THESIS REPORT

Ph.D.

Generalizations and Properties of the Multiscale Maxima and Zero-Crossings Representations

by Z. Berman
Advisor: J. Baras

Ph.D. 92-9

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Generalizations and Properties of the Multiscale Maxima and Zero-Crossings Representations

by

Zeev Berman

Dissertation submitted to the Faculty of The Graduate School
of The University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
August 1992

Advisory Committee:

Professor John Baras, Chairman/ Co-Advisor
Professor Carlos Berenstein, Co-Advisor
Professor P. S. Krishnaprasad
Professor John Benedetto
Associate Professor Prakash Narayan
Associate Professor Nariman Farvardin

ABSTRACT

Title of Dissertation: **Generalizations and Properties of the Multiscale Maxima and Zero-Crossings Representations**

Zeev Berman, Doctor of Philosophy, 1992

Dissertation directed by: Professor John Baras
Department of Electrical Engineering
Professor Carlos Berenstein
Department of Mathematics

The analysis of a discrete multiscale edge representation is considered. A general signal description, called an inherently bounded Adaptive Quasi Linear Representation (AQLR), motivated by two important examples, namely, the wavelet maxima representation, and the wavelet zero-crossings representation, is introduced. This thesis addresses the questions of uniqueness, stability, and reconstruction. It is shown, that the dyadic wavelet maxima (zero-crossings) representation is, in general, nonunique. Nevertheless, these representations are always stable. Using the idea of the inherently bounded AQLR, two stability results are proven. For a general perturbation, a global BIBO stability is shown. For a special case, where perturbations are limited to the continuous part of the representation, a Lipschitz condition is satisfied. Two reconstruction algorithms, based on the minimization of an appropriate cost function, are proposed. The first is based on the integration

of the gradient of the cost function; the second is a standard steepest descent algorithm. Both algorithms are shown to converge. The last part of this dissertation describes possible modifications in the basic multiscale maxima representations. The main idea is to preserve the structure of the inherently bounded AQLR, while allowing a trade-off between reconstruction quality and amount of information required for representation. In particular, it is shown how quantization can be considered as an integral part of the representation.

DEDICATION

To my loving and devoted wife Ruti,

to my great daughters Osnat, Yael, and Galia

for their love, support, and trust.

ACKNOWLEDGMENTS

First, I would like to express my gratitude to my main advisor Dr. John Baras. He proposed my to work on the exciting field of wavelets. Due to his broad view on engineering problems I leave University of Maryland with a big collection of problems to think about. Especially, I appreciate a real academic freedom which was given to me. I enjoyed it very much and it prepared me to an independent academic activity. In addition, I am very grateful for his support for my several conference travels which kept me updated in the very dynamic field of wavelets and allowed me to meet experts in the subject.

My second advisor, Dr. Carlos Berenstein gave me a great deal of support. His constant encouragement greatly helped me to keep focused on the problem. Whatever was the problem, Dr. Carlos Berenstein was always willing to advise and help.

Perhaps the most important breakthrough in this dissertation happened due to several remarks from Dr. Sthephane Mallat about the importance of the stability problem. Important help I got from Dr. Andre Tits, who suggested key ideas for the proof of Theorem 8. Encouraging remarks from Dr. Ronald

Coifman, Dr. Martin Vetterli and Dr. Oliver Rioul appeared exactly on time to help me to focus on the right problems.

In the course of this work, I have profiled greatly from discussions with many interesting people, especially Dr. Nariman Farvardin, Dr. Armand Makowski, Dr. John Benedetto, Dr. Lorenzo Finesso. I was lucky to get help from many my colleagues, graduate students, especially I am grateful to Nicholaos Sidiripoulos and Buno Pati.

All the above was impossible without a scholarship from my permanent employer, Rafael, DoD, Israel.

Last and certainly the most important was the support that I got from my family. Their constant love, continuous assistance and unlimited confidence, even in the most difficult moments, allowed me to pursue this work with pleasure. My loving wife Ruti and my great daughters Osnat, Yael, Galia have a great, perhaps invisible at first glance, part in this work.

TABLE OF CONTENTS

List of Figures viii

Notation x

1 INTRODUCTION AND BACKGROUND **1**

1.1 Introduction 1

1.2 Previous works 5

1.3 Wavelet transform 7

1.3.1 Continuous Wavelet Transform 8

1.3.2 Dyadic continuous wavelet transform 10

1.3.3 Extrema and zero-crossings of the wavelet transform . . 11

1.3.4 Discrete dyadic wavelet transform 12

1.3.5 Finite scale wavelet series 14

1.3.6 Fast wavelet algorithms 18

1.3.7 Interpretation of discrete signals in continuous context . 19

1.4 Overview and Contributions of the Thesis 21

2 THE BASIC REPRESENTATIONS **25**

2.1 Introduction 25

2.2 The Multiscale Maxima Representation 26

2.3 The Multiscale Zero-Crossings Representation 36

3 THE GENERALIZATION: AQLR's AND THEIR PROPERTIES **41**

3.1 Uniqueness Characterization 41

3.2 The structure of the reconstruction set 43

3.3 Bounds on the reconstruction set 45

4 UNIQUENESS **50**

4.1 Description by Discrete Fourier Transform 50

4.2 The nonuniqueness theorem 52

4.3 Test for uniqueness 58

4.4 Examples 63

4.4.1 An example of a unique maxima representation 64

4.4.2 An example of a nonunique maxima representation . . . 67

4.4.3 An example of zero-crossings nonunique representation . 78

5 STABILITY **81**

5.1 BIBO stability 81

5.2 A Lipschitz condition 82

5.3 Discussion 88

6 RECONSTRUCTION **90**

6.1 General theory for an inherently bounded AQLR 90

6.2 The detailed algorithm for a multiscale representation 99

6.3 The rate of convergence 107

7 MODIFICATIONS IN MAXIMA REPRESENTATIONS **114**

7.1 Introduction 114

7.2 Improvement in the representation quality 115

7.3 Reductions in $S_J f$ 117

7.4 Quantization 120

8 CONCLUSIONS **129**

8.1 Summary 129

8.2 General remarks 130

8.3 Further research 132

A APPENDIX - A **134**

A.1 The proof of Lemma 4 134

B BIBLIOGRAPHY **136**

LIST OF FIGURES

4.1 The signal f and its wavelet decomposition. 65

4.2 The reconstruction errors. 68

4.3 The signal $f(k)$ and its wavelet decomposition. 69

4.4 The signals and their first level wavelet transforms. 73

4.5 The second and the third level wavelet transforms. 74

4.6 The boundary of the set $\mathcal{A}$ on the plane (a_1, a_2) 75

4.7 All sequences belonging to $\Gamma(R_m f)$. 76

4.8 All $W_1\eta$ such that $\eta \in \Gamma(R_m f)$. 77

4.9 All $W_2\eta$ such that $\eta \in \Gamma(R_m f)$. 78

4.10 The sequences f and f_a 79

4.11 The sequences $W_1 f$ and $W_1 f_a$ 80

6.1 Cost function $v(\eta_k)$ as a function of k 111

6.2 The noise to signal ratio, $\frac{N}{S}$, as a function of k 112

6.3 The original and reconstructed signals 113

7.1 $\frac{N}{S}$ for the reconstruction from $R^{\Delta} f$ 119

7.2 $\frac{N}{S}$ for reconstruction from $R^q f$ 127

7.3 $\frac{N}{S}$ for reconstruction from $R^{\Delta,q} f$ 128

NOTATION

AQLR	Acronym for Adaptive Quasi Linear Representation
$\mathcal{L}$	Linear space of real, finite sequences
$\Re$	The set of real numbers
X	The operator providing the set of local maximizers
Y	The operator providing the set of local minimizers
Z	The operator providing the set of zero-crossings
W_j	Linear operator
S_j	Linear operator
R_m	Maxima representation operator
R_z	Operator associated with zero-crossings representation
R	Signal representation operator
Γ	Reconstruction set
T_{mf}	Linear operator associated with maxima representation
$t_j^{mf}(k)$	Type of k
$(\cdot)^c$	Complement of $(\cdot)$.

$\overline{(\cdot)}$	Closure of the set $(\cdot)$.
$\|(\cdot)\|$	Euclidean norm
$*$	Convolution operator
$\widehat{(\cdot)}$	Fourier Transform of $(\cdot)$.
$\mathcal{N}(\cdot)$	Null space of $(\cdot)$
$\nabla(\cdot)$	The gradient of $(\cdot)$ (as a column vector)

CHAPTER
ONE

INTRODUCTION AND BACKGROUND

An interesting and promising approach to signal representation is to make explicit important features in the data. The first example, taught in elementary calculus, is a "sketch" of a function based on extrema of a signal and possibly of its first few derivatives. The second instance, widely used in computer vision, is an edge representation of an image. If the size of expected features is a priori unknown, the need for a multiscale analysis is apparent. Therefore, it is not surprising that multiscale sharp variation points (edges) are meaningful features for many signals, and they have been applied, for example, in edge detection [10, 34], signal compression [31], pattern matching [29], detection of transient signals [18, 24] and speech analysis [46].

1.1 Introduction

Traditionally, multiscale edges are determined either as extrema of Gaussian-filtered signals [45] or as zero-crossings of signals convolved with the Laplacian of a Gaussian (see e.g. [22] for a comprehensive review).

S. Mallat in a series of papers [33, 29, 30] (the last joint with S. Zhong) introduced zero-crossings and extrema of the wavelet transform as a multiscale edge representation. Two important advantages of this method are low algorithmic complexity and flexibility in choosing the basic filter. Moreover, [29] and [30] propose reconstruction procedures and show accurate numerical reconstruction results from zero-crossings and maxima representations. In [29, 30], as in many other works in this area, the basic algorithms were developed using continuous variables. The continuous approach gives an excellent background to motivate and justify the use of either local extrema or zero-crossings as important signal features. Unfortunately, in the continuous framework, analytic tools to investigate the information content of the representation are not yet available. The knowledge about properties of the representations is mainly based on empirical reconstruction results. From the theoretical point of view, there are still important open problems, e.g. stability, uniqueness, and structure of a reconstruction set (a family of signals having the same representation).

Our objective is to analyze these theoretical questions using a model of an actual implementation. The main assumption is that the data is discrete and finite. The discrete multiscale maxima and zero-crossings representations are defined in the general set-up of a linear filter bank, however, the main goal is to consider a particular case where the filter bank describes the wavelet transform. Since reconstruction sets of both maxima and zero-crossings representations have a similar structure, a general form is introduced and named Adaptive

Quasi Linear Representation (AQLR). Moreover, many generalizations of the basic maxima and zero-crossings representations fit into the framework of the AQLR. This thesis uses the idea of the AQLR to investigate rigorously three fundamental questions: uniqueness, stability, and reconstruction.

Regarding the uniqueness question, first, conditions for uniqueness are presented. By applying these conditions to the wavelet transform-based representation, a conclusive result is obtained. It turns out, that neither the wavelet maxima representation nor the wavelet zero-crossings representation is, in general, unique. The proof is based on constructing a sinusoidal sequence, whose maxima (zero-crossings) representations cannot be unique for any dyadic wavelet transform.

The next subject is stability of the representation. This issue is of great importance because there are many known examples of unstable zero-crossings representations. In order to improve stability properties, Mallat [29] has included additional sums in the standard zero-crossings representation and, together with Zhong [30], they have introduced the wavelet maxima representation, as a stable alternative to the zero-crossings representation. Indeed, very good numerical results have been reported, but stability analysis has not been pursued. Using the idea of the inherently bounded AQLR, we are able to prove stability results. For a general perturbation, global BIBO (bounded input, bounded output) stability is shown. For a special case, where perturbations are limited to the continuous part of the representation, a Lipschitz condition is satisfied.

One of the most important practical problems is the need for an effective reconstruction scheme. Mallat and Zhong [30] and Mallat [29] have used an algorithm based on alternate projections. In this dissertation, an alternative reconstruction scheme is proposed. The procedure is valid for any inherently bounded AQLR and is based on an appropriate cost function, whose minimum is achieved at the reconstruction set. Specifically, we focus on an algorithm which is based on the integration of the gradient of the cost function. It is shown that this algorithm approaches the reconstruction set. This method yields efficient, parallel algorithms, especially promising in the case of the wavelet transform. In particular, the analog-hardware implementation, which is similar to a neural network, may lead to a very efficient and fast scheme.

In addition, in a general set-up of the inherently bounded AQLR, a standard steepest descent algorithm is described and its convergence is shown. Using this approach for the case of the wavelet maxima representation, an efficient digital reconstruction algorithm is developed and several examples of reconstruction are presented.

The last part of this dissertation describes possible modifications in the basic multiscale maxima representations. The main idea is to preserve the structure of the inherently bounded AQLR, while allowing a trade-off between reconstruction quality and amount of information required for representation. From a wealth of possible modifications, some examples of reasonable modifications have been chosen, one of them shows how quantization can be considered as an integral part

of the representation and how the reconstruction from an interval successfully replaces the standard approach of reconstructing from an approximating point.

1.2 Previous works

The multiscale edge representation has mainly been investigated in the zero-crossings case. The best-known result concerning the reconstruction of a signal from zero-crossings is the Logan Theorem [27]. This theorem basically states that zero-crossings uniquely define the signal within the family of band-pass signals having the property that the width of the band is smaller than the lower frequency of the band. Proving this theorem, Logan made an analytic extension of the signal and used standard properties of zeros of analytic functions. These tools are known as unstable and Logan has noticed that the reconstruction from zero-crossings appears to be very difficult and impractical. Under certain restrictions on the class of signals, usually polynomial data have been assumed, several additional proofs that zero-crossings form a complete (unique) signal representation have been published. All known proofs do not provide any stability results since they are based on unstable characterizations of analytic functions. The reader is referred to [22] for more details and further references.

In addition, in the case of general initial data, the restriction to polynomial data or even to band-limited signals may provide a poor approximation of the original signal. The situation is similar to the fact that a polynomial is determined by its zeros, but any nonzero value of a continuous function cannot be

determined from zero-crossings of the function.

In spite of the last remark, there have been a number of attempts to reconstruct signals from multiscale zero-crossings, especially in image processing, e.g. [14, 48, 39]. They have been based on the belief that the restriction of a given reconstruction scheme into "natural" image data will be sufficiently stable and precise. Although good reconstruction results have been shown, stability results have not been proven.

R. Hummel and R. Moniot [22] have exhibited the stability problem by showing two significantly different signals having almost the same multiscale zero-crossings representations. In order to stabilize the reconstruction of a function from its zero-crossings, the authors have included the gradient along each zero-crossing. In fact, improved numerical results have been reported but stability has not been analyzed. The reconstruction algorithm in [22] is based on the solution of a Heat Equation; this approach is valid only for the Laplacian of a Gaussian filter and it is required to record the zero-crossings on a dense sequence of scales.

Aware of the above problems, S. Mallat [29] proposed to use the wavelet zero-crossings representation as a complete and stable signal description. In order to overcome the apparent instability of zero-crossings, he has included the values of the wavelet transform integral calculated between two consecutive zero-crossings. Using a reconstruction algorithm based on alternate projections, very accurate reconstruction results have been shown. In subsequent work, S. Mallat

together with S. Zhong [31] introduced the wavelet maxima representation as an alternative to the wavelet zero-crossings representation. As in the zero-crossings case, they have demonstrated very accurate reconstruction results. But, in both papers, neither uniqueness nor stability has been proven.

Recently two independent counterexamples for uniqueness have been published. One example was given in a continuous context by Y. Meyer [35]. His example is based on two functions :

$$f_o(x) = \begin{cases} 1 + \cos(x) & \text{if } |x| < \pi \\ 0 & \text{otherwise} \end{cases} .$$

and

$$f_\alpha(x) = \begin{cases} \sum_0^\infty \alpha_k \left(1 + \cos\left((2k+1)x\right)\right) & \text{if } |x| < \pi \\ 0 & \text{otherwise} \end{cases} .$$

Meyer [35] proves that, for a particular wavelet transform, a family of functions $f_0(x) + f_\alpha(x)$, for a suitable set of sequences $\alpha = \{\alpha_k\}_0^\infty$, has the same wavelet zero-crossings representation. The second counterexample was given by Berman [3] and will be described in Section 4.4.2

1.3 Wavelet transform

Since the motivation of this work is to investigate the wavelet maxima and zero-crossings representations, this section gives a brief review on some properties of the wavelet transform.

The wavelet transform is based on analyzing a signal (function) by dilations

and translations of a single analyzing function. This idea may be traced back to the beginning of the century [21] or more recently to the work by A. Calderon [9] in functional analysis and to [44] in quantum field theory and statistical analysis. In engineering, multiresolution signal processing, used in computer vision, and subband coding, developed for speech and image compression, have been recently recognized as different views of discrete wavelet transforms. Therefore, the wavelet transform should be viewed as a well established mathematically, unified framework for many signal analysis techniques. Recently, the subject has drawn much attention, mostly due to establishing connection with multiresolution analysis by S. Mallat [33], and due to constructing orthonormal bases of compactly supported wavelets by I. Daubechies [15]. This paper does not attempt to survey the tremendous number of works recently published. The paper [38] by O. Rioul and M. Vetterli is a very good comprehensive tutorial on wavelets in signal processing and provides an extensive bibliography on the subject. Other general references for wavelet theory and its applications are [1, 11, 12, 17, 13, 16]. In this section, only subjects related to this work will be presented. Our presentation follows essentially the approach described in [29].

1.3.1 Continuous Wavelet Transform

Let $\Psi_s(\xi)$ denote the dilation of a function $\Psi(\xi)$ by a factor (scale) s :

$$\Psi_s(\xi) \triangleq \frac{1}{s}\Psi\left(\frac{\xi}{s}\right)$$

The continuous wavelet transform of a (continuous) function $f(\xi)$ at the scale s and position ξ can be written as the following convolution:

$$W_s f(\xi) = f * \Psi_s(\xi).$$

Let us denote by $\widehat{\Psi}(\omega)$ the Fourier transform of $\Psi(\xi)$. Morlet and Grossman [20] showed that if $\widehat{\Psi}(\omega)$ satisfies:

$$\int_{-\infty}^{\infty} \frac{|\widehat{\Psi}(\omega)|^2}{|\omega|} d\omega = C_\Psi < \infty, \tag{1.3.1}$$

then $f(\xi)$ can be reconstructed from its wavelet transform. The precise statement assumes $f \in L^2(\Re)$ (Hilbert space of measurable, square-integrable real functions), and notices that $f(\xi)$ can be reconstructed only for ξ at which f is continuous. The condition (1.3.1) is called the admissibility condition and it implies that:

$$\widehat{\Psi}(0) = 0 \tag{1.3.2}$$

and

$$\int_{-\infty}^{+\infty} \Psi(\xi) d\xi = 0.$$

The term "wavelet" is used for any function from $L^2(\Re)$ which satisfies the admissibility condition. From (1.3.2) we see that the wavelet $\Psi(\xi)$ can be interpreted as the impulse response of a band-pass filter. Then the wavelet transform of a function f can be viewed as the response of a family of dilated band-pass filters to a signal f.

The main idea of multiscale analysis is that different properties of the signal

f can be detected from $W_s f$ for different scales s. Large s is suitable for low frequency properties, while decreasing s increases sensitivity to small details.

1.3.2 Dyadic continuous wavelet transform

In practical implementations, both parameters, ξ and s, of the continuous wavelet transform $W_s(\xi)$ have to be discretized. The relationship between continuous and discrete versions of wavelet transform will be described carefully. At this point, a discrete set of scales is considered. It turns out, that if

$$\sum_{-\infty}^{\infty} \left|\widehat{\Psi}\left(2^j\omega\right)\right|^2 = 1, \tag{1.3.3}$$

then the scale parameter can be sampled along the dyadic [1] sequence $\{2^j\}_{j\in Z}$ while preserving the reconstruction property. Any wavelet satisfying equation (1.3.3) is called a dyadic wavelet. The sequence of functions:

$$\{W_{2^j} f(\xi)\}_{j\in Z}$$

is called the dyadic wavelet transform.

Let $\tilde{\Psi}_{2^j}(\xi) \triangleq \Psi_{2^j}(-\xi)$. The function f is reconstructed from its dyadic wavelet transform by the following formula:

$$f(\xi) = \sum_{j=-\infty}^{\infty} W_{2^j} f * \tilde{\Psi}_{2^j}(\xi). \tag{1.3.4}$$

[1] Z denotes the set of all integers.

1.3.3 Extrema and zero-crossings of the wavelet transform

In the context of dyadic wavelet transforms, extrema and zero-crossings have an interesting interpretation. First, let us introduce $\theta(\xi)$, a smoothing function. We assume that a smoothing function is the impulse response of a low-pass filter, i.e. it is a function whose Fourier transform has energy concentrated in the low-frequencies. A classical example used in many applications is the Gaussian. Let us assume that the first and the second order derivatives of $\theta(\xi)$ are dyadic wavelets, denoted $\Psi^1(\xi)$ and $\Psi^2(\xi)$, respectively

$$\Psi^1(\xi) \triangleq \frac{d\theta(\xi)}{d\xi} \tag{1.3.5}$$

and

$$\Psi^2(\xi) \triangleq \frac{d^2\theta(\xi)}{d\xi^2}. \tag{1.3.6}$$

The wavelet transforms, defined with respect to each of these wavelets, are given by:

$$W^1_{2^j} f(\xi) = f * \Psi^1_{2^j}(\xi) \tag{1.3.7}$$

and

$$W^2_{2^j} f(\xi) = f * \Psi^2_{2^j}(\xi). \tag{1.3.8}$$

Observe that:

$$W^1_{2^j} f(\xi) = f * \left(2^j \frac{d\theta_{2^j}}{d\xi}\right)(\xi) = 2^j \ \frac{d}{d\xi}\left(f * \theta_{2^j}\right)(\xi) \tag{1.3.9}$$

and

$$W^2_{2^j} f(\xi) = f * \left(2^{2j} \frac{d^2\theta_{2^j}}{d\xi^2}\right)(\xi) = 2^{2j} \ \frac{d^2}{d\xi^2}\left(f * \theta_{2^j}\right)(\xi). \tag{1.3.10}$$

Therefore, in this case, the wavelet transforms $W^1_{2^j} f(\xi)$ and $W^2_{2^j} f(\xi)$ are proportional respectively to the first and the second derivatives of $f(\xi)$ smoothed by $\theta_{2^j}(\xi)$. As a result, the inflection points of $f * \theta_{2^j}(\xi)$ correspond to the local extrema of $W^1_{2^j} f(\xi)$ and to the zero-crossings of $W^2_{2^j} f(\xi)$. Observe that if one uses $W^1_{2^j} f(\xi)$ and detects local extrema, the result is essentially equivalent to Canny's edge detection [10]. Likewise, the zero-crossings detection of $W^2_{2^j} f(\xi)$ can be viewed as a Marr-Hildreth edge-detection algorithm [34]. An inflection point of $f * \theta_{2^j}$ can either be a maximum or a minimum of the absolute value of its first derivative. Sharp variation points, which appear to be the most important features of many signals, are exactly the maxima of the absolute value of the first derivative. The minima correspond to the slow variation points. These two different types of inflection points can be distinguished from the extrema of $W^1_{2^j} f(\xi)$ but cannot be differentiated from the zero-crossings of $W^2_{2^j} f(\xi)$.

1.3.4 Discrete dyadic wavelet transform

The dyadic wavelet transform, defined in the previous section, has infinite number of scales and cannot be implemented in this form. In practice, due to finite view or interest, we are limited by a finite largest scale. Due to finite resolution of any measurement equipment, the finest scale is bounded from below as well. For normalization purposes, let us assume that the finest scale is 1 and 2^J is the largest scale.

For finite number of scales, we need to understand what kind of information

is included in

$$\{W_{2^j} f(\xi)\}_{1 \le j \le J}. \tag{1.3.11}$$

First, let us introduce a function $\Phi(\xi)$ whose Fourier transform satisfies:

$$\left|\widehat{\Phi}(\omega)\right|^2 = \sum_{j=1}^{\infty} \left|\widehat{\Psi}\left(2^j\omega\right)\right|^2. \tag{1.3.12}$$

The function $\Phi(\xi)$ is called the scaling function. Using (1.3.12) one can show that:

$$\left|\widehat{\Phi}(\omega)\right|^2 = \sum_{j=1}^{J} \left|\widehat{\Psi}\left(2^j\omega\right)\right|^2 + \left|\widehat{\Phi}\left(2^J\omega\right)\right|^2 \tag{1.3.13}$$

Using $\sum_{-\infty}^{\infty} \left|\widehat{\Psi}\left(2^j\omega\right)\right|^2 = 1$, it can be shown that $\lim_{\omega \to 0} \left|\widehat{\Phi}(\omega)\right| = 1$. Therefore $\Phi(\omega)$ can be viewed as the impulse response of a low-pass filter or equivalently as a smoothing function. Let us define the smoothing operator S_{2^j} by:

$$S_{2^j} f(\xi) = f * \Phi_{2^j}(\xi) \tag{1.3.14}$$

where $\Phi_{2^j}(\xi) = \frac{1}{2^j}\Phi\left(\frac{\xi}{2^j}\right)$. The larger the scale 2^j, the more details are removed by the smoothing operator S_{2^j}.

Notice that one cannot expect to reconstruct the original signal $f(\xi)$ from $\{W_{2^j} f(\xi)\}_{j=1}^{\infty}$. It turns out, that by disregarding the part $\{W_{2^j} f(\xi)\}_{j=-\infty}^{0}$ we introduce the same error as by smoothing the function $f(\xi)$ by S_1. In other words, one can reconstruct exactly the signal $S_1 f(\xi)$ from $\{W_{2^j} f(\xi)\}_{j=1}^{\infty}$. Moreover, $S_1 f(\xi)$ can be calculated from the following finite collection of functions:

$$\left\{S_{2^J} f(\xi), \{W_{2^j} f(\xi)\}_{j=1}^{J}\right\}.$$

That is, $S_{2^J} f(\xi)$ contains the same information about $S_1 f(\xi)$ as $\{W_{2^j} f(\xi)\}_{j=J+1}^{\infty}$.

Energy conservation considerations give an additional view on the above facts. The Fourier transform of $S_1 f(\xi)$, $S_J f(\xi)$ and $W_{2^j} f(\xi)$ are respectively given by:

$$\widehat{S_1 f}(\omega) = \hat{\Phi}(\omega)\hat{f}(\omega) \tag{1.3.15}$$

$$\widehat{S_{2^J} f}(\omega) = \hat{\Phi}\left(2^J\omega\right)\hat{f}(\omega) \tag{1.3.16}$$

$$\widehat{W_{2^J} f}(\omega) = \hat{\Psi}\left(2^J\omega\right)\hat{f}(\omega) \tag{1.3.17}$$

Using Parseval's theorem and equations (1.3.15), (1.3.16) and (1.3.17), it can be shown that:

$$\|S_1 f(\xi)\|^2 = \sum_{j=1}^{J} \|W_{2^j} f(\xi)\|^2 + \|S_{2^J} f(\xi)\|^2 . \tag{1.3.18}$$

The functions $\left\{S_{2^J} f(\xi), \{W_{2^j} f(\xi)\}_{j=1}^{J}\right\}$ are called the finite scale wavelet transform of $S_1 f(\xi)$. The last step toward introduction of a practical signal representation it to establish a sampling scheme for ξ.

1.3.5 Finite scale wavelet series

Let $\{S_1 f(n)\}_{n\in Z}$ be samplings at integers (for normalization purposes) of the function $S_1 f(\xi)$. Similarly, $\{S_{2^J} f(n)\}_{n\in Z}$, $\{W_{2^j} f(n)\}_{n\in Z}$ are samplings at integers of the functions $S_{2^J} f(\xi)$ and $W_{2^j} f(\xi)$. In this section, $\{S_1 f(n)\}_{n\in Z}$ is assumed to be known and the question of how to calculate $\{S_{2^J} f(n)\}_{n\in Z}$ and $\{W_{2^j} f(n)\}_{n\in Z}$ is discussed.

First, let us consider $J = 1$. Let H be a discrete filter whose convolution with $\{S_1 f(n)\}_{n\in Z}$ produces $\{S_{2^1} f(n)\}_{n\in Z}$. In other words, the Fourier series

of $\{S_{2^1} f(n)\}_{n\in Z}$ is equal to the Fourier series of $\{S_1 f(n)\}_{n\in Z}$ multiplied by a 2π-periodic function $H(\omega)$. The Fourier series of these signals are respectively

$$\sum_{n=-\infty}^{\infty} (f * \Phi)(n) e^{-in\omega} \qquad \text{and} \qquad \sum_{n=-\infty}^{\infty} (f * \Phi_2)(n) e^{-in\omega}.$$

These two series, using the Poisson formula, can be rewritten as:

$$\sum_{n=-\infty}^{\infty} \widehat{f}(\omega + 2n\pi)\widehat{\Phi}(\omega + 2n\pi) \qquad \text{and} \qquad \sum_{n=-\infty}^{\infty} \widehat{f}(\omega + 2n\pi)\widehat{\Phi}\left(2(\omega + 2n\pi)\right).$$

In order that, for all $\widehat{f}(\omega)$, the right series be equal to the left series multiplied by $H(\omega)$ the following condition has to be satisfied:

$$\widehat{\Phi}(2\omega) = H(\omega)\widehat{\Phi}(\omega). \tag{1.3.19}$$

Notice that $\left|\widehat{\Phi}(0)\right| = 1$ implies $|H(0)| = 1$. Thus, the function $H(\omega)$ can be interpreted as the transfer function of a discrete low-pass filter.

By cascading equation (1.3.19), a necessary condition on $\widehat{\Phi}(\omega)$ is obtained.

$$\widehat{\Phi}(\omega) = \prod_{p=1}^{\infty} H\left(2^{-p}\omega\right). \tag{1.3.20}$$

A following sufficient condition is shown in [32]. It turns out that if the 2π-periodic function $H(\omega)$ satisfies

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 \leq 1, \tag{1.3.21}$$

then the function $\Phi(\xi)$ whose Fourier transform is defined by equation (1.3.20) is a function in $L^2(\Re)$.

However, the resulting limit (1.3.20) sometimes happens to be an "ugly", discontinuous, fractal function. In recent years, there has been a great interest

in finding conditions on $H(\omega)$ such that the underlying $\Phi(\xi)$ would have desired regularity properties. The classical result on the subject is due to I. Daubechies [15].

Lemma 1 *Let us assume that $H(\omega)$ has $N_\pi \geq 1$ zeros at $\omega = \pi$ and define $F(\omega)$ and K in the following way.*

$$H(\omega) = \left(\frac{(1+e^{i\omega})}{2}\right)^{N_\pi} F(\omega) \tag{1.3.22}$$

$$K = \sup_{\omega \in [0,2\pi]} |F(\omega)| . \tag{1.3.23}$$

If $K < 2^{N_\pi - 1}$ then the product $\prod_{p=1}^{\infty} H(2^{-p}\omega)$ converges pointwise to $\widehat{\Phi}(\omega)$, Fourier transform of a continuous function $\Phi(\xi)$.

In view of these considerations, it was accepted as a standard technique, in the design of wavelet-based discrete filters, to have at π as many as possible zeros of $H(\omega)$ (see [38, 43]). Recently, O. Rioul [37] has shown that $H(\pi) = 0$ is a necessary condition to have a resulting $\Phi(\xi)$ continuous.

For a given scaling function $\Phi(\xi)$, the corresponding wavelet $\Psi(\xi)$ is obtained from equation (1.3.13) by substituting $J = 1$.

$$\left|\widehat{\Psi}(2\omega)\right|^2 = \left|\widehat{\Phi}(\omega)\right|^2 - \left|\widehat{\Phi}(2\omega)\right|^2 \tag{1.3.24}$$

Using equations (1.3.19) and (1.3.24), we can define a discrete filter $G(\omega)$ such that:

$$\widehat{\Psi}(2\omega) = G(\omega)\widehat{\Phi}(\omega) \tag{1.3.25}$$

where

$$|G(\omega)|^2 + |H(\omega)|^2 = 1 \tag{1.3.26}$$

The function $G(\omega)$ is 2π-periodic and it can be interpreted as the transfer function of a high-pass filter.

One of the best known properties of wavelet transforms is the possibility that a wavelet $\Psi(\xi)$ can generate an orthogonal basis. It can be obtained by changing the inequality (1.3.21) to an equality and then choosing $G(\omega) = e^{-i\omega}\overline{H(\omega+\pi)}$ (for details and proofs see Section 2.4 in [33]). However, this condition does not bring any significant benefit in the context of local maxima or zero-crossings.

For the maxima representation, it is advantageous to have a wavelet $\Psi(\xi)$ equal to the first derivative of a smoothing function θ. This implies that $\hat{\Psi}(\omega)$ must have a zero of order 1 at $\omega = 1$. Because $\left|\hat{\Phi}(0)\right| = 1$, equation (1.3.25) yields that $G(\omega)$ must have a zero of order 1 at $\omega = 0$. Since $G(\omega)$ and $H(\omega)$ are related by equation (1.3.26), the latter condition is equivalent to

$$H(\omega) = 1 + O(\omega^2), \tag{1.3.27}$$

where $O(\nu)$ is an arbitrary function such that:

$$\lim_{\nu \to 0} \frac{O(\nu)}{\nu} = K < \infty. \tag{1.3.28}$$

For the zero-crossings representation, we want to build a wavelet $\Psi(\xi)$ equal to a second-order derivative of a smoothing function $\theta(\xi)$. This implies that $G(\omega)$ must have a zero of order 2 at $\omega = 0$.

1.3.6 Fast wavelet algorithms

As was shown in the previous section, if the discrete filters H, G satisfy conditions (1.3.21) and (1.3.26) then $\{S_{2^j}f(n)\}_{n\in Z}$ and $\{W_{2^j}f(n)\}_{n\in Z}$ can be calculated from $\{S_1 f(n)\}_{n\in Z}$. It turns out that these calculations involve neither the wavelet function $\Psi(\xi)$ nor the scaling function $\Phi(\xi)$; they are solely based on the discrete filters H, G.

Before presenting the actual algorithm, some notations need to be introduced. Let H_p, G_p denote the discrete filter whose transfer functions are $H(2^p\omega)$, $G(2^p\omega)$, respectively. The impulse response of H_p (G_p) is obtained by putting $2^p - 1$ zeros between every two adjacent coefficients of the impulse response of the filter H (G). We also denote $\widetilde{H}_p$ and $\widetilde{G}_p$ the filters whose transfer functions are respectively $\overline{H(2^p\omega)}$ and $\overline{G(2^p\omega)}$ (complex conjugates of $H(2^p\omega)$ and $G(2^p\omega)$). Finally, let

$$W^d_{2^j} f \triangleq \{W_{2^j}f(n)\}_{n\in Z} \tag{1.3.29}$$

and

$$S^d_{2^j} f \triangleq \{S_{2^j}f(n)\}_{n\in Z}\,. \tag{1.3.30}$$

Proposition 1 *Let g_j and h_j denote the impulse responses of filters G_j and H_j. Then the discrete signals $W^d_{2^j}f$ and $S^d_{2^j}f$ are obtained by the following recursion formulas*

$$W^d_{2^{j+1}} f = S^d_{2^j} f * g_j \tag{1.3.31}$$

$$S^d_{2^{j+1}} f = S^d_{2^j} f * h_j \tag{1.3.32}$$

for $j = 0, 1, \ldots, J-1$.

The proof is a straightforward consequence of (1.3.19) and (1.3.25).

The complexity of this algorithm is of order NKJ, where N is a number of nonzero samples in the original signal $\{S_1 f(n)\}_{n \in Z}$, K is a number of nonzero coefficients in impulse responses of the filters H and G, and J is of course the number of decomposition levels. The number of levels cannot exceed $\log(N)$, and usually K is a small number, therefore the algorithm complexity is assumed to be of order $(N \log N)$. Thus, in similarity to fast Fourier transform, this algorithm is called the fast wavelet transform.

The reconstruction or inverse wavelet transform is the calculation of $S_1^d f$ from $\left\{W_{2^j}^d f\right\}_{j=1}^J$ and $S_{2^J}^d f$.

Proposition 2 *The discrete signal* $S_1^d f$ *can be calculated by the following recursion formula.*

$$S_{2^{j-1}}^d f = W_{2^j}^d f * \tilde{g}_{j-1} + S_{2^j}^d f * \tilde{h}_{j-1} \tag{1.3.33}$$

for $j = J, J-1, \ldots, 2, 1$. *Naturally,* $\tilde{g}_j, \tilde{h}_j$ *are impulse responses of filters* $\tilde{G}_j, \widetilde{H}_j$, *respectively.*

The proof is again a consequence of (1.3.19) and (1.3.25).

1.3.7 Interpretation of discrete signals in continuous context

In the previous sections, it has been shown that for a given continuous function $f(\xi)$, one can calculate $S_1 f(\xi)$ and then, by sampling it at integers, obtain the

discrete signal $S_1^d f$. Now, using the fast wavelet algorithm, one can calculate $S_{2^J}^d f$ and $W_{2^j}^d f$ which are indeed samples at integers of the continuous functions $S_{2^J} f(\xi)$ and $W_{2^j} f(\xi)$, respectively.

At this point, one can ask what would happen if an arbitrary discrete sequence is applied to the fast wavelet algorithm. The answer can be found in the following representation property.

Lemma 2 *[30]*

Let $D = \{d_n\}_{n \in Z}$ be a discrete signal of finite energy: $\sum_{n=-\infty}^{\infty} |d_n|^2 < \infty$. Let us suppose that there exist $K_1 > 0$ and $K_2 > 0$ such that for all $\omega \in \Re$ the Fourier transform $\widehat{\Phi}(\omega)$ satisfies:

$$K_1 \leq \sum_{n=-\infty}^{\infty} \left|\widehat{\Phi}(\omega + 2n\pi)\right|^2 \leq K_2. \tag{1.3.34}$$

Then there exists a function $f(\xi) \in L^2(\Re)$ such that

$$\forall n \in Z \qquad S_1 f(n) = d_n \tag{1.3.35}$$

In sampling theory terminology, the above characteristic can be interpreted as a modified interpolation property (see e.g. [40]).

This well established relationship between continuous and discrete signals is a very important property of wavelet transforms. Consequently, it enables us to interpret continuous and discrete results as complementary views on corresponding objects. In essence, the approach in [29] and [30] is based on giving continuous interpretation to discrete signals. As was mentioned earlier, our

method is based on the analysis of discrete signals. Nevertheless, at least in the wavelet transform context, these works might be regarded as different views of corresponding problems.

1.4 Overview and Contributions of the Thesis

Our point of view is as follows. Using a continuous variable approach, several very promising representations and reconstruction algorithms have been developed. Especially algorithms based on wavelets deserve particular attention, because of low complexity of the fast wavelet transform and because of possible flexibility in choosing the basic filter. These algorithms provide accurate numerical reconstruction results, but their basic properties have not been analyzed yet. The reason seems to be, that the analysis of a continuous multiscale representation is a very difficult mathematical problem. On the other hand, even if the continuous analysis is given, the conclusions about the discrete realization are not obvious.

This thesis can be viewed as an attempt to analyze rigorously the numerical reconstruction results from the wavelet maxima and zero-crossings representations. This objective leads to the discrete and finite data assumption. In essence, using this assumption, we were able to obtain several new results. The following chapters describe original contributions of this dissertation.

It turns out, that the discrete implementation of a continuous framework is a delicate procedure and many details should be worked out. Even, in the case

where the discretization of the linear transform is straightforward [2], the maxima and zero-crossings representations should be redefined and the investigated problems should be restated. The basic definitions are given in Chapter 2. The first observation is that the structure of the wavelet transform is not essential for the analysis and can be generalized to any linear filter bank. After introducing precise definitions of the multiscale maxima (zero -crossings) representations, the exact conditions that an arbitrary sequence has to satisfy in order to belong to the reconstruction set are presented. As a generalization of these conditions, the structure of the Adaptive Quasi Linear Representation (AQLR) is defined.

The next observation is that, from the stability and reconstruction point of view, the most important property of the wavelet maxima and zero-crossings representation, as introduced by S. Mallat and S. Zhong [30, 29], is the fact that the reconstruction set is always bounded. This finding leads to the definition of the inherently bounded AQLR.

Chapter 3 uses some concepts from convex analysis and linear parametric programming in order to establish foundations for the subsequent discussion on uniqueness, stability, and reconstruction. This chapter only assumes the structure of AQLR, especially the inherently bounded AQLR.

Chapter 4 investigates the uniqueness question. The general theorem about nonuniqueness of the wavelet maxima (zero-crossings) representation is proven.

[2] An additional advantage of the wavelet transform is the correspondence between the continuous and the discrete transforms.

The proof is constructive in the sense that a sequence which has a nonunique representation is shown. In view of additional counterexamples by Y. Meyer [35], the importance of our result is mostly related to its generality. In addition, the question of how a particular representation can be tested for uniqueness is discussed. Several examples of both unique and nonunique representations are given.

Stability results are described in Chapter 5. It turns out, that BIBO stability is closely related to the definition of the inherently AQLR. For a special case, where perturbations are limited to the continuous part of the representation, a Lipschitz condition is satisfied. To the best of our knowledge, these are the first rigorous stability results established in the context of multiscale maxima (zero-crossings) representations.

Chapter 6 is devoted to the reconstruction issue. In the set-up of the inherently bounded AQLR, a new reconstruction scheme is proposed. The proposed algorithm is based on a cost function which does not have local extrema outside the reconstruction set. The convergence of two algorithms is shown: the first is based on the integration of the gradient of the cost function and can be implemented by analog hardware; the second is a standard steepest descent algorithm which is used in digital simulations.

The idea to minimize a cost function in order to reconstruct a signal from the multiscale edge representation has appeared in many works, e.g. [22, 48, 39]. The comparison reveals the following advantages of the proposed algorithm.

- This algorithm is based on a continuously differentiable cost function.
- It does not apply approximations.
- It is adapted for both unique and nonunique cases.
- Its validity and convergence are guaranteed.

In short, we obtain new theoretical results because we benefit from the well established representation structure, especially from its boundness property.

Chapter 7 describes possible modifications in the basic multiscale maxima representation which preserve the structure of the inherently bounded AQLR. From a wealth of possible modifications, several examples of reasonable modifications have been chosen. Perhaps the most interesting example is the one describing how quantization can be considered as an integral part of the representation and how the reconstruction from an interval successfully replaces the standard approach to reconstruct from an approximating point.

Chapter 8 concludes this thesis with a summary, some general remarks and, as usual, the hardest questions are left for further research.

CHAPTER
TWO

THE BASIC REPRESENTATIONS

2.1 Introduction

This chapter consists of two similar sections, the first describes the discrete multiscale maxima representation and the second is devoted to the discrete multiscale zero-crossings representation. The main goal of the chapter is to state precise definitions and to characterize the reconstruction set, which is an inverse image of the representation. It turns out, that both the multiscale maxima representation and the multiscale zero-crossings representation induce a very similar structure in their reconstruction sets. Therefore a general signal description called an Adaptive Quasi Linear Representation (AQLR) is introduced. In this thesis, a subclass, named an inherently bounded AQLR, is of main interest. In subsequent chapters these structures will be studied extensively. This chapter also develops sufficient conditions for a signal representation to be an inherently bounded AQLR. As the main result, it is shown that both the wavelet maxima representation and the wavelet zero-crossings representation are inherently

bounded AQLR's.

Both sections are based on the definitions introduced by S. Mallat and S. Zhong [30], and S. Mallat [29]. However, there are two noticeable differences between the approach of [30, 29] and ours. First, we assume no knowledge about the (time) continuous version of signals, namely, the original signal is assumed to be (time) discrete and all representation properties (uniqueness, stability, reconstruction) are considered with respect to this signal. Second, the wavelet transform, used in [30, 29], is generalized here to an arbitrary collection of linear operators, i.e. the general structure consists of a linear filter bank.

The adjective "basic" in the tittle has been introduced in order to distinguish between representations described here and further generalizations which will be introduced in chapter 7.

2.2 The Multiscale Maxima Representation

As was mentioned earlier, the main assumption in this thesis is that signals are (time) discrete and of finite duration. In other words, all signals under consideration belong to $\mathcal{L}$, a linear space of real, finite sequences:

$$\mathcal{L} \triangleq \left\{ f : f = \{f(n)\}_{n=0}^{N-1}, \quad f(n) \in \Re \right\}.$$

In the sequel we will use concepts like convex set, basis, linear operator, its null and range space, linear operator representation by a matrix, norm, etc. which are usually related to vector spaces. In these cases, we will regard this space

as a standard $\Re^N$. However, due to the need of introducing the ideas of local maxima, minima and zero-crossings the form of a linear space of finite sequences appears to be appropriate.

The definitions of local maxima and minima sets are introduced as follows. Let X and Y denote operators on $\mathcal{L}$ which provide the sets of local maximizers and minimizers, respectively, of a sequence $f \in \mathcal{L}$. The formal definitions are :

$$Xf = \{k : f(k+1) \leq f(k) \text{ and } f(k-1) \leq f(k) \quad k = 0, 1, 2, \ldots, N-1\} \quad (2.2.1)$$

$$Yf = \{k : f(k+1) \geq f(k) \text{ and } f(k-1) \geq f(k) \quad k = 0, 1, 2, \ldots, N-1\}. \quad (2.2.2)$$

In this thesis, in order to avoid boundary problems, an N-periodic extension of finite sequences is assumed.

As was already mentioned, the basic-set up is an arbitrary linear filter bank. Namely, in this thesis, the term "multiscale" refers to any finite collection of linear operators. Let $W_1, W_2, \ldots, W_J, S_J$ denote linear operators on $\mathcal{L}$. The collection $\{W_1, W_2, \ldots, W_J, S_J\}$ will be called the multiscale linear operator.

The sets XW_jf, YW_jf are local maxima and minima points of the sequence W_jf. The values of W_jf at extreme points are denoted by $\{W_jf(k)\}_{k \in XW_jf \cup YW_jf}$. Using the above notation, we can define the multiscale local extrema representation, R_mf as:

$$R_mf \triangleq \left\{ \left\{ XW_jf, YW_jf, \{W_jf(k)\}_{k \in XW_jf \cup YW_jf} \right\}_{j=1}^{J}, S_Jf \right\}. \quad (2.2.3)$$

Observe that the whole signal S_Jf is included in the representation. Usually, S_J is a low pass filter, and S_Jf can be described in a more compact way. This

point will be discussed further in Chapter 7. At this stage the main goal is to give a precise and general definition. This definition matches the basic definition used in [30]. Mallat and Zhong [30] have further modified this transformation to include only local maxima of absolute values. They have used the term "maxima representation" for this signal description. Following [31], $R_m f$, even in the version (2.2.3), will be called the multiscale maxima representation. In the context of [30] $W_1 f, W_2, \ldots, W_J f, S_J f$ describes the J-level wavelet decomposition of a signal f [1]. In this case $R_m f$ will be called the wavelet maxima representation.

Perhaps the main drawback of this representation is the fact that its analysis is not easy, mainly because R_m is a nonlinear operator. Our approach is to separate its linear and nonlinear components. The determination of the extreme point sets is a highly nonlinear operation on f. However, for the given extrema sets, $XW_j f$ and $YW_j f$, the remaining data are obtained by a linear operation of sampling an image of a linear operator at fixed points. This observation is the motivation for considering $R_m f$ as consisting of two parts: the sampling information and the maxima information. The sampling information is the sequence $S_J f$ and the values of $W_j f$ at the points $XW_j f \cup YW_j f$ (j=1,2,...,J). The maxima information consists of the sets $XW_j f, YW_j f$ and the fact that the elements of $XW_j f$ and $YW_j f$ are the local maximizers and minimizers of $W_j f$.

Let T_{mf} denote the linear operator associated with the sampling information.

[1] The shorter notation $W_1 f, W_2, \ldots, W_J f, S_J f$ is used instead of the standard $W_1^d f, W_{2^1}^d, \ldots, W_{2^J}^d f, S_{2^J}^d f$

The following is its precise definition.

$$T_{mf} : \mathcal{L} \rightarrow \mathcal{L}^e$$

such that for all $h \in \mathcal{L}$

$$T_{mf}h = \{S_J h, \{W_1 h(k)\}_{k \in XW_1 f \cup YW_1 f}, \ldots, \{W_J h(k)\}_{k \in XW_J f \cup YW_J f}\}. \quad (2.2.4)$$

$\mathcal{L}^e$ is the linear space of finite, real sequences of length N^e, where

$$N^e = N + \sum_{j=1}^{J} (\mid XW_j f \mid + \mid YW_j f \mid).$$

For any set A, $\mid A \mid$ denotes its cardinal number.

Now, $R_m f$ is written in an alternative way as:

$$R_m f = \left\{ \{XW_j f, YW_j f\}_{j=1}^{J}, T_{mf} f \right\}. \quad (2.2.5)$$

This form will lead to a definition of a general family of signal descriptions having reconstruction sets with common structure. For a given representation Rf, the reconstruction set $\Gamma(Rf)$ is defined as the set of all sequences satisfying this representation, i.e.

$$\Gamma(Rf) \triangleq \{h \in \mathcal{L} : Rh = Rf\}. \quad (2.2.6)$$

In other words, the reconstruction set is the inverse image of the representation.

At this point, the structure of the reconstruction set of the multiscale maxima representation is considered. It is clear that in order to satisfy a given maxima representation, a sequence $h \in \mathcal{L}$, in addition to obeying the sampling

information $T_{mf}h = T_{mf}f$, needs to meet the requirement that W_jh has local extrema at the points of XW_jf and YW_jf.

Suppose that $T_{mf}h = T_{mf}f$ and for a moment let us dwell upon the latter condition. Loosely speaking, we have to assure that W_jh is increasing after a minimum and before a maximum and it is decreasing otherwise. In order to make it rigorous we need to introduce several definitions. For any $k \in XW_jf \cup YW_jf$, the segment of k with respect to extrema of f at level j, $P_j^{mf}(k)$ is defined as:

$$P_j^{mf}(k) \triangleq \{k, k+1, \ldots, k+r\} \tag{2.2.7}$$

such that:

$$r \geq 1$$

$$k+r \in XW_jf \cup YW_jf$$

$$k+1, \ldots, k+r-1 \in (XW_jf \cup YW_jf)^c$$

where $(XW_jf \cup YW_jf)^c$ denotes the complement of the set $XW_jf \cup YW_jf$ with respect to $\{0, 1, \ldots, N-1\}$. Note that due to the N-periodic extension employed, $k+i$ is defined modulo N.

The desired monotonic property can be achieved by enforcing an appropriate constraint on $W_jf(k+1) - W_jf(k)$ $(> 0, \geq 0, < 0, \leq 0)$. If one of the points $k+1, k$ is not an extremum, such a constraint is a function of k and will be defined by the type of k, $t_j^{mf}(k)$. If both k and $k+1$ are extreme points, the specific constraint cannot be defined solely either by k or by $k+1$. However, in the latter case, the sampling information assures the right relationship between

$W_j f(k+1)$ and $W_j f(k)$. Consequently, the regular subset of $XW_j f \cup YW_j f$ is defined by:

$$(XW_j f \cup YW_j f)^r \triangleq \{k \in XW_j f \cup YW_j f : k+1 \in (XW_j f \cup YW_j f)^c\}. \quad (2.2.8)$$

For all $k \in (XW_j f \cup YW_j f)^r$, the type of k with respect to $XW_j f, YW_j f$, $t_j^{mf}(k)$ is defined by:

$$t_j^{mf}(k) \triangleq \begin{cases} -1 & \text{if } k \in XW_j f \\ 1 & \text{otherwise.} \end{cases}$$

For all $i \in (XW_j f \cup YW_j f)^c$, the type of i with respect to $XW_j f, YW_j f$, $t_j^{mf}(i)$ is introduced by:

$$t_j^{mf}(i) \triangleq \begin{cases} -1 & \text{if } i \in P_j^{mf}(k) \text{ and } k \in XW_j f \\ 1 & \text{otherwise.} \end{cases}$$

This is a valid definition because it is easy to show that for given j and $i \in (XW_j f \cup YW_j f)^c$ there exists exactly one $k \in XW_j f \cup YW_j f$ such that $i \in P_j^{mf}(k)$.

The essence of the above definitions is that $t_j^{mf}(k) = 1$ if and only if $W_j f(k+1) > W_j f(k)$. A similar statement is true for $t_j^{mf}(k) = -1$. Since the maxima representation preserves these monotonic properties, the following theorem is easily verified.

Theorem 1 *Let $R_m f$ be a given multiscale maxima representation. Then $h \in \Gamma(R_m f)$ if and only if*

$$T_{mf} h = T_{mf} f \quad (2.2.9)$$

$$t_j^{mf}(k) \cdot (W_j h(k+1) - W_j h(k)) \;\; > \;\; 0. \tag{2.2.10}$$

The last inequality should be satisfied for $j = 1, 2, \ldots, J$ *and for all*

$$k \in (XW_j f \cup YW_j f)^c \cup (XW_j f \cup YW_j f)^r.$$

The maxima representations can be cast into the form $Rf = \{Vf, Tf\}$, where Vf consists of sets of integers from $\{0, 1, 2, \ldots, N-1\}$ and T is a linear operator which may depend on Vf. However, the key feature of the maxima representation is the fact that the set Vf yields additional constraints in the form of linear inequalities, which do not appear directly in Rf. Stimulated by this observation, we define the following general family of signal representations.

Definition 2.2.1 *The representation* $Rf = \{Vf, Tf\}$ *is called an Adaptive Quasi Linear Representation (AQLR) if there exists a linear operator* $\mathbf{C}$ *and a sequence* $\mathbf{c}$ *such that:*

$$x \in \Gamma(Rf) \Leftrightarrow Tx = Tf \;\; and \;\; \mathbf{C}x > \mathbf{c}. \tag{2.2.11}$$

$\mathbf{C}, \mathbf{c}$ *may depend on* Vf*, but they must be independent of* Tf.

The reasoning behind the name "Adaptive Quasi Linear Representation" (AQLR) is the following. This representation is adaptive since $T, \mathbf{C}, \mathbf{c}$ depend on the sequence f (via the set Vf). It is quasi linear because it is based on a set of linear equalities and inequalities. In addition, the AQLR appears to be

the simplest nonlinear multiscale representation, at least from the point of view of the structural complexity of the reconstruction set.

In the sequel, instead of "sequence" notation e.g. $x \in \mathcal{L}$, sometimes we will use "vector" notation", then x will be a column vector, $x \in \Re^N$ and linear operators, like $\mathbf{C}$, will appear as corresponding matrices. The interpretation of (refeq:mat1) in matrix and vector context is obvious.

Clearly, the following is true.

Proposition 3 *Any multiscale maxima representation is an AQLR.*

The next definition is a generalization of an essential boundedness property of the wavelet maxima representation which says that the reconstruction set is bounded by the linear part of the representation.

Definition 2.2.2 *An AQLR is called inherently bounded if there exists a real* $K > 0$ *such that*

$$x \in \Gamma(Rf) \Rightarrow \|x\| \leq K\|Tf\|. \tag{2.2.12}$$

In this thesis, $\|\cdot\|$ denotes the Euclidean norm in an appropriate, finite dimensional linear space. The coefficient K can depend on the parameters of the representation e.g. $N, J, W_1, \ldots, W_J, S_J$ but it must be independent of Vf and Tf.

Condition (2.2.12) is very similar to a well known condition for stability in regular or irregular sampling theory. See, for example, a comprehensive paper by J. Benedetto [2]. Our situation is slightly dissimilar because for different

signals, different linear operators are obtained. Nevertheless, condition (2.2.12) implies stability [2].

The vast majority of the results in this thesis are developed in the framework of the inherently bounded AQLR's. In order to have a good idea about what kind of a representation matches this definition, the following definitions are introduced.

Definition 2.2.3 *Let* $Rf = \{Vf, Tf\}$ *be based on the linear filter bank* $\{W_1, \ldots, W_J, S_J\}$. *This representation is called bounding if there exists* $K > 0$, *such that for all* $h \in \Gamma(Rf)$ *and* $j = 1, 2, \ldots, J$.

$$\|W_j h\| \leq K\|Tf\| \tag{2.2.13}$$

and

$$\|S_J h\| \leq K\|Tf\|. \tag{2.2.14}$$

Proposition 4 *Any multiscale maxima representation is bounding.*

Proof: Let $h \in \Gamma(R_m f)$. $S_J h$ is included in $T_{m_f} h$, hence:

$$\|S_J h\| \leq \|T_{m_f} f\|. \tag{2.2.15}$$

Consider:

$$\mid W_j h(n) \mid \leq \max_n \mid W_j h(n) \mid = \max_n \mid W_j f(n) \mid \leq \|T_{m_f} f\|. \tag{2.2.16}$$

[2] see Chapter 5

The middle equality holds because $W_j h$ has the same local extrema as $W_j f$, in particular it has the same global extrema as $W_j f$. The right inequality is valid since $\max_n \mid W_j h(n) \mid$ appears (with its original sign) as a component of $T_{m_f} f$. Therefore we conclude

$$\|W_j h\| \leq \sqrt{N} \|T_{m_f} f\|. \tag{2.2.17}$$

□

One can think about multiscale maxima representations as consisting of two phases: the first is a multiscale linear operator which provides $\{W_1 f, W_2 f, \ldots, W_J f, S_J f\}$; the second phase is a nonlinear operation on $W_j f$ $(j = 1, 2, \ldots, J)$. In this context, the bounding property is related to the nonlinear part of the representation.

A bounding multiscale representation is inherently bounded if it is based on a linear filter bank which fulfills the following condition.

Definition 2.2.4 *A multiscale linear operator* $\{W_1, W_2, \ldots, W_J, S_J\}$ *is called complete if there exists* $K > 0$ *such that for all* $x \in \mathcal{L}$

$$\|x\| \leq K \left(\left(\sum_{j=1}^{J} \|W_j x\| \right) + \|S_J x\| \right) \tag{2.2.18}$$

Proposition 5 *Let* $Rf = \{Vf, Tf\}$ *be based on the linear filter bank* $\{W_1, \ldots, W_J, S_J\}$. *If* Rf *is a bounding AQLR and the linear filter bank is complete then this representation is an inherently bounded AQLR.*

Proof It is a straightforward consequence of the definitions. Indeed,

$$\|x\| \leq K_1 \left(\left(\sum_{j=1}^{J} \|W_j x\| \right) + \|S_J x\| \right) \leq K_1 \cdot K_2 \cdot (J+1) \cdot \|Tf\|. \qquad (2.2.19)$$

□

Since $\{W_1, W_2, \ldots, W_J, S_J\}$ are linear operators, the underlying multiscale representation is complete if and only if the composite matrix $[\mathcal{W}_1' \vdots \mathcal{W}_2' \vdots \cdots \vdots \mathcal{W}_J' \vdots \mathcal{S}_J']$ is of a full rank. $\mathcal{W}_j'$ denotes the transpose of the matrix corresponding to the linear operator W_j. In other words, the word complete which is associated with uniqueness appears to be proper, the multiscale linear representation is complete if and only if $\{W_1 f, W_2 f, \ldots, W_J f, S_J f\}$ is a unique representation of a signal f. Due to the perfect reconstruction property, it is evident, that the wavelet decomposition is a complete multiscale representation. Therefore:

Proposition 6 *The wavelet maxima representation is an inherently bounded AQLR.*

The next section describes a very similar treatment for the multiscale zero-crossings representation. The main observation is that the wavelet zero-crossings representation also is an inherently bounded AQLR.

2.3 The Multiscale Zero-Crossings Representation

In defining the multiscale zero-crossings representation, we essentially follow [29], but minor changes are necessary due to our basic assumption that only a

(time) discrete signal version is available. Let Z be an operator which provides the set of zero-crossings of a given sequence $f \in \mathcal{L}$, i.e.

$$Zf \triangleq \{k : f(k-1) \cdot f(k) \leq 0 \quad k = 0, 1, \ldots, N-1\}. \tag{2.3.1}$$

Mallat in [29] has stabilized the zero-crossings representation by including the values of the wavelet transform integral calculated between consecutive zero-crossing points. For the purpose of the precise discrete definition of these values, the segment of k, with respect to zero-crossings of f at level j, is introduced. It is denoted by $P_j^{zf}(k)$ and defined, for all $k \in ZW_j f$, as follows:

$$P_j^{zf}(k) \triangleq \{k, k+1, \ldots, k+r\} \tag{2.3.2}$$

such that

$$r \geq 0$$

$$k + r + 1 \in ZW_j f$$

$$k+1, k+2, \ldots, k+r \in (ZW_j f)^c .$$

The sequence of sums of $h(n)$ between consecutive zero-crossing points of f at level j, $U_j^{zf} h$ is given by:

$$U_j^{zf} h \triangleq \left\{ \sum_{i \in P_j^{zf}(k)} W_j h(i) \right\}_{k \in ZW_j f} . \tag{2.3.3}$$

The multiscale zero-crossings representation, $R_z f$, is defined as:

$$R_z f \triangleq \left\{ \{ZW_j f, U_j^{zf} f\}_{j=1}^{J}, S_J f \right\}. \tag{2.3.4}$$

As in the maxima representation case, for fixed sets ZW_jf, the remaining data $U_j^{zf}f$ and S_Jf are obtained by a linear operator, denoted by T_{zf}.

$$T_{zf} : \mathcal{L} \to \mathcal{L}^o$$

such that:

$$T_{zf}h = \{S_Jh, U_1^{zf}h, \ldots, U_J^{zf}h\}. \tag{2.3.5}$$

$\mathcal{L}^o$ is a linear space of finite, real sequences of length N^o.

$$N^o = N + \sum_{j=1}^{J} \mid ZW_jf \mid .$$

The zero-crossings representation becomes:

$$R_zf = \left\{\{ZW_jf\}_{j=1}^J, T_{zf}f\right\}. \tag{2.3.6}$$

The above form is helpful in the study of the structure of the reconstruction set. Note, that in order to have $h \in \Gamma(R_zf)$, in addition to obeying

$$T_{zf}h = T_{zf}f,$$

W_jh has to satisfy sign constraints yielding zero-crossings exactly at ZW_jf points. For the purpose of stating precisely the latter constraint, the set $(ZW_jf)^r$ is defined.

$$(ZW_jf)^r \triangleq \{k \in ZW_jf : \left(U_j^{zf}f\right)(k) \neq 0\}. \tag{2.3.7}$$

Observe that $(ZW_jf)^r$ consists of "proper" zero-crossing points, namely only points k for which $W_jf(k) \neq 0$ are taken into account.

Theorem 2 *Let $R_z f$ be a given multiscale zero-crossings representation. $h \in \Gamma(R_z f)$ if and only if*

$$T_{zf} h = T_{zf} f \tag{2.3.8}$$

$$sgn\left(\left(U_j^{zf} f\right)(k)\right) \cdot W_j h(i) > 0. \tag{2.3.9}$$

The last inequality should be satisfied for $j = 1, 2, \ldots, J$ and for all $i \in (ZW_j f)^c \cup (ZW_j f)^r$ where k satisfies $i \in P_j^{zf}(k)$.

Proof: By a straightforward application of the above definitions.

□

As an immediate consequence of Theorem 2 we have:

Proposition 7 *The multiscale zero-crossings representation is an AQLR.*

The following characteristic of the zero-crossings representation is achieved due to the additional information, proposed by S. Mallat, of the sums of elements $W_j f(k)$ calculated between consecutive zero-crossings.

Theorem 3 *The multiscale zero-crossings representation is a bounding representation.*

Proof: Let $h \in \Gamma(R_z f)$. Tf will be an abbreviated notation for $T_{zf} f$. We need to find a constant $K > 0$ such that $\|W_j h\| \leq K\|Tf\|$ and $\|S_J h\| \leq K\|Tf\|$.

Let j and $k \in ZW_j f$ be arbitrary and fixed. It follows from the sampling information constraint (2.3.8) that:

$$\sum_{l \in P_j^{zf}(k)} W_j h(l) = \sum_{l \in P_j^{zf}(k)} W_j f(l) = \left(U_j^{zf} f\right)(k). \tag{2.3.10}$$

Since $W_j h$ has the same zero-crossing points as $W_j f$, for all $l \in P_j^{zf}(k)$ the values of $W_j h(l)$ have the same, fixed sign. Therefore

$$\sum_{l \in P_j^{zf}(k)} \mid W_j h(l) \mid = \left|\left(U_j^{zf} f\right)(k)\right|. \tag{2.3.11}$$

Applying

$$\sum x_i^2 \leq \sum x_i^2 + 2 \cdot \sum x_i x_j = (\sum x_i)^2$$

for nonnegative x_i's, we obtain:

$$\sum_{l \in P_j^{zf}(k)} \mid W_j h(l) \mid^2 \leq \left|\left(U_j^{zf} f\right)(k)\right|^2. \tag{2.3.12}$$

Now consider:

$$\|W_j h\|^2 = \sum_{k \in ZW_j f} \sum_{l \in P_j^{zf}(k)} \mid W_j h(l) \mid^2 \leq$$

$$\leq \sum_{k \in ZW_j f} \left|\left(U_j^{zf} f\right)(k)\right|^2 \leq \|Tf\|^2.$$

In addition

$$\|S_J h\| = \|S_J f\| \leq \|Tf\|$$

because $S_J h = S_J f$ and $S_J f$ is included in Tf. Using Proposition 5 we verify that:

Proposition 8 *The wavelet zero-crossings representation is an inherently bounded AQLR.*

CHAPTER
THREE

THE GENERALIZATION: AQLR's AND THEIR PROPERTIES

After two important examples of inherently bounded AQLR's, the wavelet maxima representation and the wavelet zero-crossings representation, have been described, several basic properties of AQLR's are presented. The following results are introduced: uniqueness characterization, description of the reconstruction set by its vertices, and bounds on the reconstruction set. The first result is valid for any AQLR, while the remaining two are valid only for inherently bounded AQLR. They are based on convex analysis and parametric linear programming. There are many relevant sources for the subject; we have mostly used [42, 19]. The primary objective of this section is to establish foundations for the subsequent discussion about uniqueness, stability and reconstruction.

3.1 Uniqueness Characterization

A representation $Rf = \{Vf, Tf\}$ is said to be unique, if the reconstruction set $\Gamma(Rf)$ consists of exactly one element. We have the following uniqueness characterization for AQLR's.

Lemma 3 *Let $Rf = \{Vf, Tf\}$ be an AQLR. Then Rf is unique if and only if the kernel of the operator T is trivial, i.e. $\mathcal{N}T = \{0\}$.*

Proof: The lemma becomes obvious by topological arguments. Nevertheless, an elementary but constructive proof will be given. Initially, let us assume that the representation is not unique. Then there exists $h \neq f$ such that $Rh = Rf$. In particular, $Th = Tf$, but then $0 \neq h - f \in \mathcal{N}T$.

Next, consider the case where the kernel of T, $\mathcal{N}T$ is not trivial. Let $h \neq 0$ be such that $Th = 0$. Suppose $\alpha > 0$ and consider $f_\alpha \triangleq \alpha h + f$, as a candidate to belong to $\Gamma(Rf)$. Of course $Tf_\alpha = Tf$, therefore $f_\alpha \in \Gamma(Rf)$ if and only if $\mathbf{C}f_\alpha > \mathbf{c}$ (see Definition 2.2.1). The latter is equivalent to:

$$\alpha \cdot \mathbf{C}h > \mathbf{c} - \mathbf{C}f. \tag{3.1.1}$$

Let $(\mathbf{c}-\mathbf{C}f)_i$ be the i-th component on the vector $\mathbf{c}-\mathbf{C}f$. Observe that $(\mathbf{c}-\mathbf{C}f)_i$ is negative for all i. Define:

$$\alpha_0 \triangleq min\left\{\frac{(\mathbf{c}-\mathbf{C}f)_i}{(\mathbf{C}h)_i} : (\mathbf{C}h)_i < 0\right\}. \tag{3.1.2}$$

Note that $\alpha_0 > 0$. It is easy to show that for all α such that $0 \leq \alpha < \alpha_0$:

$$\mathbf{C}f_\alpha > \mathbf{c} \tag{3.1.3}$$

Consequently, the representation Rf is not unique.

□

This claim has some significant consequences. Using the above lemma, an algorithm which tests for uniqueness can be developed. One option is to derive

it from a rank test of the operator T. Another, more ambitious, approach is to characterize, for a particular application, all sets Vf giving rise to a unique representation. Perhaps the most important consequence of Lemma 3 is the fact that uniqueness of the representation Rf is equivalent to uniqueness of the underlying irregular sampling Tf. In other words, in the unique case, all the information about the signal is already contained in Tf. Additional constraints $\mathbf{C}f > \mathbf{c}$ are redundant. On the other hand, from the signal compression, understanding and interpretation point of view, it seems to be desirable that little information would be specified explicitly by Tf and as much as possible information about a signal should be described implicitly by $\mathbf{C}f > \mathbf{c}$. Therefore, in our opinion, the most important and interesting features of AQLR's appear in the nonunique case.

3.2 The structure of the reconstruction set

At this point, the structure of the reconstruction set is described. Let $Rf = \{Vf, Tf\}$ be an AQLR, its reconstruction set is given as: [1]

$$\Gamma = \{x : Tx = Tf, \ \mathbf{C}x > \mathbf{c}\}. \tag{3.2.1}$$

The closure of the reconstruction set, $\overline{\Gamma}$ is the following convex polyhedron.

$$\overline{\Gamma} = \{x : Tx = Tf, \ \mathbf{C}x \geq \mathbf{c}\}. \tag{3.2.2}$$

[1]The abbreviated notation Γ is used instead of $\Gamma(Rf)$.

Since every equality of the form $x_i = t_i$ can be replaced by two inequalities $x_i \geq t_i$, $-x_i \geq -t_i$, without loss of generality, we can assume that

$$\overline{\Gamma} = \{x : \mathbf{B}x \geq \mathbf{b}\}$$

for a given $p \times N$ matrix $\mathbf{B}$ and a p-dimensional vector $\mathbf{b}$.

For an inherently bounded AQLR's, the associated set Γ^c is bounded. Therefore as a special case of the theorem of Krein and Milman [26], the following holds.

Theorem 4 *For an inherently bounded AQLR, the closure of the reconstruction set is the convex hull of its finitely many vertices.*

In the sequel, the following property of a polyhedron vertex will be used. Let $\{x : \mathbf{B}x \geq \mathbf{b}\}$ be a polyhedron and v^i its vertex. Then, there exist N rows of $\mathbf{B}$, which constitute a regular matrix $[\mathbf{B}]^i$ such that:

$$v^i = \left([\mathbf{B}]^i\right)^{-1} \cdot [\mathbf{b}]^i \tag{3.2.3}$$

where $[\mathbf{b}]^i$ is a subvector of $\mathbf{b}$ corresponding to these N rows. By inserting zero columns to the matrix $([\mathbf{B}]^i)^{-1}$, the matrix D^i is obtained, such that:

$$v^i = D^i\mathbf{b}. \tag{3.2.4}$$

Since the closure of the reconstruction set is the convex hull of its vertices, the above equation can characterize the changes in the reconstruction set due to perturbations in either the matrix $\mathbf{B}$ or the vector $\mathbf{b}$. Accordingly, it will be used to prove the stability results.

3.3 Bounds on the reconstruction set

The last part of this section addresses the problem of finding bounds for the set $\Gamma^c = \{x : \mathbf{B}x \geq \mathbf{b}\}$. Especially, we will focus on the bound in which the dependency on the matrix $\mathbf{B}$ and on the vector $\mathbf{b}$ will appear in different factors.

Consider the following characterization of a bounded polyhedron.

Theorem 5 ([42] pp 65)

The polyhedron $\overline{\Gamma} = \{x : \mathbf{B}x \geq \mathbf{b}\}$ *is bounded if and only if it contains no halfline. If* $\overline{\Gamma} \neq \emptyset$ *the latter statement holds if and only if the associated homogeneous system of inequalities:*

$$\mathbf{B}x \geq 0 \tag{3.3.1}$$

admits no nonzero solution.

Notice, that the homogeneous system of inequalities is independent of $\mathbf{b}$. Therefore, if there exists one $\mathbf{b}_0$ yielding a bounded polyhedron, then $\{x : \mathbf{B}x \geq \mathbf{b}\}$ is bounded for all $\mathbf{b}$. Let us assume that the matrix $\mathbf{B}$ is fixed and arbitrary, but there exists $\mathbf{b}_0$ such that the set $\{x : \mathbf{B}x \geq \mathbf{b}_0\}$ is nonempty and bounded. Then, from the second statement of the theorem, for all $x \neq 0$ there exists at least one index i such that $(\mathbf{B}x)_i < 0$. Let us define

$$\mathcal{I}(x) \triangleq \{i : (\mathbf{B}x)_i < 0\}. \tag{3.3.2}$$

Observe, that the following function is well defined for all $x \neq 0$.

$$\lambda_m(x) \triangleq \min\left\{\frac{\mathbf{b}_i}{(\mathbf{B}x)_i} : \quad i \in \mathcal{I}(x)\right\}. \tag{3.3.3}$$

Next, consider Γ_U, the projection of $\overline{\Gamma}$ on the unit ball:

$$\Gamma_U \triangleq \{\hat{x} : \|\hat{x}\| = 1, \text{ and } \exists \lambda(\hat{x}) > 0 \text{ such that } \lambda(\hat{x}) \cdot \hat{x} \in \overline{\Gamma}\}. \tag{3.3.4}$$

Proposition 9 *$\lambda_m(\hat{x})$ is a positive, continuous function for all $\hat{x} \in \Gamma_U$.*

Proof: If $\hat{x} \in \Gamma_U$ then $\lambda(\hat{x})(\mathbf{B}x)_i \geq \mathbf{b}_i$ $(i = 1, 2, \ldots, p)$. Since for $i \in \mathcal{I}(\hat{x})$, $(\mathbf{B}\hat{x})_i$ is negative, therefore, in this case, $\mathbf{b}_i$ has to be negative and

$$\frac{\mathbf{b}_i}{(\mathbf{B}\hat{x})_i} > 0$$

for all $i \in \mathcal{I}(\hat{x})$. Thus, in particular, $\lambda_m(\hat{x}) > 0$ for all $\hat{x} \in \Gamma_U$. Consequently:

$$\beta_m(x) \triangleq \frac{1}{\lambda_m(x)} = \max\left\{\frac{(\mathbf{B}x)_i}{\mathbf{b}_i} : i \in \mathcal{I}(x)\right\}. \tag{3.3.5}$$

To proceed we need to get rid of the set $\mathcal{I}(x)$ inside the above maximum. Let us define :

$$\mathcal{I}_b \triangleq \{i : \mathbf{b}_i < 0\}. \tag{3.3.6}$$

Notice, that $\mathcal{I}(\hat{x}) \subseteq \mathcal{I}_b$ for all $\hat{x} \in \Gamma_U$. Thus, $\beta_m(\hat{x})$ can be written as:

$$\beta_m(\hat{x}) = max\left\{max\left\{\frac{(\mathbf{B}x)_i}{\mathbf{b}_i}, 0\right\} : i \in \mathcal{I}_b\right\}. \tag{3.3.7}$$

From the above formula, it is clear that $\beta_m(\hat{x})$ is a continuous function for all $\hat{x} \in \Gamma_U$. Therefore, since $\beta_m(\hat{x}) > 0$ for all $\hat{x} \in \Gamma_U$,

$$\lambda_m(\hat{x}) = \frac{1}{\beta_m(\hat{x})}$$

is a continuous function for all $\hat{x} \in \Gamma_U$ as well.

□

Because Γ_U is a compact set, the following maximum is well defined.

$$\lambda_m \triangleq \max\{\lambda_m(\widehat{x}) : \widehat{x} \in \Gamma_U\}. \tag{3.3.8}$$

In view of the above considerations, it is easy to show that λ_m is a tight bound on $\overline{\Gamma}$, namely:

$$\forall x \in \overline{\Gamma} \quad \|x\| \leq \lambda_m \tag{3.3.9}$$

$$\exists x \in \overline{\Gamma} \text{ such that } \|x\| = \lambda_m. \tag{3.3.10}$$

The bound λ_m is clearly the best possible. However, it has two important disadvantages. The first is the need to know Γ_U, although it can be determined independently of calculating $\overline{\Gamma}$, it may involve complex computations. The second disadvantage is that the effects on the bound of $\mathbf{B}$ and $\mathbf{b}$ are not separated. In what follows, a less accurate bound, but without the above drawbacks, will be calculated. Consider:

$$\lambda_m(x) = \min\left\{\frac{\mathbf{b}_i}{(\mathbf{B}x)_i} : i \in \mathcal{I}(x)\right\} \leq$$

$$\leq \max\{|\ \mathbf{b}_i\ |: i = 1, 2, \ldots, p\} \cdot \min\{\left\{\frac{-1}{(\mathbf{B}x)_i} : i \in \mathcal{I}(x)\right\} \leq$$

$$\leq \|\mathbf{b}\| \cdot \min\{\left\{\frac{-1}{(\mathbf{B}x)_i} : i \in \mathcal{I}(x)\right\}.$$

Let us define

$$\lambda_o(x) \triangleq \min\{\left\{\frac{-1}{(\mathbf{B}x)_i} : i \in \mathcal{I}(x)\right\}. \tag{3.3.11}$$

Proposition 10 *$\lambda_o(x)$ is a positive and continuous function for all $x \neq 0$.*

Proof: As a consequence of the definition of $\mathcal{I}(x)$, $\lambda_o(x)$ is positive for all $x \neq 0$. To show continuity, let us consider:

$$\beta_o(x) \triangleq \frac{1}{\lambda_o(x)} = \max\{-(\mathbf{B}x)_i : i \in \mathcal{I}(x)\}. \tag{3.3.12}$$

Which can also be written as:

$$\beta_o(x) = \max\left\{\max\left\{-(\mathbf{B}x)_i, 0\right\} : i = 1, 2, \ldots, p\right\}. \tag{3.3.13}$$

It is apparent from (3.3.13) that $\beta_o(x)$ is a continuous function of x, and thus $\lambda_o(x)$ is continuous as well.

□

From the latter form one can see that $\beta_o(x)$ depends on $\mathbf{B}$ and x but it is b-independent. Consider any compact set $\overline{U}$ containing nonzero elements such that $\Gamma_U \subseteq \overline{U}$. Then for all $\widehat{x} \in \Gamma_U$:

$$\lambda_m(\widehat{x}) \leq \|b\| \cdot \frac{1}{\beta_o(\widehat{x})} \leq$$

$$\leq \|b\| \cdot \max\left\{\frac{1}{\beta_o(x)} : x \in \overline{U}\right\}. \tag{3.3.14}$$

For example, the unit ball $U = \{x : \|x\| = 1\}$ is used as a set $\overline{U}$. Then, using only the matrix $\mathbf{B}$, the coefficient β_U is calculated as:

$$\beta_U \triangleq \min\left\{\max\left\{\max\{-(\mathbf{B}x)_i, 0\} : i = 1, 2, \ldots, p\right\} : x \in U\right\}. \tag{3.3.15}$$

Combining together (3.3.14), (3.3.9), (3.3.8) the following result is obtained.

$$\|x\| \leq \frac{\|\mathbf{b}\|}{\beta_U} \quad \forall x \in \overline{\Gamma}. \tag{3.3.16}$$

The above bound will be used to prove the convergence of the reconstruction algorithm.

CHAPTER
FOUR

UNIQUENESS

The main result established in this chapter is that, in general, the wavelet maxima (zero-crossings) representation is not unique. In addition, a uniqueness test for a particular representation is discussed. This chapter concludes with several examples of both unique and nonunique signal representations. Results described here are based on some specific properties of the discrete wavelet transform which, together with an associated notation, are introduced in the first section.

4.1 Description by Discrete Fourier Transform

As was mentioned in Section 1.3.5, discrete wavelet transform is based on two discrete filters H and G. Using equations (1.3.31,1.3.32), the discrete wavelet transform can be calculated by the following recursion ($j = 1, 2, \ldots, J-1$):

$$S_{j+1}f = h_j * S_j f \tag{4.1.1}$$

$$W_{j+1}f = g_j * S_j f. \tag{4.1.2}$$

Since all described signals are discrete, the superscript d used in Section 1.3.5 is omitted here. $S_0 f$ is equal to the original signal f. In addition, instead of writing subscripts 2^j, related to discrete sampling of the continuous wavelet transform at scale 2^j, we simply use the subscript j.

Further clarifications are needed due to our assumption regarding finite sequence length. In this case, a standard approach is to define the convolution operation as N-circular or equivalently to consider all signals as N-periodic (for details see e.g. [41]). Consequently, the impulse responses of the filters H and G are denoted by $\{h(k)\}_{k=0}^{N-1}$ and $\{g(k)\}_{k=0}^{N-1}$, respectively. The Discrete Fourier Transform (DFT), $\widehat{f} = \{\widehat{f}(k)\}_{k=0}^{N-1}$ of a sequence $f = \{f(k)\}_{k=0}^{N-1}$ is defined as:

$$\widehat{f}(k) = \sum_{n=0}^{N-1} f(n) exp(-2\pi i \frac{nk}{N}) \quad k = 0, 1, \ldots, N-1. \tag{4.1.3}$$

From the definition we see that the DFT's of impulse responses of discrete filters H, G are given as:

$$\widehat{h}(k) = H\left(2\pi \frac{k}{N}\right) \quad \text{and} \quad \widehat{g}(k) = G\left(2\pi \frac{k}{N}\right) \tag{4.1.4}$$

i.e. the DFT is equivalent to sampling the continuous Fourier transform at $\omega_k = 2\pi \frac{k}{N}$.

Since convolution corresponds to multiplication of DFT's, equations (4.1.1) and (4.1.2) can be rewritten as:

$$\widehat{(S_{j+1} f)}(k) = \widehat{(S_j f)}(k) \cdot \widehat{h_j}(k) \quad k = 0, 1, \ldots, N-1 \tag{4.1.5}$$

$$\widehat{(W_{j+1} f)}(k) = \widehat{(S_j f)}(k) \cdot \widehat{g_j}(k) \quad k = 0, 1, \ldots, N-1 \tag{4.1.6}$$

where $\widehat{h}_j$, $\widehat{g}_j$ are DFT's of impulse responses of filters H_j, G_j respectively. They are calculated from from $H(\omega)$ and $G(\omega)$ in the following way:

$$\widehat{h_j}(k) = H\left(2^j 2\pi \frac{k}{N}\right) \tag{4.1.7}$$

$$\widehat{g_j}(k) = G\left(2^j 2\pi \frac{k}{N}\right). \tag{4.1.8}$$

By cascading equations (4.1.5) and (4.1.6), we can obtain:

$$(\widehat{S_J f})(k) = \hat{f}(k) \cdot \prod_{p=0}^{J-1} \hat{h}_p(k) \tag{4.1.9}$$

$$(\widehat{W_j f})(k) = \hat{f}(k) \cdot \hat{g}_{j-1} \cdot \prod_{p=0}^{j-2} \hat{h}_p(k). \tag{4.1.10}$$

The latter is valid for $j = 1, 2, \ldots, J$ with the standard convention that product over an empty set of indices is equal to 1. Using equations (4.1.9) and (4.1.10), we can define the discrete transfer functions $\widehat{S_J}$ and $\widehat{W_j}$ of the linear operators S_J and W_j, respectively, as

$$(\widehat{S_J})(k) = \prod_{p=0}^{J-1} \hat{h}_p(k) \tag{4.1.11}$$

$$(\widehat{W_j})(k) = \hat{g}_{j-1}(k) \cdot \prod_{p=0}^{j-2} \hat{h}_p(k). \tag{4.1.12}$$

4.2 The nonuniqueness theorem

This section aims to show that, in general, the discrete dyadic wavelet maxima (zero-crossings) representation is not unique. The precise statement of the nonuniqueness theorem is as follows.

Theorem 6 *Consider a discrete dyadic wavelet maxima (zero-crossings) representation based on a discrete low pass filter $H(\omega)$. If $H(\pi) = 0$, $J \geq 3$, and N is a multiple of 2^J then there exists a sequence f which has a nonunique maxima (zero-crossings) representation.*

Let us point out that, although the hypothesis of the theorem may seem to be demanding, it is just a technical condition. Usually the number of levels, J, satisfies $J \geq 3$. In order to benefit from the fast wavelet transform, N has to be a multiple of 2^J. Since $H(\omega)$ is a low pass filter, it is natural to assume that $|H(\omega)|$ reaches its minimum at π. If this minimum is nonzero, then essentially $S_J f$ contains all information about f and the maxima (zero-crossings) information is redundant. Moreover $H(\pi) = 0$ is a well known condition for the regularity of the underlying scaling function $\Phi(\xi)$ (for more information, see Section 1.1.3). Indeed, all filters used by Mallat, Zhong and many others fulfill the conditions of Theorem refth:t1.

Most of the section describes the proof of the theorem, which will be divided to proofs of several propositions. The result is a consequence of Lemma 3, which relates uniqueness of the representation to the set $\mathcal{N}T$, the kernel of the sampling information. The main idea is to construct a sequence f such that the set $\mathcal{N}T$ corresponding to the representation Rf cannot be $\{0\}$. The construction of the counter example is based on the set $\mathcal{B}$, defined as follows:

$$\mathcal{B} = \{y_p\}_{p=1}^{2^J-1} \tag{4.2.1}$$

where

$$y_{2p-1}(k) = cos(\frac{2\pi pk}{2^J}) \quad p = 1, 2, \ldots, 2^{J-1} \tag{4.2.2}$$

$$y_{2p}(k) = sin(\frac{2\pi pk}{2^J}) \quad p = 1, 2, \ldots, 2^{J-1} - 1. \tag{4.2.3}$$

Proposition 11 *The set $\mathcal{B}$ is included in $\mathcal{N}S_J$, the kernel of the operator S_J.*

Proof: Let us consider:

$$m(p) = \frac{pN}{2^J} \quad p = \pm 1, \pm 2, \ldots, \pm 2^{J-1}. \tag{4.2.4}$$

Since N is a multiple of 2^J, $m(p)$ is an integer. Notice that p can be written as $p = 2^l p_1$ where $0 \leq l \leq J-1$ and p_1 is an odd number. Observe that

$$\widehat{h}_{J-1-l}(m(p)) = H\left(2^{J-1-l} 2\pi \frac{2^l p_1 N}{N 2^J}\right) = H(\pi p_1) = 0. \tag{4.2.5}$$

Therefore, using (4.1.11), we obtain:

$$\widehat{S_J}\,(m(p)) = 0. \tag{4.2.6}$$

The integers $m(p)$'s, as zeros of the transfer function $\widehat{S_J}$, will be used to define sequences belonging to the null space of S_J. Let e_p be the following exponential sequence

$$e_p(k) = exp\left(2\pi i \frac{pk}{N}\right) \quad k = 0, 1, \ldots, N-1. \tag{4.2.7}$$

Its Discrete Fourier Transform, $\widehat{e_p}$ is given by:

$$\widehat{e_p}(k) = N\delta_p(k) \tag{4.2.8}$$

where

$$\delta_p(k) = \begin{cases} 1 & \text{if } k = p \\ 0 & \text{otherwise.} \end{cases}$$

Combining together (4.2.6) and (4.2.8), one can conclude that $(\widehat{S_J e_{m(p)}}) = 0$, thus

$$S_J e_{m(p)} = 0. \tag{4.2.9}$$

The sequences y_p are expressed by $e_{m(p)}$'s in the subsequent way:

$$y_{2p-1}(k) = \cos(\frac{2\pi pk}{2^J}) = \frac{1}{2}(e_{m(p)} + e_{m(-p)})$$

$$y_{2p}(k) = \sin(\frac{2\pi pk}{2^J}) = \frac{1}{2i}(e_{m(p)} - e_{m(-p)}).$$

Therefore $S_J y_p = 0$ for $p = 1, 2, \ldots, 2^J - 1$.

□

Notice, that y_{2^J} does not appear in the set $\mathcal{B}$. The reason is that $y_{2^J} \equiv 0$ and in the next proposition the independence of the set $\mathcal{B}$ is asserted.

Proposition 12 *The set $\mathcal{B}$ is linearly independent.*

Proof: It is a well known fact, which can easily be proven by showing that the set $\mathcal{B}$ is orthogonal and does not contain zero.

□

As a generic example of nonuniqueness the following sequence is proposed.

$$f(k) = \cos(2\pi \frac{k}{2^J}) \quad k = 0, 1, \ldots, N-1. \tag{4.2.10}$$

Observe that the same sequence is proposed for all dyadic wavelet transforms and for both the maxima representation and the zero-crossings representation.

The representation $R_m f$ $(R_z f)$ is unique if and only if $\mathcal{N}T_{mf} = \{0\}$ $(\mathcal{N}T_{zf} = \{0\})$. Consequently, the nonuniqueness of $R_m f$ $(R_m f)$ is easily deduced from the following proposition.

Proposition 13 *The equation*

$$T_{mf}\eta = 0 \quad (T_{zf}\eta = 0) \qquad \eta \in span(\mathcal{B}) \tag{4.2.11}$$

has a nontrivial solution.

Proof: Consider an arbitrary $\eta \in span(\mathcal{B})$.

$$\eta = \sum_{p=1}^{2^J-1} \alpha_p y_p \tag{4.2.12}$$

The dimension of $span(\mathcal{B})$ is $2^J - 1$. The idea is to show that the set of equations $T_{mf}\eta = 0$ $(T_{zf}\eta = 0)$ yields less than $2^J - 1$ independent equations with unknowns $\{\alpha_p\}$. Recall that:

$$T_{mf}\eta = \{S_J\eta, \{W_1\eta(k)\}_{k \in XW_1 f \cup YW_1 f}, \ldots, \{W_J\eta(k)\}_{k \in XW_J f \cup YW_J f}\}.$$

and

$$T_{zf}\eta = \{S_J\eta, U_1^{zf}\eta, \ldots, U_J^{zf}\eta\}.$$

From Proposition 11 we see that $S_J\eta = 0$ for all $\{\alpha_p\}$. Let j be fixed. Consider the equations:

$$W_j\eta(k) = 0 \qquad k \in XW_j f \cup YW_j f \tag{4.2.13}$$

or

$$U_j^{zf}\eta(k) = 0 \quad k \in ZW_jf. \tag{4.2.14}$$

Recognize that f is a 2^J-periodic sinusoid, therefore W_jf is a 2^J-periodic sinusoid as well. Moreover y_p and W_jy_p for $p = 1, 2, \ldots, 2^J - 1$ are also 2^J-periodic. Therefore solving equation (4.2.13), it suffices to consider

$$k \in \{0, 1, \ldots, 2^J - 1\} \bigcap (XW_jf \cup YW_jf).$$

But W_jf has only two local extreme points in a 2^J period, consequently (4.2.13) contains only two different equations with unknowns $\{\alpha_p\}$! Similarly for zero-crossings, (4.2.14) has only two different equations. There are J levels ($j = 1, 2, \ldots, J$) so the set of equations $T_{mf}\eta = 0$ ($T_{zf}\eta = 0$) consists of, at most, $2J$ independent equations. But

$$2^J - 1 > 2J \quad \forall J \geq 3. \tag{4.2.15}$$

Accordingly, the equation (4.2.11) has a nontrivial solution and the representation is not unique.

□

Some remarks need to be made at this point. From the proof, it turns out, that it is relatively easy to produce more examples of nonunique dyadic wavelet maxima (zero-crossings) representations using 2^p-periodic signals, where p is an integer. For example, consider $J = 5$ and let f be a 2^5-periodic signal. Then W_jf ($j = 1, 2, \ldots, 5$) are 2^5-periodic as well. In this case, if $2^J - 1 = 31$ is greater

than the total number of local extrema (zero-crossings) of $W_1 f, W_2 f, \ldots, W_5 f$ per one 2^J period, then the representation is not unique. In other words, if $W_j f$'s have, in the mean, less than $\frac{31}{5} = 6.2$ local extrema (zero-crossings) in one period, than $R_m f$ ($R_z f$) cannot be unique.

Hummel with Moniot [22], Mallat [29], and Mallat with Zhong [31] have reported that high frequency errors may occur in the discrete maxima (zero-crossings) representation. For these 2^J-periodic signals, components of the reconstruction error can appear as 2^p-periodic signals for $p = 1, 2, \ldots, J$. Most of them cannot be related as high frequency errors.

From our simulations and from Mallat's results it turns out that for the vast majority of signals, the representation is unique. We even conjecture that the wavelet maxima (zero-crossings) representation is unique for a generic family of signals, but we are not able to prove it.

4.3 Test for uniqueness

From the previous section we have learned that uniqueness is signal dependent. The next natural question to ask is: What are the characteristics of a family of signals having a unique representation ? This problem appears to be difficult and, unfortunately, we are not yet able to answer this question. Nevertheless, a given representation can be tested, quite efficiently, whether it is unique or it is not.

Let $R_m f$, the wavelet maxima representation of a signal f, be given. Recall

that:

$$R_m f = \left\{ \{XW_j f, YW_j f\}_{j=1}^{J}, T_{mf} f \right\}. \tag{4.3.1}$$

From Lemma 3, we already know that uniqueness depends only on the linear operator T_{mf} which, in general, is signal dependent. T_{mf} can be divided into two parts, the first is the operator S_J, which is signal independent, and the second will be denoted by T^w. For all $\eta \in \mathcal{L}$, $T^w \eta$ is defined as:

$$T^w \eta \triangleq \left\{ \{W_1 \eta(k)\}_{k \in XW_1 f \cup YW_1 f}, \ldots, \{W_J \eta(k)\}_{k \in XW_J f \cup YW_J f} \right\}. \tag{4.3.2}$$

Consequently we can write:

$$T_{mf} \eta = \{S_J \eta, T^w \eta\} \tag{4.3.3}$$

for all $\eta \in \mathcal{L}$.

Since our approach is based on Lemma 3, the null space of T_{mf} is investigated. The definition of T^w implies:

$$\mathcal{N}T_{mf} = \mathcal{N}S_J \cap \mathcal{N}T^w \tag{4.3.4}$$

where $\mathcal{N}T_{mf}, \mathcal{N}S_J, \mathcal{N}T^w$ are the null spaces of operators T_{mf}, S_J, T_w, respectively. Let $\mathcal{B} = \{y_p\}_{p=1}^{P}$ be a basis of $\mathcal{N}S_J$. As a consequence of the fact that $\mathcal{N}T_{mf} \subseteq \mathcal{N}S_J$, every $\eta \in \mathcal{N}T_{mf}$ can be written as:

$$\eta = \sum_{p=1}^{P} \alpha_p y_p \tag{4.3.5}$$

Using this representation, we can say that:

$$\begin{aligned} \eta \in \mathcal{N}T_{mf} &\Leftrightarrow T^w \left(\textstyle\sum_{p=1}^{P} \alpha_p y_p \right) = 0 \\ &\Leftrightarrow \textstyle\sum_{p=1}^{P} \alpha_p T^w y_p = 0 \end{aligned} \tag{4.3.6}$$

In other words, $\mathcal{N}T_{mf} = \{0\}$ if and only if zero is the only vector $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_P)'$ solving $\sum_{p=1}^{P} \alpha_p T^w y_p = 0$.

In essence, the remaining steps are to describe the set $\mathcal{B}$, the basis of $\mathcal{N}S_J$, and to give a matrix interpretation to equation (4.3.6).

First, let us consider closely a particular wavelet transform on which the maxima representation is based. In order to be able to compare results, we use the same wavelets as [31]. Two cases were described in [31]. One corresponds to the cubic spline wavelet with:

$$H(\omega) = \left(cos\left(\frac{\omega}{2}\right)\right)^4$$

and the second corresponds to a Haar wavelet with:

$$H(\omega) = exp(-i\frac{\omega}{2})cos(\frac{\omega}{2}).$$

$G(\omega)$ is chosen to satisfy:

$$\mid G(\omega) \mid^2 + \mid H(\omega) \mid^2 = 1.$$

For the cubic spline wavelet, following [31], the transfer function $G(\omega)$ was chosen as:

$$G(\omega) = \left(1 - \left(cos\left(\frac{\omega}{2}\right)\right)^8\right)^{\frac{1}{2}} exp(-\frac{i\omega}{2} - \frac{\pi}{2}).$$

Sampling $H(\omega)$, $G(\omega)$ at $\omega_k = \frac{2\pi k}{N}$ for $k = 0, 1, \ldots, N-1$ gives the Discrete Fourier Transforms (DFT's) $\widehat{h}$ and $\widehat{g}$:

$$\widehat{h}(k) = \left(cos\left(\frac{\pi k}{N}\right)\right)^4 \qquad k = 0, 1, \ldots, N-1 \tag{4.3.7}$$

$$\hat{g}(k) = \left(1 - \left(cos\left(\frac{\pi k}{N}\right)\right)^8\right)^{\frac{1}{2}} exp(-\frac{i\pi k}{N} - \frac{\pi}{2}) \quad k = 0, 1, \ldots, N-1 \quad (4.3.8)$$

Eventually sequences h and g are obtained by the inverse DFT.

$$h(k) = \frac{1}{N}\sum_{n=0}^{N-1} \hat{h}(n) exp(\frac{i\pi kn}{N}) \quad (4.3.9)$$

$$g(k) = \frac{1}{N}\sum_{n=0}^{N-1} \hat{g}(n) exp(\frac{i\pi kn}{N}) \quad (4.3.10)$$

For N=256, the following numerical results were obtained:

$$\begin{aligned} h(0) &= 0.375 & g(1) = -g(0) &= 0.5907 \\ h(1) = h(-1) &= 0.250 & g(2) = -g(-1) &= 0.1107 \\ h(2) = h(-2) &= 0.0625 & g(3) = -g(-2) &= 0.0145 \end{aligned}$$

There are small differences with the coefficients which appear in [31]; for larger N (e.g $N = 4096$) one gets exactly the numbers given in [31].

Recall that the set $\mathcal{B}$, used in the proof of Theorem 6, is an independent set which is included in $\mathcal{N}S_J$. This set was constructed based on the fact that $H(\pi) = 0$. It turns out, that if $\omega = \pi$ is the only zero of $H(\omega)$ then the set $\mathcal{B}$ is a basis for the space $\mathcal{N}S_J$. Since for both the cubic spline wavelet and the Haar wavelet, the only zeros of $H(\omega)$ appear at $\omega = \pi$ the set $\mathcal{B}$ is a basis for $\mathcal{N}S_J$ in both cases.

Recall that $\mathcal{B}$ consists of:

$$y_{2p-1}(k) = cos(\frac{2\pi pk}{2^J}) \quad p = 1, 2, \ldots, 2^{J-1} \quad (4.3.11)$$

$$y_{2p}(k) = sin(\frac{2\pi pk}{2^J}) \quad p = 1, 2, \ldots, 2^{J-1} - 1. \quad (4.3.12)$$

A function η, which belongs to the null space of S_J can be written as:

$$\eta(k) = \sum_{p=1}^{2^J-1} \alpha_p y_p(k).$$

By the linearity properties we can write:

$$W_j\eta(k) = \sum_{p=1}^{2^J-1} \alpha_p W_j y_p(k).$$

Let us define the column vector of free coefficients:

$$\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{2^J-1})'.$$

For every $j = 1, 2, \ldots, J$ and every $k = 0, 1, \ldots, N-1$, let $\underline{\mathcal{W}}_j(k)$ denote the following row vector:

$$\underline{\mathcal{W}}_j(k) = (W_j p_1(k), W_j p_2(k), \ldots, W_j p_{2^J-1}(k)).$$

The matrix $\mathcal{W}$ is defined as consisting of rows $\underline{\mathcal{W}}_j(k)$ for $k \in XW_jf \cup YW_jf$ and $j = 1, 2, \ldots, J$. According to Lemma 3, f has a unique maxima representation if and only if the only solution for:

$$\mathcal{W} \cdot \underline{\alpha} = 0 \tag{4.3.13}$$

is the vector $\underline{\alpha} = 0$. The latter condition is equivalent to:

$$rank(\mathcal{W}) = 2^J - 1.$$

Theorem 7 *The wavelet maxima representation $R_m f$ is unique if and only if the rank of the matrix $\mathcal{W}$ is $2^J - 1$.*

Conclusion 1 *If the number of the extrema points is less than* $2^J - 1$, *the representation has to be nonunique. On the other hand, if the number of extrema points is equal or greater than* $2^J - 1$ *then uniqueness of the representation can be deduced from the rank*$(\mathcal{W})$. *In the latter case, there may be situations in which analysis of the rank*$(\mathcal{W})$ *can allow to eliminate some extrema from the representation.*

4.4 Examples

During this study, we decomposed many signals and usually only a part of extrema points sufficed to yield uniqueness. One of the reasons is that many extrema points are obtained in the regions where a decomposed signal is close to zero. Signal reconstruction from plenty of small extrema points is expected to be numerically unstable. The parametric multiscale maxima representation, introduced in Chapter 7, attempts to overcome this problem. Meantime, in examples of the basic wavelet maxima, only signals with very clear maxima points are considered.

Our first example illustrates the common case of a unique maxima representation. In addition to showing uniqueness, a reconstruction algorithm, based on the efficient solution of linear equalities, is described. The second example shows a nonunique representation. For this particular maxima representation, using Theorem 1, the exact reconstruction set is calculated. The third example is a nonunique wavelet zero-crossings representation.

4.4.1 An example of a unique maxima representation

Let us assume $N = 256$ and $J = 3$ and consider the following sequence:

$$f(k) = sin(\frac{2\pi k}{256}(3(\frac{\mid k \mid}{128})^5 + 1)).$$

This sequence was chosen in order to exhibit different frequency components without evoking too many extreme points. Figure 4.1 shows its wavelet decomposition, based on the cubic spline wavelet described in the previous section.

The extrema sets are as follows:

$$XW_1 f \cup YW_1 f = \{0, 92, 110, 121, 129, 136, 147, 165\}$$

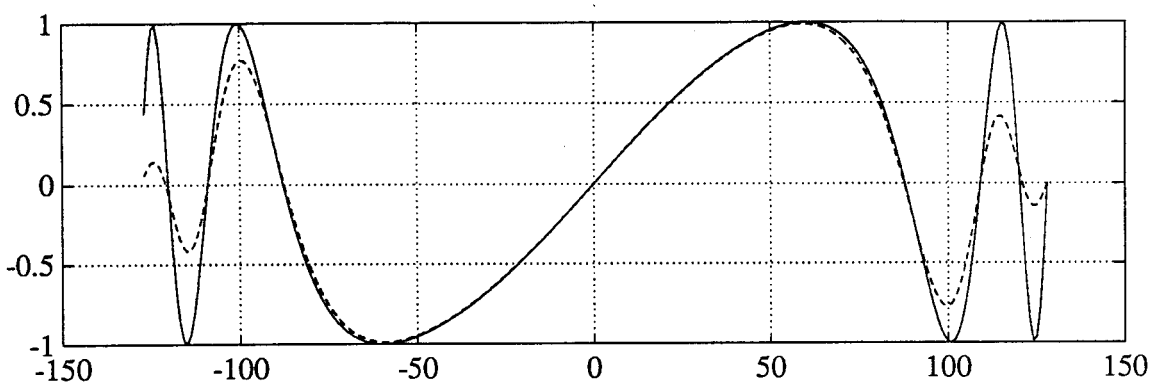
$$XW_2 f \cup YW_2 f = \{1, 92, 111, 121, 129, 137, 147, 166\}$$

$$XW_3 f \cup YW_3 f = \{2, 92, 111, 122, 130, 138, 149, 168\}$$

All together we have 24 extrema points, while the dimension of the null space of S_3 is 7. We can pick up 7 rows of the matrix $\mathcal{W}$ and check for regularity (non-singularity). We have chosen:

$$\mathcal{W}_s = \begin{pmatrix} \underline{\mathcal{W}_1}(0) \\ \underline{\mathcal{W}_1}(92) \\ \underline{\mathcal{W}_1}(110) \\ \underline{\mathcal{W}_1}(121) \\ \underline{\mathcal{W}_2}(1) \\ \underline{\mathcal{W}_2}(92) \\ \underline{\mathcal{W}_3}(2) \end{pmatrix}$$

The signals f and $S_3 f$ (dashed).



The signals $W_1 f$, $W_2 f$(dashed), $W_3 f$(dotted).

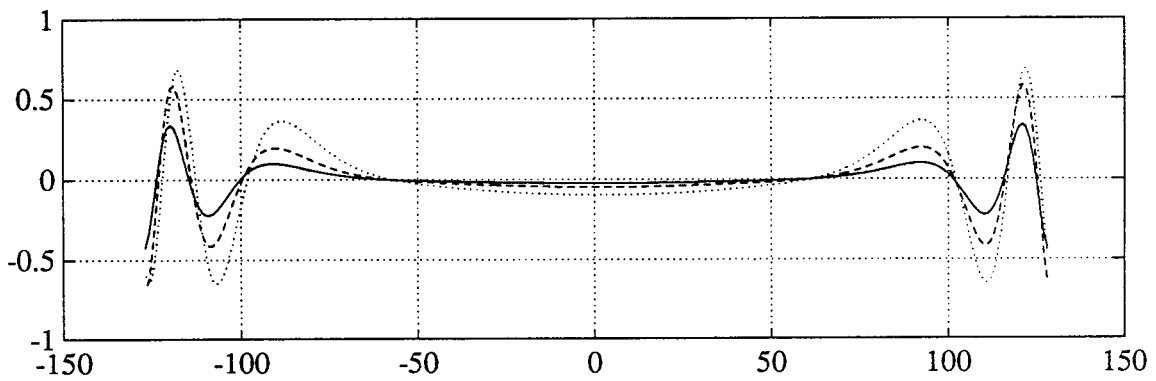


Figure 4.1: The signal f and its wavelet decomposition.

which is equal to:

$$
\begin{array}{rrrrrrr}
-0.2621 & -0.6328 & -0.6847 & -0.6847 & -0.9237 & -0.3826 & -1.0000 \\
0.2621 & 0.6328 & -0.6847 & -0.6847 & 0.9237 & 0.3826 & -1.0000 \\
-0.6328 & 0.2621 & 0.6847 & 0.6847 & 0.3826 & -0.9237 & -1.0000 \\
0.2621 & -0.6328 & 0.6847 & -0.6847 & 0.9237 & -0.3826 & 1.0000 \\
-0.0000 & -0.7054 & -0.0000 & -0.2500 & 0.0000 & -0.0208 & 0 \\
0.4988 & 0.4988 & -0.2500 & 0.0000 & 0.0147 & -0.0147 & 0 \\
0.0000 & -0.1821 & 0.0000 & 0.0000 & 0.0000 & 0.0054 & 0
\end{array}
$$

This matrix is regular, therefore we can conclude that: **The maxima representation $R_m f$ of the sequence $f(k)$, defined here, is unique.** Furthermore, essentially, one can give up 17 extrema points out of 24, and still have a unique maxima representation.

The inverse of the matrix $\mathcal{W}_s$ is equal to:

$$\begin{array}{rrrrrrr}
-1.3151 & -0.3815 & 0.5082 & -1.1883 & 9.2922 & 2.7834 & -20.2619 \\
-0.0512 & 0.0000 & 0.0000 & -0.0512 & 0.2804 & -0.0000 & -6.2206 \\
-2.5894 & -0.7229 & 1.0055 & -2.3068 & 18.1422 & 1.5071 & -50.1899 \\
0.2889 & -0.0000 & -0.0000 & 0.2889 & -5.5825 & -0.0000 & 19.6132 \\
0.5874 & 0.6496 & -0.1442 & 1.0927 & -6.7751 & -0.7899 & 20.2882 \\
-1.7392 & 0.0000 & 0.0000 & -1.7392 & 9.5260 & 0.0000 & -24.8083 \\
1.0750 & -0.0050 & -0.6884 & 1.3816 & -8.5990 & -1.0318 & 20.9344
\end{array}$$

This matrix does not have large entries, therefore, it can yield an apparently stable signal reconstruction. In this case, the reconstructed signal is calculated in the following way. Let us define:

$$\widehat{f_s}(k) = \begin{cases} \frac{\widehat{(S_J f)}(k)}{\widehat{(S_J)}(k)} & \widehat{(S_J)}(k) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.4.1}$$

The sequence f_s is calculated from $\widehat{f_s}$ by the inverse DFT.

Let $\mathcal{W}_e$ denote the sampling operator at the chosen extreme points , i.e., in our case,

$$\mathcal{W}_e\eta = (W_1\eta(0), W_1\eta(92), W_1\eta(110), W_1\eta(121), W_2\eta(1), W_2\eta(94), W_3\eta(2))'$$

Then the reconstructed signal is given by:

$$f_r = f_s + \sum_{p=1}^{7} \alpha_p y_p$$

where the vector $\underline{\alpha}$ is calculated by:

$$\underline{\alpha} = \mathcal{W}_s^{-1}(\mathcal{W}_e f - \mathcal{W}_e f_s).$$

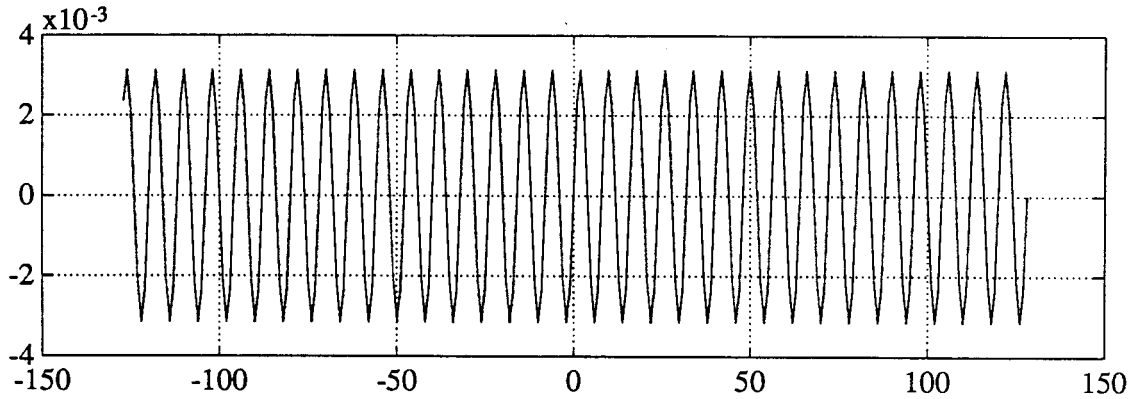
Figure 4.2 describes $f - f_s$, the error from the deconvolution from $S_3 f$ which is not large in this case. This error is harmonic, since it has to belong to the null space of S_3. The bottom part of the figure describes the error from the complete reconstruction, $f - f_r$ which is less than 10^{-7}.

One important remark is in order. The weak point in this reconstruction is the division by $(\widehat{S_J})(k)$ in the definition (4.4.1). Small values of $(\widehat{S_J})(k)$ introduce large sensitivity to numerical or approximation errors in $S_J f$. Our precise reconstruction results are due to high accuracy of the floating point number representation and due to a small number of levels. Another possible way to overcome this problem, is to reconstruct a signal using more maxima points than required by the uniqueness test.

4.4.2 An example of a nonunique maxima representation

In this section a sequence which has a nonunique maxima representation is described. As in the previous section, we assume $N = 256$, $J = 3$, and the cubic spline wavelet.

The error from the reconstruction from $S_3 f$.



The error from the complete reconstruction.

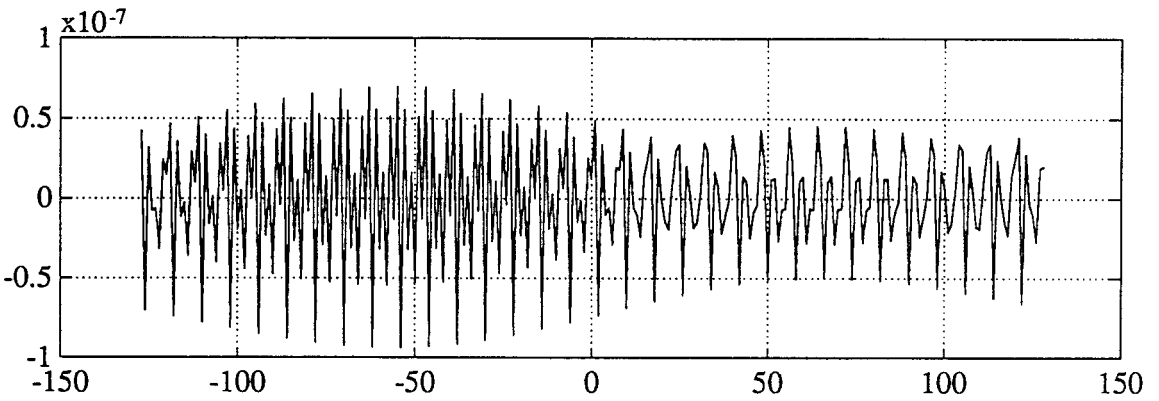


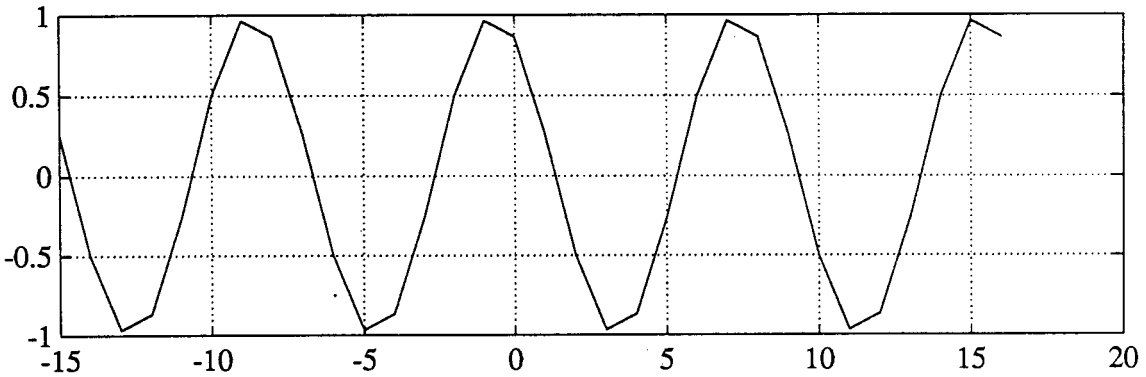
Figure 4.2: The reconstruction errors.

Consider:

$$f(k) = cos(\frac{2\pi k}{8} + \frac{\pi}{6}).$$

This sequence is from the null space of S_3. Figure 4.3 describes its wavelet decomposition($S_3 f = 0$ is omitted). Here at every level we have 64 extrema points, they appear at regular distances. The sets of the extrema are given

Signal f



Signals $W_1 f$, $W_2 f$(dashed), $W_3 f$(dotted).

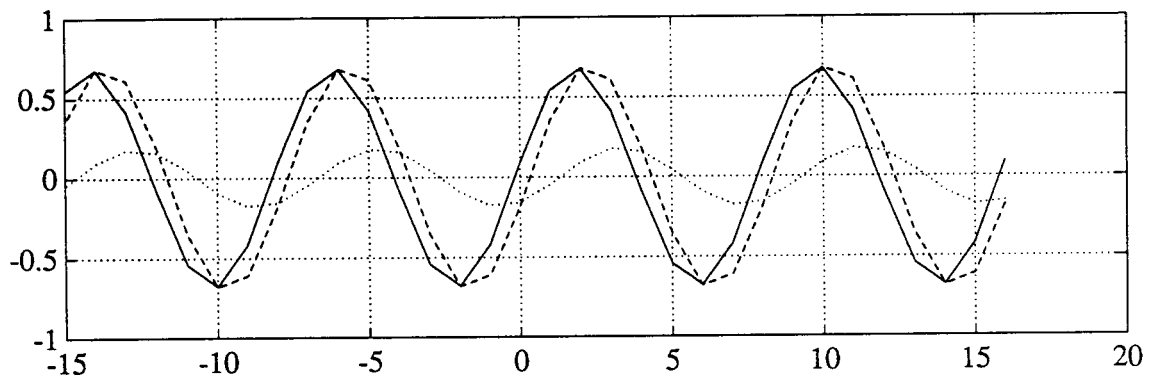


Figure 4.3: The signal $f(k)$ and its wavelet decomposition.

below:

$$XW_1 f = \{2 + 8 \cdot k : k \ \ integer\}$$

$$YW_1 f = \{6 + 8 \cdot k : k \ \ integer\}$$

$$XW_2 f = \{2 + 8 \cdot k : k \ \ integer\}$$

$$YW_2 f = \{6 + 8 \cdot k : k \ \ integer\}$$

$$XW_3 f = \{3 + 8 \cdot k : k \ \ integer\}$$

$$YW_3 f = \{7 + 8 \cdot k : k \ \ integer\}.$$

The basis for the null space of S_3 is:

$$\begin{aligned} y_1(k) &= cos(\tfrac{2\pi k}{8}) \qquad y_2(k) = sin(\tfrac{2\pi k}{8}) \\ y_3(k) &= cos(\tfrac{4\pi k}{8}) \qquad y_4(k) = sin(\tfrac{4\pi k}{8}) \\ y_5(k) &= cos(\tfrac{6\pi k}{8}) \qquad y_6(k) = sin(\tfrac{6\pi k}{8}) \\ y_7(k) &= cos(\tfrac{8\pi k}{8}) \end{aligned}$$

All the functions y_p are 8-periodic. Linear operators preserve this periodicity. Therefore the rows $\underline{\mathcal{W}}_j(k)$ are also 8-periodic in the sense:

$$\underline{\mathcal{W}}_j(k) = \underline{\mathcal{W}}_j(k+8) \ \ k = 0, 1, \ldots, N-1 \qquad j = 1, 2, 3.$$

Thus every level contributes only 2 different rows to the matrix $\mathcal{W}$, and then there can be only 6 different rows in $\mathcal{W}$! The ultimate conclusion is that, in this case, the maxima representation is **not unique**. In essence, till now, we have just restored, for this particular case, the proof of Theorem 6. We continue to study this case in order to calculate exactly the reconstruction set of this representation.

The different rows of $\mathcal{W}$ are as follows:

$$\begin{array}{rrrrrrr}
0.6328 & -0.2621 & 0.6847 & 0.6847 & -0.3826 & 0.9237 & -1.0000 \\
-0.6328 & 0.2621 & 0.6847 & 0.6847 & 0.3826 & -0.9237 & -1.0000 \\
0.4988 & -0.4988 & 0.2500 & 0.0000 & 0.0147 & 0.0147 & 0 \\
-0.4988 & 0.4988 & 0.2500 & -0.0000 & -0.0147 & -0.0147 & 0 \\
0.1288 & -0.1288 & 0.0000 & 0.0000 & -0.0038 & -0.0038 & 0 \\
-0.1288 & 0.1288 & -0.0000 & 0.0000 & 0.0038 & 0.0038 & 0
\end{array}$$

The rank of this matrix is only 5. One can observe that the last two rows are dependent. Consider:

$$\mathcal{W}_s = \mathcal{W}(1:5, [2,3,5,6,7]).$$

$\mathcal{W}_s$ is the submatrix of $\mathcal{W}$ consisting of the elements from the rows 1,2,3,4,5 and the columns 2,3,5,6,7. It is a regular matrix, with an inverse matrix not having large entries. Therefore we will use it for the representation of a general sequence satisfying $S_3 f = 0$ and giving zero samples at extreme points (i.e. from the space $\mathcal{N}T_{mf}$).

Let, α_1^p and α_2^p be defined as:

$$\alpha_1^p = -\mathcal{W}_s^{-1} \cdot \mathcal{W}^{c1}$$

$$\alpha_2^p = -\mathcal{W}_s^{-1} \cdot \mathcal{W}^{c4}$$

where $\mathcal{W}^{c1}$ and $\mathcal{W}^{c4}$ are the first and the fourth columns of the first five rows of $\mathcal{W}$. The space $\mathcal{N}T_{mf}$ is spanned by the two following sequences:

$$y_1^w = y_1 + \alpha_1^p(1) \cdot y_2 + \alpha_1^p(2) \cdot y_3 + \alpha_1^p(3) \cdot y_5 + \alpha_1^p(4) \cdot y_6 + \alpha_1^p(5) \cdot y_7$$

$$y_2^w = y_4 + \alpha_2^p(1) \cdot y_2 + \alpha_2^p(2) \cdot y_3 + \alpha_2^p(3) \cdot y_5 + \alpha_2^p(4) \cdot y_6 + \alpha_2^p(5) \cdot y_7.$$

Therefore, the general solution of $T_{mf}\eta = T_{mf}f$ is:

$$\eta = f + a_1 \cdot y_1^w + a_2 \cdot y_2^w.$$

Conditions for $W_j\eta$ to be monotonic between extrema of W_jf introduce 24 linear inequalities in a_1 and a_2. Elementary analysis of this system leads to the following seven dominant inequalities:

$$\begin{aligned}
1.0000a_1 &+ 1.3059a_2 > -0.4330 \\
1.0000a_1 &- 1.3059a_2 > -0.4330 \\
0.0000a_1 &+ 1.0000a_2 > -0.1849 \\
0.0000a_1 &+ 1.0000a_2 < 0.1849 \\
1.0000a_1 &+ 0.3574a_2 < 0.1007 \\
1.0000a_1 &- 0.3574a_2 < 0.1007 \\
1.0000a_1 &+ 0.0000a_2 < 0.0991
\end{aligned} \tag{4.4.2}$$

First, we pick up three pairs (0, 0), (0, -0.4), (-0.2, 0.16) which satisfy the system of inequalities (4.4.2). In order to visualize different sequences with the same wavelet maxima representation, let us define three sequences. The first is the original f and the next two are defined as:

$$f_a = f - 0.4 \cdot y_1^w$$

$$f_b = f - 0.2 \cdot y_1^w + 0.16 \cdot y_2^w.$$

Figure 4.4 and Figure 4.5 show these sequences and their wavelet transforms. From the graphs one can indeed see that all have the same discrete wavelet

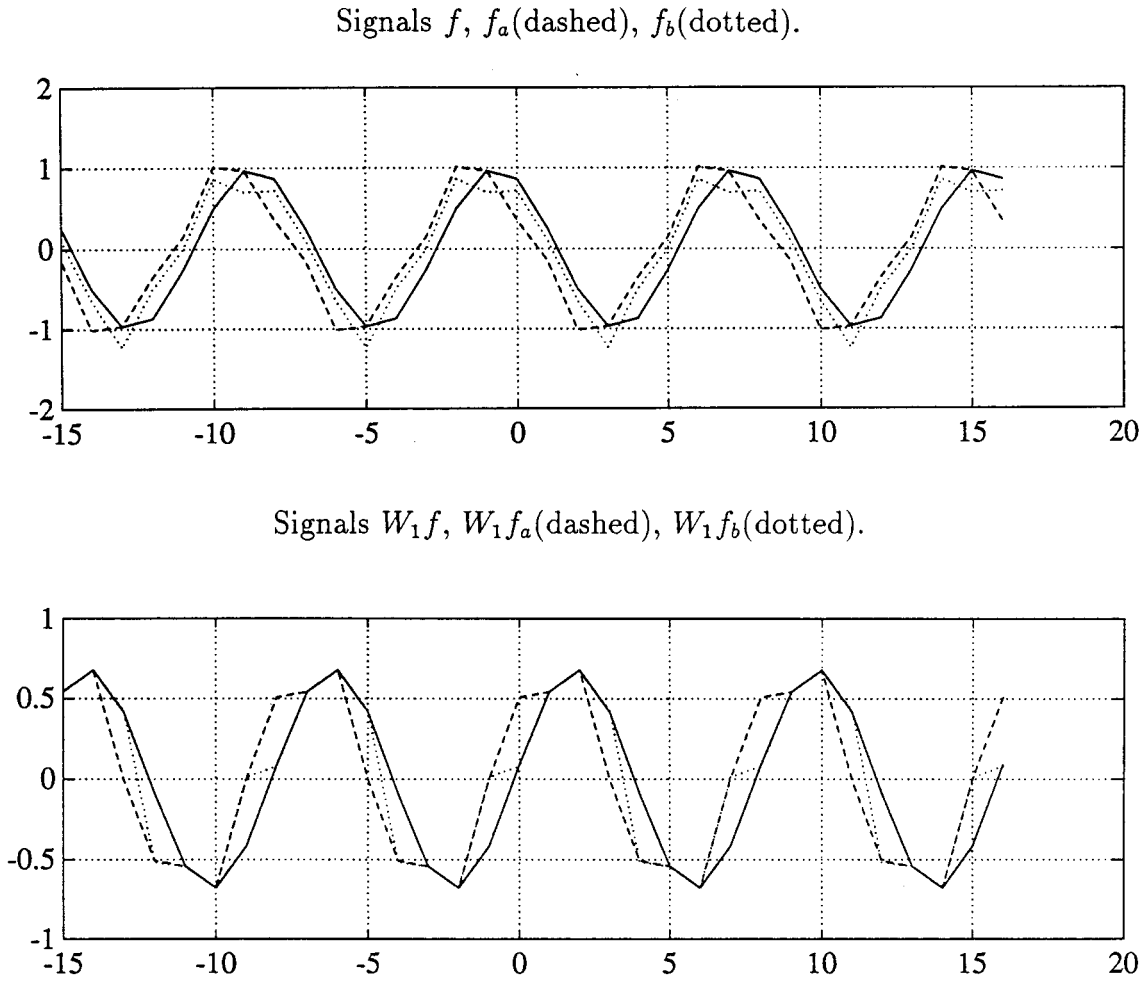


Figure 4.4: The signals and their first level wavelet transforms.

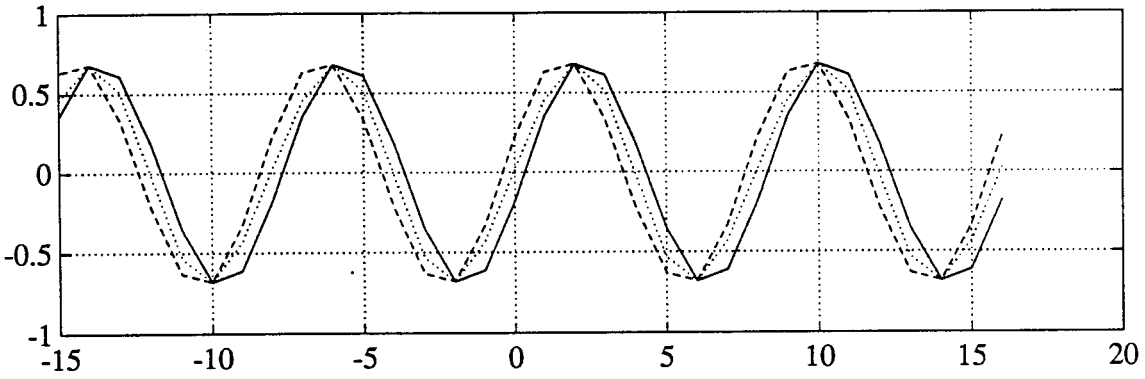
maxima representation. Let us assume that f_a was reconstructed from the representation of f, in this case the noise to signal ratio is defined as

$$\frac{\|f - f_a\|^2}{\|f\|^2}$$

and is equal to 0.345. In spite of the high $\frac{N}{S}$ ratio, these signals have a very similar shape.

Set of inequalities (4.4.2) can be solved precisely. Figure 4.6 describes this

Signals $W_2 f$, $W_2 f_a$(dashed), $W_2 f_b$(dotted).



Signals $W_3 f$, $W_3 f_a$(dashed), $W_3 f_b$(dotted).

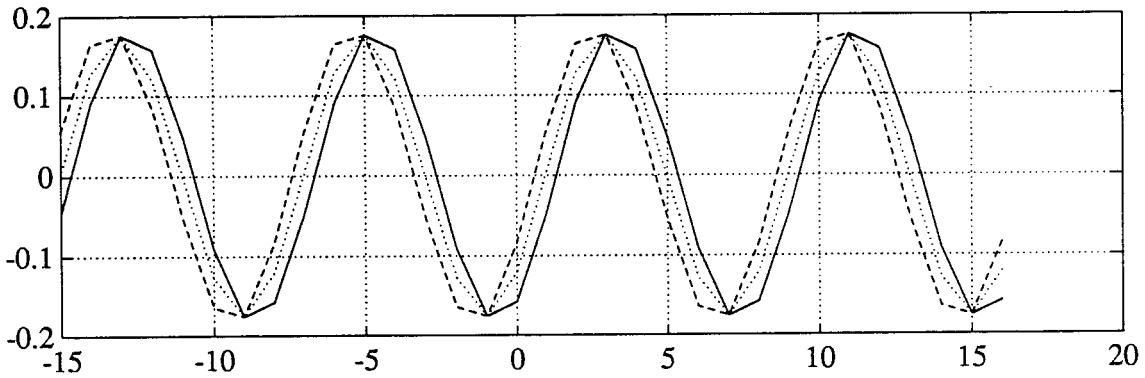


Figure 4.5: The second and the third level wavelet transforms.

solution by showing the boundary of set $\mathcal{A}$,the set of all pairs (a_1, a_2) satisfying set of inequalities (4.4.2).

Now, the reconstruction set of this representation is given as:

$$\Gamma(R_m f) = \{\eta : \eta = f + a_1 \cdot y_1^w + a_2 \cdot y_2^w \ \ (a_1, a_2) \in \mathcal{A}\} \tag{4.4.3}$$

Figures 4.7, 4.8, 4.9 were obtained by plotting a family of functions on the same graph. Figure 2 describes the set $\Gamma(R_m f)$, Figure 3 and 4 show the sets

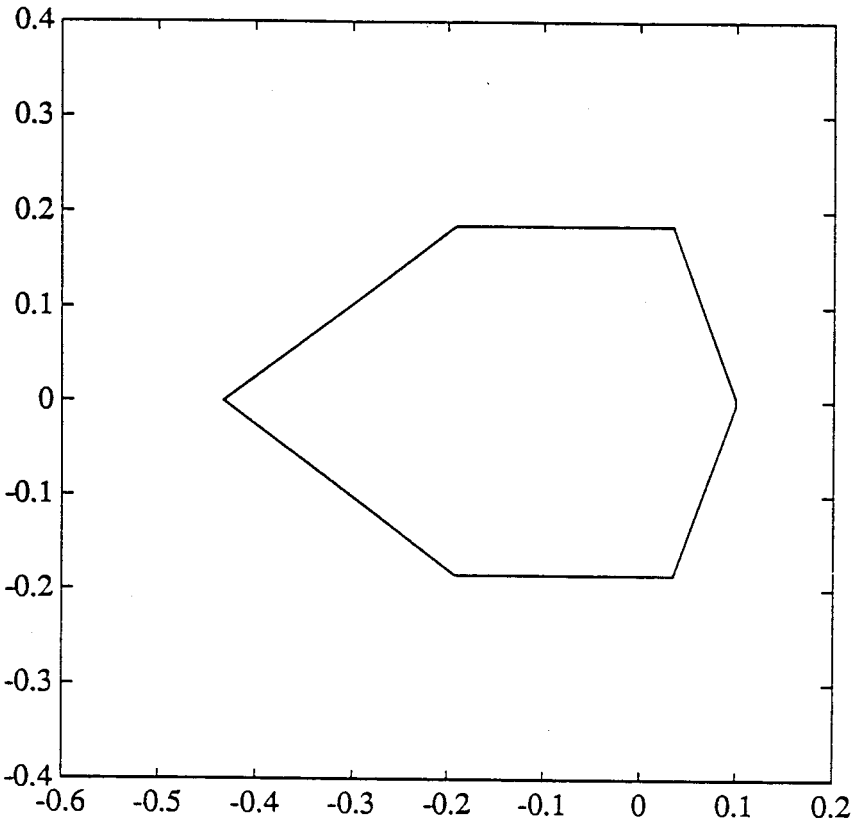


Figure 4.6: The boundary of the set $\mathcal{A}$ on the plane (a_1, a_2)

sets $\{W_1\eta : \eta \in \Gamma(R_m f)\}$ and $\{W_2\eta : \eta \in \Gamma(R_m f)\}$. Of course all functions in Figure 4.7 and 4.8 have the same extreme points.

As was already mentioned, only a nonunique case may exhibit distinctive properties of the multiscale maxima representation. Therefore, it is worthwhile to examine carefully the reconstruction set $\Gamma(R_{mf})$. Since the cubic spline wavelet transform is based on wavelet $\Psi(\xi)$ which is a derivative of a smoothing function, all sequences in $\Gamma(R_{mf})$ have the same multiscale sharper variation points. Moreover, they all appear to have a very similar shape. This "shape"

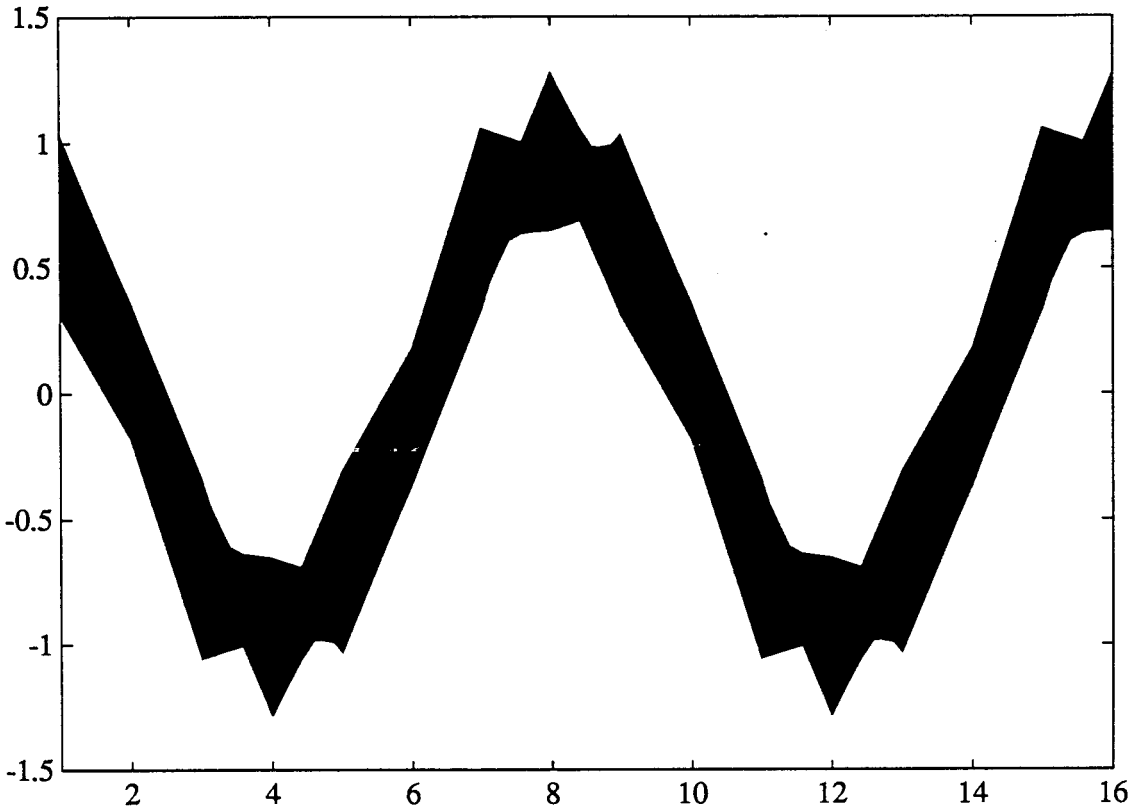


Figure 4.7: All sequences belonging to $\Gamma(R_m f)$.

preserving property, perhaps expected, but not easily formulated and proved, is apparently the most interesting and promising feature of the multiscale maxima representation.

A nonunique representation can be viewed as an approximation, in this case we can, at least for the above example, make the following observation. The reconstruction set of the multiscale maxima representation, as a subset of $\Re^N$, appears to be much less directional homogeneous than reconstruction sets based on other standard approximations techniques like quantization or truncation.

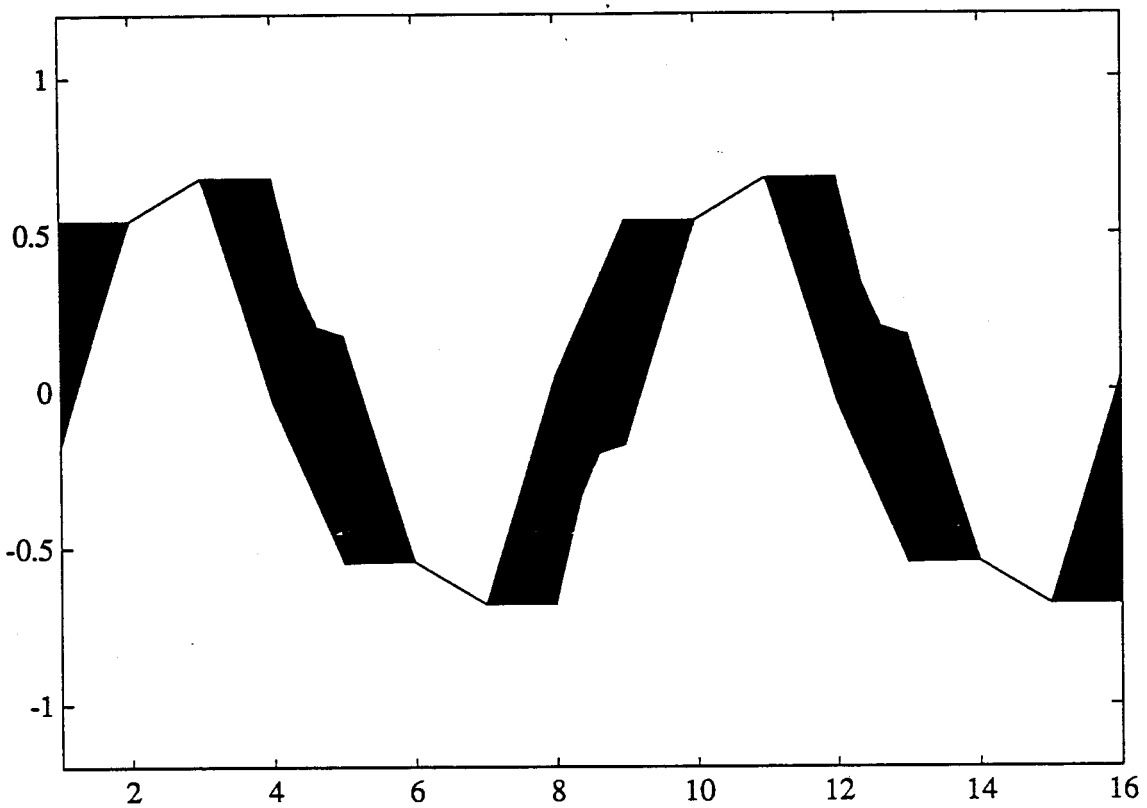


Figure 4.8: All $W_1\eta$ such that $\eta \in \Gamma(R_m f)$.

The reconstruction set $\Gamma(R_{mf})$ was calculated for $J = 3$. It turns out that, in this particular case, $\Gamma(R_{mf}) \subseteq \mathcal{N}S_3$. But then, for all $\eta \in \Gamma(R_{mf})$ and for $j > 3$ the following is true:

$$W_j\eta = 0. \tag{4.4.4}$$

Therefore all $\eta \in \Gamma(R_{mf})$ have the same wavelet maxima representation regardless J as long $J \geq 3$.

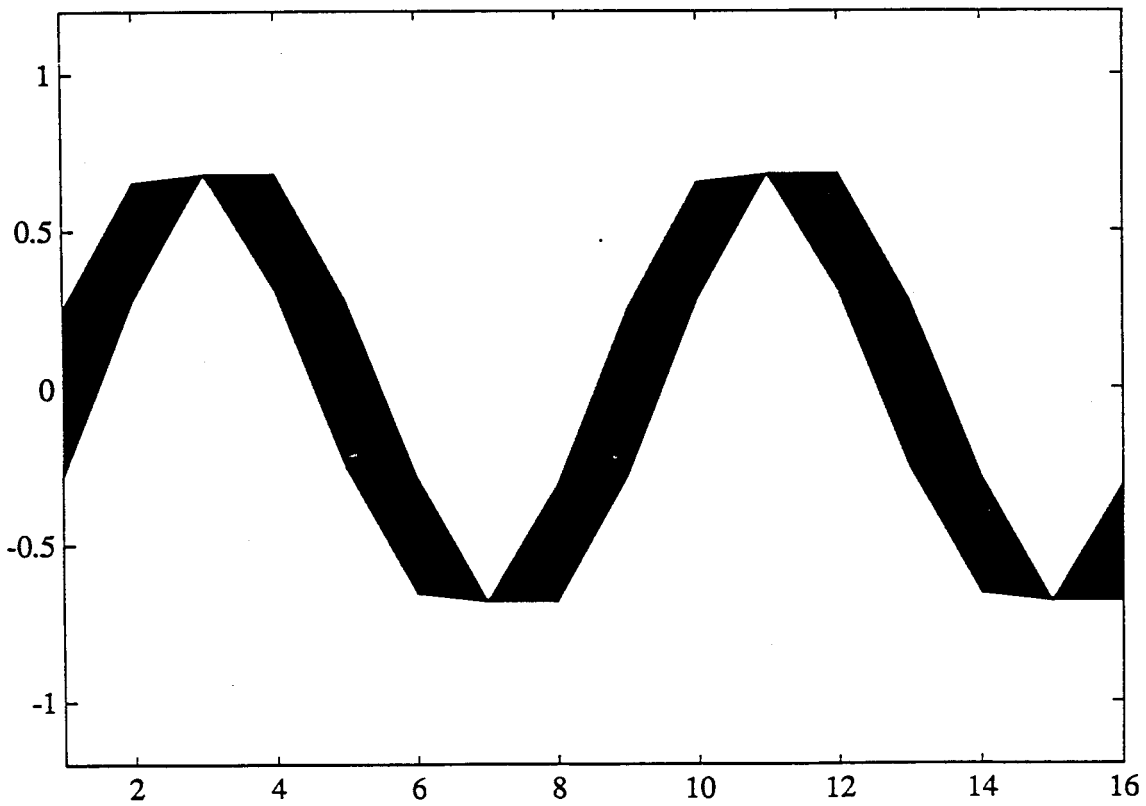


Figure 4.9: All $W_2\eta$ such that $\eta \in \Gamma(R_m f)$.

4.4.3 An example of zero-crossings nonunique representation

Now, let us consider the zero-crossings representation based on the wavelet transform defined in [29]. Let $J = 5$ and $N = 256$. By essentially the same analysis as the one described in the previous sections we can show many signals with a unique representation and a few with a nonunique representation. We skip the description of detailed analysis and just define two signals having a nonunique wavelet zero-crossings representation.

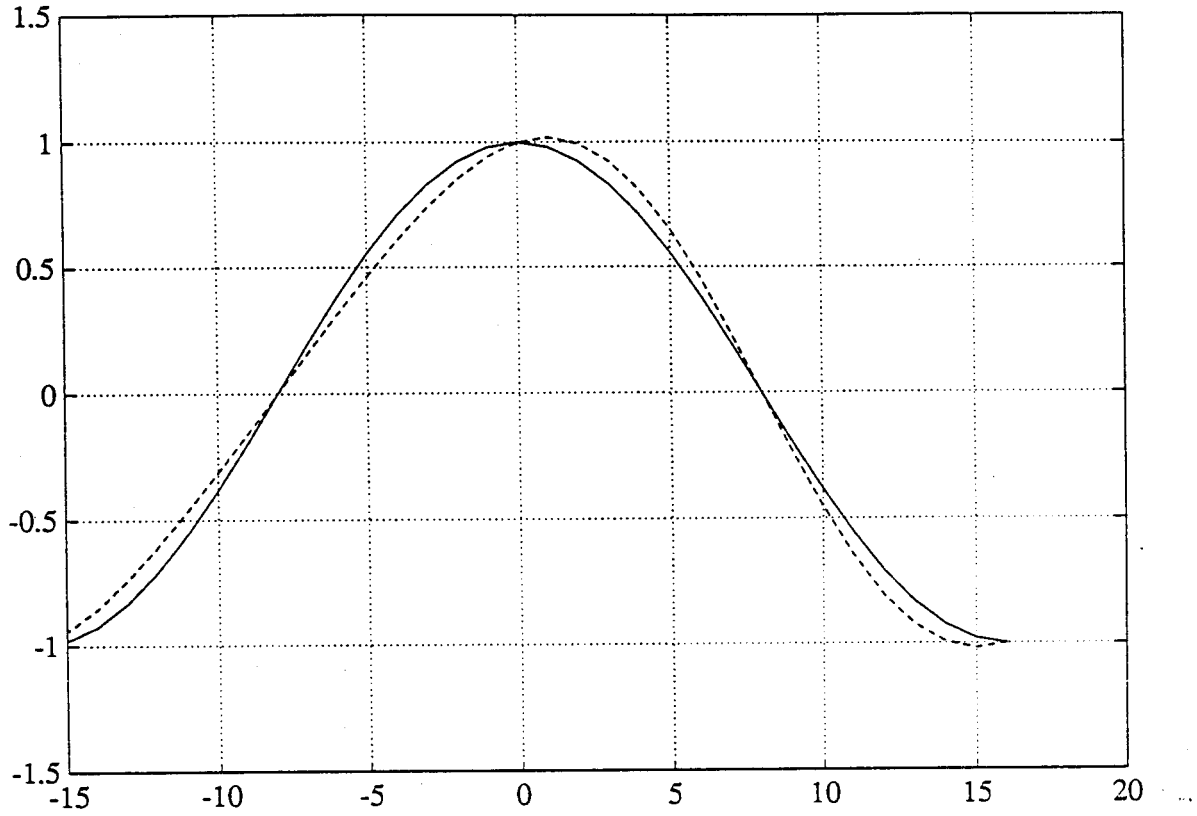


Figure 4.10: The sequences f and f_a

Consider

$$f(k) = \cos\left(\frac{2\pi k}{32}\right) \qquad k = 0, 1, \ldots, N-1$$

$$f_a(k) = \cos\left(\frac{2\pi k}{32}\right) + 0.1 \cdot \sin\left(\frac{2\pi k}{16}\right) \qquad k = 0, 1, \ldots, N-1.$$

Using directly Theorem 2, it can be shown that they have the same zero-crossings representation. Figure 4.10 describes these sequences, while Figure 4.11 presents their first level wavelet transforms (other levels wavelet transforms are very similar).

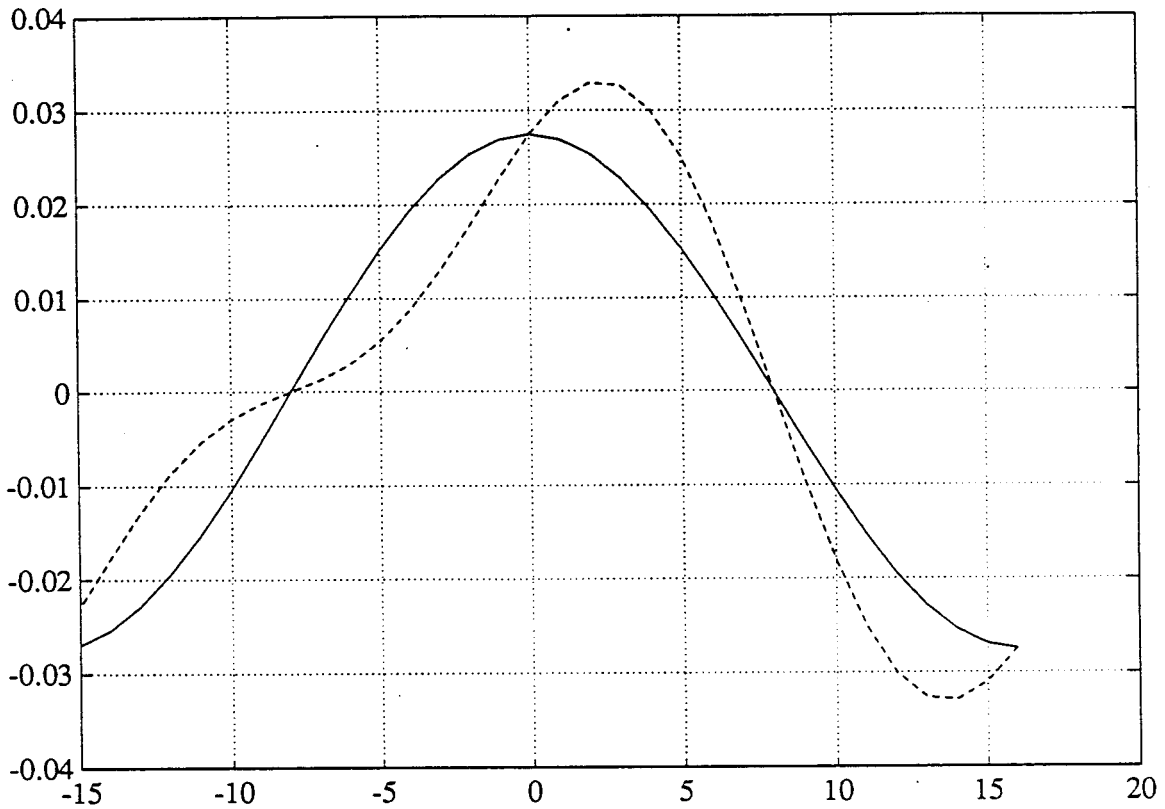


Figure 4.11: The sequences $W_1 f$ and $W_1 f_a$

CHAPTER
FIVE

STABILITY

The problem of bounding possible variations in the reconstruction set due to perturbations in the representation is considered. After spending some time, in the previous chapter, with the wavelet maxima (zero-crossings) representations, we return to the general case of inherently bounded AQLR's.

5.1 BIBO stability

To address the stability issue, the standard approach is to introduce the notion of perturbation of the representation, and of the reconstruction set. In addition, a distance measure for distinct representations and for different reconstruction sets should be defined. In general, it is not an easy task. Observe that Vf, Tf may have different sizes for different representations. Fortunately, for inherently bounded representations, the following characterization of BIBO (bounded input, bounded output) stability is easily verified.

Proposition 14 *Let* $Rf_i = \{Vf_i, T_i f_i\}$, $i = 1,2$ *be inherently bounded AQLR's. Then for all* $K_I > 0$ *there exists* K_O *such that:*

$$\|T_i f_i\| \leq K_I \ \ (i = 1,2) \Rightarrow \|x_1 - x_2\| \leq K_O \ \ \forall x_i \in \Gamma(Rf_i)$$

Proof: This claim is an immediate consequence of the definition of an inherently bounded AQLR. Indeed, using the definition of an inherently bounded AQLR and the hypothesis:

$$x_i \in \Gamma(Rf_i) \Rightarrow \|x_i\| \leq K \cdot \|T_i f_i\| \leq K \cdot K_I$$

But then:

$$\|x_1 - x_2\| \leq \|x_1\| + \|x_2\| \leq 2K \cdot K_I$$

□

The above result is strong in the sense that it is valid regardless of the sets Vf_1, Vf_2. It is weak in view of the fact that the bound on $\|x_1 - x_2\|$ is achieved by the bounds on absolute values of x_1, x_2. In this case, a small perturbation in the representation does not necessary yield a small bound of $\|x_1 - x_2\|$. The next result is complimentary in the sense that a certain structure of the perturbation is assumed, but a bound, proportional to the size of the perturbation, is given.

5.2 A Lipschitz condition

In many applications, the reasons for perturbations in a representation are arithmetic or quantization errors in a reconstruction algorithm. This kind of pertur-

bations may change the continuous values of Tf but it preserves the discrete values of Vf. Therefore the perturbed representation, $(Rf)_p$, can be written as:

$$(Rf)_p = \{Vf, Tf + \Delta(Tf)\}. \tag{5.2.1}$$

Let Γ_p be the corresponding reconstruction set. Our results are related to the following measure defined on (Γ, Γ_p):

$$d(\Gamma, \Gamma_p) \triangleq \sup\{\|\gamma - \gamma_p\| : \gamma \in \Gamma, \gamma_p \in \Gamma_p\}.$$

Observe, that for inherently bounded AQLR's, $d(\Gamma, \Gamma_p)$ is always finite. The measure of the perturbation in the reconstruction set is the difference between $d(\Gamma, \Gamma_p)$ and the size of Γ which is defined as follows:

$$s(\Gamma) \triangleq d(\Gamma, \Gamma) = \sup\{\|\gamma_1 - \gamma_2\| : \gamma_1, \gamma_2 \in \Gamma\}. \tag{5.2.2}$$

$s(\Gamma)$ and $d(\Gamma, \Gamma_p)$ describe the largest possible Euclidean norm of a reconstruction error, from the original representation and from a perturbed one, respectively.

One remark is in order. In general, for an arbitrary $\Delta(Tf)$, the associated reconstruction set may be empty and then $d(\Gamma, \Gamma_p)$ would not be defined. In the sequel, it is assumed that this problem is treated by a reconstruction algorithm and hence $\Delta(Tf)$ yields a nonempty Γ_p. In this case, the following Lipschitz condition is satisfied.

Theorem 8 *For all inherently bounded AQLR, there exists $K > 0$ such that:*

$$d(\Gamma, \Gamma_p) \leq K \cdot \|\Delta(Tf)\| + s(\Gamma). \tag{5.2.3}$$

Proof: Let $\overline{\Gamma}$ and $\overline{\Gamma_p}$ be the closures of the sets Γ and Γ_p, respectively. Since $\|v - \bar{v}\|$ is a continuous function on $\overline{\Gamma} \times \overline{\Gamma_p}$, which is a compact set, then there exist $v \in \overline{\Gamma}$ and $\bar{v} \in \overline{\Gamma_p}$ such that:

$$\|v - \bar{v}\| = d(\overline{\Gamma}, \overline{\Gamma_p}) = d(\Gamma, \Gamma_p). \tag{5.2.4}$$

Moreover, v and $\bar{v}$ have to be vertices of $\overline{\Gamma}$ and $\overline{\Gamma_p}$. Indeed, if, for example, v is not a vertex, then there exists $\epsilon > 0$ such that $v + \epsilon\,(v - \bar{v}) \in \overline{\Gamma}$, and:

$$\|v + \epsilon\,(v - \bar{v}) - \bar{v}\| = (1 + \epsilon) \cdot \|v - \bar{v}\| > \|v - \bar{v}\|. \tag{5.2.5}$$

This contradicts the fact that $\|v - \bar{v}\| = d(\overline{\Gamma}, \overline{\Gamma_p})$.

Let $\Delta(Tf)$ be fixed and arbitrary, such that Γ_p is nonempty. We define

$$(Rf)_p^\lambda \triangleq \{Vf, Tf + \lambda \cdot \Delta(Tf)\} \quad 0 \leq \lambda \leq 1 \tag{5.2.6}$$

with the underlying reconstruction set denoted by Γ_p^λ. From the definition of an Adaptive Quasi Linear Representation (AQLR):

$$\Gamma_p^\lambda = \{x : Tx = Tf + \lambda \cdot \Delta(Tf) \text{ and } \mathbf{C}x > \mathbf{c}\}. \tag{5.2.7}$$

The above formula yields the following observation: if $x_0 \in \Gamma = \Gamma_p^0$ and $x_1 \in \Gamma_p = \Gamma_p^1$ then

$$x_0 + \lambda \cdot (x_1 - x_0) \in \Gamma_p^\lambda \quad 0 \leq \lambda \leq 1. \tag{5.2.8}$$

Therefore Γ_p^λ is nonempty for $0 \leq \lambda \leq 1$ and $d(\Gamma, \Gamma_p^\lambda)$ is well defined.

Next, notice that the closure of Γ_p^λ is given by:

$$\left(\Gamma_p^\lambda\right)^c = \{x : Tx = Tf + \lambda \cdot \Delta(Tf),\ \ \mathbf{C}x \geq \mathbf{c}\} = \{x : \mathbf{B}x \geq \mathbf{b} + \lambda\Delta b\}. \quad (5.2.9)$$

where $\mathbf{B}$ is a $p \times N$ matrix and $\mathbf{b}, \Delta b$ are p-dimensional vectors (see Section 3.2). Since every equality of $Tx = Tf + \lambda \cdot \Delta(Tf)$ appears in two rows in $\mathbf{B}x \geq \mathbf{b} + \lambda\Delta b$:

$$\|\Delta b\| = 2\|\Delta(Tf)\|. \quad (5.2.10)$$

We know that

$$d(\Gamma, \Gamma_p^\lambda) = \|v^\lambda - \bar{v}^\lambda\| \quad (5.2.11)$$

where v^λ is a vertex of $\overline{\Gamma}$ and $\bar{v}^\lambda$ is a vertex of $\overline{\Gamma_p^\lambda}$. Using (3.2.4) we can write:

$$v^\lambda = D^i b \quad \text{and} \quad \bar{v}^\lambda = \bar{D}^i(\mathbf{b} + \lambda\Delta b). \quad (5.2.12)$$

Both matrices D^i and $\bar{D}^i$ are obtained from an inverse of a regular submatrix of $\mathbf{B}$. Note that $\|D\mathbf{b} - \bar{D}(\mathbf{b} + \lambda\Delta b)\|$ is a continuous function of λ for any two matrices $D, \bar{D}$. Therefore, if

$$\|v^\lambda - \bar{v}^\lambda\| < \|v_o - \bar{v}_o\| \quad (5.2.13)$$

for all pairs $v_o, \bar{v}_o$ of vertices of $\overline{\Gamma}, \overline{\Gamma_p^\lambda}$, respectively, which are different from $v^\lambda, \bar{v}^\lambda$, then there exists a segment $[\lambda_i, \lambda_{i+1}]$ such that:

$$d(\Gamma, \Gamma_p^\lambda) = \|D^i\mathbf{b} - \bar{D}^i(\mathbf{b} + \lambda\Delta b)\| \quad \forall\lambda \in [\lambda_i, \lambda_{i+1}]. \quad (5.2.14)$$

Furthermore, there exists another pair of vertices, with associated matrices D^{i-1} and $\bar{D}^{i-1}$, such that:

$$d\left(\Gamma, \Gamma_p^{\lambda_i}\right) = \|D^i\mathbf{b} - \bar{D}^i(\mathbf{b} + \lambda_i\Delta b)\| = \|D^{i-1}\mathbf{b} - \bar{D}^{i-1}(\mathbf{b} + \lambda_i\Delta b)\|. \quad (5.2.15)$$

Next observe that since the number of regular submatrices of $\mathbf{B}$ is finite, the number of possible pairs $D, \bar{D}$ is finite as well. Consider $\|D\mathbf{b} - \bar{D}(\mathbf{b} + \lambda\Delta b)\|$ and $\|D_o\mathbf{b} - \bar{D}_o(\mathbf{b} + \lambda\Delta b)\|$ as two functions of λ. As square roots of quadratic forms, these expressions may coincide or be equal for at most two values of λ. Therefore all possible pairs of these functions intersect at finitely many points. Consequently, there exist L points:

$$0 = \lambda_0 < \lambda_1 \ldots < \lambda_{L-1} = 1 \tag{5.2.16}$$

and L pairs of matrices $(D^i, \bar{D}^i)$ $i = 0, 1, \ldots, L-1$ such that $D^i b$ is a vertex of Γ^c and $\bar{D}^i(\mathbf{b} + \lambda\Delta b)$ is a vertex of $\left(\Gamma_p^\lambda\right)^c$ for all $\lambda \in [\lambda_i, \lambda_{i+1}]$. Moreover

$$d(\Gamma, \Gamma_p^\lambda) = \|D^i\mathbf{b} - \bar{D}^i(\mathbf{b} + \lambda\Delta b)\| \quad \forall\lambda \in [\lambda_i, \lambda_{i+1}]. \tag{5.2.17}$$

$$d\left(\Gamma, \Gamma_p^{\lambda_i}\right) = \|D^i\mathbf{b} - \bar{D}^i(\mathbf{b} + \lambda_i\Delta b)\| = \|D^{i-1}\mathbf{b} - \bar{D}^{i-1}(\mathbf{b} + \lambda_i\Delta b)\|. \tag{5.2.18}$$

Proposition 15

$$d\left(\Gamma, \Gamma_p^{\lambda_i}\right) \leq d(\Gamma, \Gamma) + \sum_{k=0}^{i-1}(\lambda_{k+1} - \lambda_k) \cdot \|\bar{D}^k\Delta b\|. \tag{5.2.19}$$

Proof: By induction on i. Let $i = 1$.

$$d(\Gamma, \Gamma_p^{\lambda_1}) = \|D^1\mathbf{b} - \bar{D}^1(\mathbf{b} + \lambda_1\Delta b)\| =$$

$$= \|D^0\mathbf{b} - \bar{D}^0(\mathbf{b} + \lambda_1\Delta b)\| =$$

$$= \|D^0\mathbf{b} - \bar{D}^0\mathbf{b} + \bar{D}^0\mathbf{b} - \bar{D}^0(\mathbf{b} + \lambda_1\Delta b)\| \leq$$

$$\leq \|D^0\mathbf{b} - \bar{D}^0\mathbf{b}\| + \|\bar{D}^0\mathbf{b} - \bar{D}^0(\mathbf{b} + \lambda_1\Delta b)\| =$$

$$= d(\Gamma, \Gamma) + \lambda_1 \cdot \|\bar{D}\Delta b\|.$$

Since $\lambda_0 = 0$, the above is exactly the claim for $i = 1$. By induction, let us assume that the proposition holds for $i - 1$. Consider:

$$d(\Gamma, \Gamma_p^{\lambda_i}) = \|D^{i-1}\mathbf{b} - \bar{D}^{i-1}(\mathbf{b} + \lambda_i \Delta b)\| =$$

$$= \|D^{i-1}\mathbf{b} - \bar{D}^{i-1}(\mathbf{b} + \lambda_{i-1}\Delta b) + \bar{D}^{i-1}(\mathbf{b} + \lambda_{i-1}\Delta b) - \bar{D}^{i-1}(\mathbf{b} + \lambda_i \Delta b)\| \leq$$

$$\leq \|D^{i-1}\mathbf{b} - \bar{D}^{i-1}(\mathbf{b} + \lambda_{i-1}\Delta b)\| + \|\bar{D}^{i-1}(\mathbf{b} + \lambda_{i-1}\Delta b) - \bar{D}^{i-1}(\mathbf{b} + \lambda_{i-1}\Delta b)\| =$$

$$= d(\Gamma, \Gamma_p^{\lambda_{i-1}}) + (\lambda_i - \lambda_{i-1})\|\bar{D}^{i-1}\Delta b\| \leq$$

(using the induction assumption)

$$\leq d(\Gamma, \Gamma) + \sum_{k=0}^{i-2}(\lambda_{k+1} - \lambda_k)\|\bar{D}^k \Delta b\| + (\lambda_i - \lambda_{i-1})\|\bar{D}^{i-1}\Delta b\| =$$

$$= d(\Gamma, \Gamma) + \sum_{k=0}^{i-1}(\lambda_{k+1} - \lambda_k)\|\bar{D}^k \Delta b\|.$$

This concludes the proof of the proposition.

□

Using the proposition we deduce that the distance between Γ and Γ_p satisfies:

$$d(\Gamma, \Gamma_p) \leq d(\Gamma, \Gamma) + \sum_{k=0}^{L-1}(\lambda_{k+1} - \lambda_k)\|\bar{D}^k \Delta b\|. \tag{5.2.20}$$

Let $\|\bar{D}^i\|$ be the induced matrix norm of $\bar{D}^i$. Then

$$\|\bar{D}^i \Delta B\| \leq \|\bar{D}^i\| \cdot \|\Delta b\|. \tag{5.2.21}$$

Since the number of possible matrix $\bar{D}^i$ is finite, there exists $K_D > 0$ such that:

$$\|\bar{D}^i\| \leq K_D \tag{5.2.22}$$

for all valid $\bar{D}^i$. Combining together (5.2.22),(5.2.21), (5.2.20) we show that

$$d(\Gamma, \Gamma_p) \leq d(\Gamma, \Gamma) + K_D \|\Delta b\|. \tag{5.2.23}$$

By taking $K = 2K_D$ and using (5.2.10), the desired relation is obtained:

$$(\Gamma, \Gamma_p) \leq d(\Gamma, \Gamma) + K \cdot \|\Delta(Tf)\|. \tag{5.2.24}$$

□

Observe that the above result is global in the sense that as long as $\Delta(Tf)$ gives rise to a nonempty reconstruction set, the theorem holds regardless of the size of $\Delta(Tf)$.

5.3 Discussion

One can ask whether the kind of stability just established is indeed the property that has been desired to achieve. The answer has several different aspects and let us dwell a while upon this subject. First consider the following citation from Hummel and Moniot [22] "stability of the representation concerns continuity of the inverse map ". Theorem 8 is exactly of this type. Another citation, from [29] is as follows: "a representation is said to be unstable if a small perturbation of the representation may correspond to an arbitrary large perturbation of the original function." This definition refers to BIBO stability, which was given by Proposition 14. In view of these considerations, stability, as presented in this thesis, is indeed a necessary property of a multiscale edge representation. But it

does not mean that an inherently bounded AQLR will always provide accurate reconstruction results. Somehow, perhaps because of partial uniqueness results obtained by unstable tools, poor reconstruction results are often regarded as evidences of instability. For instance, such was the case in the example given in [22], mentioned earlier. A careful discrete analysis may point out different possible reasons for inadequate reconstruction results, e.g. nonuniqueness of the discrete representation, instability or high sensitivity of the reconstruction algorithm. Therefore, stability should not be viewed as a sufficient condition of a desired signal description. In our opinion, every practical signal representation has to be tested quantitatively with respect to the size and the structure of reconstruction sets and with respect to sensitivity of the reconstruction algorithm.

CHAPTER
SIX

RECONSTRUCTION

6.1 General theory for an inherently bounded AQLR

For a nonunique representation, there are several ways to define a reconstruction algorithm. One can require to find all elements from the reconstruction set, on the other hand, sometimes it is desired to determine the smallest element satisfying a given representation. In this work, a reconstruction algorithm is defined as a procedure to find an arbitrary element x belonging to the closure of the reconstruction set, $\overline{\Gamma}$. As mentioned earlier, we propose a reconstruction algorithm based on an appropriate potential function $v(x)$. This function should satisfy:

$$v(x) = 0, \quad \forall x \in \overline{\Gamma} \tag{6.1.1}$$

$$v(x) > 0, \quad \forall x \in \left(\overline{\Gamma}\right)^c. \tag{6.1.2}$$

where $\left(\overline{\Gamma}\right)^c$ denotes the complement of $\overline{\Gamma}$ in $\mathcal{L}$. Furthermore, it will be shown that the proposed $v(x)$ does not have any local extremum outside $\overline{\Gamma}$, i.e.

$$\|\nabla v(x)\| > 0 \quad \forall x \in \left(\overline{\Gamma}\right)^c. \tag{6.1.3}$$

$\nabla v(x)$ denotes the gradient of $v(x)$ with respect to x, namely it is a column vector of derivatives of v with respect to components of x. With such a potential function, the reconstruction is achieved by any minimization algorithm operating on $v(x)$. We will focus on the reconstruction algorithm based on the differential equation:

$$\dot{x}(t) = -\nabla\left(v\left(x(t)\right)\right) \tag{6.1.4}$$

whose analog hardware implementation gives rise to a very fast algorithm. In addition, a steepest descent algorithm, appropriate for a digital simulation, is described and its convergence is shown.

In this section, a general inherently bounded Adaptive Quasi Linear Representation (AQLR) is considered. As mentioned in Section 3.2, the closure of the reconstruction set, $\overline{\Gamma}$, can be written as:

$$\overline{\Gamma} = \{x : \mathbf{B}x \geq \mathbf{b}\} \tag{6.1.5}$$

for a given $p \times N$ matrix $\mathbf{B}$ and a p-dimensional vector $\mathbf{b}$. The function $v(x)$ is derived from this representation in the following way.

$$v(x) \triangleq \sum_{i=1}^{p} \rho(\mathbf{B}x - \mathbf{b})_i \tag{6.1.6}$$

where $(\mathbf{B}x - \mathbf{b})_i$ denotes the i-th component of the vector $\mathbf{B}x - \mathbf{b}$. The function $\rho(\cdot)$ is defined by:

$$\rho(\xi) \triangleq \begin{cases} \xi^2 & \text{if } \xi \leq 0 \\ 0 & \text{otherwise} \end{cases} .$$

Using the above definitions, it is easy to verify that (6.1.1) and (6.1.2) hold.

Observe that $\rho(\xi)$ is continuously differentiable. Therefore $v(x)$ is continuous and continuously differentiable. By elementary calculations, the gradient of $v(x)$ can be shown to be:

$$\nabla v(x) = 2\mathbf{B}'Z(\mathbf{B}x - \mathbf{b})$$

where Z is a $p \times p$ diagonal matrix defined by:

$$Z(i,i) \triangleq \begin{cases} 1 & \text{if } (\mathbf{B}x - \mathbf{b})_i \leq 0 \\ 0 & \text{otherwise} \end{cases} .$$

Naturally, $\mathbf{B}'$ denotes the transpose of the matrix $\mathbf{B}$.

The following theorem states that $v(x)$ does not have local extrema outside the set $\overline{\Gamma}$.

Theorem 9 *Let Γ be nonempty. Then $\nabla v(x) = 0$ if and only if $x \in \overline{\Gamma}$.*

Proof: If $x \in \overline{\Gamma}$ then $Z(i,i) = 0$ for $i = 1, 2, \ldots, p$ and clearly $\nabla v(x) = 0$. Let us assume $\nabla v(x) = 2\mathbf{B}'Z(\mathbf{B}x - \mathbf{b}) = 0$. Since $x \in \overline{\Gamma}$ if and only if $Z(\mathbf{B}x - \mathbf{b}) = 0$, we need to show that $Z(\mathbf{B}x - \mathbf{b}) = 0$. Consider the following decomposition of $Z\mathbf{b}$:

$$Z\mathbf{b} = Z\mathbf{B}y + \mathbf{b}_o \tag{6.1.7}$$

such that $\mathbf{b}_o \perp \mathcal{R}(Z\mathbf{B})$, namely $\mathbf{b}_o' Z\mathbf{B}x = 0 \ \ \forall x$, or equivalently:

$$(Z\mathbf{B})'\mathbf{b}_o = 0. \tag{6.1.8}$$

Substituting this decomposition into the hypothesis yields:

$$0 = 2\mathbf{B}'Z(\mathbf{B}x - \mathbf{b}) = 2\mathbf{B}'(Z\mathbf{B}x - Z\mathbf{B}y - \mathbf{b}_o) = 2\mathbf{B}'(Z\mathbf{B}x - Z\mathbf{B}y).$$

Using $Z = Z'Z$ we see that

$$2(Z\mathbf{B})'(Z\mathbf{B})(x - y) = 0$$

which implies

$$(x - y)'(Z\mathbf{B})'(Z\mathbf{B})(x - y) = \|Z\mathbf{B}(x - y)\| = 0.$$

Therefore

$$Z\mathbf{B}x = Z\mathbf{B}y. \tag{6.1.9}$$

Consequently, in this case:

$$Z(\mathbf{B}x - \mathbf{b}) = Z\mathbf{B}x - Z\mathbf{B}y - \mathbf{b}_o = -\mathbf{b}_o. \tag{6.1.10}$$

Hence, it suffices to prove that $\mathbf{b}_o = 0$. This will be based on the following statement of the Farkas' Lemma ([36],p 472-474).

Theorem 10 *Exactly one of the two alternatives holds:*

1. $\exists x$ *s.t.* $Z\mathbf{B}x \geq Zb$.

2. $\exists \mathbf{b}_o$ *such that* $(Z\mathbf{B})'\mathbf{b}_o = 0 \ \ \mathbf{b}_o \geq 0 \ \ (Z\mathbf{B})'\mathbf{b}_o > 0$.

Equation (6.1.8) already states $(Z\mathbf{B})'\mathbf{b}_o = 0$. Observe that from the definition of Z, $Z\mathbf{B}x - Z\mathbf{b} \leq 0$, therefore $\mathbf{b}_o \geq 0$. At this point, let us assume by contradiction that $\|\mathbf{b}_o\| > 0$. Consider

$$(Z\mathbf{b})'\mathbf{b}_o = (ZBy + \mathbf{b}_o)'\mathbf{b}_o = \|\mathbf{b}_o\|^2 > 0.$$

Therefore the second alternative holds. For the first alternative take any $x \in \overline{\Gamma}$. Then $\mathbf{B}x \geq \mathbf{b}$, and for any matrix Z with nonnegative entries: $Z\mathbf{B}x \geq Z\mathbf{b}$. Hence the first alternative holds as well. This is the contradiction we were after, and eventually we have $\mathbf{b}_o = 0$.

□

In view of these considerations, a reconstruction scheme can be implemented as:

$$\arg\min \{v(x) : x \in \mathcal{L}\}. \tag{6.1.11}$$

The minimization is significantly facilitated by the property that local extrema of $v(x)$ appear only in $\overline{\Gamma}$. We are going to focus on the algorithm based on the differential equation (6.1.4). The desired property is that for all $x(0)$, $x(t)$ will approach the set $\overline{\Gamma}$ as $t \to \infty$. In other words, $x(t)$ should approximate an element from $\overline{\Gamma}$ for t large enough. The convergence result is based on La Salle's Theorem.

Theorem 11 *(La Salle)*

Let Ω be a compact set with the property that every solution of $\dot{x}(t) = f(x)$ which

starts in Ω remains for all future time in Ω. Let $v : \Omega \to \Re$ be a continuously differentiable function such that $\dot{v}(x) \leq 0$ in Ω. Let E be the set of all points in Ω where $\dot{v}(x) = 0$. Let M be the largest invariant set in E. Then every solution starting in Ω approaches M as $t \to \infty$.

An invariant set M is defined by:

$$x(0) \in M \Rightarrow x(t) \in M \quad \forall t \geq 0.$$

We say that $\{x(t)\}$ approaches a set M if for all $\epsilon > 0$ there exists T, such that:

$$dist(x(t), M) < \epsilon, \quad \forall t > T$$

where $dist(x(t), M)$ denotes the smallest distance between $x(t)$ and a point from M, i.e.

$$dist(x(t), M) \triangleq \inf_{\xi \in M} \|x(t) - \xi\|.$$

The proof of Theorem 11 can be found, for example, in [25].

At this point we are able to prove the following convergence result.

Theorem 12 *Let $\overline{\Gamma}$ be the closure of the reconstruction set of the inherently bounded AQLR. Then for all $x(0)$, the solution of*

$$\dot{x}(t) = -\nabla \left(v \left(x(t) \right) \right). \tag{6.1.12}$$

will approach $\overline{\Gamma}$ as $t \to \infty$.

Proof: Let $x(0)$ be arbitrary and fixed. Define:

$$\Omega = \left\{ x : v(x) \leq v(x(0)) \right\}. \tag{6.1.13}$$

Since $\dot{v}(x) = -\|\nabla(v(x))\|^2 \leq 0$, every solution of (6.1.12) which starts in Ω remains there. $\overline{\Gamma} \subseteq \Omega$ because for all $x \in \overline{\Gamma}$ $v(x) = 0$. As a consequence of Theorem 9 $E = \overline{\Gamma}$. But $\overline{\Gamma}$ is an invariant set, therefore $M = \overline{\Gamma}$. By showing that Ω is compact we will get the desired convergence result. Ω is closed because $v(x)$ is continuous. Boundness of Ω is based on the fact that the representation is inherently bounded. Let $x \in \Omega$ be arbitrary. Define the vector $\mathbf{b}_x$ by:

$$(\mathbf{b}_x)_i \triangleq \begin{cases} (\mathbf{B}x)_i & \text{if } (\mathbf{B}x)_i < \mathbf{B}_i \\ \\ \mathbf{b}_i & \text{otherwise.} \end{cases}$$

Proposition 16 *The norm of* $\mathbf{b}_x$ *is bounded in the following way*

$$\|\mathbf{b}_x\|^2 \leq \|\mathbf{b}\|^2 + v(x). \tag{6.1.14}$$

Proof: For any vector y we define:

$$\|y\|_s^2 \triangleq \sum_{\{i:(\mathbf{B}x)_i \geq \mathbf{b}_i\}} y_i^2$$

$$\|y\|_{ns}^2 \triangleq \sum_{\{i:(\mathbf{B}x)_i < \mathbf{b}_i\}} y_i^2.$$

Note that $\|y\|_s, \|y\|_{ns}$ are norms of appropriate projections of y. Therefore, we can write:

$$\|\mathbf{b}_x\|^2 = \|\mathbf{b}_x\|_s^2 + \|\mathbf{b}_x\|_{ns}^2 =$$

$$= \|\mathbf{b}_x\|_s^2 + \|\mathbf{b}_x - \mathbf{b} + \mathbf{b}\|_{ns}^2 \leq$$

$$\leq \|\mathbf{b}_x\|_s^2 + \|\mathbf{b}_x - \mathbf{b}\|_{ns}^2 + \|\mathbf{b}\|_{ns}^2$$

Using $v(x) = \|\mathbf{b}_x - \mathbf{b}\|_{ns}^2$ and $\|\mathbf{b}_x\|_s^2 = \|\mathbf{b}\|_s^2$, the claim of the proposition is shown.

□

To conclude the proof of boundness of Ω observe that

$$\mathbf{B}x \geq \mathbf{b}_x.$$

Using the result (3.3.16) we see see that:

$$\|x\| \leq K_1 \|\mathbf{b}_x\|. \tag{6.1.15}$$

From the definition of Ω and from the proposition it can be shown that for all $x \in \Omega$

$$\|x\| \leq K_1 \sqrt{\|\mathbf{b}\|^2 + v(x(0))} \tag{6.1.16}$$

namely, Ω is bounded and the proof is completed.

□

Theorem 12 enables us to use a very fast analog-hardware implementation to reconstruct signals. However, before acquiring a costly and not flexible hardware, an ability to perform digital simulations is required. The last part of these section is devoted to this issue. First, we define a steepest descent algorithm, based on the cost function $v(x)$.

Definition 6.1.5 *([28]p215)*

Let Rf be a given inherently bounded AQLR, and $v(x)$ be the corresponding cost function. The steepest descent algorithm is calculated by the following iterative formula. For any $x_0 \in \mathcal{L}$, we define the sequence $\{x_k\}$.

$$x_{k+1} = x_k - \alpha_k \cdot \nabla v(x_k) \quad k = 0, 1, 2, \ldots \tag{6.1.17}$$

where α_k is a nonnegative scalar minimizing $v\left(x_k - \alpha_k \cdot \nabla v(x_k)\right)$.

The convergence of this algorithm is described in [28], let us cite (with minor changes in notation) from page 215.

Theorem 13 *If $v(x)$ has continuous partial derivatives and $\{x_k\}$ is bounded, then the limit of any convergent subsequence of $\{x_k\}$ belongs to $\overline{\Gamma}$.*

The proof is based on the Global Convergence Theorem, proposed by Zangwill [47], and appears in [28] pages 209-215 and pages 187-188.

Note, that the conclusion of Theorem 13 is written in apparently different terms than Theorem 12. However,

Lemma 4 *Let $\{x_k\}$ be a bounded sequence. If the limit of any convergent subsequence of $\{x_k\}$ belongs to $\overline{\Gamma}$ then the sequence $\{x_k\}$ approaches $\overline{\Gamma}$ as $k \to \infty$.*

The proof is a standard $\epsilon - \delta$ consideration and is given in Appendix A.

Observe that without the property that $\{x_k\}$ is bounded, we can not deduce the convergence of this algorithm. Here again, in addition to several times in the past, we can benefit from the powerful properties of inherently bounded AQLR's. The boundedness of $\{x_k\}$ can be shown from $v(x_k) \leq v(x_0)$ by the same technique used in the proof of Theorem 12 (see equation (6.1.16)). Consequently, for any x_0, the steepest descent algorithm, based on the cost function $v(x)$, will approach the closure of the reconstruction set. In the next sections we will show how an efficient reconstruction algorithm can be based on this idea.

6.2 The detailed algorithm for a multiscale representation

As was shown in the previous section, to define reconstruction in the context of solving a system of inequalities $\mathbf{B}x \geq \mathbf{b}$ is advantageous for analysis. However, a direct implementation of this form may yield unnecessary high complexity, related to the use of the "large" matrix $\mathbf{B}$. In this section, using the structure of the multiscale maxima representation, an efficient algorithm to calculate the cost function $v(x)$, and its gradient $\nabla v(x)$ will be described. We will conclude with the discussion of possible implementations of reconstruction using iterative descent algorithms.

The cost function $v(x)$ is calculated by taking into account all conditions that x should satisfy in order to belong to the closure of the reconstruction set, $\overline{\Gamma}$. For the multiscale representation, these conditions can be clustered according to the different scales. To be more specific, let us consider the multiscale maxima representation:

$$R_m f = \left\{ \left\{ XW_j f, YW_j f, \{W_j f(k)\}_{k \in XW_j f \cup YW_j f} \right\}_{j=1}^{J}, S_J f \right\}. \tag{6.2.1}$$

The j-th part of this representation, $R_j f$ is defined as the following triple:

$$R_j f \triangleq \left\{ XW_j f, YW_j f, \{W_j f(k)\}_{k \in XW_j f \cup YW_j f} \right\} \tag{6.2.2}$$

In order to keep a uniform notation, $R_{J+1} f$ refers to $S_J f$ and is defined as the following triple.

$$R_{J+1} f \triangleq \left\{ \{k\}_{k=0}^{N-1}, \{k\}_{k=0}^{N-1}, S_J f \right\}. \tag{6.2.3}$$

The conditions for a sequence η to belong to the closure of the reconstruction set, $\overline{\Gamma(R_{mf})}$ are easily derived from Theorem 1. Theorem 1 can be rewritten as follows. An arbitrary $\eta \in \mathcal{L}$ belongs to $\overline{\Gamma(R_{mf})}$ if and only if

$$S_J\eta(k) = S_Jf(k) \qquad k = 0, 1, \ldots, N-1$$

and for $j = 1, 2, \ldots, J$

$$W_j\eta(k) = W_jf(k), \quad k \in XW_jf \cup YW_jf \tag{6.2.4}$$

$$t_j^{mf}(k) \cdot (W_j\eta(k+1) - W_j\eta(k)) \geq 0 \tag{6.2.5}$$

$$k \in (XW_jf \cup YW_jf)^c \cup (XW_jf \cup YW_jf)^r.$$

Recall that the type of k, $t_j^{mf}(k)$, is calculated from the sets XW_jf, YW_jf. With the objective of shortening the forthcoming notation, let us define

$$\tilde{t}(k, R_jf) \triangleq \begin{cases} t_j^{mf}(k) & \text{if } k \in (XW_jf \cup YW_jf)^c \cup (XW_jf \cup YW_jf)^r. \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.6}$$

Observe that if $XW_jf = YW_jf = \{k\}_{k=0}^{N-1}$ then

$$(XW_jf \cup YW_jf)^c \cup (XW_jf \cup YW_jf)^r = \emptyset$$

and therefore $\tilde{t}(k, R_{J+1}f) = 0$ for all $k = 0, 1, \ldots, N-1$.

In addition, we need to define the characteristic function of k with respect to the set of local extrema, $\chi(k, R_jf)$.

$$\chi(k, R_jf) \triangleq \begin{cases} 1 & \text{if } k \in XW_jf \cup YW_jf \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.7}$$

Using the above definitions, we see that $\eta \in \overline{\Gamma(R_m f)}$ in and only if

$$\chi(k, R_{J+1} f) \cdot (S_J \eta(k) - S_J f(k)) = 0 \quad k = 0, 1, \ldots, N-1 \tag{6.2.8}$$

and for $j = 1, 2, \ldots, J$

$$\chi(k, R_j f) \cdot \left(W_j \eta(k) - \breve{W}_j f(k)\right) = 0 \quad k = 0, 1, \ldots, N-1 \tag{6.2.9}$$

$$\tilde{t}(k, R_j f) \cdot (W_j \eta(k+1) - W_j \eta(k)) \geq 0 \quad k = 0, 1, \ldots, N-1 \tag{6.2.10}$$

where $\breve{W}_j f(k)$ is introduced to justify the use of values which do not appear in the maxima representation. The specific values of $\breve{W}_j f(k)$ for $k \in (XW_j f \cup YW_j f)^c$ are not important, because they are multiplied by $\chi(k, R_j f)$, which is zero in this case. For a formal definition, let us write:

$$\breve{W}_j f(k) \triangleq \begin{cases} W_j f(k) & \text{if } k \in XW_j f \cup YW_j f \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.11}$$

The cost function $v(\eta)$ is calculated from conditions (6.2.8), (6.2.9), (6.2.10). Observe that, for a given j, the related conditions can be calculated using $W_j \eta$ and $R_j f$, which will be called the local information. Toward the objective of defining rigorously the jth part of the cost function, let $\breve{R} = (\breve{X}, \breve{Y}, \breve{T})$ be a generic one-level maxima representation. In other words, $\breve{X}, \breve{Y}$ are sets of numbers from $\{0, 1, \ldots, N-1\}$ and $\breve{T}$ is a sequence of $|\breve{X}|+|\breve{Y}|$ real numbers. Observe that $\chi(k, \breve{R})$, $\tilde{t}(k, \breve{R})$ are well defined for any such $\breve{R}$. Let $\breve{T} = \{\breve{T}(k)\}_{k=0}^{N-1}$ be a sequence obtained from $\breve{T}$ by filling missing values with zeros. Its formal definition is as follows:

$$\breve{T}(k) \triangleq \begin{cases} \breve{T}\left(order(k, XW_j f \cup YW_j f)\right) & \text{if } k \in XW_j f \cup YW_j f \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.12}$$

Where $order(k, XW_j f \cup YW_j f)$ is the index of k in ordered list consisting of elements from $XW_j f \cup YW_j f$.

For any $y \in \mathcal{L}$, we define the local cost function $\nu(y, \breve{R})$ as follows.

$$\nu(y, \breve{R}) \triangleq \sum_{k=0}^{N-1}$$

$$\left(\chi(k, \breve{R}) \cdot \left(y(k) - \breve{T}(k)\right)^2 + \chi_n\left(\tilde{t}(k, \breve{R})\left(y(k+1) - y(k)\right)\right) \cdot \left(y(k+1) - y(k)\right)^2\right) \tag{6.2.13}$$

where $\chi_n(\xi)$ is a characteristic function of $\xi \leq 0$, i.e.

$$\chi_n(\xi) \triangleq \begin{cases} 1 & \text{if } \xi \leq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6.2.14}$$

Recall that we note interchangeably column vectors from $\Re^N$ and sequences from $\mathcal{L}$. The coefficient $\chi(k, \breve{R})$ is y-independent and $\chi_n\left(\tilde{t}(k, \breve{R})\left(y(k+1) - y(k)\right)\right)$ is a piecewise constant function of $y \in \Re^N$. Therefore, the local cost function, $\nu(y, \breve{R})$, can be referred to as a piecewise quadratic form of $y \in \Re^N$.

In the sequel we will use a uniform notation for S and W operators. Since J is considered constant in this thesis, we will note $W_{J+1}\eta \triangleq S_J\eta$. After this long introduction, we are now able to write a short definition of the cost function.

Proposition 17 *The cost function of an arbitrary $\eta \in \mathcal{L}$, with regard to the*

multiscale maxima representation $R_m f$, is given by the following formula:

$$v(\eta) = \sum_{j=1}^{J+1} \nu(W_j\eta, R_j f) \tag{6.2.15}$$

Notice that $v(\eta)$ is a piecewise quadratic form in $\eta \in \Re^N$.

Recall that $v(\eta)$ accounts, with quadratic cost, for all constraints required to have $\eta \in \overline{\Gamma(R_m f)}$. Equation (6.2.15) is a straightforward consequence of conditions (6.2.8),(6.2.9), (6.2.10) and notations and definitions introduced afterwards. In the case that the reader is confused by the amount of notations and definitions recently introduced, we would suggest to track back from equation (6.2.15) to conditions (6.2.8),(6.2.9),(6.2.10).

From equation (6.2.15) we can see that the calculation of the cost function consists of three steps:

- calculate the multiscale decomposition $\{W_j\eta\}_{j=1}^{J+1}$.
- calculate the local cost function $\nu(W_j\eta, R_j f)$ by scanning the sequence $\{W_j\eta(k)\}_{k=0}^{N-1}$ against the j-th level maxima representation $R_j f$.
- sum up the resulting local cost functions

The algorithmic complexity of the last two steps is $O(NJ)$, therefore it does not exceed the algorithmic complexity of calculating the multiscale decomposition. We will show that a similar statement is true for the calculation of the cost function gradient, $\nabla v(\eta)$.

Equation (6.2.15) yields

$$\nabla v(\eta) = \sum_{j=1}^{J+1} \nabla \nu(W_j, R_j f). \tag{6.2.16}$$

Let $\mathbf{W}_j$ be a matrix corresponding to the linear operator W_j, namely:

$$W_j \eta = \mathbf{W}_j \cdot \eta \tag{6.2.17}$$

Now, let $\nu_y(y, \breve{R})$ denote the column vector of derivatives of ν with respect to components of the vector y, i.e.

$$\nu_y(y, \breve{R})(k) = \frac{\partial \nu(y, \breve{R})}{\partial y(k)} \tag{6.2.18}$$

$\nu_y(y, \breve{R})$ will be called the local gradient.

The k-th component of the gradient $\nabla \nu(W_j \eta, \breve{R})$ (with respect to η) is given by:

$$\nabla \nu(\eta)(k) = \frac{\partial \nu(W_j \eta, \breve{R})}{\partial \eta(k)} = \sum_{p=0}^{N-1} \frac{\partial \nu(y, \breve{R})}{\partial y(p)} \cdot \frac{\partial y(p)}{\partial \eta(k)} \tag{6.2.19}$$

Since $y = \mathbf{W}_j \eta$, $\frac{\partial y(p)}{\partial \eta(k)} = \mathbf{W}_j(p, k)$, where $\mathbf{W}_j(p, k)$ denotes the (p, k) entry of the matrix $\mathbf{W}_j$. Substituting this term into (6.2.18) provides:

$$\nabla \nu(\eta)(k) = \sum_{p=0}^{N-1} \mathbf{W}_j(p, k) \cdot \frac{\partial \nu(y, \breve{R})}{\partial y(p)}$$

and then we obtain the following column vector notation.

$$\nabla \nu(\eta) = \mathbf{W}_j' \cdot \nu_y(y, \breve{R}) \tag{6.2.20}$$

The local gradient which is the derivative of ν with respect to y is calculated directly from equation (6.2.13). The result is:

$$\nu_y(y, \breve{R})(k) = 2\chi(k, \breve{R}) \cdot \Big(y(k) - \breve{T}(k)\Big) -$$

$$- 2\chi_n \left(\tilde{t}\left(k, \check{R}\right)(y(k+1) - y(k))\right) \cdot (y(k+1) - y(k)) +$$

$$+ 2\chi_n \left(\tilde{t}\left(k-1, \check{R}\right)(y(k) - y(k-1))\right) \cdot (y(k) - y(k-1)) \qquad (6.2.21)$$

Summarizing, the gradient of the cost function $v(\eta)$ is given by the following formula:

$$\nabla(\eta) = \sum_{j=1}^{J+1} \mathbf{W}_j' \cdot \nu_y(W_j\eta, R_j f). \qquad (6.2.22)$$

The gradient calculation consists of four steps:

- calculate the decomposition $\{W_j\eta\}_{j=1}^{J+1}$
- calculate local gradients $\nu_y(W_j\eta, R_j f)$
- calculate $\mathbf{W}_j' \cdot \nu_y(W_j\eta, R_j f)$
- sum up results for $j = 1, 2, \ldots, J+1$

Let W_j' be the linear operator corresponding to the matrix $\mathbf{W}_j'$. Of course, if the linear operator W_j is implemented by a matrix operation , the complexity of calculating $W_j'\eta$ is the same as of $W_j\eta$. Moreover, if W_j is implemented by convolution with, say, a filter $h(k)$, then W_j' can be implemented by convolution with the filter $\tilde{h}(k) = h(-k)$ and again the same complexity is involved. In the wavelet transform case, W_j' corresponds to the inverse wavelet transform and then the gradient calculation has an attractive interpretation:

- calculate the transform
- calculate the local gradients

- calculate the inverse transform

In view of the above considerations, we can conclude that the algorithmic complexity of calculating the gradient of the cost function is of the same order as calculating the decomposition $\{W_j\eta\}_{j=1}^{J+1}$.

As was already mentioned, any minimization of the cost function can be considered as a reconstruction algorithm. For large and nonlinear systems, the standard approach is to use an iterative algorithm. Most of the relevant algorithms can be identified as belonging to one of the following categories: Basic Descent, Conjugate Directions and Quasi-Newton methods. See [28] for an excellent description of the most important methods. Basically, every such algorithm performs two-phase steps:

- find a direction to move (usually it is based on the gradient)
- line search - find a relative minimum (or its approximation) of the cost function along the chosen direction

Due to the "piecewise" quadratic structure of the function $v(\eta)$, many interesting methods for both precise and approximate search can be considered. However, we have learned from the simulations that different methods provide very similar performances.

We have started with the precise search along the gradient, i.e. we have used the following recursion formula:

$$\eta_{k+1} = \eta_k - \alpha_k \cdot \nabla v\left(\eta_k\right) \tag{6.2.23}$$

where

$$\alpha_k = \arg \min_{0 \leq \alpha < \infty} v\left(\eta_k - \alpha_k \cdot \nabla v(\eta_k)\right). \tag{6.2.24}$$

Then using coarser and coarser approximations for line search along the gradient, surprisingly, we concluded with the following algorithm:

$$\eta_{k+1} = \eta_k - \alpha \cdot \nabla v\left(\eta_k\right) \tag{6.2.25}$$

where α is chosen from $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\}$ as a largest number satisfying:

$$v\left(\eta_k - \alpha_k \cdot \nabla v(\eta_k)\right) \leq v(\eta_k).$$

Since development of an efficient, reliable, and generic reconstruction algorithm is not the major concern of this dissertation and since the above scheme performs satisfactory in the study cases, an algorithm based on equation (6.2.25) is used in the following section. In practice, the search for α was extremely simple, because in all experiments reported here α was either 0.5 or 0.25.

6.3 The rate of convergence

Although we have shown convergence results of the reconstruction algorithm, we are not able to analyze the rate of convergence. In spite of the fact that $v(x)$ is a piecewise quadratic form and for quadratic forms there are well known results about the bounds of the rate of convergence, applicability of these results to this dissertation is really problematic. First, for different regions we have different quadratic forms, and the algorithm usually switches between them in a way

which is difficult to predict. Second, even for one region, the exact structure of the quadratic form depends on the particular representation and we did not find any related general characteristic. Moreover, for a reconstruction algorithm, the problem of rate of convergence is twofold: first we can ask the standard question of how fast does the cost function $v(\eta_k)$ decrease. The second, and more important, question is how fast and in what manner does the sequence $\{x_k\}$ approach the set $\overline{\Gamma}$. Unfortunately, bounds, used in the proof of Theorem 12, are too loose for practical implementations. In view of these considerations, the only source to learn about the convergence rate is experimental data.

We have tested many analytical signals and it was apparent that a convergence pattern is very similar for different signals. In this section, we will present two signals whose wavelet maxima representation is already known to the reader. Recall the example of a unique wavelet maxima representation, described in Section 4.4.1. Let us denote it by f_u.

$$f_u(k) = sin(\frac{2\pi k}{256}(3(\frac{\mid k \mid}{128})^5 + 1)) \qquad k = 0, 1, \ldots, 255. \tag{6.3.1}$$

The signal $f_n(k)$ is the well-investigated example of nonunique wavelet maxima representation (see Section 4.4.2).

$$f_n(k) = cos(\frac{2\pi k}{8} + \frac{\pi}{6}) \qquad k = 0, 1, \ldots, 255. \tag{6.3.2}$$

We use the cubic spline wavelet, described in Section 4.3 and consider reconstruction from both $R_m f_u$ and $R_m f_n$. Figure 6.1 describes the behavior of the cost function $v(\eta_k)$ as a function of the number of iterations k. First, observe

that for the first 100 steps, both cases look very similar and achieve approximately $v(\eta) = 10^{-3}$. At this point, the relative improvement for $R_m f_n$ becomes extremely slow, while for $R_m f_u$ the improvement is still significant.

For the next graph we need to define the noise to signal ratio, $\frac{N}{S}$.

$$\frac{N}{S} \triangleq \frac{\|\eta_k - f\|^2}{\|f\|^2}$$

Figure 6.2 describes the behavior of $\frac{N}{S}$ as a function of k. As expected, the reconstruction from $R_m f_u$ is much more precise, and after about 100 steps we achieve $\frac{N}{S} = 3 \cdot 10^{-3}$, observe that in order to have $\frac{N}{S} = 10^{-2}$, it suffices to use only 30 steps. On the other hand, for $R_m f_n$ we hardly reach $\frac{N}{S} = 5 \cdot 10^{-2}$. However, the latter should be considered a satisfactory result because for the signal f_a, described in Section 4.4.2 and belonging to the reconstruction set, the noise to signal ratio is 0.3458. In order to visualize better the quality of the reconstruction, Figure 6.3 describes the reconstructed signals, after 100 iterations and after 4000 iterations. We can see that all signals are similarly shaped and appear to belong to the reconstruction set. For this case, the reconstruction set was calculated precisely, hence we can find the distance from a given reconstructed signal to the set $\Gamma(R_m f_n)$. It turns out that after 100 steps, the ratio $\frac{N}{S}$, where N corresponds to the component of the reconstructed sequence which is outside $\Gamma(R_m f_n)$ is below 0.01.

Our conclusion from the experimental data is twofold:

- it is very easy, robust, and fast to get results corresponding to $\frac{N}{S} \approx 10^{-2}$

- it is extremely difficult, sometimes even impossible, to get very accurate (say $\frac{N}{S} \approx 10^{-7}$) results .

Therefore, we would recommend to use this type of representation in cases where the required quality corresponds to $\frac{N}{S}$ in the range $10^{-1} \div 10^{-3}$. Fortunately, most engineering applications in important areas like speech analysis, computer vision, data fusion, etc. require this sort of quality.

In the next chapter, we will discuss how one can make multiscale maxima representation much more flexible and more compact while preserving the ability to reconstruct signals with $\frac{N}{S} = 10^{-1} \div 10^{-3}$.

continuous line - case of $R_m f_n$
dashed line - the case of $R_m f_u$

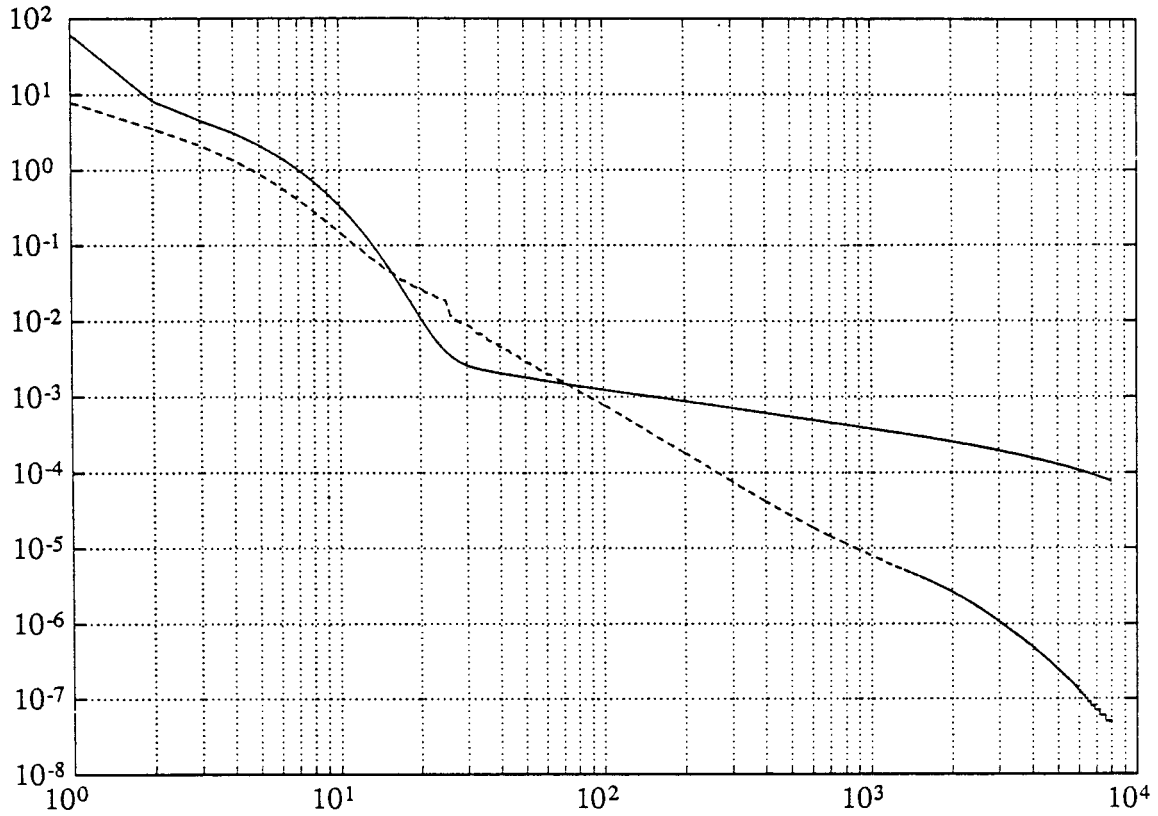


Figure 6.1: Cost function $v(\eta_k)$ as a function of k

continuous line - case of $R_m f_n$
dashed line - the case of $R_m f_u$

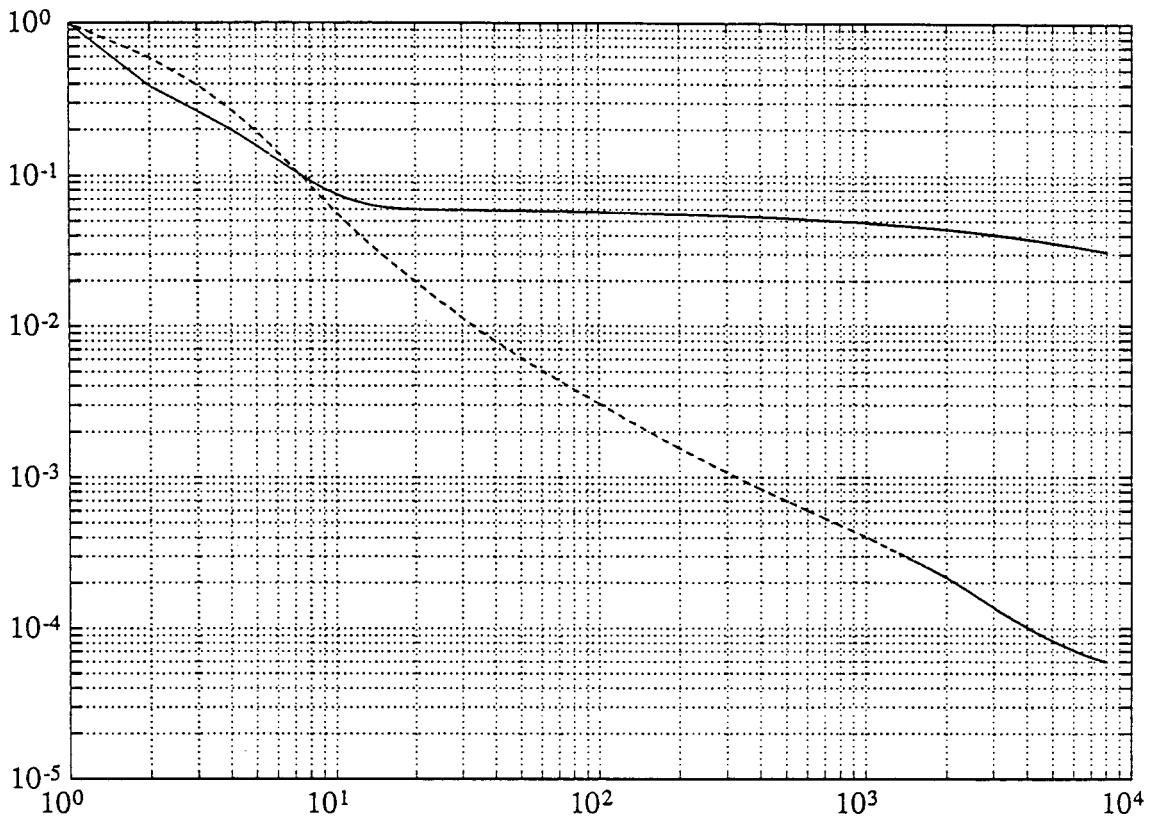


Figure 6.2: The noise to signal ratio, $\frac{N}{S}$, as a function of k

continuous line	-	original signal f_n
dashed line	-	the reconstructed signal after 100 iterations
dotted line	-	the reconstruction signal after 4000 iterations

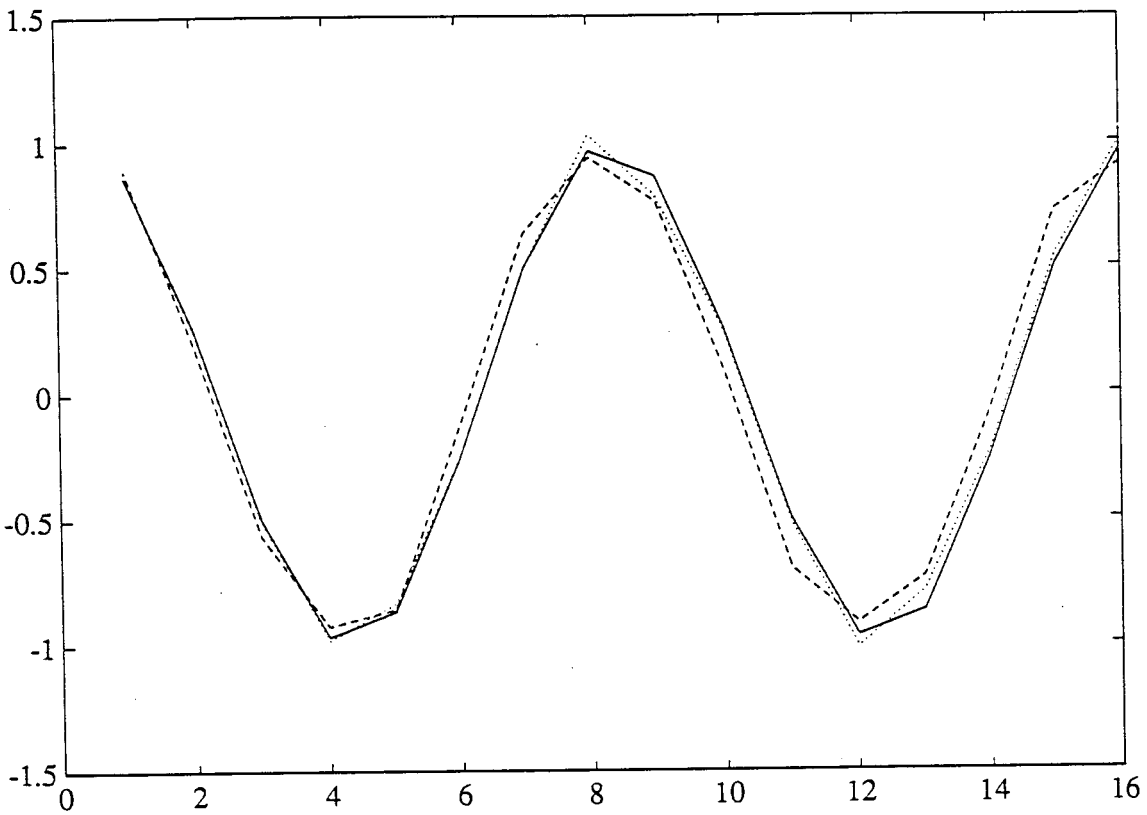


Figure 6.3: The original and reconstructed signals

CHAPTER
SEVEN

MODIFICATIONS IN MAXIMA REPRESENTATIONS

7.1 Introduction

All previous chapters concerned with analysis of the basic form of signal representations proposed by S. Mallat and S. Zhong [30] and S. Mallat [29]. As a side benefit of this mathematical analysis, structural attributes required to attain the described stability and reconstruction characteristics have been well understood. This knowledge enables us to introduce many modifications while preserving the desired properties within the framework of inherently bounded AQLR's.

Our main objective is to create a structure allowing a trade-off between the amount of information required for representation and the reconstruction quality. We have followed the subsequent guidelines.

- **Create**
  - the ability to add and delete information.

- **Preserve**
  - the structure of the inherently bounded AQLR;
  - the multiscale structure;
  - the algorithmic complexity.

It turns out that there is an abundance of possible modifications. The aim of this chapter is to describe several examples of reasonable modifications to exhibit the scope of flexibility in the resulting representations and to encourage creativity for further modifications.

7.2 Improvement in the representation quality

Usually, the basic multiscale maxima representation provides an excellent reconstruction quality, but there are rare cases, like the nonunique example from Section 4.4.2, that some improvement might be required. In this case, one may try to use the basic maxima representation based on an alternative filter bank $\{\acute{W}_1, \acute{W}_2, \ldots, \acute{W}_{J'}, \acute{S}_{J'}\}$. The interesting and important issue of how to find the "best", or even an "appropriate", filter bank is beyond the scope of this dissertation. In this chapter, we assume that the multiscale operator $\{W_1, W_2, \ldots, W_J, S_J\}$ is fixed and we focus on structural modifications in the basic maxima representation.

In order to improve the quality of the representation we propose to add new sampling points. For $j = 1, 2, \ldots, J$, let $\mathcal{F}_j$ be a subset of $\{0, 1, \ldots, N-1\}$. We

define the following representation:

$$R^e f = \left\{R_j^e f\right\}_{j=1}^{J+1} \tag{7.2.1}$$

where

$$R_j^e f \triangleq \left\{XW_j f, YW_j f, \{W_j f(k)\}_{k\in XW_j f \cup YW_j f}, \{W_j f(k)\}_{k\in\mathcal{F}_j}\right\} \tag{7.2.2}$$

Notice that the definition of the representation is not complete without defining the reconstruction set. In this case the definition is as follows. An arbitrary $\eta \in \mathcal{L}$ belongs to $\Gamma(R^e f)$ if and only if:

$$S_J\eta(k) = S_J f(k) \quad k = 0, 1, \ldots, N-1 \tag{7.2.3}$$

and for $j = 1, 2, \ldots, J$

$$W_j\eta(k) = W_j f(k), \quad k \in XW_j f \cup YW_j f \cup \mathcal{F}_j \tag{7.2.4}$$

$$t_j^{mf}(k) \cdot (W_j\eta(k+1) - W_j\eta(k)) \geq 0 \tag{7.2.5}$$

$$k \in (XW_j f \cup YW_j f)^c \cup (XW_j f \cup YW_j f)^r.$$

From the definitions we see that:

Proposition 18 *The multiscale representation $R^e f$ is an inherently bounded AQLR.*

Observe that from the above definition, it is straightforward to define the local cost function $\nu(y, R_j^e f)$ (with $y = W_j\eta$). From the local cost function, the local gradient $\nu_y(y, R_j^e f)$ can easily be calculated. All the remaining components

of the reconstruction algorithm are the same as of the algorithm described in Section 6.2.

In general, the question of how to choose the "best" sets $\mathcal{F}_j$ may lead to very exciting research. The reason that we did not pursue further the investigatation of this problem is twofold:

- its analysis seems to be very difficult;
- the need for improvement in the representation is very exceptional.

7.3 Reductions in $S_J f$

In the basic maxima representation, usually, a greater part of samples belong to the sequence $\{S_J f(k)\}_{k=0}^{N-1}$. In the wavelet transform case, $S_J f$ is a significantly blurred version of f and the whole sequence $\{S_J f(k)\}_{k=0}^{N-1}$ appears to contain redundant information. Some reduction in the number of samples devoted to $S_J f$ should not reduce the reconstruction quality.

It turns out, that using downsampling of $S_J f$ we can still attain the desired properties. A version of $S_J f$, downsampled at rate Δ, is defined as follows:

$$S_J^{\Delta} f \triangleq \{S_J f(k)\}_{k=0,\Delta,2\Delta,\ldots} \tag{7.3.1}$$

Let $R^{\Delta} f$ be a basic representation in which $S_J f$ was replaced by $S^{\Delta} f$, namely:

$$R^{\Delta} f \triangleq \left\{\{R_j f\}_{j=1}^{J}, S_J^{\Delta} f\right\} \tag{7.3.2}$$

Certainly, $R^{\Delta} f$ is an AQLR. From the discussion in Section 2.2, we know that if $\{W_1, W_2, \ldots, W_J, S_J^{\Delta}\}$ is a complete multiscale operator then $R^{\Delta} f$ is an

inherently bounded AQLR.

Fact

Let J=5, N=256, and consider the discrete wavelet transform based on the cubic spline. Then $\{W_1, W_2, \ldots, W_J, S_J^\Delta\}$ is a complete multiscale operator for $\Delta = 1, 2, 4, 8, 16, 32$.

Proof

It suffices to check, for $\Delta = 32$, the rank of the composite matrix

$$[\mathcal{W}_1' \vdots \mathcal{W}_2' \vdots \cdots \vdots \mathcal{W}_J' \vdots (\mathcal{S}_J^\Delta)']$$

A numerical test shows that , indeed, for $\Delta = 32$ the above composite matrix has a full rank.

□

All examples in this chapter were calculated using the cubic spline wavelet, J=5, and N=256. We will show several reconstructions from different modifications, all with the same original sequence f defined as follows:

$$f(k) = \sin\left(\frac{6\pi k}{256}\right) \cdot e^{-\frac{k}{128}} \quad k = 0, 1, \ldots, 255. \tag{7.3.3}$$

Figure 7.1 describes the behavior of the noise to signal ratio, during the reconstruction from $R^\Delta f$, for two different cases: $\Delta = 1$, which correspond to the basic representation, and $\Delta = 32$. We see that for the first 20 steps and $\frac{N}{S} \approx 2 \cdot 10^{-2}$, both cases are very similar. For a better $\frac{N}{S}$ ratio, the case of $S_J^\Delta f$ converges slower.

continuous line - $\Delta = 1$
dashed line - $\Delta = 32$

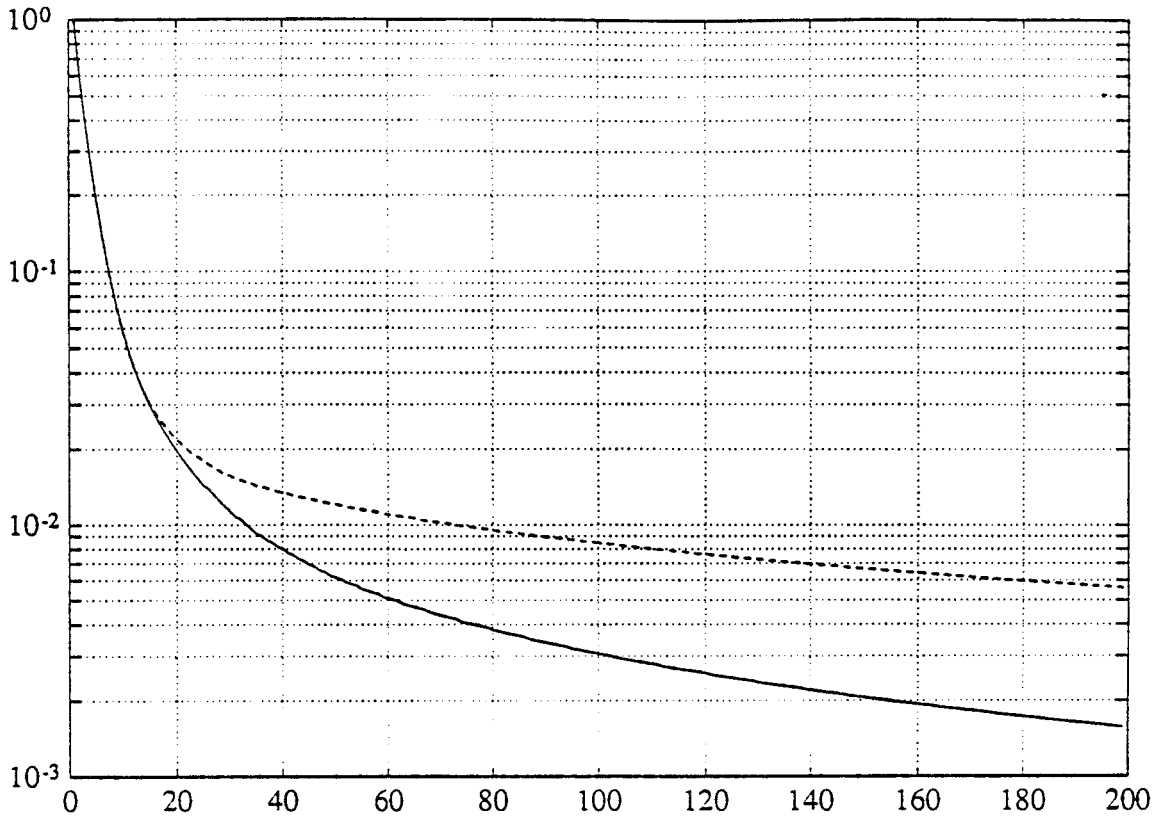


Figure 7.1: $\frac{N}{S}$ for the reconstruction from $R^{\Delta}f$

An alternative approach to reduce the amount of the information related to $S_J f$ is to describe this sequence by its local extrema.

Proposition 19 *Let*

$$R^x f \triangleq \left\{ \{R_j f\}_{j=1}^{J}, XS_J f, YS_J f, \{S_J f(k)\}_{k \in XS_J f \cup YS_J f} \right\} \tag{7.3.4}$$

then $R^x f$ is an inherently bounded AQLR.

Proof:

$R^x f$ is a standard maxima representation for $\{W_1, W_2, \ldots, W_J, S_J, \emptyset\}$ where $\emptyset$ represents an "empty" operator e.g.

$$\emptyset\eta(k) = 0 \quad k = 0, 1, \ldots, N-1.$$

Since $\{W_1, W_2, \ldots, W_J, S_J\}$ is complete, the multiscale operator $\{W_1, W_2, \ldots, W_J, S_J, \emptyset\}$ is complete as well.

□

For the tested cases, the performance of reconstruction algorithms, for $R^x f$ and for $R^\Delta f$ are similar.

7.4 Quantization

Up to this point we have assumed that we are able to preserve the exact values of the linear part of the representation. In practice, only floating point data types match this assumption. In many situations, for example in data communication or in data storage, an efficient data compression is required. In these cases, continuous values are approximated with a relatively low precision.

Quantization is a standard technique for this problem. It is based on a set $\mathcal{D}$ defined as follows.

$$\mathcal{D} \triangleq \{d(0), d(1), d(2), \ldots, d(L), d(L+1)\} \tag{7.4.1}$$

such that

$$-\infty = d(0) < d(1) < d(2) < \cdots < d(L) < d(L+1) = \infty. \tag{7.4.2}$$

A quantizer $(\cdot)_q$ is defined as a mapping from $\Re$ to the set $\{0, 1, \ldots, L, L+1\}$ such that, for r in $\Re$

$$(r)_q \triangleq \begin{cases} i & \text{if } r \leq d(i) \text{ and } r > d(i-1) \\ 0 & \text{if } r < d(1). \end{cases} \tag{7.4.3}$$

Now, let us assume that instead of having $Rf = \{R_j f\}_{j=1}^{J+1}$ where

$$R_j f = \left\{ XW_j f, YW_j f, \{W_j f(k)\}_{k \in XW_j f \cup YW_j f} \right\},$$

we are given

$$(Rf)_q = \left\{ (R_j f)_q \right\}_{j=1}^{J+1} \tag{7.4.4}$$

with:

$$(R_j f)_q = \left\{ XW_j f, YW_j f, \left\{ d\left((W_j f(k))_q \right) \right\}_{k \in XW_j f \cup YW_j f} \right\}.$$

In this case, a standard approach would be to consider $(Rf)_q$ as an approximation of Rf and then simply use the original reconstruction scheme with $(Rf)_q$ as the input data. In other words, one can search for $\eta \in \mathcal{L}$ such that

$$R\eta = (Rf)_q \tag{7.4.5}$$

But, such η may not exist ! Nevertheless, one can propose to search for η whose multiscale maxima representation best approximate, in an appropriate sense, the given $(Rf)_q$. Although some of the desired shape preserving characteristics

may be lost, this is a feasible approach. It turns out that, without increasing complexity of the representation, a much better method can be developed. The idea is to change the definition of the representation so that the quantized version will always belong to the range of the representation.

The main observation is that, in our case, we can replace one equality constraint by two inequality constraints without changing significantly the complexity of reconstruction. Therefore, instead of requiring

$$W_j\eta(k) = d\left((W_jf(k))_q\right),$$

the constraints are:

$$W_j\eta(k) \leq d\left((W_jf(k))_q\right)$$

and

$$W_j\eta(k) \geq d\left((W_jf(k))_q - 1\right).$$

In words, as an alternative for reconstruction from approximate values, reconstruction from an interval, to which a precise value belongs, is proposed.

In order to define it rigorously, first let us denote this representation by R^qf. The structure of the representation is as follows:

$$R^qf \triangleq \left\{(R_jf)_q\right\}_{j=1}^{J+1}. \tag{7.4.6}$$

This representation is characterized by the following reconstruction set. An arbitrary $\eta \in \mathcal{L}$ belongs to $\Gamma(R^qf)$ if and only if

$$S_J\eta(k) \leq d\left((S_Jf(k))_q\right) \qquad k = 0, 1, \ldots, N-1$$

$$S_J\eta(k) \geq d\left((S_Jf(k))_q - 1\right) \qquad k = 0, 1, \ldots, N-1$$

$$W_j\eta(k) \leq d\left((W_jf(k))_q\right) \quad k \in XW_jf \cup YW_jf \tag{7.4.7}$$

$$W_j\eta(k) \geq d\left((W_jf(k))_q - 1\right) \quad k \in XW_jf \cup YW_jf \tag{7.4.8}$$

$$t_j^{mf}(k) \cdot (W_j\eta(k+1) - W_j\eta(k)) \geq 0 \tag{7.4.9}$$

$$k \in (XW_jf \cup YW_jf)^c \cup (XW_jf \cup YW_jf)^r.$$

From the above definition, we conclude:

Proposition 20 *The representation R^qf is an AQLR.*

In the case of underflow or overflow in data representation, the bounding property might be lost. Therefore, for the property of the inherently bounded AQLR, we arrive at the following result.

Proposition 21 *If*

$$d(1) \leq \min\{W_jf(k), S_Jf(k) : k = 0, 1, \ldots, N-1 \;\; j = 1, 2, \ldots, J\}$$

and

$$d(L) \geq \min\{W_jf(k), S_Jf(k) : k = 0, 1, \ldots, N-1 \;\; j = 1, 2, \ldots, J\},$$

then R^qf is an inherently bounded AQLR.

The problem of underflow and overflow might be treated much better by allowing an adaptive set $\mathcal{D}$. In fact, $\mathcal{D}$ was assumed fixed in order to keep the notation as simple as possible. It is clear that $\mathcal{D}$ may depend on any information

which is available during reconstruction, for example on the level j, on whether a given extremum is a maximum or a minimum, etc. Mallat and Zhong [30] pointed out a different significance of maxima and minima points. Accordingly it seems to be very reasonable to use different quantizers for maxima and minima points. In [30], small extrema values are disregarded. Observe that this operation also can be viewed as a part of the quantization. In addition the set $\mathcal{D}$ may depend on the original signal f, in this case, some information about the chosen set should be added to the representation.

In view of the above considerations we are naturally led to the issue of how to design an appropriate set $\mathcal{D}$. General analysis appears to be very difficult and we can only propose to focus on a specific application and to perform experimental studies.

In this dissertation we show the basic feasibility of the algorithm. We have tested the following version of the representation. Let $R^q f = \left\{(R_j f)_q\right\}$ be the given representation with the following reconstruction set:

$$\eta \in \mathcal{L} \quad \text{belongs to} \quad \Gamma(R^q f)$$

if and only if:

$$S_J\eta(k) \leq S_J f(k) + q \cdot \alpha_{J+1} \qquad k = 0, 1, \ldots, N-1 \tag{7.4.10}$$

$$S_J\eta(k) \geq S_J f(k) - q \cdot \alpha_{J+1} \qquad k = 0, 1, \ldots, N-1 \tag{7.4.11}$$

$$W_j\eta(k) \leq W_j f(k) + q \cdot \alpha_j \quad k \in XW_j f \cup YW_j f \tag{7.4.12}$$

$$W_j\eta(k) \geq W_jf(k) - q \cdot \alpha_j \quad k \in XW_jf \cup YW_jf \tag{7.4.13}$$

$$t_j^{mf}(k) \cdot (W_j\eta(k+1) - W_j\eta(k)) \geq 0 \tag{7.4.14}$$

$$k \in (XW_jf \cup YW_jf)^c \cup (XW_jf \cup YW_jf)^r.$$

The normalizing factors α_j are numbers taken from $\{\ldots, 0.2, 0.5, 1.0, 2.0, 5.0, \ldots\}$. They are chosen to be as large as possible but smaller then the global maxima of the absolute value of the decomposed signal $W_jf(k)$. The coefficient q describes the available number of bits per sample. In the following examples, the values $q = 0, 0.01, 0.04$ are used. Loosely speaking, $q = 0.01$ corresponds to 8 bits per sample, while $q = 0.04$ refers to 6 bits per sample. Figure 7.2 describes reconstruction from R^qf by showing the noise to signal ratio as a function of iteration number. As expected, if q increases then the ratio $\frac{N}{S}$ increases as well. For a given requirement for $\frac{N}{S}$, one can find the appropriate q. Recall that due to shape preserving properties, probably, relatively large noise to signal ratios are acceptable.

The last graph, described here, refers to the reconstruction from the representation $R^{\Delta,q}f$, in which, first S_J was reduced by the factor 32 and then quantization was performed with $q = 0.01$. Also here we see that degradation in performances, for some cases, may be acceptable.

The quantization scheme described above can be viewed as modifications of equality constraints related to the reconstruction set. Similarly, inequality

constraints of the form

$$t_j^{mf}(k) \cdot (W_j\eta(k+1) - W_j\eta(k)) \geq 0$$

can be modified. For example, one can consider threshold-like conditions on a sequence variation on a given interval. However, since we do not have tools to analyze quality of these representations, we leave these ideas to applications oriented experimental studies.

continuous line	-	$q = 0$
dashed line	-	$q = 0.01$
'+' line	-	$q = 0.04$

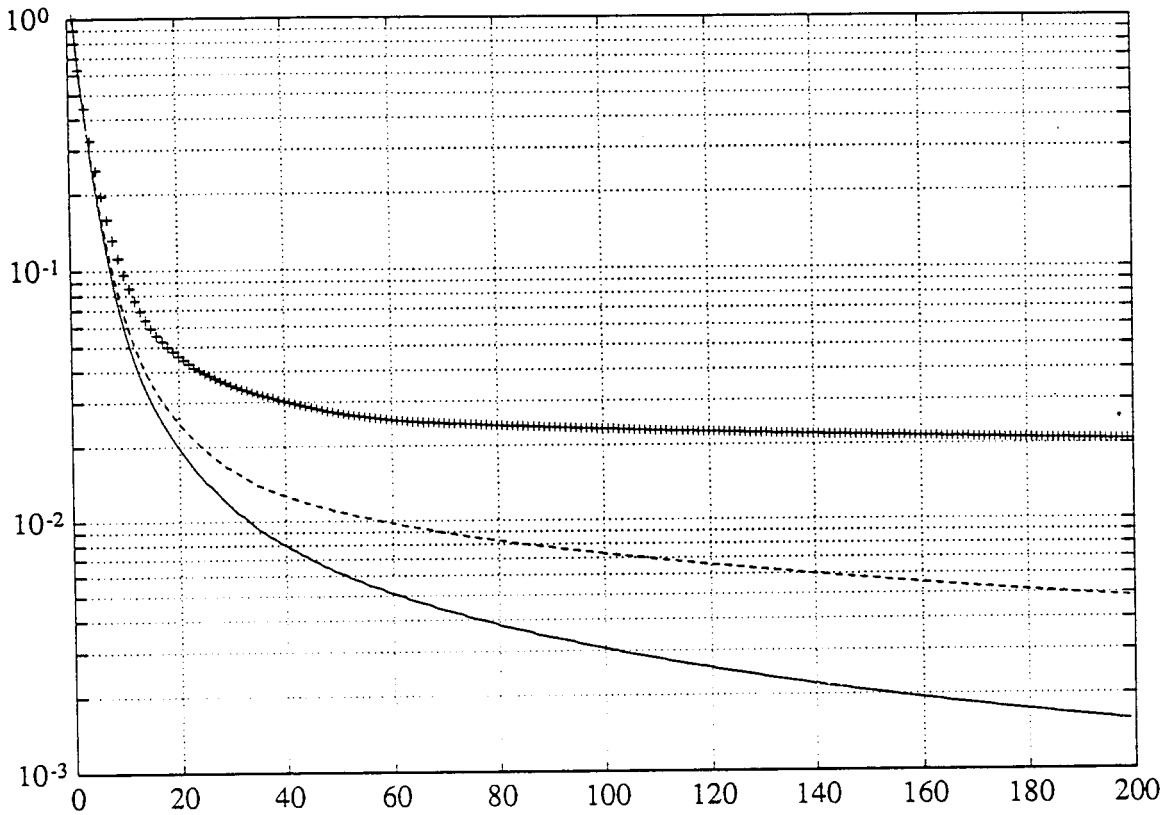


Figure 7.2: $\frac{N}{S}$ for reconstruction from $R^q f$

continuous line - $q = 0$ and $\Delta = 1$
dashed line - $q = 0.01$ and $\Delta = 32$

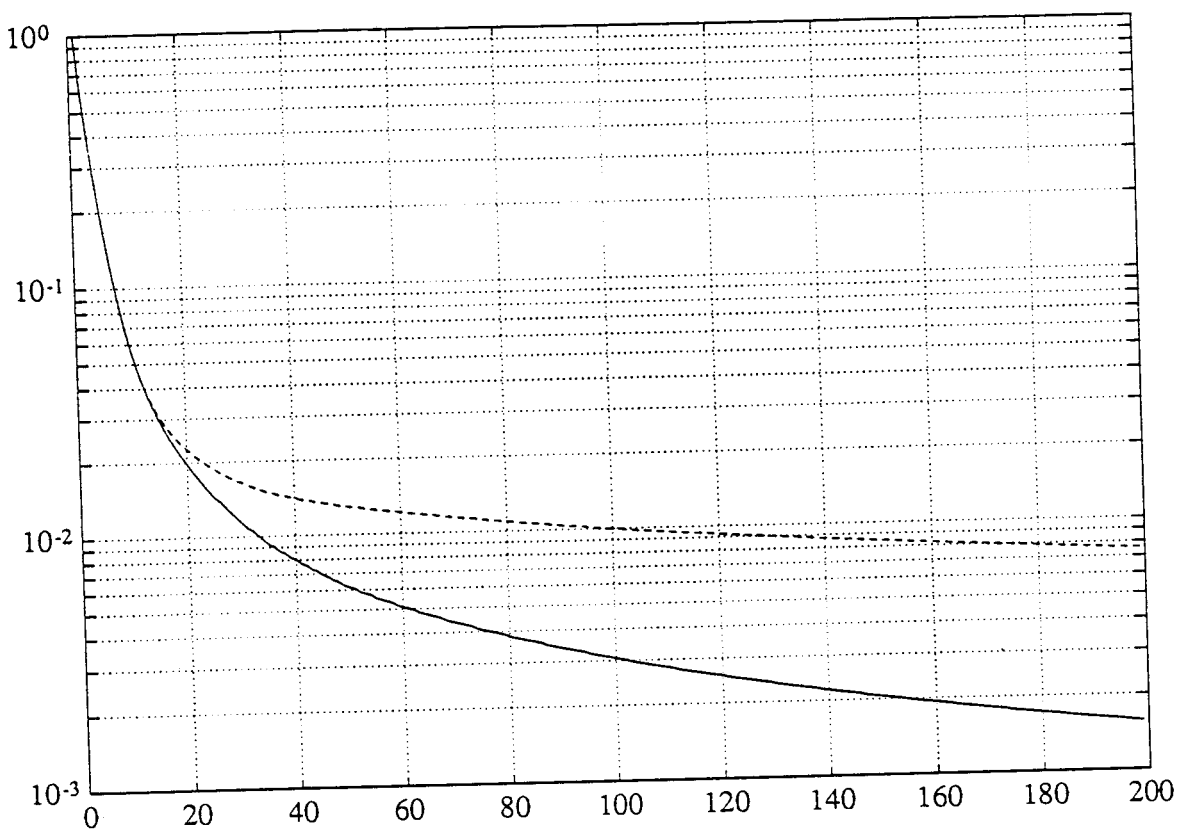


Figure 7.3: $\frac{N}{S}$ for reconstruction from $R^{\Delta,q} f$

CHAPTER
EIGHT

CONCLUSIONS

8.1 Summary

New theoretical results[1], regarding uniqueness and stability of the multiscale (wavelet) maxima and zero crossings representations have been presented. The concluding result, which states that the wavelet maxima (zero-crossings) representation is stable but nonunique, provides a new consideration of these signal descriptions. The standard multiscale zero-crossings (without any additional information) representation was assumed, at least for some family of signals, unique but unstable. Perhaps, this instability was the main obstacle to achieve, in spite of the first enthusiasm for zero-crossings in multiresolution representations (in the early 1980's), many engineering applications of this technique.

The new reconstruction scheme, based on the descent algorithm, appears to be less complex than alternate convex projections while it provides very similar experimental results. If, for a specific application, building a special purpose

[1]See Section 1.4 for a detailed list of results.

hardware is justified then an efficient, neural network-like, analog hardware circuit might yield a very fast reconstruction algorithm.

Summarizing, a new analysis, accompanied with analytical tools for a promising family of representations, has been presented. However, it is premature to predict how much of this promise can indeed be accomplished. The critical question of characterizing the class of real engineering problems that can adequately be solved by this approach still requires additional research.

8.2 General remarks

In the foregoing discussion, several general comments are pointed out. In our opinion, in addition to the actual results, there are three important consequences of this work.

The first is to show feasibility and capability of discrete analysis. In general, the discrete approach described here may be applied for a variety of representations and reconstruction algorithms, providing new insights into their properties. We believe that, even for complex algorithms, testing for uniqueness and computing a precise reconstruction set, even for a few examples, is worth the effort.

The second is the conclusion that, in order to benefit from novel characteristics, beyond properties of an adaptive irregular sampling, the multiscale maxima (zero-crossings) representation should be considered in the nonunique context. Signal processing based on a unique representation has the advantage of possible separation between different processing units. This separation facilitates

significantly analysis and design, but it does not improve system performance. Therefore, since the use of the multiscale maxima (zero-crossing) representation will require joint analysis and design of a whole system, one might expect involved analysis and difficult design with possible, as usual when local optimization is replaced by a global one, improvement in performance. However, in the signal processing community, the core of theoretical studies has been developed in the framework of unique representations. In our opinion, the need to develop more analytical tools and applications for nonunique representations is apparent.

The third is the framework of inherently bounded AQLR's. Observe, that this structure makes possible to define, analyze, and reconstruct a wide family of signal representations. Important examples are modifications of the basic maxima representation, some of them were shown in Chapter 7. Design and analysis methods for these modifications require further research.

Let us conclude with a citation from [22].

"The general methodology of studying a representation in terms of its mathematical properties, and developing reconstruction methods to evaluate the stability and variations in the fibers, in analogy with the study undertaken here, is highly recommended."

Hummel and Moniot meant, by using the word "here", their work, but we hope that they would agree to use it in the context of our dissertation as well.

8.3 Further research

From the theoretical point of view, there are still many interesting open questions concerning the discrete analysis of the multiscale edge representation. Consider the following, partial, list of problems requiring further research.

- What is the family of signals for which the wavelet maxima (zero-crossings) representation is unique ?
- What kind of information should be added to the dyadic wavelet maxima (zero-crossings) representation in order to assure general uniqueness ?
- Is multiscale zero-crossings points [2] representation indeed unstable ?
- If it is, what is the minimal additional information which stabilizes it ?

As a first step in the undergoing research, this work dealt only with one dimensional signals. The reason is twofold: firstly, we thought that in the simpler case the basic properties would be better recognizable, secondly, the one dimensional multiscale edge representation has its own variety of applications. One of the most promising application areas is speech analysis, for example, pitch detection [23] or modeling signal transformations in the auditory nervous system [46]. On the other hand, up to this point, the vast majority of multiscale edge representations has been implemented in computer vision. Therefore, it is

[2] without any additional information

advisable to extend these results for two dimensional signals. Surprisingly, there is an essential difference between maxima and zero-crossings representations. Two dimensional multiscale zero-crossings representation can easily be cast into the structure of inherently bounded AQLR, thus the related results are valid in this case. However, a two dimensional maxima representation appears to have a different structure. It is because the absolute value of the gradient, which is a nonlinear operation on an original signal, is involved in the two dimensional maxima representation. In order to proceed with a similar analysis one has to choose between:

- to extend the framework of the AQLR
- to change the definition of the two dimensional maxima representation to match the structure of the AQLR.

In both cases, further research is required.

As was already mentioned, there is a substantial need to study, first experimentally, many practical aspects of multiscale maxima and zero-crossings representations. In our opinion, one should consider, in the context of a specific application, the problem of how to design and analyze suitable modifications in the basic representations. Perhaps, in such a set-up, the shape preserving characteristics of the multiscale maxima representation might be better defined, understood, and analyzed.

A

A.1 The proof of Lemma 4

We need to show that for all ϵ there exists K such that

$$k > K \Rightarrow dist\left(x_k, \overline{\Gamma}\right) < \epsilon$$

By contradiction, let us assume that there exists $\epsilon > 0$ such that for all K there exists $k' > K$ such that $dist(x_{k'}, \overline{\Gamma}) \geq \epsilon$. We define iteratively the following sequence. Let k_0 be an arbitrary index such that $dist(x_0, \overline{\Gamma}) \geq \epsilon$. For a given k_i, we define k_{i+1} to satisfy:

$$k_{i+1} > k_i \quad \text{and} \quad dist(x_{k_{i+1}}, \overline{\Gamma}) \geq \epsilon \tag{1.1.1}$$

Since $\{x_k\}$ is bounded, $\{x_{k_i}\}_{i=0}^{\infty}$ is bounded as well. Therefore it has a convergent subsequence, let us denote it as $\{x_{k_i}\}_{i \in I}$. By the hyphothesis, $\{x_{k_i}\}_{i \in I}$ converges to a limit in $\overline{\Gamma}$, say $x \in \overline{\Gamma}$. But since $dist(x, \overline{\Gamma})$ is a continuous function of x, $\left\{dist(x_{k_i}, \overline{\Gamma}) - dist(x, \overline{\Gamma})\right\}_{i \in I}$ converges to zero. But this is a contradiction

since $dist(x, \overline{\Gamma}) = 0$ and $dist(x_{k_i}, \overline{\Gamma}) \geq \epsilon$. Thus, we conclude that indeed $\{x_k\}$ approaches $\overline{\Gamma}$.

BIBLIOGRAPHY

[1] Special issue on wavelet transforms and multiresolution signal analysis. IEEE Transactions on Information Theory, Vol 38, March 1992.

[2] J. J. Benedetto. Irregular sampling and frames. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications.* Academic Press, 1992.

[3] Z. Berman. The uniqueness question of discrete wavelet maxima representation. Technical Report TR 91-48r1, University of Maryland, System Research Center, April 1991.

[4] Z. Berman. A reconstruction set of a discrete wavelet maxima representation. In *Proc. IEEE International Conference on Acoust., Speech, Signal Proc.*, San Francisco, 1992.

[5] Z. Berman and J. S. Baras. An analysis of discrete zero-crossings and maxima wavelet representation. In *International Conference Wavelets and Applications*, Toulouse, 1992.

[6] Z. Berman and J. S. Baras. A study on discrete multiscale edge representations. In *Proc. 26th Annual Conference on Information Sciences and Systems*, Princeton, 1992.

[7] Z. Berman and J. S. Baras. Uniqueness and stability of discrete zero-crossings and maxima wavelet representations. In *Proc. Sixth European Signal Processing Conference*, Brussels, 1992.

[8] Z. Berman, J. S. Baras, and C. Berenstein. The parametric wavelet maxima representation. In *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, submitted*, Victoria, 1992.

[9] A. P. Calderon. Intermediate spaces and interpolation, the complex method. *Studia Math.*, 24:113–190, 1964.

[10] J. Canny. A comptutational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[11] C. K. Chui. *An Introduction to Wavelets.* Academic Press, 1992.

[12] C. K. Chui, editor. *Wavelets: A Tutorial in Theory and Applications.* Academic Press, 1992.

[13] J. M. Combes, A. Grossmann, and Ph. Tchamitchian, editors. *Wavelets.* Springer-Verlag, 1988.

[14] S. Curtis and A. Oppenheim. Reconstruction of multidimensional signals from zero crossings. *J. Opt. Soc. Amer.*, 4:221–230, 1987.

[15] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. on Pure and Applied Mathematics*, XLI:909–996, 1988.

[16] I. Daubechies. *Ten Lectures on Wavelets, CBMS-NSF Series in Applied Mathematics.* SIAM, 1992.

[17] M. B. Ruskai et al., editor. *Wavelets and Their Applications.* Jones and Bartlett Publishers, 1992.

[18] M. Frish and H. Messer. The use of the wavelet transform in the detection of an unknown transient signal. *IEEE Trans. on Information Theory*, 38(2):892–897, March 1992.

[19] T. Gal. *Postoptimal Analyses, Parametric Programming, and Related Topics.* McGraw-Hill, 1979.

[20] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.*, 15:723–736, 1984.

[21] A. Haar. Zur theorie der orthogonalen funktionen-systeme. *Math. Annal.*, 69:331–371, 1910.

[22] R. Hummel and R. Moniot. Reconstruction from zero crossing in scale space. *IEEE Trans. on ASSP*, 37(12):2111–2130, December 1989.

[23] S. Kadambe and G. F. Boudreaux-Bartels. A comparison of wavelet functions for pitch detection of speech signals. In *Proc. IEEE International Conference on Acoust., Speech, Signal Proc.*, Toronto, 1991.

[24] S. Kadambe and G. F. Boudreaux-Bartels. Application of the wavelet transform for pitch detection of speech signals. *IEEE Trans. on Information Theory*, 38(2):917–924, March 1992.

[25] H. S. Khalil. *Nonlinear Systems.* Macmillan Publishing Company, 1992.

[26] M. Krein and D. Milman. On extreme points of regularly convex sets. *Studia Math.*, 9:133–138, 1940.

[27] B. Logan. Information in the zero-crossings of band pass signals. *Bell Systems Tech. Journ.*, 56:510, 1977.

[28] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, 1989.

[29] S. Mallat. Zero-crossing of a wavelet transform. *IEEE Trans. on Information Theory*, 37(4):1019–1033, July 1991.

[30] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. Courant Institute of Mathematical Sciences, Technical Report 592, November 1991, to appear in IEEE Trans. on PAMI.

[31] S. Mallat and S. Zhong. Complete signal representation with multiscale edges. Courant Institute of Mathematical Sciences, Technical Report 483, December 1989.

[32] S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$. *Trans. of the AMS*, 315(1):69–87, September 1989.

[33] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on PAMI*, 11(7):674–693, July 1989.

[34] D. Marr and E. Hildreth. Theory of edge detection. *Proc. of the Royal Society of London*, 207:187–217, 1980.

[35] Y. Meyer. Ondelettes et applications. Ceremade et Institut Universitaire de France, preprint, 1992.

[36] B. Noble and J. W. Daniel. *Applied Linear Algebra.* Prentice-Hall, 1988.

[37] O. Rioul. Dyadic up-scaling schemes: simple criteria for regularity. *submitted to SIAM J. Math. Anal.*, 1991.

[38] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4):14–38, October 91.

[39] J. Sanz and T. Huang. Image representation by sign information. *IEEE Trans. Pattern. Anal. Machine Intell.*, 11:729–738, 1989.

[40] K. Seip. Wavelets in $H^2(\mathbf{R})$: Sampling, interpolation, and phase space density. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications.* Academic Press, 1992.

[41] S. Stearns and D. Hush. *Digital Signal Analysis.* Prentice Hall, New Jersey, 1990.

[42] J. Stoer and C. Witzgall. *Convexity and Optimization in Finite Dimensions I.* Springler Verlag, Berlin, 1970.

[43] M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *to appear IEEE Trans. on Signal Processing*, September 1992.

[44] K. G. Wilson and J. B. Kogut. *Physics reports*, 12C:77, 1974.

[45] A. Witkin. Scale-space filtering. In *Proc. 8th Int. Joint Conf. Artificial Intell.*, 1983.

[46] X. Yang, K. Wang, and S. A. Shamma. Auditory representation of acoustic signals. *IEEE Trans. on Information Theory*, 38(2):824–839, March 1992.

[47] W. I. Zangwill. *Nonlinear Programming: A Unified Approach.* Prentice-Hall, 1969.

[48] Y. Y. Zeevi and D. Rotem. Image reconstruction from zero crossings. *IEEE Trans. Acoust., Speech, Signal Processing*, 34:1269–1277, 1986.