverse discrete wavelet transform.

Step 6: To reduce the noise, it is passed through the Wiener filter.

Step 7: To evaluate the performance of the proposed technique mean square error (MSE) between original image and estimated image is computed, which is given by

$$MSE = \frac{1}{MN} \sum_{x_1=0}^{M-1} \sum_{x_2=0}^{N-1} [f(x_1, x_2) - \tilde{f}(x_1, x_2)]^2$$

where $f(x_1, x_2)$ is the original image and $\tilde{f}(x_1, x_2)$ is the denoised image obtained by the proposed method.

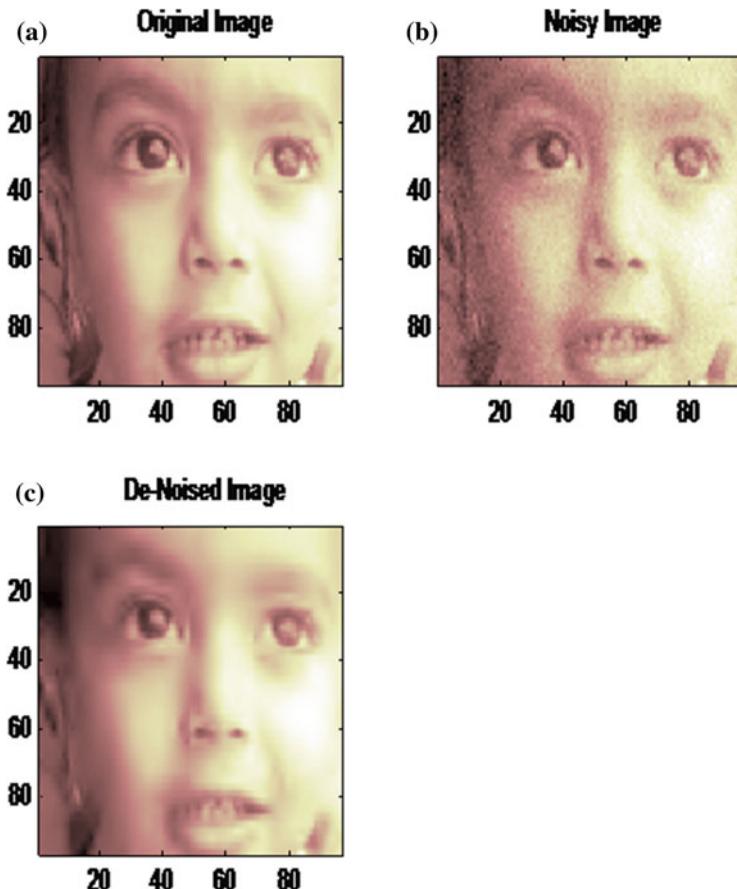


Fig. 6.1 (a) Original image (b) Noisy image (c) De-noised image

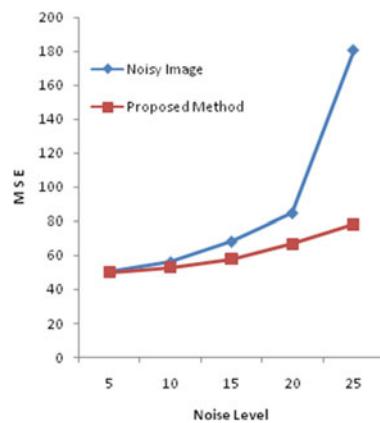
Simulation and Results

For the simulation of the proposed method, an original image of size 96×96 has been loaded, shown in Fig. 6.1a. Original image is now contaminated by additive white Gaussian noise (AWGN) with zero mean and variance σ^2 . This noisy image, shown in Fig. 6.1b, is used as the test image for the simulation of proposed method. Noise is now reduced using proposed method which is shown in Fig. 6.1c. From the figure, it is clear that the noise has been greatly reduced. Daubechies (db4) wavelet is used to decompose the signal up to third level.

To evaluate the performance of the proposed method MSE is computed for different noise levels. The results obtained, from the simulation, are given in Table 6.1.

Table 6.1 Computation of MSE

S.No.	Mean square error			
	Noise level	Noisy image	Donoho's method	Proposed method
1	5	50.2475	49.5858	50.0311
2	10	55.9363	53.9819	50.7093
3	15	68.0344	59.6310	56.0102
4	20	84.4674	68.8841	64.2903
5	25	180.4070	81.0344	75.3636

Fig. 6.2 Performance of proposed method

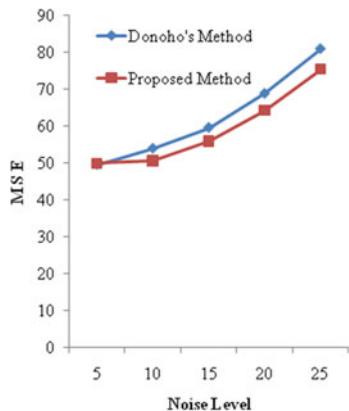
For the graphical representation, a plot of mean square error versus noise level is shown in Fig. 6.2. From the plot, it is observed that MSE of denoised image has drastically reduced as compared to the MSE of noisy image.

For the comparison purpose, the result obtained from the proposed method is compared with result obtained using Donoho's thresholding [97]. Here the wavelet transform used in both cases is db4. A comparative plot is shown in Fig. 6.3. From the plot, it is observed that the proposed method gives better result as compared to Donoho's thresholding.

Conclusions

In the present work, an image denoising method based on wavelet decomposition of the signal and Wiener filter as post filtering is proposed. An optimum threshold value is estimated by computing the minimum error between detailed coefficients of noisy image and the original noise-free image. A very popular wavelet transform proposed by Daubechies is used. To evaluate the performance of the proposed method mean square error is computed for different noise levels. The result obtained is compared with Donoho's thresholding method, and it is found that the proposed method gives better result.

Fig. 6.3 Comparison of proposed method



6.3 Applications of Wavelet Packets in Image Denoising

Wavelets with vanishing moments are very effective for representing piecewise smooth images. However, wavelets are ill-suited to represent long oscillatory patterns in images with abundant textures. These oscillatory variations of intensity can only be represented by small-scale wavelet coefficients. As a result, those small-scale coefficients are quantized to zero in the low bit rate image compression and are thresholded or shrunk to zero in image denoising, which degrades compression and denoising performance. Overcomplete wavelet packet [79] contains a mass of libraries of waveforms, from which the best wavelet packet base can be found to efficiently represent long oscillatory patterns. For image compression, the best wavelet packet base can be obtained by pruning the full wavelet packet decomposition of the image in terms of predefined cost function. Various cost functions have been developed, including the vector entropy [79], the cost functions involved in the rate-distortion and quantizers [170, 198, 227], and the cost functions dependent on the coding strategy [226]. In the aforementioned methods, the filter bank at each node of the wavelet packet tree is fixed and what is adaptively selected is the tree structure of wavelet packet decompositions. It has been shown that the best wavelet packet bases obtain better performance than the standard wavelet bases. Additionally, adaptive subband decomposition [122] and nonlinear wavelet transforms [66] are also used to improve compression performance in a different adaptive mode and what are adaptively selected are the filter's coefficients at the nodes of wavelet decompositions.

In this section, a novel wavelet packet denoising method has been proposed for image denoising. It can effectively reduce the noise from the noisy image that is contaminated by additive white Gaussian noise. To evaluate the denoising performance, we compare our method with existing method [20]. The results obtained show improved performances over existing method.

Proposed Methodology

Let the contaminated image signal I be as follows:

$$I(r, c) = I_0(r, c) + I_n(r, c)$$

where $I_0(r, c)$ is the original image and $I_n(r, c)$ is the noise with following two conditions:

- (i) The energy of $I_0(r, c)$ is captured, to a high percentage, by transform values whose magnitudes are greater than a threshold $T_s > 0$.
- (ii) The noise signal's transform values all have magnitudes which lie below a noise threshold $T_n < T_s$.

Then, the noise in $I(r, c)$ has been removed by wavelet packet thresholding. All the values of its transform whose magnitudes lie below the noise threshold T_n are set to zero and an inverse transform, providing a good approximation of $I(r, c)$ [267].

Implementation

The proposed method can be implemented by the following steps:

- Step 1: An image of size 96×96 is loaded.
- Step 2: This image is taken as an original signal $I_0(r, c)$.
- Step 3: It is now contaminated by white Gaussian noise $I_n(r, c)$ to get noisy image $I(r, c)$.
- Step 4: As we have already described that wavelet packet is a generalization of wavelets, so fourth-order Daubechies wavelet is used as mother wavelet, i.e., db4.
- Step 5: The noisy image signal is treated with wavelet packet transform along with proper threshold value. To choose optimal threshold value we have considered a basic assumption. Since noise in image signal is additive white Gaussian noise, it has a variance with invariant of time. Also, most of the noise variance lies on detailed coefficients of first level of decomposition. So, the fixed global threshold is employed when noisy signal is analyzed in each scale.

Threshold value is estimated with the coefficients as

$$\text{thr} = \sigma \sqrt{2 \log(N)/N}$$

where the variance σ can be estimated using median estimator as

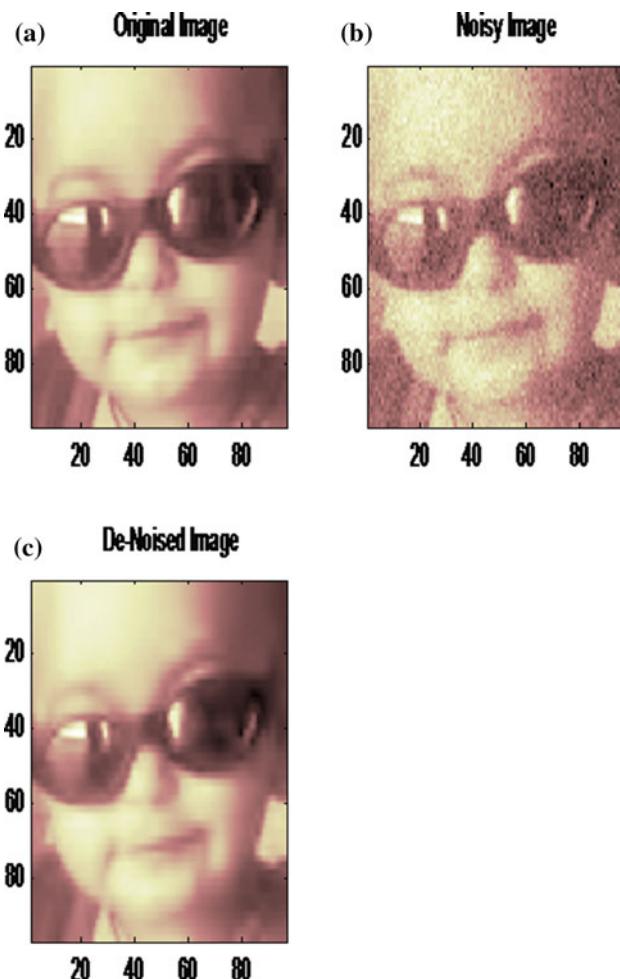
$$\sigma = \frac{\text{median}(c_i)}{0.6745}$$

where c_i are the high-frequency wavelet coefficients. The factor 0.6745 rescales the numerator so that σ is a suitable estimator for the standard deviation of Gaussian white noise. A soft thresholding is used to shrinkage the detailed coefficients of the noisy image signal.

Step 6: Results obtained by this method are tabulated in Table 6.2.

Table 6.2 Computation of MSE

SNR (in dB)	Mean square error	
	Existing method	Wavelet packet transform
-5	30.9011	29.3223
0	41.5663	35.0091
5	51.5735	41.9211
10	59.6338	48.3221
15	69.6111	54.9320
20	90.3465	65.1133
25	180.4070	78.1897

**Fig. 6.4** (a) Original image (b) Noisy image (c) De-noised image

Step 7: The denoised image $I(r, c)$ is reconstructed by applying inverse wavelet packet transform by keeping all approximated coefficients and all thresholded detailed coefficients.

Step 8: To evaluate the performance of the proposed method the MSE values are found using the formula

$$MSE = \frac{1}{MN} \sum_{r=0}^{M-1} \sum_{c=0}^{N-1} [I(r, c) - I_0(r, c)]^2$$

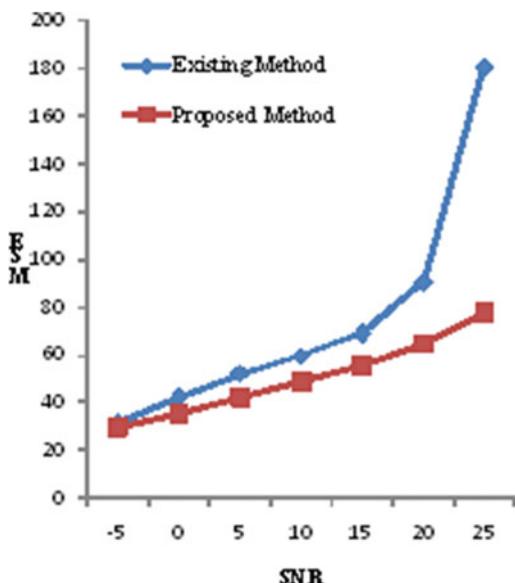
where $M = N = 0, 1, \dots, 95$.

Simulation and Results

For the simulation of the proposed method, an original image of size 96×96 has been loaded, shown in Fig. 6.4a. Original image is now contaminated by additive white Gaussian noise (AWGN) with zero mean and variance σ^2 . This noisy image, shown in Fig. 6.4b, is used as the test image for the simulation of proposed method. Noise is now reduced using proposed method which is shown in Fig. 6.4c. From the figure, it is clear that the noise has been greatly reduced. Daubechies (db4) wavelet is used to decompose the signal up to third level.

To evaluate the performance of the proposed method MSE is computed for various values of SNR, i.e., $-5, 0, 5, 10, 15, 20, 25$ dB. The results obtained, from the simulation, are given in Table 6.2. These noisy signals are then filtered using the method described in [20] and the proposed method. The comparison between exist-

Fig. 6.5 Existing method versus proposed method



ing method and proposed method is shown in Fig. 6.5. The proposed technique is simulated using MATLAB 7.0.

Conclusions

In the present work, an image denoising method based on wavelet packets decomposition is proposed. To choose optimal threshold value a basic assumption is considered. Since noise in the image signal is additive white Gaussian noise it has a constant power spectral density (PSD) for all frequencies. Hence, the fixed global threshold is employed when noisy image is analyzed in each scale. A very popular wavelet proposed by Daubechies (db4) is used as mother wavelet. To evaluate the performance of the proposed method mean square error is computed for different values of SNR. The result obtained is compared with Anjum Khan's method [20], and it is found that the proposed method gives better result.

References

1. I. Abdallah, S. Montrsor, M. Baudry, *Speech signal detection in noisy environment using a local entropic criterion*, in *Eurospeech* (Rhodes, Greece, 1997)
2. Abdullah, Characterizations of dimension functions of wavelet packets. *Jordan J. Math. Stat.* **5**(3), 151–167 (2012)
3. Abdullah, Tight wave packet frames for $L^2(\mathbb{R})$ and $\mathcal{H}^2(\mathbb{R})$. *Arab J. Math. Sci.* **19**(2), 151–158 (2013)
4. Abdullah, S. Ali, Non-existence of regular wavelet packets in Hardy space $\mathcal{H}^2(\mathbb{R})$. *Int. J. Mod. Math. Sci.* **8**(1), 36–47 (2013)
5. Abdullah, Vector-valued multiresolution analysis on local fields. *Analysis* **34**(4), 415–428 (2014)
6. F.A. Abdullah, Shah, Wave packet frames on local fields of positive characteristic. *Appl. Math. Comput.* **249**, 133–141 (2014)
7. R.A. Adams, *Sobolev Spaces* (Academic Press, New York, 1975)
8. V.X. Afonso et al., Comparing stress ECG enhancement algorithms. *IEEE Eng. Med. Biol. Mag.* **15**, 37–44 (1996)
9. Z.A. Jaffery, Afroz, K. Ahmad, ECG baseline drift elimination using wavelet packet transform. *Australian J. Basic. Appl. Sci.* **8**(3), 459–466 (2014)
10. K. Ahmad, R. Kumar, Band-limited wavelet packets, in *Mathematics and its Applications in Industry and Business*, ed. by A.H. Siddiqi, K. Ahmad (Narosa Publishing House, New Delhi, 2000), pp. 1–16
11. K. Ahmad, R. Kumar, Pointwise convergence of wavelet packet series. *Atti Sem. Fis. Univ. Modena* **48**, 107–120 (2000)
12. K. Ahmad, R. Kumar, L. Debnath, On Fourier transforms of wavelet packets. *Zeit. Anal. Anwend.* **20**, 579–588 (2001)
13. K. Ahmad, R. Kumar, L. Debnath, Existence of unconditional wavelet packet bases for the spaces $L^p(\mathbb{R})$ and $\mathcal{H}^1(\mathbb{R})$, *Taiwanese J. Math.* **10**, 851–863 (2006)
14. K. Ahmad, R. Kumar, A. Kumar, Convolution bounds and convergence of wavelet packet series, in *Recent Trends in Mathematics and its Applications*, ed. by M.R. Khan, et al. (World Education Publishers, New Delhi, 2012), pp. 55–69
15. K. Ahmad, Abdullah, Certain results on wavelet packets. *J. Anal. Appl.* **4**, 179–199 (2006)
16. K. Ahmad, F.A. Shah Abdullah, Characterization of Sobolev Spaces $L^{p,s}(\mathbb{R})$ using wavelet packets. *The Aligarh Bull. Maths.* **27**(1), 23–36 (2008)
17. K. Ahmad, F.A. Shah, Abdullah, Behavior of shrinkage operators on wavelet packet series. *Southeast Asian Bull. Math.* **35**, 1–12 (2011)

18. K. Ahmad, F.A. Shah, *Introduction to Wavelet Analysis with Applications* (Real World Education Publishers, New Delhi, 2013)
19. J.B. Allen, L.R. Rabinar, A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE* **65**(11), 1558–1564 (1977)
20. A. Khan, B. Ismail, Image denoising with a new threshold value using wavelets. *J. Data Sci.* **10**, 259–270 (2012)
21. M. Antonini, M. Barlaud, M. Mathieu, I. Daubechies, Image coding using wavelet transform. *IEEE Trans. Image Process.* **2**, 205–220 (1992)
22. I. Antunes, P.M.S. Burt, Speech denoising by soft thresholding. *IEEE Int. Sympos. Ind. Electron.* **1**, 532–536 (2006)
23. J.P. Areanas, Combining adaptive filtering and wavelet techniques for vibration signal enhancement. *Acustica, Paper ID- 99*, 1–8 (2004)
24. S.P. Arunachalam, L.F. Brown, Real-time estimation of the ECG derived respiration signal using a new algorithm for baseline wander noise removal, in *IEEE Conference of Engineering in Medicine and Biology Society* (2009), pp. 5681–5684
25. P. Auscher, Solution of two problems on wavelets. *J. Geom. Anal.* **5**, 181–236 (1995)
26. M. Bahaura, J. Rouat, Wavelet speech enhancement based on the teager energy operator. *IEEE Signal Process. Lett.* **8**(1), 10–12 (2001)
27. G. Battle, *Wavelets and Renormalization* (World Scientific Publishing Co., Singapore, 1997)
28. A. Benassi, S. Jaffard, D. Roux, Analyse multi-échelle des champs gaussiens markoviens d'ordre p indexés par (0,1). *Stoch. Process. Appl.* **47**, 275–297 (1993)
29. J.J. Benedetto, M.W. Frazier (eds.), *Wavelets: Mathematics and Applications* (CRC Press, London, 1993)
30. J.E. Berbari, Principles of electrocardiography, in *The Biomedical Engineering Handbook*, 2nd edn. ed. by J.D. Bronzino (CRC Press LLC, Boca Raton, 2000). <https://doi.org/10.1201/9781420049510.ch13>
31. J. Bercher, C. Vignat, Estimating the entropy of a signal with applications. *IEEE Trans. Signal Process.* **48**(6), 1687–1694 (2000)
32. J. Berger, R.R. Coifman, M.J. Goldberg, Removing noise from music using local trigonometric bases and wavelet packets. *J. Audio Eng. Soc.* **42**(10), 808–818 (1994)
33. E. Bernstein, W. Evans, Wavelet based noise reduction for speech recognition, in *Proceedings of the workshop on Robust Speech Recognition for Unknown Communication Channels* (1997), pp. 111–114
34. M. Berouti, R. Schwartz, J. Makhoul, Enhancement of speech corrupted by acoustic noise, in *Proceedings of the IEEE ICASSP* (Washington DC, 1979), pp. 208–211
35. N. Bershad, O. Macchi, Comparison of RLS and LMS algorithms for tracking a chirped signal, in *Proceedings of the International Conference on Acoustic Speech and Signal Processing ICASSP-89* (1989), pp. 896–899
36. L. Birgé, P. Massart, *From Model Selection to Adaptive Estimation*, ed. by D. Pollard, Festchrift for L. Le Cam (Springer, Berlin, 1997), pp. 55–88
37. M. Blanco-Velasco, B. Weng, K.E. Barner, A new ECG enhancement algorithm for stress ECG tests, in *Proceeding of the Computer in Cardiology Conference* (2006), pp. 17–20
38. S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process. ASSP* **27**(2), 113–120 (1979)
39. A. Bonami, F. Soria, G. Weiss, Band-limited wavelets. *J. Geomet. Anal.* **3**(6), 544–578 (1993)
40. M. Bownik, Z. Rzeszotnik, D. Speegle, A characterization of dimension functions of wavelets. *Appl. Comput. Harmon. Anal.* **10**, 71–92 (2001)
41. R. Bracewell, *The Fourier Transform and its Applications*, 2nd edn. (McGraw-Hill, New York, 1986)
42. D.J. Bronzino, *The Biomedical Engineering Handbook*, 2nd edn. (CRC Press, London, 1999)
43. F.C.A. Brooks, L. Hanzo, A multiband excited waveform-interpolated 2.35-kbps speech code for band limited channels. *IEEE Trans. Vehicul. Technol.* **49**(3), 766–777 (2000)
44. A. Bruce, H.-Y. Gao, *Applied Wavelet Analysis with S-PLUS* (Springer, New York, 1996)
45. A. Bruce, D. Donoho, H.-Y. Gao, Wavelet analysis. *IEEE Spectr.* 26–35 (1996)

46. C.S. Burrus, R.A. Gopinath, G. Haitao, *Introduction to Wavelets and Wavelet Transforms: A Primer* (Prentice Hall, Upper Saddle River, 1998)
47. C.S. Burrus, R.A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms* (Prentice-Hall, Upper Saddle River, 1998), p. 07458
48. B. Cammero, A. Drygajlo, Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms. *IEEE Trans. Signal Process.* **47**(6), 1622–1635 (1999)
49. L. Carleson, On convergence and growth of partial sums of Fourier series. *Acta Math.* **116**, 135–195 (1966)
50. S. Carlsson, Sketch based coding of grey level images. *Signal Process.* **15**(1), 57–83 (1988)
51. J.M. Carnicer, W. Dahmen, J.M. Peña, *Local decomposition of refinable spaces* (Technical report, Institut für Geometric und angewandte Mathematik, RWTH Aachen, 1994)
52. S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.* **9**(9), 1135–1151 (2000)
53. S. Chang, Y. Know, S. Yang, I. Kim, Speech enhancement for non-stationary noise environment by adaptive wavelet packet. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* **1**, 561–564 (2002)
54. S.H. Chen, J.F. Wang, Speech enhancement using perceptual wavelet packet decomposition and teager energy operator. *J. VLSI Signal Processing Syst.* **36**(2), 125–139 (2004)
55. M. Chendeb, K. Mohamad, D. Jacques, Methodology of wavelet packet selection for event detection. *Signal Process. Arch.* **86**(12), 3826–3841 (2006)
56. D. Cho, T.D. Bui, Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Process. Image Commun.* **20**(1), 77–89 (2005)
57. G.F. Choeiter, J.R. Glass, A wavelet and filter bank framework for phonetic classification. *Proceedings of the International Conference of Acoustic, Speech, and Signal Processing* **1**, 933–936 (2005)
58. S.A. Chouakri, F. Berekci-Reguig, Wavelet denoising of the electrocardiogram signal based on the corrupted noise estimation. *Comput. Cardiol.* **32**, 1021–1024 (2005)
59. J. Chuan, X. Bo, An improved entropy-based endpoint detection algorithm, in *ICSLP'02* (Beijing, 2002)
60. C.K. Chui, *An Introduction to Wavelets* (Academic Press, New York, 1992)
61. C.K. Chui (ed.), *Wavelet: A Tutorial in Theory and Applications* (Academic Press, New York, 1992)
62. C.K. Chui, On cardinal spline-wavelets, in *Wavelets and their Applications*, ed. by M.B. Ruskai, et al. (Jones and Bartlett Publishers, Boston, 1992), pp. 419–438
63. C.K. Chui, E. Quak, Wavelets on a bounded interval, in *Numerical Methods of Approximation Theory*, vol. 9, ed. by D. Braess, L.L. Schumaker (Basel, Birkhäuser, 1992), pp. 53–75
64. C.K. Chui, C. Li, Non-orthogonal wavelet packets. *SIAM J. Math. Anal.* **24**, 712–738 (1993)
65. C.K. Chui, L. Montefusco, L. Puccio (eds.), *Wavelets: Theory, Algorithms and Applications* (Academic Press, New York, 1994)
66. R.L. Claypoole, G.M. Davis, R.G. Barabik, Nonlinear wavelet transforms for image coding via lifting. *IEEE Trans. Image Process.* **12**(12), 1449–1459 (2003)
67. G.D. Clifford, F. Azuaje, P. McCrory, *Advanced Methods and Tools for ECG Data Analysis* (MA Norwood, Artech House, 2006)
68. L. Cohen, Time-frequency distributions- a review. *Proc. IEEE* **77**(7), 941–981 (1989)
69. A. Cohen, Wavelets and digital signal processing, in *Wavelets and their Applications*, ed. by M.B. Ruskai, et al. (Jones and Bratlett Publishers, Boston, 1992), pp. 105–121
70. A. Cohen, I. Daubechies, Orthonormal bases of compactly supported wavelets III. Better frequency resolution. *SIAM J. Math. Anal.* **24**(2), 520–527 (1993)
71. A. Cohen, I. Daubechies, On the instability of arbitrary biorthogonal wavelet packets. *SIAM J. Math. Anal.* **24**, 1340–1354 (1993)
72. R.R. Coifman, A real variable characterization of H^p . *Studia. Math.* **51**, 269–274 (1974)
73. R.R. Coifman, Y. Meyer, E.M. Stein, Some new function spaces and their applications to harmonic analysis. *J. Funct. Anal.* **62**, 304–335 (1985)

74. R.R. Coifman, Y. Meyer, S. Quake, M.V. Wickerhauser, Signal processing and compression with wavelet packets, Technical report, Yale University (1990)
75. R.R. Coifman, Y. Meyer, *Orthonormal Wavelet Packet Basis, Preprint* (Yale University, New Haven, CT, 1990)
76. R.R. Coifman, Y. Meyer, Remarques sur l'analyse de Fourier à fenêtre. C.R. Acad. Sci. Paris, **312**, 259–261 (1991)
77. R.R. Coifman, Y. Meyer, M.V. Wickerhauser, Wavelet analysis and signal processing, in *Wavelets and their Applications*, ed. by M.B. Ruskai, et al. (Jones and Bartlett Publishers, Boston, 1992), pp. 153–178
78. R.R. Coifman, Y. Meyer, M.V. Wickerhauser, Size properties of wavelet packets, in *Wavelets and their Applications*, ed. by M.B. Ruskai, et al. (Jones and Bartlett Publishers, Boston, 1992), pp. 453–470
79. R.R. Coifman, Y. Meyer, M.V. Wickerhauser, Entropy-based algorithms for best basis selection. IEEE Trans. Inform. Theory **38**(2), 713–718 (1992)
80. R.R. Coifman, M.V. Wickerhauser, Wavelets and adapted waveform analysis, in *Proceedings of Symposia in Applied Mathematics*, vol. 47 (Providence, Rhode Island, 1993), pp. 119–153
81. L. Couvreur, C. Couvreur, Wavelet-based method for nonparametric estimation of HMMs. IEEE Signal Process. Lett. **7**(2), 25–27 (2000)
82. R.M. Dansereau, W. Kinsnea, V. Cleverh, Wavelet packet best basis search using generalized Renyi entropy, in *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering* (2002), pp. 1005–1008
83. K. Daqrouq, ECG baseline wandering reduction using discrete wavelet transform. Asian J. Inf. Technol. **4**(11), 989–995 (2005)
84. I. Daubechies, A. Grossmann, Y. Meyer, Painless non-orthogonal expansions. J. Math. Phys. **27**(5), 1271–1283 (1986)
85. I. Daubechies, Orthonormal basis of compactly supported wavelets. Commun. Pure Appl. Math. **41**, 909–996 (1988)
86. I. Daubechies, The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf. Theory **36**, 961–1005 (1990)
87. I. Daubechies, *Ten Lectures on Wavelets, CBS-NSF Regional Conferences in Applied Mathematics*, vol. 61 (SIAM, Philadelphia, PA, 1992)
88. I. Daubechies, Orthonormal basis of compactly supported wavelets II, Variations on a theme. SIAM J. Math. Anal. **24**(2), 499–519 (1993)
89. I. Daubechies, Where do wavelets come from? Proc. IEEE **84**(4), 510–513 (1996)
90. R. DeVore, B. Jawerth, B. Lucier, Image compression through wavelet transform coding. IEEE Trans. Inf. Theory **38**, 719–747 (1992)
91. R. DeVore, B. Lucier, Wavelets. Acta Numerica **1**, 1–61 (1992)
92. R. DeVore, B. Jawerth, V. Popov, Compression of wavelet decompositions. Am. J. Math. **114**, 737–785 (1992)
93. R. DeVore, S. Konjagin, V. Temlyakov, Hyperbolic wavelets approximation. Constr. Approx. **14**, 126 (1998)
94. R. DeVore, G. Kyriazis, D. Leviatan, V.M. Tikhovmirov, Wavelet compression and nonlinear n -widths. Adv. Comput. Math. **1**, 197–214 (1993)
95. Z.G. Ding, F.Q. Yu, Improved speech denoising algorithm based on discrete fractional Fourier transform. J. Electron. Sci. Technol. (China) **6**(1), 29–31 (2008)
96. D.L. Donoho, *Interpolating wavelet transforms* (Department of Statistics, Stanford University, Preprint, 1992)
97. D.L. Donoho, De-noising by soft-thresholding. IEEE Trans. Inf. Theory **41**(3), 613–627 (1995)
98. D.L. Donoho, I.M. Johnstone, *Threshold selection for wavelet shrinkage of noisy data, in Proceedings 16th Annual Conference of the IEEE Engineering in Medicine and Biology Society* (Maryland, USA, 1994), pp. 24a–25a
99. D.L. Donoho, I.M. Johnstone, Ideal spatial adaption by wavelet shrinkage. Biometrika **81**, 425–455 (1994)

100. D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**, 1200–1224 (1995)
101. D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage. *Ann. Stat.* **26**(3), 879–921 (1998)
102. K. Dubowik, *Automated arrhythmia analysis - An expert system for an intensive care unit, Master's Thesis* (Lund University, Lund, Dept. of Information Technology, 1999)
103. R.J. Duffin, A.C. Schaeffer, A class of non-harmonic Fourier series. *Trans. Am. Math. Soc.* **72**, 341–366 (1952)
104. G. Evangelista, Pitch-synchronous wavelet representations of speech and music signals. *IEEE Trans. Signal Process.* **41**(12), 3313–3330 (1993)
105. G. Evangelista, The coding gain of multiplexed wavelet transforms. *IEEE Trans. Signal Process.* **44**(7), 1681–1692 (1996)
106. X. Fang, X. Wang, Construction of minimally supported frequency wavelets. *J. Fourier Anal. Appl.* **2**(4), 315–327 (1996)
107. O. Farooq, S. Datta, Mel filter-like admissible wavelet packet structure for speech recognition. *IEEE Signal Process. Lett.* **8**(7), 196–198 (2001)
108. O. Farooq, S. Datta, Wavelet-based denoising for robust feature extraction for speech recognition. *Electron. Lett.* **39**(1), 163–165 (2003)
109. C. Fefferman, E.M. Stein, Some maximal inequalities. *Am. J. Math.* **93**, 137–193 (1971)
110. C. Fefferman, E.M. Stein, H^p spaces of several variables. *Acta Math.* **129**, 137–193 (1972)
111. P. Flandrin, Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inf. Theory - Part 2*, **38**(2), 910–917 (1992)
112. M.V. Frazier, B. Jawerth, Decompositions of Besov spaces. *Indiana Univ. Math. J.* **34**, 777–799 (1985)
113. M.V. Frazier, B. Jawerth, A discrete transform and decomposition of distribution spaces. *J. Funct. Anal.* **93**, 34–170 (1990)
114. M.V. Frazier, B. Jawerth, Applications of the φ and wavelet transforms to the theory of function spaces, in *Wavelets and their Applications*, ed. by M.B. Ruskai, et al. (Jones and Bartlett Publishers, Boston, 1992), pp. 377–417
115. G.A. Friesen, T.C. Jannet, M.A. Jadallah, A comparison of the noise sensitivity of nine QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **37**, 85–98 (1990)
116. F. Gabbanini, M. Vannucci, Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi dome. *J. Comput. Graph. Stat.* (2004), pp. 187–190
117. D. Gabor, Theory of communication. *J. IEE (London)* **93**, 429–457 (1946)
118. W. Gao, H. Li, Z. Zhuang, T. Wang, De-noising of ECG signal based on stationary wavelet transform. *Acta Electronica Sinica* **32**, 238–240 (2003)
119. T.K. Garg, Abdullah, K. Ahmad, Characterization of Lebesgue spaces $L^p(\mathbb{R})$ using wavelet packets in *Modern Mathematical Models, Methods and Algorithms for Real World System*, ed. by A.H. Siddiqi, I. Duff, O. Christensen, (Anamaya Publishers, New Delhi, 2006 and Anshan Ltd. U.K., 2006), pp. 302–317
120. T.K. Garg, Abdullah, K. Ahmad, Characterization of Hardy space $\mathcal{H}^1(\mathbb{R})$ using wavelet packets, in *Modern Methods in Analysis and its Applications*, ed. by Mursaleen, (Anamaya Publishers, New Delhi, 2011), pp. 274–288
121. J. Garcia-Cuerva, J.L. Rubio de Francia, *Weighted Norm Inequalities and Related Topics* (Elsevier Science Publishers, North Holland, 1985)
122. O.N. Gerek, A.E. Cetin, Adaptive polyphase subband decomposition structures for image compression. *IEEE Trans. Image Process.* **9**(10), 1649–1660 (2000)
123. Y. Ghanbari, M.R. Karami, Spectral subtraction in the wavelet domain for speech enhancement. *IJSIT* **1**, 26–30 (2004)
124. Y. Ghanbari, M.R. Karami, A new approach for speech enhancement based on adaptive thresholding of wavelet packets. *Speech Commun.* **48**, 927–940 (2006)
125. Y. Ghanbari, M.R. Karami, B. Amelifard, Improved multi-band spectral subtraction method for speech enhancement, in *Proceedings of the 6th ISTED International Conference Signal and Image Processing* (Honolulu, Hawaii, USA, 2004), pp. 225–230

126. J.N. Gowdy, Z. Tufekci, Mel-scaled discrete wavelet coefficients for speech recognition. Proceedings of the International Conference on Acoustics, Speech and Signal Processing **3**, 1351–1354 (2004)
127. A. Graps, Introduction to wavelets. IEEE Comput. Sci. Eng. **2**(2), 50–61 (1995), <http://www.cis.udel.edu/~amer/CISC651/IEEEwavelet.pdf>
128. G. Gripenberg, Unconditional bases of wavelets for Sobolev spaces. SIAM J. Math. Anal. **24**(2), 1030–1042 (1993)
129. G. Gripenberg, Wavelet bases in $L^p(\mathbb{R})$. Studia Math. **106**(2), 175–187 (1993)
130. G. Gripenberg, A necessary and sufficient condition for the existence of a father wavelet. Studia Math. **114**, 207–226 (1995)
131. A. Grossmann, J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Anal. **15**(4), 723–736 (1984)
132. T. Gulzow, A. Engelsberg, U. Heute, Comparison of a discrete wavelet transformation and nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement. Signal Process **64**, 5–19 (1998)
133. M. Gupta, A. Gilbert, Robust speech recognition using wavelet coefficient features, in *Proceedings of the International Workshop on Automatic Speech Recognition and Understanding* (2001)
134. P.S. Hamilton, A comparison of adaptive and non-adaptive filters for reduction of power line interference in the ECG. IEEE Trans. Biomed. Eng. **43**, 105–109 (1996)
135. D.J. Hamilton, D.C. Thomson, W.A. Sandham, ANN compression of morphologically similar ECG complexes. Med. Biol. Eng. Comput. **33**(6), 841–843 (1995)
136. G.H. Hardy, J.E. Littlewood, A maximal theorem with function theoretic applications. Acta Math. **54**, 81–116 (1930)
137. S. Hargittai, Efficient and fast ECG baseline wander reduction without distortion of important clinical information, in *IEEE Conferences on Computers in Cardiology* (2008), pp. 841–844
138. S. Haykin, *Adaptive Filter Theory*, 3rd edn. (Prentice Hall Inc, Upper Saddle River, 1996)
139. J. Heart, Common electrocardiographic artifacts mimicking arrhythmias in ambulatory monitoring. Am. Heart J. **144**(2), 187–197 (2002)
140. E. Hernández, G. Weiss, *A First Course on Wavelets* (CRC Press, New York, 1996)
141. E. Hernández, X. Wang, G. Weiss, Smoothing minimally supported frequency wavelets: part I. J. Fourier Anal. Appl. **2**(4), 329–340 (1996)
142. E. Hernández, X. Wang, G. Weiss, Smoothing minimally supported frequency (MSF) wavelets: part II. J. Fourier Anal. Appl. **3**, 23–41 (1997)
143. F. Hlawatsch, G.F. Boudreaux-Bartles, Linear and quadratic time-frequency signal representations. IEEE Signal Proc. Mag. **9**(2), 21–67 (1992)
144. Y.H. Hu, S. Palreddy, W.J. Tompkins, A patient-adaptable ECG beat classifier using a mixture of experts approach. IEEE Trans. Biomed. Eng. **44**, 891–900 (1997)
145. Y.H. Hu, W.J. Tompkins, J.L. Urrusti, V.X. Afonso, Applications of artificial neural networks for ECG signal detection and classification. J. Electrocardiol. **26**, 66–73 (1993)
146. H. Inoue, A. Miyazaki, A noise reduction method for ECG signals using the dyadic wavelet transform. IEICE Trans. Fund. **6**, 1001–1007 (1998)
147. E. Jafer, A.E. Mahdi, Wavelet-based perceptual speech enhancement using adaptive threshold estimation, in *Proceedings of Eurospeech* (2003), pp. 569–572
148. Z.A. Jaffery, P. Sharma, K. Ahmad, Noise reduction in speech signal using thresholding in wavelet transform domain. Int. J. Appl. Eng. Res. **14**, 2503–2509 (2010)
149. Z. A. Jaffery, P. Sharma and K. Ahmad, Estimation of speech signal in the presence of white noise using wavelet transform, in *Proceedings of the International Conference on control, Communication and Power Engineering, 2010 (ACEEE)* (2010), pp. 320–322
150. Z.A. Jaffery, Afroz, K. Ahmad, Performance comparison of wavelet threshold estimators for ECG signal denoising in *International Conference on Advances in Recent Technologies in Communication and Computing, 2010* (IEEE, 2010), pp. 248–251
151. Z.A. Jaffery, P. Sharma, K. Ahmad, Applications of wavelet packets in speech denoising. JMI Int. J. Math. Sci. **1**(2), 44–53 (2010)

152. Z.A. Jaffery, K. Ahmad Afroz, Denoising of ECG signal using wavelet transform. *JMI Int. J. Math. Sci.* **2**(1), 26–33 (2011)
153. Z.A. Jaffery, P. Sharma, K. Ahmad, Selection of optimal decomposition level based on entropy for speech denoising using wavelet packets. *J. Bioinf. Intell. Control (U.S.A.)* **1**(2), 196–202 (2013)
154. A.M. Jarrah, R. Kumar, K. Ahmad, Certain characterizations of wavelet packets. *Atti Sem. Fis. Univ. Modena* **53**, 147–164 (2005)
155. R. Javaid, R. Besar, F.S. Abas, Performance evaluation of percent root mean square difference for ECG signals compression. *Signal Process. Int. J.* **2**(2), 1–9 (2006)
156. B. Jawerth, W. Sweldens, An overview of wavelet based multiresolution analysis. *SIAM Rev.* **36**(3), 377–412 (1994)
157. B. Jawerth, W. Sweldens, Bi-orthogonal smooth local trigonometric bases. *J. Fourier Anal. Appl.* **2**(2), 109–133 (1995)
158. I.M. Johnstone, B.W. Silverman, Wavelet threshold estimator for data with correlated noise. *J. R. Stat. Soc. B.* **59**, 319–351 (1997)
159. S. Kadambé, G.F. Boudreault-Bartels, Application of the wavelet transform for pitch detection of speech signals. *IEEE Trans. Inf. Theory* **38**(2), 917–924 (1992)
160. G. Kaiser, *A Friendly Guide to Wavelets* (Birkhäuser, Boston, 1994)
161. S. Kamath, P. Lizou, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, in *Proceedings of the ICASSP-2002* (Orlando, 2002)
162. M. Kania, M. Fereniec, R. Maniewski, Wavelet denoising for multi-lead high resolution ECG signals. *Meas. Sci. Rev.* **7**, 30–33 (2007)
163. P.P. Kanjilal, S. Palit, G. Saha, Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. *IEEE Trans. Biomed. Eng.* **44**, 51–59 (1997)
164. J.R. Karam, W.J. Philips, W. Robertson, M.M. Artimy, New wavelet packet model for automatic speech recognition system. *Canadian Conf. Eletr. Comput. Eng.* **1**, 511–514 (2001)
165. Y. Katznelson, *An Introduction to Harmonic Analysis* (Wiley, New York, 1968)
166. F. Keinert, *Wavelets and Multiwavelets* (Chapman & Hall, CRC, London, 2004)
167. S.E. Kelly, M.A. Kon, L.A. Raphael, Pointwise convergence of wavelet expansions. *Bull. Am. Math. Soc.* **30**, 87–94 (1994)
168. S.E. Kelly, M.A. Kon, L.A. Raphael, Local convergence for wavelet expansions. *J. Funct. Anal.* **126**, 102–138 (1994)
169. K. Khan, P. Sharma, K. Ahmad, Application of wavelets in image denoising. *J. Bioinf. Intell. Control (U.S.A.)* **1**(2), 171–174 (2013)
170. W.H. Kim, T. Nguyen, Y.H. Hu, Adaptive wavelet packet based image coding with optimal entropy-constrained lattice vector quantizer. *IEEE Trans. Signal Process.* **64**(7), 2019–2026 (1998)
171. M. Kobayashi, M. Sakamoto, T. Saito, Y. Hashimoto, M. Nishimura, K. Suzuki, Wavelet analysis used in text-to-speech synthesis. *IEEE Trans. Circuits Syst. II Analog. Digital Signal Process.* **45**(8), 1125–1129 (1998)
172. B. Kotnik, Z. Kacic, B. Horrat, The usage of wavelet packet transformation in automatic noisy speech recognition systems. *Proceedings of the International Conference on Computer as a Tool* **2**, 131–134 (2003)
173. B. Kotnik, Robust speech parameterization based on joint wavelet packet decomposition and autoregressive modeling. Ph.D. thesis, Maribor University, Slovenia (2004)
174. R. Kronland-Martinet, J. Morlet, A. Grossmann, Analysis of sound patterns through wavelet transforms. *Int. J. Pattern Recognit. Artif. Intell.* **1**(2), 273–302 (1987)
175. A. Kufner, O. John, S. Fucik, *Function Spaces* (Academia, Publishing House of the Czechoslovak, Academy of Sciences, Prague, 1977)
176. C.T. Lu, H.C. Wang, Enhancement of single channel speech based on masking property and wavelet transform. *Speech Commun.* **41**(3), 409–427 (2003)
177. A. Lallouni, Wavelet based speech enhancement using two different threshold-based denoising algorithms, in *Proceedings of the Canadian Conference on Electrical and Computer Engineering* (2004), pp. 315–318

178. P. Lander, Time-frequency plane wiener filtering of the high resolution ECG: development and application. *IEEE Trans. Biomed. Eng.* **44**, 256–265 (1997)
179. W.M. Lawton, Necessary and sufficient conditions for constructing orthonormal wavelet bases. *J. Math. Phys.* **32**(1), 57–61 (1991)
180. K.Y. Lee, B.G. Lee, S. Ann, Adaptive filtering for speech enhancement in colored noise. *IEEE Trans. Signal Process. Lett.* **4**, 277–279 (1997)
181. P.G. Lemarié, Y. Meyer, Ondelettes et bases Hilbertiennes. *Rev. Mat. Iberoamericana* **2**, 1–18 (1986)
182. P.G. Lemarié-Rieusset, La propriété de support minimal dans les analyses multi-résolution. *C.R. Acad. Sci. Paris Sér. I Math.* **312**(11), 773–776 (1991)
183. P.G. Lemarié-Rieusset, Existence de “fonction-pere” pour les ondelettes support compact. *C.R. Acad. Sci. Paris Sér. I Math.* **314**, 17–19 (1992)
184. P.G. Lemarié-Rieusset, Sur l’existence des analyses multi-résolutions en théorie des ondelettes. *Rev. Mat. Iberoamericana* **8**, 457–474 (1992)
185. L.R. Litwin, Speech coding with wavelets. *IEEE Potentials* **17**(2), 38–41 (1998)
186. M. Lounsbury, T.D. Deroose, J. Warren, Multiresolution surfaces of arbitrary topological type, Preprint, Department of Computer Science and Engineering 93-10-05, University of Washington (1993)
187. P. Maab, H.-G. Stark, Wavelets and digital image processing. *Surv. Math. Ind.* **4**, 195–235 (1994)
188. W.R. Madych, Some elementary properties of multiresolution analysis of $L^2(\mathbb{R}^n)$, in *Wavelets: A Tutorial in Theory and Applications*, ed. by C.K. Chui (Academic Press, New York, 1992), pp. 259–294
189. S.G. Mallat, Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Am. Math. Soc.* **315**, 69–87 (1989)
190. S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern. Anal. Mach. Intell.* **11**, 674–693 (1989)
191. S.G. Mallat, W.L. Hwang, Singularity detection and processing with wavelets. *IEEE Trans. Inf. Theory* **38**(2), 617–643 (1992)
192. S.G. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, New York, 1998)
193. P. Manchanda, K. Ahmad, A.H. Siddiqi (eds.), *Current Trends in Industrial and Applied Mathematics* (Anamaya Publishers, New Delhi, 2002)
194. Y. Meyer, *Wavelets and Operators* (Cambridge University Press, Cambridge, 1992)
195. Y. Meyer, *Wavelets: Algorithms and Applications* (SIAM, Philadelphia, 1993)
196. Y. Meyer, S. Roques, *Progress in Wavelet Analysis and Applications* (Frontières, 1993)
197. C.R. Meyer, H.N. Keiser, Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques. *Comput. Biomed. Res.* **10**, 459–470 (1977)
198. F.G. Meyer, A.Z. Averbach, J.O. Stromberg, Fast adaptive wavelet packet image compression. *IEEE Trans. Image Process.* **9**(5), 792–800 (2000)
199. A. Mikhled, D. Khaled, ECG signal denoising by wavelet transform thresholding. *Am. J. Appl. Sci.* **5**(3), 276–281 (2008)
200. M. Dai, S.-L. Liana, Removal of baseline wander from dynamic electrocardiogram signals, in *IEEE Conference on Image and Signal Processing* (2009), pp. 1–4
201. A.T. Mohammadad, M. Behzad, A wavelet packets approach to electrocardiograph baseline drift cancellation. *Int. J. Biomed. Imaging*, Article ID. 97157, 1–9 (2006)
202. A. Mohammad, D. Jafari, Entropy in signal processing. *Traitement du signal*, 87–116 (1994)
203. H. Monson Hayes, *Statistical Digital Signal Processing and Modeling* (Wiley Inc, New Jersey, 2002)
204. G.B. Moody, R.G. Mark, A.L. Goldberger, Physionet: a research resource for studies of complex physiologic and biomedical signals. *Comput. Cardiol. Proc.* 179–182 (2000)
205. N. Hojung, H.-S. Kim, Kwon, S.-I. Yang, Speaker verification system using hybrid model with pitch detection by wavelets in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-scale Analysis* (1998), pp. 153–156

206. V. Namias, The fractional order Fourier transform and its application to quantum mechanics. *J. Inst. Math. Appl.* **25**(3), 241–265 (1980)
207. N.H. Nielsen, M.V. Wickerhauser, Wavelets and time-frequency analysis. *Proc. IEEE* **84**(4), 523–540 (1996)
208. R.C. Nongpiur, *Impulse Noise Removal in Speech Using Wavelets, ICASSP* (Las Vegas, USA, 2008), pp. 1593–1597
209. A.V. Oppenheim, R.W. Schafer, *Digital Signal Processing* (Prentice-Hall Inc, Englewood Cliffs, 2001)
210. J. Pan, W.J. Tompkins, A real time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **31**, 715–718 (1985)
211. S.V. Pandit, ECG baseline drift removal through STFT, in *Proceedings of the 18th International Conference of the IEEE Engineering in Medicine and Biology Society* (Amsterdam, Netherlands, 1996), pp. 1405–1406
212. J. Peetre, Sur les espaces de Besov. *C.R. Acad. Sci. Paris. Ser. A-B* **264**, 281–283 (1967)
213. J. Peetre, *New Thoughts on Besov Spaces, Duke Mathematics Series* (Durham, N.C., 1976)
214. T.V. Pham, G. Kubin, DWT-based classification of acoustic-phonetic classes and phonetic units, in *Proceedings of International Conference on Spoken Language Processing* (2004), pp. 985–988
215. T.V. Pham, G. Kubin, WPD-based noise suppression using nonlinearily weighted threshold quantile estimation and optimal wavelet shrinking, in *Proceedings of Interspeech* (2005), pp. 2089–2092
216. T.V. Pham, G. Kubin, DWT-based phonetic groups classification using neural networks. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* **1**, 401–404 (2005)
217. F. Phan, E. Micheli-Tzanakou, S. Sideman, Speaker identification using neural networks and wavelets. *IEEE Eng. Med. Biol. Mag.* **19**(1), 92–101 (2000)
218. J. Pitas, A.N. Venetsanopoulos, Order statistics in digital image processing. *Proc. IEEE* **12**, 1893–1921 (1992)
219. M.R. Portnoff, Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Trans. ASSP* **28**, 55–69 (1980)
220. L. Prasad, S.S. Iyenger, *Wavelet Analysis with Applications to Image Processing* (CRC Press, New York, 1997)
221. E. Quak, N. Weyrich, Decomposition and reconstruction algorithms for spline wavelets on a bounded interval. *Appl. Comput. Harmon. Anal.* **1**, 217–231 (1994)
222. E. Quak, N. Weyrich, Algorithms for trigonometric wavelet packets. *Appl. Comput. Harmon. Anal.* **4**, 74–96 (1997)
223. E. Quak, N. Weyrich, Algorithms for spline wavelet packets on an interval. *BIT* **37**(1), 76–95 (1997)
224. L.R. Rabiner, B. Gold, *Theory and Application of Digital Signal Processing* (Prentice-Hall Inc, Englewood Cliffs, 1995)
225. L.R. Rabiner, R.W. Schafer, *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, 1978)
226. N.M. Rajpoot, R.G. Wilson, F.G. Meyer, R.R. Coifman, Adaptive wavelet packet basis selection for zero tree image coding. *IEEE Trans. Image Process.* **12**(12), 1460–1472 (2003)
227. K. Ramchandran, M. Vetterli, Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. Image Process.* **2**(2), 160–174 (1993)
228. G.U. Reddy, M. Muralidhar, S. Varadarajan, ECG de-noising using improved thresholding based on wavelet transforms. *Int. J. Comput. Sci. Netw. Secur.* **9**, 221–225 (2009)
229. A.M. Reza, *From Fourier Transform to Wavelet Transform* (Spire Lab, UWM, 1999)
230. O. Rioul, M. Vetterli, Wavelets and signal processing. *IEEE Signal Process. Mag.* 14–38 (1991)
231. O. Rioul, M. Vetterli, Wavelets and time-frequency analysis. *Proc. IEEE* **4**, 523–540 (1996)
232. M.B. Ruskai, et al. (eds.), *Wavelets and their Applications* (Jones Bartlett Publishers, Boston, 1992)

233. H. Sadabadi, A.Ghaffari Mghasemi, A, mathematical algorithm for ECG signal denoising using window analysis. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.* **1**, 73–78 (2007)
234. R. Sarikaya, B. Pellom, J. Hasan, Wavelet packet transform features with application to speaker identification, in *Proceedings of the Nordic Signal Processing Symposium* (1998), pp. 81–84
235. R. Sarikaya, J.H.L. Hasen, High resolution speech feature parametrization for monophone-based stressed speech recognition. *IEEE Signal Process. Lett.* **7**(7), 182–185 (2000)
236. O. Sayadi, B.S. Mohammad, ECG baseline correction with adaptive bionic wavelet transform, in *IEEE International Symposium on Signal Processing and Its Application* (2007), pp. 1–4
237. J.W. Seak, K.S. Bae, Speech enhancement with reduction of noise components in the wavelet domain. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* **2**, 1323–1326 (1997)
238. E. Séré, *Thesis* (University of Paris-Dauphine, CEREMADE, Paris, France, 1992)
239. F.A. Shah, Abdullah, A characterization of tight wavelet frames on local fields of positive characteristic. *J. Contemp. Math. Anal.* **49**(6), 251–259 (2014)
240. F.A. Shah, Abdullah, Nonuniform multiresolution analysis on local fields positive characteristic. *Complex Anal. Oper. Theory* **9**(7), 1589–1608 (2015)
241. C.E. Shannon, Communication in the presence of noise. *Proc. I.R.E.* **37**, 10–21 (1949)
242. C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948)
243. J.M. Shapiro, An embedded hierachial image coder using zero trees of wavelet coefficients, in *Data Compression Conference*, ed. by J.A. Storer, M. Cohn (IEEE Computer Society Press, Los Alamitos, CA, 1993), pp. 214–223
244. P. Sharma, K. Khan, K. Ahmad, Image denoising using local contrast and adaptive mean in wavelet transform domain. *Int. J. Wavelets Multiresol. Inf. Process.* **12**(6), 1450038 (15 pages) (2014)
245. P. Sharma, K. Khan, K. Ahmad, Applications of wavelet packets in image denoising. *Australian J. Basic Appl. Sci.* **9**(31), 594–599 (2015)
246. H. Sheikhzadeh, H.R. Abutalebi, An improved wavelet based speech enhancement system, in *Proceedings of the 7th European Conference on Speech Communication and Technology (Euro Speech), Aalborg, Denmark* (2001), pp. 1855–1858
247. J.L. Shen, J.W. Hung, L.S. Lee, Robust entropy-based endpoint detection for speech recognition in noisy environments, in *Proceedings of the International Conference on Spoken Language Processing* (1998)
248. S. Kishan, *Digital Signal Processing in Telecommunications* (Prentice-Hall PTR, New Jersey, 1995)
249. S.-H. Chen, J.-F. Wang, Application of wavelet transform for C/V segmentation on Mandarin speech signals. *IEEE Proc. Vis. Image Signal Process.* **148**(2), 133–139 (2001)
250. S.-H. Chen, J.-F. Wang, A noise-robust pitch detection method using wavelet transform with aliasing compensation. *IEEE Proc. Vis. Image Signal Process.* **149**(6), 327–334 (2002)
251. J.S. Shing, J.-F. Wang, C.-H. Wu, A channel-weighting method for speech recognition using wavelet decompositions, in *IEEE Asia-Pacific Conference on Circuits and Systems* (1994), pp. 519–523
252. A.H. Siddiqi, Certain problems concerning wavelets and wavelet packets, in *Industrial and Applied Mathematics*, ed. by A.H. Siddiqi, K. Ahmad (Narosa Publishing House, New Delhi, 1998), pp. 13–42
253. A.H. Siddiqi, K. Ahmad (eds.), *Industrial and Applied Mathematics* (Narosa Publishing House, New Delhi, 1998)
254. A.H. Siddiqi, I. Duff, O. Christensen (eds.) *Modern Mathematical Models, Methods and Algorithms for Real World System* (Anamaya Publishers, New Delhi, 2006 and Anshan Ltd. U.K., 2006)
255. A.H. Siddiqi, K. Ahmad, P. Manchanda, *Introduction to Functional Analysis with Applications* (Anamaya Publishers, New Delhi, 2006 and Anshan Ltd. U.K., 2006)
256. W.M. Siebert, *Circuits, Signals and Systems* (MIT Press and McGraw-Hill, Cambridge, 1986)

257. K.P. Soman, K.I. Ramachandran, *Insight into Wavelets from Theory to Practice*, vol. 2 (Prentice Hall, India, 2005)
258. L. Sommo, Time varying filtering for removal of baseline wander in exercise ECGs, in *Computer in Cardiology* (Italy, 1991), pp. 145–148
259. J. Stegmann, G. Shroder, Robust voice-activity detection based on the wavelet transform, in *IEEE Workshop on Speech Coding for Telecommunications Proceedings* (1997), pp. 99–100
260. E.M. Stein, *Topics in Harmonic Analysis: Related to the Littlewood-Paley Theory* (Princeton University Press, Princeton, 1970)
261. E.M. Stein, *Singular Integrals and Differentiability Properties of Functions* (Princeton University Press, Princeton, 1970)
262. G. Strang, Wavelets and dilation equations: a brief introduction. *SIAM Rev.* **31**, 614–627 (1989)
263. G. Strang, Wavelet transforms versus Fourier transforms. *Bull. Am. Math. Soc.* **28**(2), 288–305 (1993)
264. R.S. Strichartz, How to make wavelets. *Am. Math. Mon.* **100**, 539–556 (1993)
265. J.O. Strömberg, A modified Franklin system and higher order spline systems on \mathbb{R}^n as unconditional bases for Hardy spaces, in *Proceedings of Conference in Harmonic Analysis in Honor of Antoni Zygmund*, ed. by W. Beckner et al. (Wadsworth II, 1983), pp. 475–493
266. S. Sudha, G.R. Suresh, R. Sukanesh, Wavelet based image denoising using adaptive subband thresholding. *Int. J. Soft Comput.* **2**(5), 628–632 (2007)
267. M.G. Sumitra, K. Thanuskodi, M.R. Anitha, Modified time adaptive wavelet based approach for enhancing speech from adverse environments ICGST. *DSP J.* **9**(1), 33–40 (2009)
268. P. Sun, Q.H. Wu, A.M. Weindling, A. Finkelstein, K. Ibrahim, An improved morphological approach to background normalisation of ECG signals. *IEEE Trans. Biomed. Eng.* **50**(1), 117–121 (2003)
269. S. Chang, Y. Kwon, S.-I. Yang, Speech feature extracted from adaptive wavelet for speech recognition. *Electron. Lett.* **34**(23), 2211–2213 (1998)
270. W. Sweldens, The lifting scheme: a custom-design construction of bi-orthogonal wavelets. *Appl. Comput. Harmon. Anal.* **3**(2), 186–200 (1996)
271. M. Taibleson, On the theory of Lipschitz spaces of distributions in Euclidean n -space I, II and III. *J. Math. Mech.* **13**, 407–480 (1964); **14**, 821–840 (1965); **15**, 973–981 (1966)
272. B. Tan, R. Lang, H. Schroder, A. Spray, P. Demody, Applying wavelet analysis to speech segmentation and classification, in *Proceedings of the SPIE Conference on Wavelet Application in Signal and Image Processing* (1994), pp. 750–761
273. T. Tao, On the almost everywhere convergence of wavelet summation methods. *Appl. Comput. Harmon. Anal.* **3**, 384–387 (1996)
274. T. Tao, B. Vidakovic, Almost everywhere behavior of general wavelet shrinkage operators. *Appl. Comput. Harmon. Anal.* **9**, 72–82 (2000)
275. T. Thom et al., Heart disease and stroke statistics update: a report from the American heart association statistics committee and stroke statistics subcommittee. *Circulation* **113**(6), 85–151 (2006)
276. C. Tie, X. Wu, Wavelet-based de-noising of speech using adaptive decomposition, in *IEEE International Conference on Industrial Technology, ICIT-08* (2008), pp. 1–5
277. P.E. Tikkainen, Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. *Biol. Cybern.* **80**, 259–267 (1999)
278. H.M. Torres, H.L. Rufiner, Automatic speaker identification by means of Mel Cepstrum, wavelets and wavelet packets, in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2 (2000), pp. 978–981
279. H. Triebel, *Interpolation Theory* (Function Spaces, Differential Operators (North Holland, 1978)
280. S. Treil, A. Volberg, Wavelets and the angle between past and future. *J. Funct. Anal.* **143**, 269–308 (1997)
281. V. Veselinovic, D. Graupe, A wavelet transform approach to blind adaptive filtering of speech from unknown noises. *IEEE Trans. Circuit Syst. II Analog. Digital Signal Process.* **50**, 150–154 (2003)

282. M. Vetterli, C. Herley, Wavelets and filter banks: theory and design. *IEEE Trans. Acoust. Speech Signal Process.* **40**(9), 2207–2232 (1992)
283. J.S. Walker, Fourier analysis and wavelet analysis. *Notices Am. Math. Soc.* **44**(6), 658–670 (1997)
284. J.S. Walker, *A Primer on Wavelets and their Scientific Applications*, 2nd edn. (Chapman and Hall/CRC, Boca Raton, 2008)
285. D.F. Walnut, *An Introduction to Wavelet Analysis* (Birkhäuser Boston, 2002)
286. G.G. Walter, *Wavelets and Other Orthogonal Systems with Applications* (CRC Press, New York, 1994)
287. X. Wang, The study of wavelets from the properties of their Fourier transforms, Ph.D. Thesis, Washington University, St. Louis, MO (1995)
288. J. Weickert, J. Fröhlich, Image processing using wavelet algorithm for non-linear diffusion, Preprint, Kaiserslautern University, Germany (1994)
289. N. Whitmal, J.C. Rutledge, J. Cohen, Reducing correlated noise in digital hearing. *IEEE Med. Biol. Mag.* **15**, 88–96 (1996)
290. M.V. Wickerhauser, *Lectures on Wavelet Packet Algorithms* (Yale University, New Haven, 1991)
291. M.V. Wickerhauser, Acoustic signal compression with wavelet packets, in *Wavelets: A Tutorial in Theory and Applications*, ed. by C.K. Chui (Academic Press, Boston, 1992), pp. 679–700
292. M.V. Wickerhauser, Best-adapted wavelet packet bases, in *Proceedings of Symposia in Applied Mathematics*, vol. 47 (A.M.S., Providence, Rhode Island, 1993), pp. 155–171
293. M.V. Wickerhauser, *Adapted Waveform Analysis from Theory to Software* (A K Peters Ltd, 1994)
294. M.V. Wickerhauser, Time localization techniques for wavelet transforms. *CCACCA* **68**(1), 1–27 (1995)
295. P. Wojtaszczyk, *A Mathematical Introduction to Wavelets* (Cambridge University Press, Cambridge, 1997)
296. www.physionet.org/cgi-bin/atm/ATM
297. W. Xiaodong, L. Yongming, C. Hongyi, Multi-domain speech compression based on wavelet packet transform. *Electron. Lett.* **34**(2), 154–155 (1998)
298. X. Yang, K. Wang, S.A. Shamma, Auditory representation of acoustic signals. *IEEE Trans. Inf. Theory* **38**(2), 824–839 (1992)
299. D. Zhang, Wavelet approach for ECG baseline wander correction and noise reduction, in *Proceedings of the IEEE on Engineering in Medicine and Biology 27th Annual Conference* (2005), pp. 1212–1215
300. X.P. Zhang, M.T. Desai, Adaptive de-noising based on SURE risk. *IEEE Signal Process. Lett.* **5**(10), 265–267 (1998)
301. Z. Zhi-Dong, C. Yu-Quan, A new method for removal of baseline wander and power line interference in ECG signals, in *IEEE Conferences on Machine Learning and Cybernetics* (2006), pp. 4342–4347
302. D. Zhu-Gao, F.-Q. Yu, Improved speech denoising algorithm based on discrete fractional Fourier transform. *J. Electron. Sci. Technol. China* **6**(1) (2008)
303. C.J. Zarowski, *Notes on Orthogonal Wavelets and Wavelet Packets* (1995). <http://www.mathsoft.com/wavelet.html>
304. M.S.A. Zilany, M.K. Hasan, M.R. Khan, Efficient hard and soft thresholding for wavelet speech enhancement, in *Proceedings of XI European Signal Processing Conference (EUSIPCO 2002)* (Toulouse, France, 2002)

Index

A

Artifacts, 175
Atomic decomposition, 137

ECG signal monitoring method, 174
Electromagnetic induction, 176
Electrostatic sources, 176

B

Band-limited wavelet packets, 53
Besov spaces, 25

F

Father wavelet, 37
Fourier transform, 1, 6
Function spaces, 23

C

Choice of compression metric, 208
Choice of wavelet, 213
Clean speech signal, 138
Coiflets wavelets, 31
Compression strategies, 209
Continuous wavelet transform, 8
Convergence of wavelet packet series, 89
Convolution bounds, 85

H

Haar
basis function, 13
scaling function, 10
wavelet function, 16
Hardy–Littlewood maximal function, 3
Hardy space, 24
Hyperbolic bases, 214

D

Daubechies wavelets, 29
Decomposition, 197
Denoising, 155
Denoising of ECG signal using wavelet transform, 181
Digital signal processing, 132
Discrete time signals, 134
Discrete wavelet transform, 10

I

Image compression algorithms, 208
Inverse Fourier transform, 1

L

Lebesgue
point, 3
spaces, 23

E

ECG signal, 170
ECG signal characteristics, 195

M

Mathematical representation of wavelet, 184
Maximal function, 3
Mother wavelet, 37, 185

- Multiresolution analysis, 27
 Multiscale edge-based image coding, 206
 Multiscale edge representation of images, 204
- N**
 Nested space, 15
 Noise originating from sources outside the patient, 176
 Nonstationary signals, 136
 Normal ECG signal, 175
- O**
 Optimal wavelet selection, 185
- P**
 P wave, 172
- Q**
 QRS complex, 172
 QT interval, 174
 Quantisation, 136
 Quantisation noise, 136
- R**
 Radio frequency interference, 177
 Reconstruction, 198
- S**
 Scaled Haar wavelet functions, 17
 Scaling function, 27
 Signal processing, 133
 to noise ratio, 138
 Smoothness of images, 211
 Sobolev spaces, 24
 Speech signal, 135
 processing, 132
 Stationary signal, 136
 ST segment, 173
 Symlets wavelets, 31
- T**
 Texture-based image encoding, 207
 Thresholding, 34, 198
 T wave, 173
- U**
 U wave, 174
- W**
 Wavelet, 26
 packet functions, 37
 packet thresholding, 149
 thresholding, 185
 transform, 8
 White Gaussian noise, 139
 Wiener filter, 142