Shift-Invariant Adaptive Wavelet Decompositions and Applications

Israel Cohen

Shift-Invariant Adaptive Wavelet Decompositions and Applications

Research Thesis

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Science

by
Israel Cohen

Submitted to the Senate of the Technion - Israel Institute of Technology
Iyar, 5758 Haifa May 1998

This research was carried out at the Faculty of Electrical Engineering under the supervision of Professor Shalom Raz and Professor David Malah.

I wish to express my deep gratitude to my advisors, Professor Shalom Raz and Professor David Malah, for their dedicated guidance, warm attitude, enthusiastic discussions and enlightening suggestions throughout all stages of this research. Their continued support and involvement, even during their sabbatical leave, is highly appreciated. I was very fortunate for the opportunity to work with them.

I am grateful to Dr Moshe Porat for his assistance during the sabbatical leave of Professors Raz and Malah.

I would like to thank my colleagues at the Department of Applied Physics, Rafael Labs, in particular Dr Moti Botton, Dr Jochanan Leopold, Elisha Berdugo, Dr Itzhak Schnitzer and Dr Benny Mandelbaum for their encouragement and friendship.

Finally, special thanks to my wife, Tammy - without her love and support I could not have completed this work.

To the love of my life, Tammy

and to the memory of my father, Aharon Ben-Zion

Contents

List of Figures

List of Tables

Abstract

Adaptive representations in libraries of bases, including the wavelet-packet and local trigonometric decompositions (WPD, LTD), are widely used in various applications. Instead of representing a prescribed signal on a predetermined basis, it is useful to search for a suitable basis that would best fit a specified application. A major drawback restricting the use of such methods, particularly in statistical signal processing applications, such as detection, identification or noise removal (denoising), is the lack of shift-invariance. The expansion, as well as the information cost measuring its suitability for a particular application, may be significantly influenced by the alignment of the input signal with respect to the basis functions. Furthermore, the time-frequency tilings, produced by the best-basis expansions, do not generally conform to standard time-frequency energy distributions. The objective of this work is to develop a general approach for achieving shift-invariance, enhanced time-frequency decompositions and robust signal estimators using libraries of orthonormal bases.

The first problem we address is that of shift-invariant adaptive decompositions in libraries of wavelet packet and local trigonometric bases. We define extended libraries that are organized in binary-tree structures, and introduce corresponding best-basis search algorithms, namely *shift-invariant wavelet packet decomposition* (SIWPD) and *shift-invariant adapted-polarity local trigonometric decomposition* (SIAP-LTD). The shift-invariance is achieved by the introduction of additional degrees of freedom in the expansion trees, which optimize the time-localization of basis functions. The SIAP-LTD provides an extra degree

of freedom that incorporates an adaptive folding operator into the best decomposition tree, mainly intended to reduce the information cost and thus improve the time-frequency representation. We show that the proposed algorithms lead to advantageous best-basis representations that, when compared to conventional representations, are characterized by lower information cost, improved time-frequency resolution, and for a prescribed data set yield more stable cost functions. The computational complexities are investigated, and efficient procedures for their control at the expense of the attained information cost are presented.

A second issue investigated in this work, closely related to the problem of shift-invariance, is that of adaptive decompositions of time-frequency distributions and removal of interference terms associated with bilinear distributions. We show that utilizing the SIWPD and SIAP-LTD, various useful properties relevant to time-frequency analysis, including high energy concentration and suppressed interference terms, can be achieved simultaneously in the Wigner domain. Instead of smoothing, which broadens the energy distribution of signal components, we propose best-basis decompositions and cross-term manipulations that are adapted to the local distribution of the signal via a certain time-frequency distance measure.

A prescribed signal is expanded on its best basis and transformed into the Wigner domain. Subsequently, the interference terms are eliminated by adaptively thresholding the cross Wigner distribution of interactive basis functions, according to their amplitudes and distance in an idealized time-frequency plane. The distance measure is adapted to the local distribution of the signal, and the amplitude and distance thresholds balance the cross-term interference, the useful properties of the distribution, and the computational complexity.

The properties of the resultant *modified Wigner distribution* are investigated, its surpassing performance is compared with that of other methods, and its distinctive applicability to resolving multicomponent signals is demonstrated. Alternative selections of extended libraries are examined, implying that the visual quality of the modified Wigner distribution

generally conforms with the entropy of the best basis expansion, and thus can be quantified.

The final topic concerns the problem of translation-invariant denoising, using the *Minimum Description Length* (MDL) criterion. We define a collection of signal models based on an extended library of orthonormal bases, and derive an additive cost function, approximately representing the MDL principle. The description length of the noisy observed data is then minimized by optimizing the expansion-tree associated with the best-basis algorithm and thresholding the resulting coefficients. We show that the signal estimator can be efficiently combined with the *modified* Wigner distribution, yielding robust time-frequency representations. The proposed methods are compared to alternative existing methods, and their superiority is demonstrated by synthetic and real data examples.

List of Symbols and Abbreviations

$\mathcal{B}$	Library of orthonormal bases
$\mathcal{D}$	Overcomplete dictionary of waveforms
$\mathcal{I}$	Set of dyadic intervals
$\mathcal{L}(y)$	Description length of y
$\mathcal{L}(By)$	Description length of y expanded in the basis B
$\mathcal{M}$	Additive information cost function
$\mathcal{M}(Bf)$	Information cost of f expanded in the basis B
$\mathbb{N}$	Set of naturals $\{1, 2, 3, \ldots\}$
$\mathbb{R}$	Set of reals
$\mathbb{Z}$	Set of integers $\{0, \pm 1, \pm 2, \ldots\}$
$\mathbb{Z}_+$	Set of non-negative integers $\{0, +1, +2, \ldots\}$
$\mathbb{Z}_-$	Set of non-positive integers $\{0, -1, -2, \ldots\}$
A	The optimal basis for signal estimation
A_f	The best basis for the function f
$A_g(\theta, \tau)$	Ambiguity function for $g(t)$
$A_{\ell,n,m}$, $A^j_{\ell,n,m}$	The best set of wavelet-packets for the subspace $U_{\ell,n,m}$
$A^{\rho_0,\rho_1}_{\ell,n,m}$	The best set of trigonometric-functions for the subspace $U^{\rho_0,\rho_1}_{\ell,n,m}$
$B_{\ell,n,m}$, $B^j_{\ell,n,m}$	Set of wavelet-packets associated with the tree-node (ℓ, n, m)
$B^{\rho_0,\rho_1}_{\ell,n,m}$	Set of trigonometric-functions associated with the tree-node (ℓ, n, m)
$C^s, C^s(\mathbb{R})$	Class of s-times continuously differentiable functions
C_g	Center of energy of the high-pass filter g
$C_g(t, \omega; \phi)$	Cohen's class of distributions
C_h	Center of energy of the low-pass filter h
C_n	n-tap coiflet filters
$C^{\rho_0,\rho_1}_{I,k}(t)$	Set of modulating trigonometric functions on an interval I
$C_{\ell,n,m,d}$	Suboptimal basis for $U_{\ell,n,m}$
$C_{m,n}(\rho)$	Local information cost about the n-th end-point for the shift m
D	Distance threshold in time-frequency plane
D_n	n-tap Daubechies least asymmetric wavelet filters
E	Set of terminal nodes of an expansion tree (tree-set)
F, $F(\alpha, \rho)$	Folding operator on $L_2(\mathbb{R})$

F^*	Adjoint of the folding operator
$G(\omega)$	Fourier transform of $g(t)$
$GC(n)$	Gray code permutation of an integer n
$GC^{-1}(n)$	Inverse Gray code permutation of an integer n
$I_{\ell,n}, I_{\ell,n,m}$	Dyadic intervals
J	Finest resolution level
K, k_n	Integers
L	Number of decomposition levels
L^2	Square-integrable functions
$\hat{L}_2$	Periodic square-integrable functions
M	Maximum magnitude of the best-basis coefficients
N	Length of signal at its highest resolution level
P	Polarity associated with SIAP-LTD expansion-tree
$P(A)$	Probability of event A
P_L	Optimal polarity at the finest resolution level
Q, $Q(\alpha, \rho)$	Periodic folding operator on $\hat{L}_2[0,1]$
Q^*	Adjoint of the periodic folding operator
$R(n)$	Integer obtained by bit reversal of n in a certain binary representation
$R_g(t,\tau)$	Instantaneous auto-correlation function of a complex signal $g(t)$
S_L	Number of bases derived from a $(L+1)$-level SIWPD expansion-tree
S_n	n-tap Daubechies minimum phase wavelet filters
$T_g(t,\omega)$	Modified Wigner distribution of $g(t)$
$\hat{T}_y$, $\hat{T}_y(t,\omega)$	Time-frequency distribution estimate of y
$U_{\ell,n,m}$, $U^j_{\ell,n,m}$	Closure of the linear span of $B_{\ell,n,m}$
$U^{\rho_0,\rho_1}_{\ell,n,m}$	Closure of the linear span of $B^{\rho_0,\rho_1}_{\ell,n,m}$
V_j	Subspace of j'th resolution level
$W_g(t,\omega)$	Auto Wigner distribution of $g(t)$
$W_{\ell,m}$	Wavelet best-basis for $U_{\ell,0,m}$
$W_{g,f}(t,\omega)$	Cross Wigner distribution of $g(t)$ and $f(t)$
X	Set of indices of "significant" basis-functions
c_λ	Expansion coefficient
d	Maximum depth of subtrees for shift determination
$d(\varphi_\lambda, \varphi_{\lambda'})$	Distance in time-frequency plane between $\varphi_\lambda(t)$ and $\varphi_{\lambda'}(t)$
$\hat{f}(t)$	Estimate of $f(t)$
f_k	Expansion coefficients of the unknown signal
$g^*(t)$	Complex conjugate of $g(t)$
$\{h_k\}, \{g_k\}$	Wavelet decomposition filter banks
ℓ	Resolution-level index
ℓ^p	Sequences u such that $\{\lvert u(k)\rvert^p\}$ is summable
(ℓ, n, m)	Index of a tree-node
m	Shift index
m_ℓ	Shift index at the resolution level ℓ

m_c	Shift index of children-nodes
m_p	Shift index of a paren node
n	Wavelet-packet index
$r, r(t)$	Rising cutoff functions
$\bar{r}$	Complex conjugate of r
s_I, s_{II}	Components of a multicomponent signal s
s_L	Number of bases derived from a $(L+1)$-level WPD expansion-tree
$\bar{t}_\lambda$	Time location of $\varphi_\lambda(t)$
$y(t)$	Noisy data
y_k	Expansion coefficients of $y(t)$
$z(t)$	White Gaussian noise
z_k	Expansion coefficients of $z(t)$
$\Delta\omega_\lambda$, $\Delta\omega$	Frequency uncertainty of $\varphi_\lambda(t)$
Δt_λ, Δt	Time uncertainty of $\varphi_\lambda(t)$
Γ	Set of indices of "neighboring" basis-functions pairs
Λ	Set of indices of "significant" basis-functions
Λ_k, Λ_I, Λ_{II}	Equivalence classes
α, β	End-points of an interval
$\delta_{k,\ell}$	Kronecker delta function
ϵ	Action-region radius of the folding operator
ε	Relative amplitude threshold
$\eta_\tau(c)$	Hard-threshold of c by τ
$\pi_m(n)$	n-th Optimal polarity-bit for the shift m
$\rho(\alpha)$	Polarity index at α
ρ, ρ_0, ρ_1	Polarity indices
$\phi(t)$, $\hat{\phi}(t)$	Basis-functions
$\phi_{I,k}^{\rho_0,\rho_1}(t)$	Smooth local trigonometric function on an interval I
$\phi(\theta, \tau)$	Kernel function for Cohen's class distribution
φ	Scaling function
φ_λ, $\varphi_\lambda(t)$	Best-basis elements
σ^2	Power spectral density of white noise
χ_I	Periodic-indicator function for the interval I
ψ, ψ_1	Mother wavelets
ψ_0	Scaling function
ψ_n	n-th Wavelet packet
$\psi_{I,k}^{\rho_0,\rho_1}$, $\psi_{\ell,n,m,k}^{\rho_0,\rho_1}$	Periodic smooth local trigonometric functions
ω	Angular frequency
$\bar{\omega}_\lambda$	Frequency location of $\varphi_\lambda(t)$
$x \bmod y$	Modulus (signed remainder after division)
$\lfloor x \rfloor$	Integer part of x
$\mathrm{Re}\{f\}$	Real part of f
$\#S$, $\lvert S \rvert$	The number of element in the set S

$\|c\|$	Magnitude of a complex number c
$\|\|g\|\|$	Norm of g
$\text{clos}_{L^2(\mathbb{R})}\{S\}$	Closure of the linear span of S
$\langle f, g\rangle$	Inner product of f and g
$\mathbf{1}_I$	Indicator function for the interval I
$\sim$	Equivalence relation

Abbreviations

AF	Ambiguity function
BPDN	Basis-Pursuit denoising
DCT	Discrete Cosine Transform
DST	Discrete Sine Transform
DWT	Discrete Wavelet Transform
IFT	Inverese Fourier transform
LCD	Local Cosine Decomposition
LSD	Local Sine Decomposition
LTD	Local Trigonometric Decomposition
MDL	Minimum Description Length
MOFDN	Method-of-Frames denoising
MPDN	Matching-Pursuit denoising
MWD	Modified Wigner distribution
ON	Orthonormal
PSD	Power spectral density
QMF	Quadrature Mirror Filter
SI-LCD	Shift-Invariant Local Cosine Decomposition
SI-LSD	Shift-Invariant Local Sine Decomposition
SIAP-LTD	Shift-Invariant Adapted-Polarity Local Trigonometric Decomposition
SIWPD	Shift-Invariant Wavelet Packet Decomposition
SIWT	Shift-Invariant Wavelet Transform
SNR	Signal to noise ratio
SWP	Shifted Wavelet Packet
WD	Wigner distribution
WGN	White Gaussian noise
WP	Wavelet Packet
WPD	Wavelet Packet Decomposition

Chapter 1

Introduction

1.1 Motivation and Goals

Adaptive representations in libraries of bases have been widely used in recent years. Instead of representing a prescribed signal on a predetermined basis, it is useful to search for a suitable basis that would facilitate a desired application, such as compression [102, 150, 76, 1, 138], identification and classification [12, 129, 131] or noise removal (denoising) [60, 48, 49, 84]. In general, for a given signal we are looking for a basis that matches well the signal, in the sense that only relatively few coefficients in the expansion are dominant, while the remaining coefficients are small and their total contribution is negligible. How specifically suitable a basis is, depends on the problem at hand. For signal compression, as an example, a preferable basis accelerates the descending rate of the coefficients' amplitudes, when sorted in a decreasing magnitude order. Whereas for classification, we select a basis which most discriminates between given classes. Such a basis reduces the dimensionality of the problem and emphasizes the dissimilarity between distinct classes [128].

Practical "best basis" search procedures are necessarily confined to finite-size libraries. Such libraries are not only required to be flexible and versatile enough to describe various local features of signals, but also need to be aptly organized in a structure that facilitates

a fast search algorithm for the "best basis". Coifman and Meyer [42, 46, 102] were the first to introduce libraries of orthonormal bases whose elements are localized in time-frequency plane and structured into a binary tree where the best basis can be efficiently searched for. One of the libraries, a library of local trigonometric bases, consists of sines or cosines multiplied by smooth window functions. Their localization properties depend on the steepness of the ascending and descending parts of the window functions [65]. Another library, a library of wavelet packet bases, comprises basis functions which are translations and dilations of wavelet packets, and their localization properties in time-frequency plane depend on those of the "mother wavelet" [44, 69]. Both libraries are naturally organized in binary trees whose nodes represent subspaces that are orthogonally split into children-nodes [43]. Accordingly, the basis functions of a parent-node can be replaced by the collection of basis functions that correspond to the children-nodes. This flexibility in choosing a basis for each subspace implies adaptive representations, by a recursive comparison between the information costs of parent-nodes and their children-nodes.

Selecting a desirable information cost functional is clearly application dependent [128, 136, 150]. Entropy, for example, may be used to effectively measure the energy concentration of the generated nodes [48, 77, 143]. Statistical analysis of the best-basis coefficients may provide a characteristic time-frequency signature of the signal, potentially useful in simplifying identification and classification applications [12, 87]. A major drawback restricting the use of such methods is the lack of shift-invariance. The coefficients of a delayed signal are not a time-shifted version of those of the original signal. The information cost of the best-basis coefficients, measuring the suitability of the expansion for the particular application, may also be significantly influenced by the alignment of the input signal with respect to the basis functions. These phenomena are not unique to the standard *wavelet packet* and *local trigonometric decompositions* (WPD, LTD) of Coifman and Wickerhauser [45]. Other recently developed adaptive representations, such as the *time-varying wavelet*

packet decomposition and *time-varying modulated lapped transforms* proposed by Herley *et al.* [67, 68], are also sensitive to translations.

Shift-invariant multiresolution representations exist. However, some methods either entail high oversampling rates (*e.g.*, in [127, 9, 10, 86, 122] no down-sampling with the changing scale is allowed) or alternatively, the resulting representations are non-unique (as is the case for zero-crossing or local maxima methods, e.g., [93, 74, 94, 95, 8]). Furthermore, zero-crossing and related methods facilitate a signal reconstruction that is necessarily approximate. We also note that such methods lead to non-orthogonal representations, rendering the interpretation of the correlation properties among the expansion coefficients more difficult. Mallat and Zhang [96] have suggested an adaptive *matching pursuit* algorithm. Under this approach the retainment of shift-invariance necessitates an oversized library containing the basis functions and all their shifted versions. The obvious drawbacks of *matching pursuit* are the rather high complexity level as well as the non-orthogonality of the expansion.

The estimation of signals embedded in noise (*denoising*) using wavelet bases [61, 64] is also deficient for the shift-variance of the wavelet transforms. Coifman and collaborators [49, 7, 130] observed that denoising with the conventional wavelet transform and WPD may exhibit visual artifacts, such as pseudo-Gibbs phenomena in the neighborhood of discontinuities and artificial symmetries across segmentation points in the frequency domain. They attributed these artifacts to the lack of shift-invariance, and accordingly suggested to average the translation dependence by the following scheme: apply a range of shifts to the noisy data, denoise the shifted versions with the wavelet transform, then unshift and average the denoised data. This procedure, termed *Cycle-Spinning*, generally yields better visual performance on smooth parts of the signal. However, transitory features may be significantly attenuated [145]. Furthermore, information-theoretic arguments, such as the *Minimum Description Length* (MDL) principle [125] which has shown great applicability

for signal and image denoising [130, 84, 104, 146], are not considered.

Another issue, closely related to the problem of shift-invariance, is that of adaptive decompositions of time-frequency distributions and suppression of interference terms associated with bilinear distributions. The Wigner distribution (WD), for example, satisfies many desirable properties which are relevant to time-frequency analysis [18, 23]. However, its practical application is often restricted due to the presence of interference terms. These terms make the WD of multicomponent signals extremely difficult or impossible to interpret. The reduced-interference distributions [78, 152] attenuate the interference terms of the WD by employing some kind of smoothing kernel or windowing. Unfortunately, the smoothing operation reduces the energy concentration of the analyzed signal and dramatically affects the appearance and quality of the resulting time-frequency representation. Adaptive representations [4, 50] often exhibit performances far surpassing that of fixed-kernel representations. However, they are either computationally expensive or have a very limited adaptation capability.

The *cross-term deleted representations* and the *time-frequency distribution series*, proposed by Qian *et al.* [114, 115], employ the Gabor expansion to decompose the WD and manipulate cross-terms. Their major deficiency is the dependence of the performance on the choice of the window used in the Gabor expansion. An appropriate window depends on the data and may differ for different components of the same signal, . Moreover, distinct Gabor functions which are "close" in time-frequency plane should be often related to the same signal component (the extent of closeness varies according to the local distribution of the signal). As a result, their cross-term is not necessarily an interference term, but rather may have a significant effect on the energy concentration. In [142], the signal is decomposed into frequency bands using the WPD, and the Wigner distributions of all the subbands are superimposed. This attenuates interferences between subbands, but still suffers interferences within the subbands. Consequently this approach is merely suitable for signals that have

a single component in each subband. The exclusion of beneficial cross-terms, which join neighboring basis-functions, may degrade the energy concentration and may also artificially split a certain component of the signal into a few frequency-separated components.

In view of the above shortcomings, the objective of this thesis is to develop a general approach for achieving shift-invariance, enhanced time-frequency decompositions and robust signal estimators using libraries of orthonormal bases. In particular, to devise practical schemes for:

- Shift-invariant adaptive decompositions in libraries of wavelet packet and local trigonometric bases, retaining orthonormality, low information cost and manageable computational complexity.
- Adaptive time-frequency distributions that satisfy various useful properties relevant to time-frequency analysis, including high energy concentration, suppressed interference terms and shift-invariance.
- Translation-invariant denoising of signals and time-frequency distributions, based on the *Minimum Description Length* criterion.

1.2 Overview of the Thesis

The original contribution of this thesis starts from Chapter 2, where we present a *shift-invariant wavelet packet decomposition* (SIWPD), implemented via a recursive best-basis selection method and supplied with an inherent trade-off between the computational complexity and the information cost. First, we extend the library of wavelet packet bases and organize it in a binary-tree structure, so that any shifted version of a basis within the library is also included. Then, the shift-invariance is achieved by the introduction of an additional degree of freedom in the expansion tree, which enables to optimize the time-localization of

basis functions. The added dimension is a *relative shift* between a given parent-node and its respective children-nodes. Specifically, upon expanding a prescribed node at a resolution level ℓ ($-L \leq \ell \leq 0$), we examine and select one of two relative shift options — a zero shift or a $2^{-\ell}$ shift. The choice between these two options, enabled by the extended library, is made in accordance with minimizing the information cost. Hence, the attained representation is not only shift-invariant, but also characterized by a lower information cost when compared to the conventional WPD. The special case where, at any resolution level, only low frequency nodes are further expanded corresponds to a *shift-invariant wavelet transform* (SIWT).

An alternative view of SIWPD is facilitated via filter-bank terminology. Accordingly, each parent-node is expanded by high-pass and low-pass filters, followed by a 2:1 down-sampling. In executing WPD, down-sampling is achieved by ignoring all even-indexed (or all odd-indexed) terms [121, 139]. In contrast, when pursuing SIWPD, the down-sampling is carried out *adaptively* for the prescribed signal. We stress that owing to the orthogonality of the representation and the presumed additive nature of the cost functions (e.g., the Shannon entropy or rate-distortion), the decision at any given node is strictly local, i.e., independent of other nodes at the same resolution level.

The computational complexity of SIWPD is $O[2^d(L-d+2)N]$, where N denotes the length of the signal (at its highest resolution level), L is the number of decomposition levels ($L \leq \log_2 N$) and d is the maximum depth of a subtree used at a given parent-node to determine the shift mode of its children ($1 \leq d \leq L$). The key to controlling the complexity is the built-in flexibility in the choice of d. Lower d implies lower complexity at the expense of a higher information cost. At its lower bound, $d = 1$, the attained level of complexity, $O(NL)$, resembles that of WPD while still guaranteeing shift-invariance.

In Chapter 3, the strategy for obtaining shift-invariance with wavelet packet bases is generalized and applied to local trigonometric bases. We define an extended tree-structured library of smooth local trigonometric bases, and describe efficient search algorithms for

selecting the best basis. To further enhance the resultant representation, we introduce an adaptive-polarity folding operator which splits the prescribed signal and "folds" *adaptively* overlapping parts back into the segments. This operator is incorporated into the best decomposition tree by a fast numerical algorithm, namely *shift-invariant adapted-polarity local trigonometric decomposition* (SIAP-LTD). It is proved that the proposed algorithms lead to best-basis representations, which are shift-invariant, orthogonal and characterized by lower information cost when compared to the conventional LTD.

The stated advantages of SIWPD and SIAP-LTD, specifically the shift-invariance as well as the lower information cost, may prove crucial to signal compression, identification or classification applications. Furthermore, the shift-invariant nature of the information cost, renders this quantity a characteristic of the signal for a prescribed library. It should be possible now to quantify the relative efficiency of various libraries (*e.g.*, various scaling function selections) with respect to a given cost function. Such a measure would be rather senseless for shift-variant decompositions.

In Chapter 4, we present adaptive time-frequency decompositions using extended libraries of orthonormal bases. A prescribed signal is expanded on a basis of adapted waveforms that best match the signal components, and subsequently transformed into the Wigner domain. We show that the interference terms can be controlled by adaptively thresholding the cross WD of interactive basis functions according to their distance and amplitudes in an idealized time-frequency plane. The distance and amplitude thresholds balance the cross-term interference, the useful properties of the distribution, and the computational complexity. When the distance-threshold is set to zero, the modified Wigner distribution precludes any cross-terms, so essentially there is no interference terms but the energy concentration of the individual components is generally low. When the amplitude-threshold is set to zero and the distance-threshold goes to infinity, the modified Wigner distribution converges to the conventional WD. Appropriate threshold values yield

enhanced time-frequency decompositions, which achieve high resolution, high concentration and suppressed cross-term interference.

The distance measure between pairs of basis-functions is defined by weighing their Euclidean distance with their time and frequency uncertainties. Since the basis-functions are adapted to the signal's local distribution, the thresholding of the cross-terms is also adapted to the local distribution of the signal. This dispenses with the need for local adjustments of the associated distance-threshold. Exploiting the resultant adaptation, we define equivalence classes in the time-frequency plane and show that the components of a multicomponent signal can be resolved and recovered from the energy distribution up to a constant phase factor. It is demonstrated that "best orthonormal bases" in *extended* libraries are more advantageous to "optimal" expansions (*e.g.*, Matching Pursuit [96] and Basis Pursuit [14]) in *conventional* libraries. The extension of a library provides a fundamental flexibility in the expansion, while the restriction of the best-basis search procedure to orthonormal bases maintains a manageable computational complexity.

In Chapter 5, we propose a translation-invariant denoising method, which uses the MDL criterion and tree-structured best-basis algorithms. We define a collection of signal models based on the extended libraries of orthonormal bases, and apply the MDL principle to derive corresponding additive cost functions. The description length of the noisy observed data is then minimized by optimizing the expansion-tree associated with the best-basis algorithm and thresholding the resulting coefficients with a certain threshold. We show that the signal estimator can be efficiently combined with the modified Wigner distribution, yielding robust time-frequency representations. Synthetic and real data examples demonstrate the superiority of our approach over alternative existing methods.

We conclude in Chapter 6 with a summary and discussion on future research directions.

1.3 Organization

The organization of this thesis is as follows. In the next section we review the original wavelet packet bases and best-basis algorithm, and discuss the concept of shift-invariance within the framework of best-basis expansions. In Chapter 2, we extend the library of wavelet packet bases and describe the SIWPD. The computational complexity is investigated, and an efficient procedure for its control at the expense of the attained information cost is introduced. In Chapter 3, the strategy for obtaining shift-invariance is generalized and applied to local trigonometric bases. We define an extended library that is organized in a binary tree-structure, and introduce efficient search algorithms for selecting the best basis. The main algorithm derived is SIAP-LTD. Its suboptimal variants, entailing a reduced complexity and higher information cost, are *Shift-Invariant Local Cosine Decomposition* (SI-LCD) and *Shift-Invariant Local Sine Decomposition* (SI-LSD). In Chapter 4, we present adaptive decompositions of the WD using extended libraries of orthonormal bases. We define a distance measure in the time-frequency plane, facilitating the distinction between undesirable interference terms and beneficial cross-terms, and describe enhanced time-frequency distributions. In Chapter 5, a translation-invariant denoising approach, based on the Minimum Description Length criterion, is introduced. We derive additive cost functions for wavelet packet and local trigonometric bases, and utilize the shift-invariant best-basis decompositions. Finally, in Chapter 6 we conclude with a summary and discussion on future research directions.

We would like to note that Chapters 2 and 3 and part of Chapter 4 are the detailed and expanded version of our published materials. Chapter 2 is based on [27, 28], Chapter 3 is based on [29, 30, 31, 32], and part of Chapter 4 is based on [31, 33]. Additional manuscripts [34, 35, 36, 37, 38], based on Chapters 4 and 5, are about to be published.

1.4 Background

A natural framework for the understanding of wavelet bases, and for the construction of new examples, is provided by the multiresolution analysis (MRA) [92, 102, 53]. A *multiresolution analysis* of $L^2(\mathbb{R})$ is a chain of subspaces $\{V_j : j \in \mathbb{Z}\}$ satisfying the following conditions:

1. *Containment*: $V_j \subset V_{j+1} \subset L^2(\mathbb{R})$ for all $j \in \mathbb{Z}$.
2. *Decrease*: $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.
3. *Increase*: $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$.
4. *Dilation*: $f(x) \in V_j \iff f(2x) \in V_{j+1}$.
5. *Generator*: There is a function $\varphi \in V_0$ whose translates $\{\varphi(x-k) : k \in \mathbb{Z}\}$ form an orthonormal basis for V_0 .

The function φ is called a *scaling function*. From the containment, dilation and generator properties, it follows that $\{\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) : k \in \mathbb{Z}\}$ form an orthonormal basis for V_j for all $j \in \mathbb{Z}$. The central theorem of multiresolution analysis asserts that whenever the above properties are satisfied, there exists an orthonormal wavelet basis $\{\psi_{j,k}(x) : j,k \in \mathbb{Z}\}$ for $L^2(\mathbb{R})$, $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, such that,

$$P_{j+1} = P_j + \sum_{k\in\mathbb{Z}} \langle \cdot, \psi_{j,k}\rangle \psi_{j,k} \tag{1.1}$$

where P_j is the orthogonal projection operator onto V_j, given by

$$P_j = \sum_{k\in\mathbb{Z}} \langle \cdot, \varphi_{j,k}\rangle \varphi_{j,k}\,. \tag{1.2}$$

The function ψ is called a *mother wavelet*. If we define W_j to be the orthogonal complement of V_j in V_{j+1}, so that $V_{j+1} = V_j \oplus W_j$, then by the "increase" property, we can telescope

the union to write

$$L^2 = \bigoplus_{j\in\mathbb{Z}} W_j \,. \tag{1.3}$$

The subspaces W_j are called *wavelet subspaces*, and Eq. (1.3) is called a *wavelet decomposition* of L^2. If we restrict the number of decomposition levels to $\ell > 1$, then a wavelet decomposition of V_J is given by

$$V_J = V_{J-\ell} \oplus \bigoplus_{j=J-\ell}^{J-1} W_j \,. \tag{1.4}$$

Remark. The orthonormality condition of $\varphi(\cdot - k)$ can be relaxed by the requirement that $\{\varphi(\cdot - k) : k \in \mathbb{Z}\}$ constitute a Riesz basis. That is, the functions $\{\varphi(x-k) : k \in \mathbb{Z}\}$ are linearly independent, and there exist two strictly positive constants A and B such that for any $f \in V_0$

$$A\|f\|_2^2 \le \sum_{k\in\mathbb{Z}} |\langle f(x), \varphi(x-k)\rangle|^2 \le B\|f\|_2^2 \,. \tag{1.5}$$

Equivalently, the functions $\{\varphi(x-k) : k \in \mathbb{Z}\}$ form a Riesz basis for V_0 if and only if they span V_0, and for any $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$

$$A \sum_{k\in\mathbb{Z}} |c_k|^2 \le \left\| \sum_{k\in\mathbb{Z}} c_k \varphi(x-k) \right\|^2 \le B \sum_{k\in\mathbb{Z}} |c_k|^2 \,, \tag{1.6}$$

where $A > 0$ and $B < \infty$ are independent of the c_k. An orthonormal basis $\{\varphi(x-k) : k \in \mathbb{Z}\}$ for V_0 can be constructed from a Riesz basis $\{\theta(x-k) : k \in \mathbb{Z}\}$ by [53, p. 139]

$$\Phi(\omega) = (2\pi)^{-1/2} \left[\sum_{k\in\mathbb{Z}} |\Theta(\omega + 2\pi k)|^2 \right]^{-1/2} \Theta(\omega) \,, \tag{1.7}$$

where $\Phi(\omega)$ and $\Theta(\omega)$ are the Fourier transforms of $\varphi(x)$ and $\theta(x)$, respectively.

Recall that $\{\varphi_{1,k} \,:\, k \in \mathbb{Z}\}$ is an orthonormal basis for V_1. Then, the scaling function $\varphi \in V_0 \subset V_1$ satisfies a *two-scale equation*:

$$\varphi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2x - k) \stackrel{\triangle}{=} H\varphi(x)\,, \tag{1.8}$$

where $\{h_k\}$ is a square-summable sequence of coefficients, which defines a linear operator H. Once we have the scaling function φ, we may use it to construct the mother wavelet ψ. One possibility for the construction is

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k) \stackrel{\triangle}{=} G\varphi(x); \qquad g_k = (-1)^k \bar{h}_{1-k}\,, \tag{1.9}$$

where $\bar{h}$ denotes the complex conjugate of h. The operators H and G are called perfect reconstruction *quadrature mirror filters* (QMFs); They satisfy the perfect reconstruction conditions:

$$HG^* = GH^* = 0, \quad \text{and} \quad H^*H + G^*G = I, \tag{1.10}$$

where H^* and G^* are the adjoint operators of H and G, respectively, and I is the identity operator.

The wavelet packet decomposition [39, 45, 150] is a generalization of the wavelet transform, which allows a further decomposition of the wavelet subspaces $\{W_j\}_{j \in \mathbb{Z}}$. Using the QMFs H and G, we recursively define a sequence of wavelet packets $\{\psi_n\}_{n=0}^{\infty}$ by

$$\psi_{2n} \stackrel{\triangle}{=} H\psi_n\,; \qquad \psi_{2n}(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \psi_n(2x - k) \tag{1.11}$$

$$\psi_{2n+1} \stackrel{\triangle}{=} G\psi_n\,; \qquad \psi_{2n+1}(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \psi_n(2x - k)\,, \tag{1.12}$$

where $n \in \mathbb{Z}_+$, and $\psi_0 \stackrel{\triangle}{=} \varphi$ is the scaling function. Let us introduce the notation

$$B_{j,n} = \left\{ 2^{j/2} \psi_n(2^j x - k) \,:\, k \in \mathbb{Z} \right\}, \tag{1.13}$$

$$U_{j,n} = \operatorname{clos}_{L^2(\mathbb{R})} \{B_{j,n}\}\,, \tag{1.14}$$

$j \in \mathbb{Z}$, $n \in \mathbb{Z}_+$. It follows that $B_{j,n}$ is an orthonormal basis for $U_{j,n}$ [45], and each $U_{j,n}$ can be decomposed into two orthogonal subspaces:

$$U_{j,n} = U_{j-1,2n} \oplus U_{j-1,2n+1} \,. \tag{1.15}$$

Observe that in the special case of wavelet decomposition, $V_j \equiv U_{j,0}$, $W_j \equiv U_{j,1}$, and the decomposition relation is restricted to $n = 0$:

$$U_{j,0} = U_{j-1,0} \oplus U_{j-1,1} \,.$$

It is useful to associate with each subspace $U_{j,n}$ a *dyadic interval* $I_{j,n} = [2^j n, 2^j(n+1))$, and follow the relations

$$I_{j,n} = I_{j-1,2n} \cup I_{j-1,2n+1} \quad \text{and} \quad I_{j-1,2n} \cap I_{j-1,2n+1} = \emptyset \,. \tag{1.16}$$

Accordingly, a set E, such that $\{I_{j,n} \,:\, (j,n) \in E\}$ constitute a partition of $I_{J,0} = [0, 2^J)$, generates an orthonormal wavelet packet basis $\{B_{j,n} \,:\, (j,n) \in E\}$ for V_J. The collection of all thus generated bases is termed a *library* of wavelet packet bases.

The library comprises a large number of bases, from which we may pick and choose the *best* for the problem at hand. The efficiency of a given expansion is measured by an appropriate information cost functional $\mathcal{M}$ [45, 150, 128]. Examples of additive information cost functionals [150] include the *Entropy* (or *Shannon entropy*), defined by

$$\mathcal{M}(\{x_i\}) = -\sum_{i:x_i \neq 0} x_i^2 \ln x_i^2 \,, \tag{1.17}$$

the *Concentration in* ℓ^p $(0 < p < 2)$

$$\mathcal{M}(\{x_i\}) = \sum_i |x_i|^p \,, \tag{1.18}$$

and the *Logarithm of energy*

$$\mathcal{M}(\{x_i\}) = \sum_{i:x_i \neq 0} \ln x_i^2 \,. \tag{1.19}$$

Let $f(x)$ be a function in V_0, let $\mathcal{B}$ represent a library of wavelet packet bases, and denote by $\mathcal{M}(Bf)$ the information cost of representing f in a basis $B \in \mathcal{B}$.

Definition 1.1 [45] *The best basis for f in $\mathcal{B}$ with respect to $\mathcal{M}$ is $B \in \mathcal{B}$ for which $\mathcal{M}(Bf)$ is minimal.*

The best basis can be obtained by an efficient recursive selection process, which determines the best decomposition of $U_{j,n}$ based exclusively on a local minimization of the cost functional. Denote by $A_{j,n}$ the best basis for the subspace $U_{j,n}$. Then, the best basis $A_{0,0}$ for $f \in V_0 = U_{0,0}$ is determined by setting

$$A_{j,n} = \begin{cases} B_{j,n} \text{ if } \mathcal{M}(B_{j,n}f) \leq \mathcal{M}(A_{j-1,2n}f) + \mathcal{M}(A_{j-1,2n+1}f)\,, \\ A_{j-1,2n} \oplus A_{j-1,2n+1}, \text{ otherwise.} \end{cases} \tag{1.20}$$

The recursive sequence proceeds down to a specified level ℓ, where

$$A_{\ell,n} = B_{\ell,n}\,, \quad 0 \leq n < 2^{-\ell}\,. \tag{1.21}$$

This process for the selection of the best basis is *not* shift-invariant; The expansion coefficients are generally very influenced by translations of the analyzed signal. Best-basis expansions are shift-invariant if for any pair of signals, which are identical to within a time shift, their respective time-frequency representations are identical to within the *same* time shift. To achieve this property, when the signal is translated in time by τ, the elements of the newly generated best-basis are also translated by τ. Thus, the expansion coefficients remain the same, and the time-frequency representation is time-shifted by τ. In practice, we may consider discrete time translations of the form $\tau = 2^{-J}q$, where q takes integer values and J is as large as necessary.

Definition 1.2 *$f, g \in L^2(\mathbb{R})$ are said to be identical to within a resolution J time-shift if there exists $q \in \mathbb{Z}$, such that $g(t) = f(t - 2^{-J}q)$ for all $t \in \mathbb{R}$.*

Definition 1.3 *Bases $B_1, B_2 \in \mathcal{B}$ are said to be identical to within a resolution J time-shift if there exists $q \in \mathbb{Z}$, such that $\psi(t - 2^{-J}q) \in B_2$ if and only if $\psi(t) \in B_1$.*

Definition 1.4 *A best-basis decomposition is said to be shift-invariant up to a resolution level J if for any $f, g \in L^2(\mathbb{R})$ which are identical to within a resolution J time-shift, their respective best bases A_f and A_g are identical to within the* same *time-shift.*

Clearly, a best-basis decomposition, which is shift-invariant up to a resolution level J, is also shift-invariant up to a lower resolution level, because the translation is on a finer grid. In case of uniformly sampled discrete functions of length $N = 2^J$, an invariance to circular discrete translation is equivalent to shift-invariance up to a resolution level J.

Chapter 2

Shift-Invariant Wavelet Packet Decompositions

2.1 Introduction

The discrete wavelet transform (DWT) and wavelet packet decompositions [45] are widely used in recent years. One major drawback of these representations is their sensitivity to signal translations due to the dyadic structure of the wavelet expansions [53]. Consequently, the coefficients of a delayed signal are not time-shifted version of those of the original signal, and the cost of the expansion (as measured by some information cost functional) is significantly influenced by the alignment of the input signal with respect to the basis functions [27]. Thus, wavelet expansions, as well as generalizations proposed by Herley *et al.* [67, 68], may not perform well in statistical signal processing applications, such as detection or parameter estimation of signals with unknown arrival time.

This problem of wavelet transforms, namely their sensitivity to translations, has been addressed using different approaches. However, some methods either entail high over-sampling rates (*e.g.*, in [9, 10, 86, 122, 127] no down-sampling with the changing scale is allowed) or immense computational complexity (*e.g.*, the *matching pursuit* algorithm

[55, 96]). In some other methods, the resulting representations are non-unique and involve approximate signal reconstructions, as is the case for zero-crossing or local maxima methods [8, 74, 93, 94, 95]. Another approach has relaxed the requirement for shift-invariance, and defined a less restrictive property named *shiftability* [3, 134], which is accomplished by imposing limiting conditions on the scaling function [3, 6, 141].

In this chapter, we define a *shifted wavelet packet* (SWP) library that contains all the time shifted wavelet packet bases, and introduce a corresponding search algorithm, namely *shift-invariant wavelet packet decomposition* (SIWPD), for a "best basis" selection with respect to an additive cost function (e.g., the Shannon entropy). We prove that the proposed algorithm leads to a best-basis representation that is shift-invariant, orthogonal and characterized by a lower information cost. We also show that the computational complexity is manageable and may be controlled at the expense of the attained information cost down to $O(Nlog_2N)$.

To demonstrate the shift-invariant properties of SIWPD, compared to WPD which lacks this feature, we refer to the expansions of the signals $g(t)$ (Fig. 2.1) and $g(t - 2^{-6})$. These signals contain $2^7 = 128$ samples, and are identical to within 2 samples time-shift. For definiteness, we choose D_8 to serve as the scaling function (D_8 corresponds to 8-tap Daubechies least asymmetric wavelet filters [53, page 198] [51]) and the Shannon entropy as the cost function. Figs. 2.2 and 2.3 depict the "best-basis" expansion under the WPD and the SIWPD algorithms, respectively. A comparison of Figs. 2.2(b) and (d) readily reveals the sensitivity of WPD to temporal shifts while the best-basis SIWPD representation is indeed shift-invariant and characterized by a lower entropy (Fig. 2.3). It is worthwhile mentioning that the tiling grids in Figs. 2.2 and 2.3 do not in general correspond to actual time-frequency energy distributions. In fact, the energy distribution associated with each of the nominal rectangles may spread well beyond their designated areas [44]. However, when a proper "scaling function" is selected (*i.e.,* well localized in both time and frequency), the SIWPD

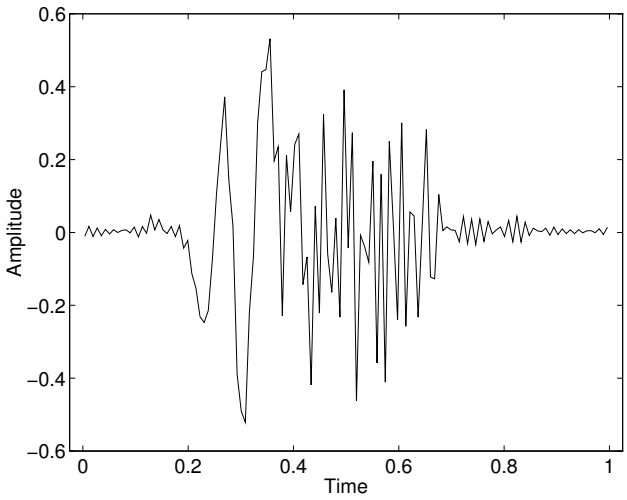


Figure 2.1: Test signal $g(t)$.

based time-frequency representation resembles shift-invariant time-frequency distributions. Fig. 2.4 displays the Wigner and smoothed Wigner distributions [23] for the signal $g(t)$. The smoothing kernel (here we chose a Gaussian) attenuates the interference terms at the expense of reduced time-frequency resolution. Obviously, the smoothed distribution (Fig. 2.4(b)) has a closer relation to the SIWPD based representation (Fig. 2.3(b)), than to the WPD based representation (Fig. 2.2(b)).

Pursuing the SIWPD algorithm, shift-invariance is achieved by the introduction of an additional degree of freedom. The added dimension is a *relative shift* between a given parent-node and its respective children-nodes. Specifically, upon expanding a prescribed node, with minimization of the information cost in mind, we test as to whether or not the information cost indeed decreases. We prove that for any given parent-node it is sufficient to examine and select one of two alternative decompositions, made feasible by the SWP library. These decompositions correspond to a zero shift and a $2^{-\ell}$ shift where ℓ ($-L \leq \ell \leq 0$) denotes the resolution level. The special case where, at any resolution level, only low frequency nodes are further expanded reduces to *shift-invariant wavelet transform* (SIWT). An alternative view

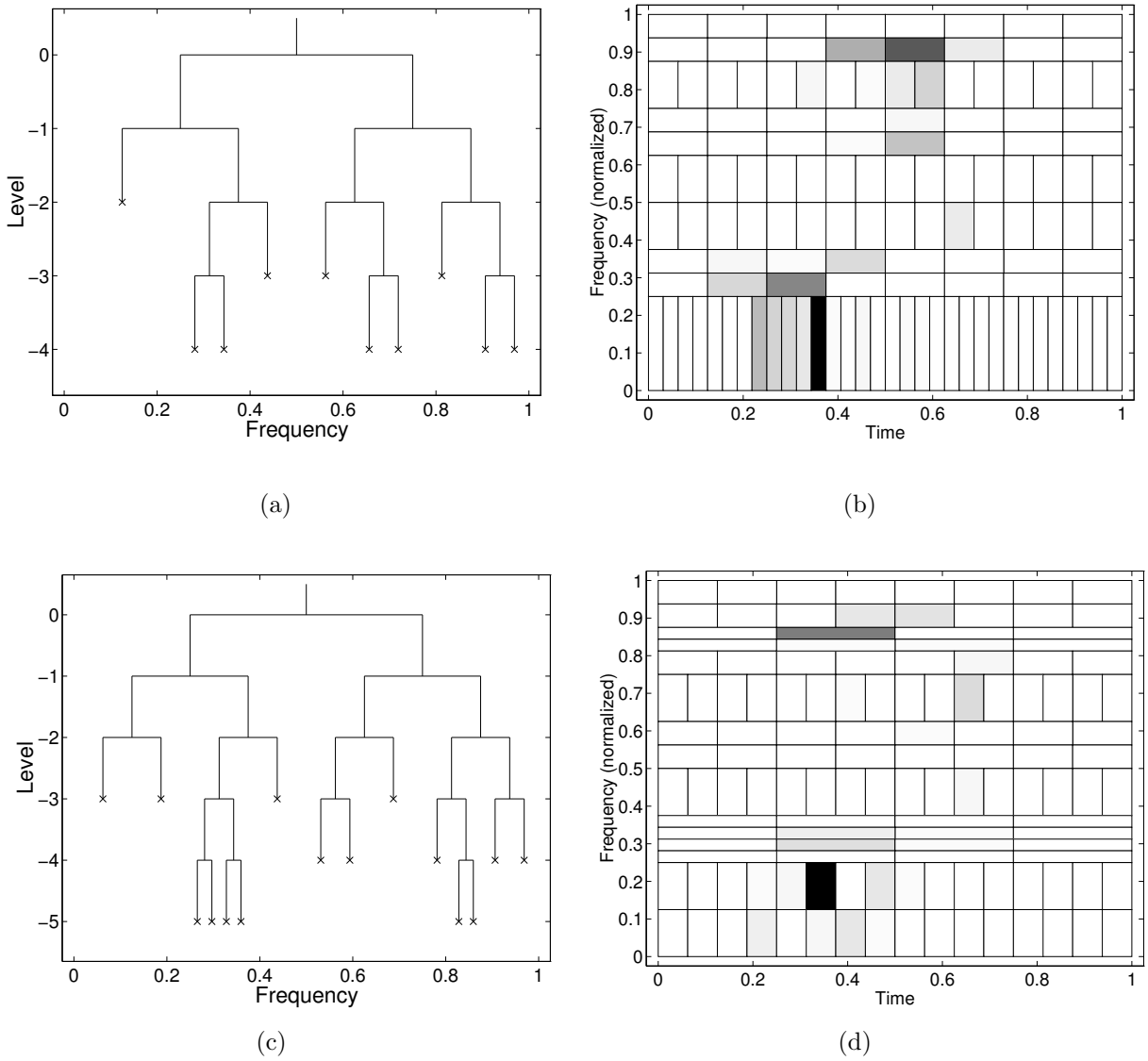


Figure 2.2: Effect of a temporal shift on the time-frequency representation using the WPD with 8-tap Daubechies least asymmetric wavelet filters: (a) The best expansion tree of $g(t)$. (b) $g(t)$ in its best basis; Entropy= 2.84. (c) The best expansion tree of $g(t - 2^{-6})$. (d) $g(t - 2^{-6})$ in its best basis; Entropy= 2.59.

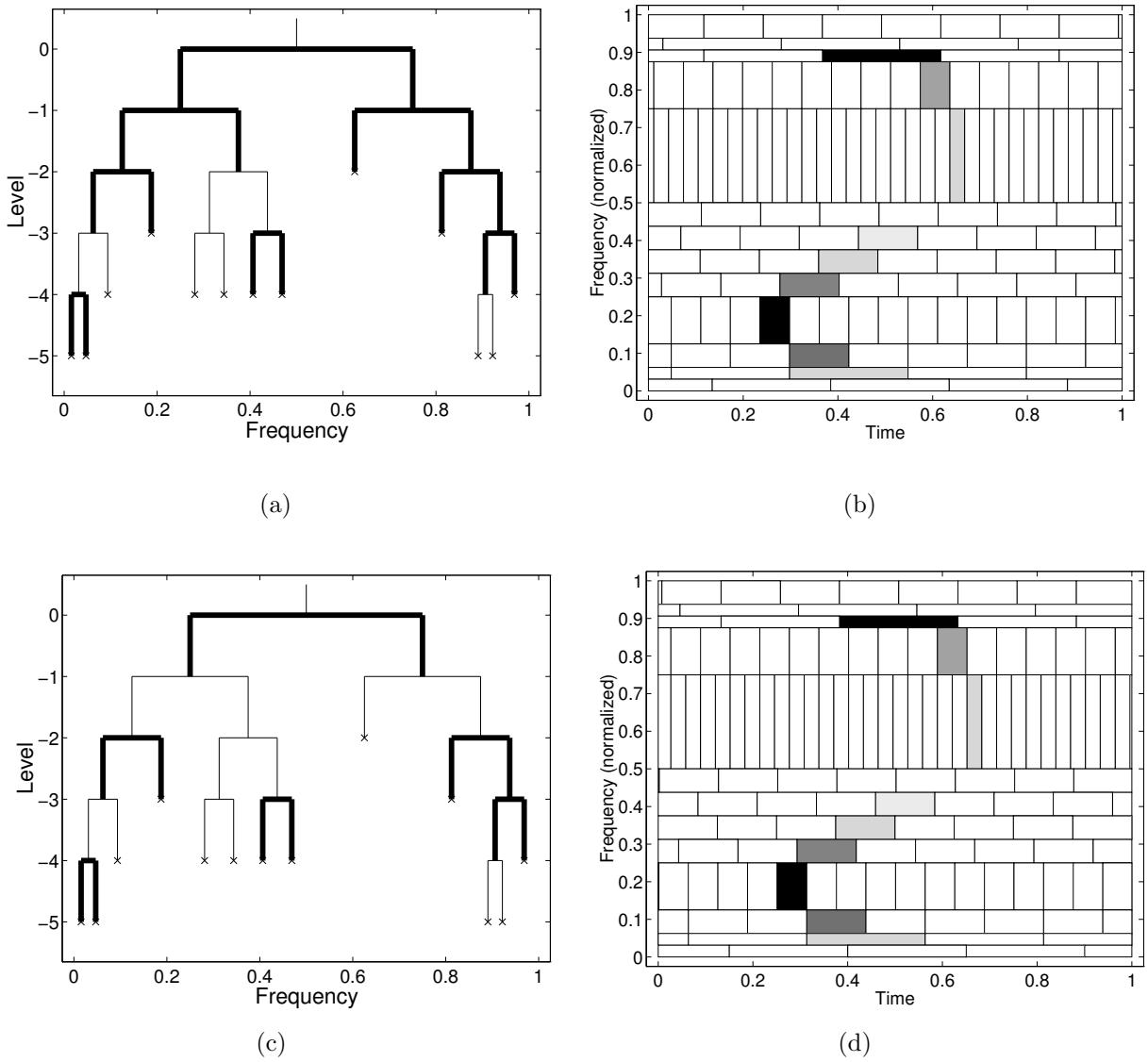
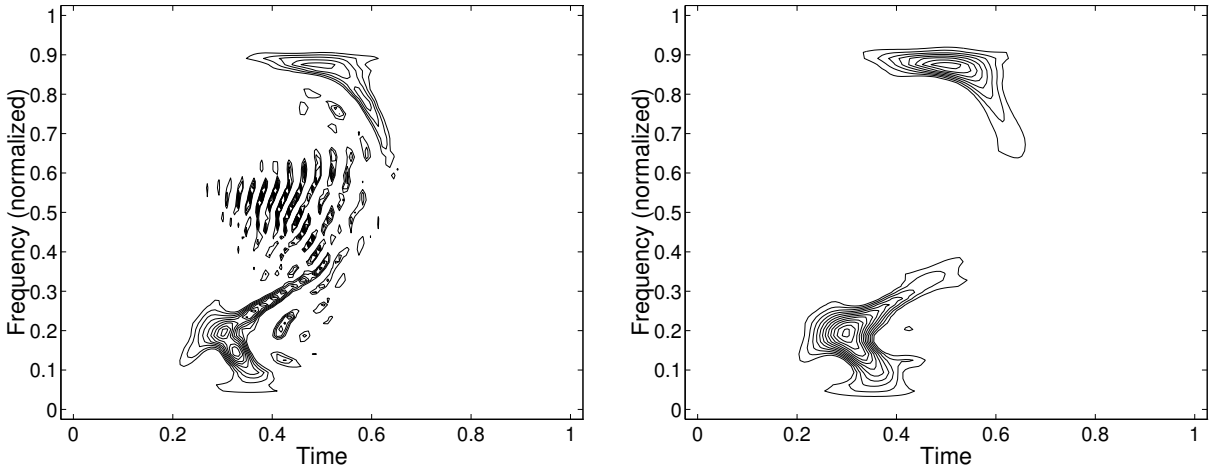


Figure 2.3: Time-frequency representation using the SIWPD with 8-tap Daubechies least asymmetric wavelet filters: (a) The best expansion tree of $g(t)$. (b) $g(t)$ in its best basis; Entropy= 1.92. (c) The best expansion tree of $g(t - 2^{-6})$. (d) $g(t - 2^{-6})$ in its best basis; Entropy= 1.92. Fine and heavy lines in the expansion tree designate alternative node decompositions. Compared with the WPD (Fig. 2.2), beneficial properties are shift-invariance and lower information cost.

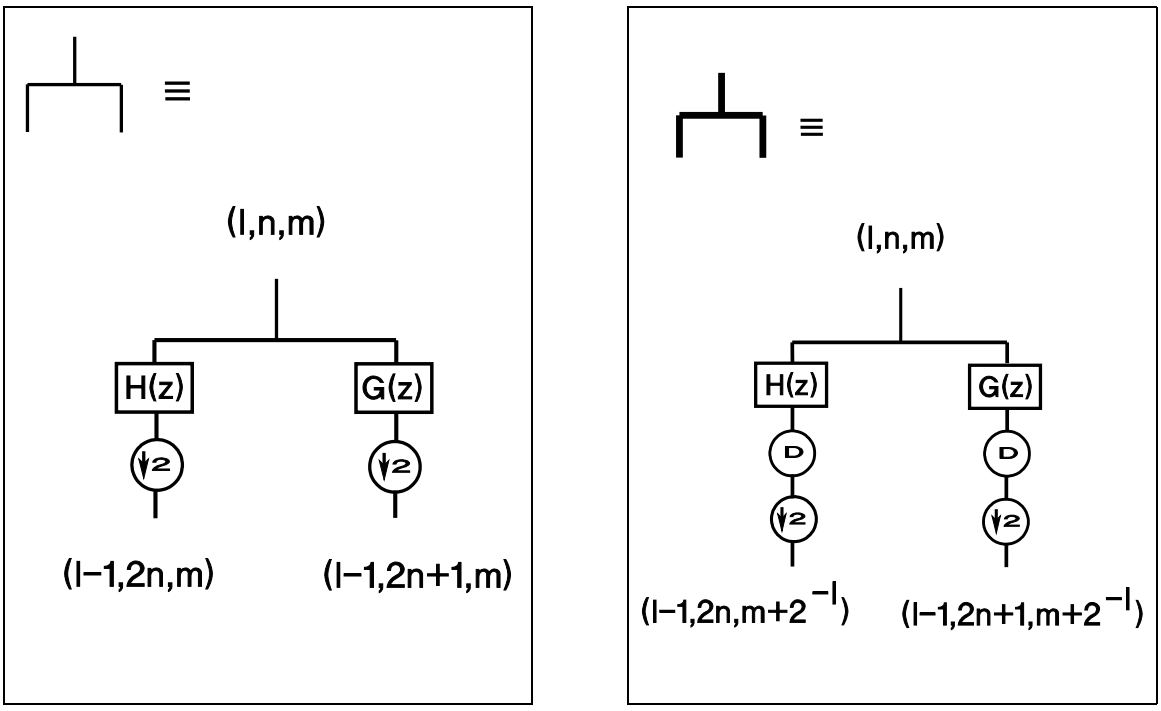


(a) (b)

Figure 2.4: Contour plots of time-frequency distributions for the signal $g(t)$: (a) Wigner distribution; (b) Smoothed Wigner distribution. Notice the close relation between the smoothed Wigner distribution and the SIWPD based time-frequency representation which is depicted in Fig. 2.3(b).

of SIWPD is facilitated via filter-bank terminology [121, 139]. Accordingly, each parent-node is expanded by high-pass and low-pass filters, followed by a 2:1 down-sampling. In executing WPD, down-sampling is achieved by ignoring all even-indexed (or all odd-indexed) terms. In contrast, when pursuing SIWPD, the down-sampling is carried out *adaptively* for the prescribed signal. We stress that owing to the orthogonality of the representation and the presumed additive nature of the cost functions (e.g., the Shannon entropy or rate-distortion), the decision at any given node is strictly local, i.e., independent of other nodes at the same resolution level.

The SIWPD expansion generates an ordinary binary tree [45]. However, each generated branch is now designated by either fine or heavy lines (Fig. 2.5) depending on the adaptive selection of the odd or the even terms, respectively. It can be readily observed that in contrast to WPD, SIWPD expansion leads to tree configurations that are independent of the time-origin. Fine and heavy lines may, however, exchange positions (e.g., compare



(a) (b)

Figure 2.5: A "parent" node binary expansion according to SIWPD: (a) High and low-pass filtering followed by a 2:1 downsampling. (b) High and low-pass filtering followed by a one sample delay (D) and subsequently by a 2:1 downsampling. Each node in the tree is indexed by the triplet (ℓ, n, m).

Figs. 2.3(a) and (c)).

The computational complexity of executing a best-basis SIWPD expansion is $O[2^d(L-d+2)N]$, where N denotes the length of the signal (at its highest resolution level), $L+1$ is the number of resolution levels ($L \leq \log_2 N$) and d is the maximum depth of a subtree used at a given parent-node to determine the shift mode of its children ($1 \leq d \leq L$). In the extreme case $d = 1$, the complexity, $O(NL)$, is similar to that associated with WPD, and the representation merges with that proposed in [57]. As a rule, the larger d and L, the larger the complexity, however, the determined best basis is of a higher quality; namely, characterized by a lower information cost.

For $d = L$ and for an identical number of resolution levels, SIWPD leads necessarily to an information cost that is lower than or equal to that resulting from standard WPD. This observation stems directly from the fact that the WP library constitutes a subset of the SWP library. In other words, WPD may be viewed as a degenerate form of SIWPD characterized by $d = 0$. In this case, the relative shift of newly generated nodes is non-adaptively set to zero and generally leads to shift-variant representations.

The best-basis expansion under SIWPD is also characterized by the invariance of the information cost. This feature is significant as it facilitates a meaningful quantitative comparison between alternative SWP libraries. Usually such a comparison between alternative libraries lacks meaning for WP, as demonstrated by the example summarized in Table 2.1.

Here, the entropies of the signals $g(t)$ (Fig. 2.1) and $g(t - 2^{-6})$ are compared. The expansions are on the best bases stemming from both the WPD and SIWPD algorithms and for D_8 and C_6 scaling functions (C_6 corresponds to 6-tap coiflet filters [53, page 261] [54]). We can readily observe the shift-invariance under SIWPD and the fact that the selection of D_8 is consistently advantageous over C_6. Just as obvious is the futility of attempting a comparison between the C_6 and D_8 based libraries under WPD. C_6 is better

	WPD		SIWPD	
	D_8	C_6	D_8	C_6
$g(t)$	2.84	**2.75**	**1.92**	2.35
$g(t-2^{-6})$	**2.59**	2.69	**1.92**	2.35

Table 2.1: Entropies of $g(t)$ (Fig. 2.1) and $g(t-2^{-6})$ represented on "best bases" obtained via WPD and SIWPD using libraries derived from D_8 and C_6 scaling functions. D_8 corresponds to 8-tap Daubechies wavelet filters, and C_6 corresponds to 6-tap coiflet filters.

for $g(t)$ while D_8 is advantageous in representing $g(t-2^{-6})$.

The organization of this chapter is as follows. In Section 2.2, we introduce a shifted wavelet packet library as a collection of orthonormal bases. Section 2.3 describes a best-basis selection algorithm. It is proved that the resultant best basis decomposition and the corresponding expansion tree are indeed shift-invariant. A shift-invariant wavelet transform is described in Section 2.4. The trade-off between computational complexity and information cost is the subject matter of Section 2.5. It presents suboptimal procedures for SIWPD, and examines their performance using both synthetic signals and a real dataset of acoustic transients. Finally, Section 2.6 briefly discusses the important extension to two-dimensional signals.

2.2 Shifted Wavelet Packet Bases

Let $\{h_n\}$ denote a real-valued quadrature mirror filter (QMF) obeying (e.g., [51, theorem (3.6), page 964],

$$\sum_n h_{n-2k} h_{n-2\ell} = \delta_{k,\ell} \tag{2.1}$$

$$\sum_n h_n = \sqrt{2}\,. \tag{2.2}$$

Let $\{\psi_n(x)\}$ be a wavelet packet family (e.g., [40, 150]) defined and generated via

$$\psi_{2n}(x) = \sqrt{2}\sum_k h_k \psi_n(2x-k) \tag{2.3}$$

$$\psi_{2n+1}(x) = \sqrt{2}\sum_k g_k \psi_n(2x-k) \tag{2.4}$$

where $g_k = (-1)^k h_{1-k}$, and $\psi_o(x) \equiv \varphi(x)$ is an orthonormal scaling function, satisfying

$$\langle \varphi(x-p), \varphi(x-q)\rangle = \delta_{p,q}, \quad p,q \in \mathbb{Z}\,. \tag{2.5}$$

Furthermore, let $f(x)$ be a function specified at the j'th resolution level, i.e. $f \in V_j$ where

$$V_j = \mathrm{clos}_{L^2(\mathbb{R})}\left\{2^{j/2}\psi_o(2^j x-k) \;:\; k \in \mathbb{Z}\right\}. \tag{2.6}$$

It is observed that the expansion of $f(x)$ on the standard basis $\{2^{j/2}\psi_o(2^j x-k) \;:\; k \in \mathbb{Z}\}$ remains invariant under $2^{-j}m$ shifts ($m \in \mathbb{Z}$). However, as $f(x) \in V_j$ is decomposed into orthonormal wavelet packets using the best-basis algorithm of Coifman and Wickerhauser [45], the often crucial property of shift-invariance is no longer valid. One way to achieve shift-invariance is to adjust the time-localization of the basis functions [111, 27, 32, 90]. That is, when an analyzed signal is translated in time by τ, a new best-basis is selected whose elements are also translated by τ compared to the former best-basis. Consequently, the expansion coefficients, that are now associated with translated basis functions, stay unchanged and the time-frequency representation is shifted in time by the same period. The ordinary construction of a wavelet packet (WP) library precludes the above procedure, since translated versions of library-bases are not necessarily included in the library. The proposed strategy in obtaining shift-invariance is based on extending the library to include all their shifted versions, organizing it in a tree structure and providing an efficient "best-basis" search algorithm.

To further pursue the stated objective we introduce the notation [27, 111]

$$B^j_{\ell,n,m} = \left\{2^{(\ell+j)/2}\psi_n\left[2^\ell(2^j x-m)-k\right] \;:\; k \in \mathbb{Z}\right\} \tag{2.7}$$

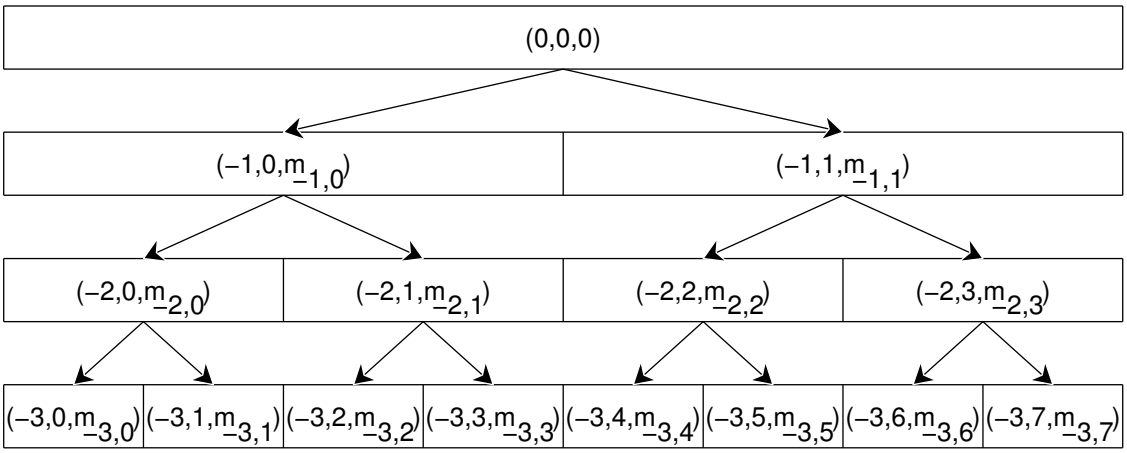


Figure 2.6: The extended set of wavelet packets organized in a binary tree structure. Each node in the tree is indexed by the triplet (ℓ, n, m) and represents the subspace $U^j_{\ell,n,m}$.

$$U^j_{\ell,n,m} = \mathrm{clos}_{L^2(\mathbb{R})} \left\{ B^j_{\ell,n,m} \right\} \tag{2.8}$$

and define *shifted-wavelet-packet* (SWP) library as a collection of all the orthonormal bases for V_j which are subsets of

$$\left\{ B^j_{\ell,n,m} \ : \ \ell \in \mathbb{Z}_-,\ n \in \mathbb{Z}_+,\ 0 \le m < 2^{-\ell} \right\} . \tag{2.9}$$

This library is larger than the WP library by a square power, but it can still be cast into a tree configuration facilitating fast search algorithms. The tree structure is depicted in Fig. 2.6. Each node in the tree is indexed by the triplet (ℓ, n, m) and represents the subspace $U^j_{\ell,n,m}$. Similar to the ordinary binary trees [45], the nodes are identified with dyadic intervals of the form $I_{\ell,n} = [2^\ell n,\ 2^\ell(n+1))$. The additional parameter m provides degrees of freedom to adjust the time-localization of the basis functions. The following proposition gives simple graphic conditions on subsets forming orthonormal bases.

Proposition 2.1 *Let $E = \{(\ell, n, m)\} \subset \mathbb{Z}_- \times \mathbb{Z}_+ \times \mathbb{Z}_+$, $0 \le m < 2^{-\ell}$, denote a collection of indices satisfying*

(i) The segments $I_{\ell,n} = [2^\ell n,\ 2^\ell(n+1))$ *are a disjoint cover of* $[0,\ 1)$.

(ii) The shift indices of a pair of nodes (ℓ_1, n_1, m_1), $(\ell_2, n_2, m_2) \in E$ *are related by*

$$m_1 \mod 2^{-\hat{\ell}+1} = m_2 \mod 2^{-\hat{\ell}+1} \tag{2.10}$$

where $\hat{\ell}$ *is the level index of a dyadic interval* $I_{\hat{\ell},\hat{n}}$ *that contains both* I_{ℓ_1,n_1} *and* I_{ℓ_2,n_2}.

Then E generates an orthonormal (ON) basis for $V_j \equiv U^j_{0,0,0}$, *i.e.* $\left\{B^j_{\ell,n,m} \,:\, (\ell,n,m) \in E\right\}$ *is an ON basis, and the set of all E as specified above generates a SWP library.*

Condition (ii) is equivalent to demanding that the relative shift between a prescribed parent-node (ℓ, n, m) and all its children-nodes is necessarily a constant whose value is restricted to either zero or to $2^{-\ell}$. In the dyadic one-dimensional case, each parent-node (ℓ, n, m) generates children-nodes $(\ell-1, 2n, m')$ and $(\ell-1, 2n+1, m'')$ where, according to condition (ii), their shift indices may take the value $m' = m'' = m$ or $m' = m'' = m + 2^{-\ell}$. The generated branches are respectively depicted by thin or heavy lines (*cf* Fig. 2.5).

The expansion tree associated with a given signal describes the signal's representation on an orthonormal basis selected from the SWP library. The index set E is interpreted as the collection of all terminal nodes. That is, all nodes beyond which no further expansion is to be carried out. A specific example of an expansion tree is shown in Fig. 2.7(a). The proposed configuration ensures that the set of terminal nodes satisfies the conditions of Proposition 2.1. In particular, refer to the terminal nodes $(-3, 0, 6)$ and $(-4, 5, 10)$. These nodes are descendants of $(-1, 0, 0)$. Hence, their related dyadic intervals $I_{-3,0} = [0, 1/8)$ and $I_{-4,5} = [5/16, 3/8)$ are contained in the dyadic interval $I_{-1,0} = [0, 1/2)$, and their shift indices are indeed related by

$$6 \bmod 2^2 = 10 \bmod 2^2 = 2\,.$$

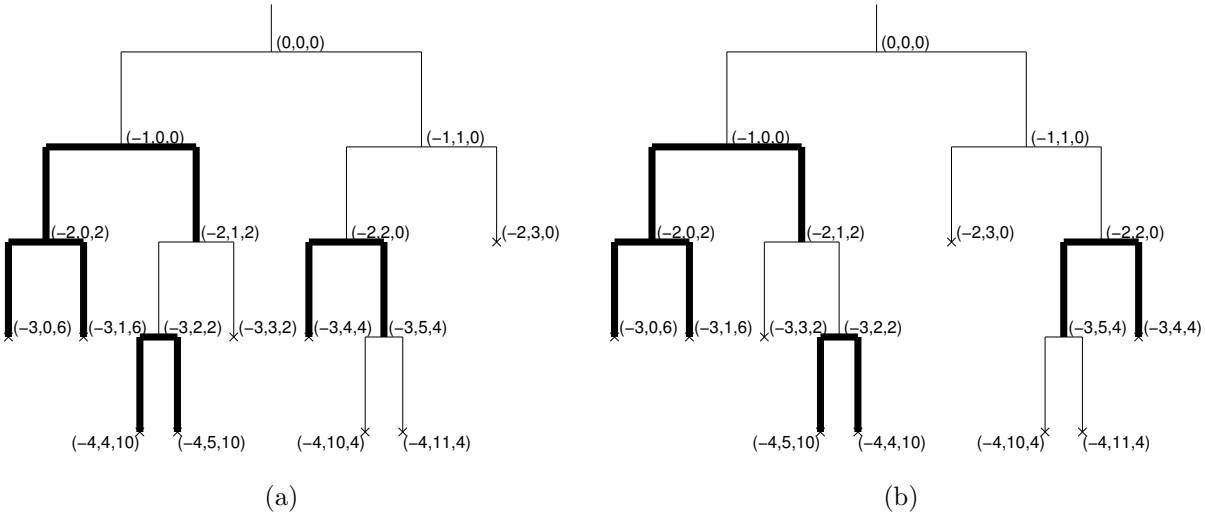


Figure 2.7: Exemplifying a SIWPD binary tree. (a) The children-nodes corresponding to (ℓ, n, m) are $(\ell - 1, 2n, \tilde{m})$ and $(\ell - 1, 2n + 1, \tilde{m})$, where $\tilde{m} = m$ (depicted by thin lines) or $\tilde{m} = m + 2^{-\ell}$ (depicted by heavy lines). (b) Rearrangement of the nodes in a *sequency* order.

The nodes of each level in this example have a natural or *Paley* order. It is normally useful to rearrange them in a *sequency* order [150], so that the nominal frequency of the associated wavelet packets increases as we move from left to right along a level of the tree. The rule to get a sequency ordered tree is to exchange the two children-nodes of each parent-node with odd sequency (inverse Gray code permutation [150, page 250]). The resultant tree is depicted in Fig. 2.7(b).

2.3 The Best-Basis Selection

Likewise the wavelet packet library [45], the tree configuration of the extended library facilitates an efficient best basis selection process. However, in contrast to the WPD, the best-basis representation is now shift-invariant.

Let $f \in V_j = U^j_{0,0,0}$, let $\mathcal{M}$ denote an additive cost function and let $\mathcal{B}$ represent a SWP

library.

Definition 2.1 [45] *The best basis for f in $\mathcal{B}$ with respect to $\mathcal{M}$ is $B \in \mathcal{B}$ for which $\mathcal{M}(Bf)$ is minimal. Here, $\mathcal{M}(Bf)$ is the information cost of representing f in the basis $B \in \mathcal{B}$.*

Let $A^j_{\ell,n,m}$ denote the best basis for the subspace $U^j_{\ell,n,m}$. Accordingly, $A^j_{0,0,0}$ constitutes the best basis for $f \in V_j$ with respect to $\mathcal{M}$. Henceforth, for notational simplicity, we omit the fixed index j. The desired best basis can be determined recursively by setting

$$A_{\ell,n,m} = \begin{cases} B_{\ell,n,m} \text{ if } \mathcal{M}(B_{\ell,n,m}f) \leq \mathcal{M}(A_{\ell-1,2n,m_c}f) + \mathcal{M}(A_{\ell-1,2n+1,m_c}f)\,, \\ A_{\ell-1,2n,m_c} \oplus A_{\ell-1,2n+1,m_c}, \text{ otherwise,} \end{cases} \tag{2.11}$$

where the shift indices of the respective children-nodes are given by

$$m_c = \begin{cases} m, \text{ if } \sum_{i=0}^{1} \mathcal{M}(A_{\ell-1,2n+i,m}f) \leq \sum_{i=0}^{1} \mathcal{M}(A_{\ell-1,2n+i,m+2^{-\ell}}f) \\ m + 2^{-\ell}, \quad \text{otherwise.} \end{cases} \tag{2.12}$$

The recursive sequence proceeds down to a specified level $\ell = -L$ ($L \leq \log_2 N$), where

$$A_{-L,n,m} = B_{-L,n,m}\,. \tag{2.13}$$

The stated procedure resembles that proposed by Coifman and Wickerhauser [45] with an added degree of freedom facilitating a relative shift (i.e., $m_c \neq m$) between a parent-node and its respective children-nodes. It is re-emphasized that the recursion considered herein restricts the shift to one of two values ($m_c - m \in \{0, 2^{-\ell}\}$). Other values are unacceptable if the orthonormality of the best basis is to be preserved. As it turns out, the generated degree of freedom is crucial in establishing time-invariance. The recursive sequence proposed in [45] may be viewed as a special case where $m_c - m$ is arbitrarily set to zero.

Lemma 2.1 *Let E_1 and E_2 denote index collections obeying Proposition 2.1, and let B_1 and B_2 be the corresponding orthonormal bases. Then B_1 and B_2 are "identical to within a*

time-shift" if and only if there exists a constant $q \in \mathbb{Z}$ *such that for all* $(\ell, n, m) \in E_1$*, we have* $(\ell, n, \tilde{m}) \in E_2$ *where* $\tilde{m} = (m + q) \bmod 2^{-\ell}$.

Proof: Bases in V_j are said to be *identical to within a time-shift* if and only if there exists $q \in \mathbb{Z}$ such that for each element in B_1 we have an identical element in B_2 that is time-shifted by $q2^{-j}$. Namely, if

$$2^{(\ell+j)/2}\psi_n[2^\ell(2^j x - m) - k] \in B_1$$

then

$$2^{(\ell+j)/2}\psi_n[2^\ell(2^j(x - q2^{-j}) - m) - k] \in B_2\,.$$

If E denotes index collection obeying Proposition 2.1 and B is its corresponding basis, then $(\ell, n, m) \in E$ is equivalent to $B^j_{\ell,n,m} \subset B$. Therefore, by observing that

$$\psi_n[2^\ell(2^j(x - q2^{-j}) - m) - k] = \psi_n[2^\ell(2^j x - \tilde{m}) - \tilde{k}]\,,$$

where $\tilde{m} = (m + q) \bmod 2^{-\ell}$ and $\tilde{k} = k + \lfloor 2^\ell(m + q)\rfloor$, the proof is concluded.

□.

Definition 2.2 *Binary trees are said to be "identical to within a time-shift" if they correspond to bases that are "identical to within a time-shift".*

Fig. 2.3(a) and (c) depict *identical to within a time-shift* trees representing the identical to within time-shift signals.

Proposition 2.2 *The best basis expansion stemming from the previously described recursive algorithm is shift-invariant.*

Proof: Let $f, g \in V_j$ be identical to within a time-shift, i.e. there exists $q \in \mathbb{Z}$ such that $g(x) = f(x - q2^{-j})$. Let A_f and A_g denote the best bases for f and g, respectively. It can

be shown (Appendix A.1) that

$$B_{\ell,n,m} \subset A_f$$

implies

$$B_{\ell,n,\tilde{m}} \subset A_g, \qquad \tilde{m} = (m+q) \bmod (2^{-\ell})$$

for all $m, n \in \mathbb{Z}_+$ and $\ell \in \mathbb{Z}_-$. Hence, A_f and A_g are *identical to within a time-shift.*

□.

The number of orthonormal bases contained in the shifted WP library can be computed recursively. Let S_L denote the number of bases associated with a $(L+1)$-level tree expansion (i.e., the expansion is to be executed down to the $\ell = -L$ level). The tree comprises a root and two L-level subtrees. Since two options exist for selecting the relative shift, we have

$$S_L = 1 + 2S_{L-1}^2, \quad S_O = 1\,. \tag{2.14}$$

Consequently, it can be shown by induction that for $L > 2$

$$0.5(2.48)^{2^L} < S_L < 0.5(2.49)^{2^L}\,. \tag{2.15}$$

A length N signal may be represented by S_L different orthonormal bases ($L \leq log_2 N$), from which the best basis is selected. While the associated complexity level is of $O(N2^{L+1})$, we demonstrate in Section 2.5 that the algorithmic complexity may be reduced substantially (down to a level of $O(NL)$) while still retaining shift-invariance. The reduced complexity, however, may lead to representations characterized by a higher cost function values.

For the sake of comparison with the established WPD algorithm [45], let s_L denote the number of bases associated with a $(L+1)$-level tree. Then

$$s_L = 1 + s_{L-1}^2, \quad s_o = 1 \tag{2.16}$$

and consequently, for $L > 2$

$$(1.50)^{2^L} < s_L < (1.51)^{2^L}\,. \tag{2.17}$$

The WPD algorithm has an attractive complexity level of $O(NL)$. However, the best basis representation is not shift-invariant. It is worthwhile stressing that despite the fact that $S_L > s_L^2$ for $L > 2$, the complexity level characterizing SIWPD is significantly below the squared WPD complexity. Specifically, $O(N2^{L+1}) \ll O(N^2L^2)$.

2.4 The Shift-Invariant Wavelet Transforms

The property of shift-invariance can also be achieved within the framework of the wavelet transform (WT) and a prescribed information cost function ($\mathcal{M}$) [90, 107]. It may be viewed as a special case whereby the tree configuration is constrained to expanding exclusively the *low frequency* nodes. The signal is expanded by introducing a scaling function (ψ_o) or a "mother-wavelet" (ψ_1). To achieve shift-invariance, we again permit the introduction of a relative shift between children-nodes and their parent-node. The shift selection is, once again, based on minimizing the cost function ($\mathcal{M}$) at hand. This procedure yields the *wavelet-best-basis* for a signal $f \in V_j$ with respect to ($\mathcal{M}$), among all the orthonormal bases generated by

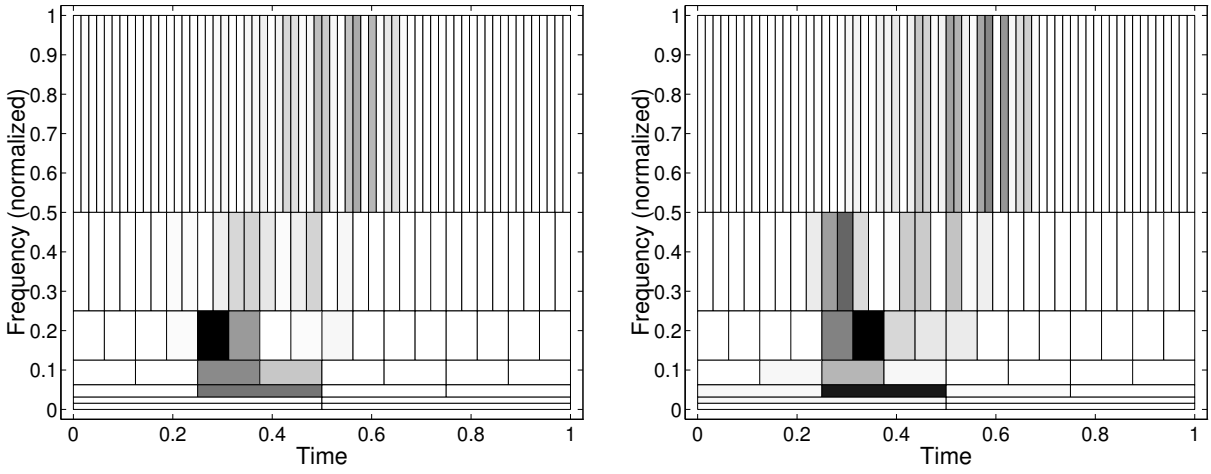
$$\left\{ B^j_{\ell,n,m} \ : \ \ell \in \mathbb{Z}_-,\ n \in (0,1),\ 0 \le m < 2^{-\ell} \right\} .$$

Let $W_{\ell,m}$ denote the *wavelet-best-basis* for $U^j_{\ell,0,m}$. The *wavelet-best-basis* for $f \in V_j \equiv U^j_{0,0,0}$ may be determined recursively via

$$W_{\ell,m} = W_{\ell-1,m_c} \oplus B^j_{\ell-1,1,m_c} \tag{2.18}$$

where

$$m_c = \begin{cases} m, \text{ if } \mathcal{M}(W_{\ell-1,m}f) + \mathcal{M}(B^j_{\ell-1,1,m}f) \le \mathcal{M}(W_{\ell-1,m+2^{-\ell}}f) + \mathcal{M}(B^j_{\ell-1,1,m+2^{-\ell}}f)\,, \\ m + 2^{-\ell}, \quad \text{otherwise.} \end{cases} \tag{2.19}$$



(a) (b)

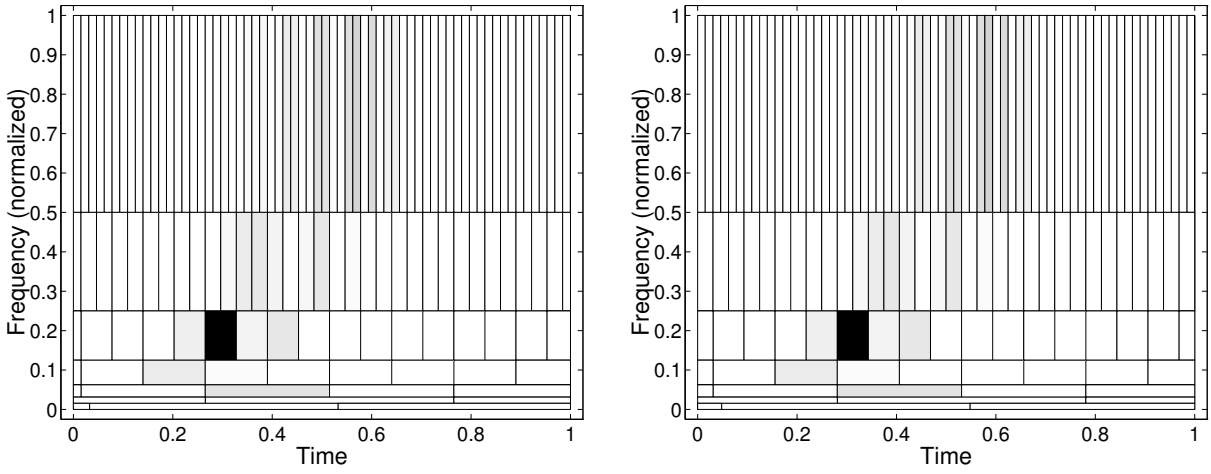
Figure 2.8: Time-frequency representation in the wavelet basis using 6-tap coiflet filters: (a) The signal $g(t)$; Entropy= 3.22. (b) The signal $g(t - 2^{-6})$; Entropy= 3.34.

The expansion is performed down to the level $\ell = -L$ ($L \leq \log_2 N$), namely

$$W_{-L,m} = B^j_{-L,0,m}\,. \tag{2.20}$$

A N-element signal may be represented by 2^L different orthonormal wavelet bases. The associated complexity level is $O(NL)$ and the resultant expansion is indeed shift-invariant.

As an example, we now refer to the signal $g(t)$, depicted in Fig. 2.1, and its translation $g(t - 2^{-6})$. The corresponding wavelet transforms, with C_6 as the scaling function [53, page 261] [54], are described in Fig. 2.8. The variations in the energy spreads of $g(t)$ and $g(t - 2^{-6})$, stemming directly from the lack of shift-invariance, are self-evident. Moreover, the transformed cost function (the Shannon entropy) is shift dependent as well. In complete contrast, the *wavelet-best-basis* decompositions depicted in Fig. 2.9, yield identical (to within a time-shift) energy distributions. The corresponding entropy is lower and independent of the time-shift.



(a) (b)

Figure 2.9: Time-frequency representation in the wavelet-best-basis using 6-tap coiflet filters: (a) The signal $g(t)$; Entropy= 3.02. (b) The signal $g(t - 2^{-6})$; Entropy= 3.02.

2.5 The Information-Cost Complexity Trade-Off

So far we have observed that WPD lacks shift-invariance but is characterized by an attractive complexity level $O(NL)$, where L denotes the lowest resolution level in the expansion tree. Comparatively, the quadratic complexity level, $O(N2^{L+1})$, associated with SIWPD is substantially higher. In return, one may achieve a potentially large reduction of the information cost, in addition to gaining the all important *shift-invariance*. However, whenever the SIWPD complexity is viewed as intolerable, one may resort to a sub-optimal SIWPD procedure entailing a reduced complexity, and higher information cost while still retaining the desirable shift-invariance.

The best basis for $f \in V_j$ with respect to $\mathcal{M}$ is, once again, obtained recursively via (2.11), but contrary to the procedure of Section 2.3, now the selection of a relative shift at a given parent-node does not necessitate tree expansion down to the lowest level. While an optimal decision on the value of a shift index is provided by (2.12), a sub-optimal shift

index may be determined by

$$m_c = \begin{cases} m, & \text{if } \sum_{i=0}^{1} \mathcal{M}(C_{\ell-1,2n+i,m,d} f) \leq \sum_{i=0}^{1} \mathcal{M}(C_{\ell-1,2n+i,m+2^{-\ell},d} f)\,, \\ m + 2^{-\ell}, & \text{otherwise,} \end{cases} \tag{2.21}$$

where $C_{\ell,n,m,d}$ denotes the best basis for $U_{\ell,n,m}$ subject to constraining the decomposition to d $(1 \leq d \leq L)$ resolution levels. Accordingly, the shift indices are estimated using subtrees of d_ℓ resolution levels depth $(d_\ell \leq d)$, where

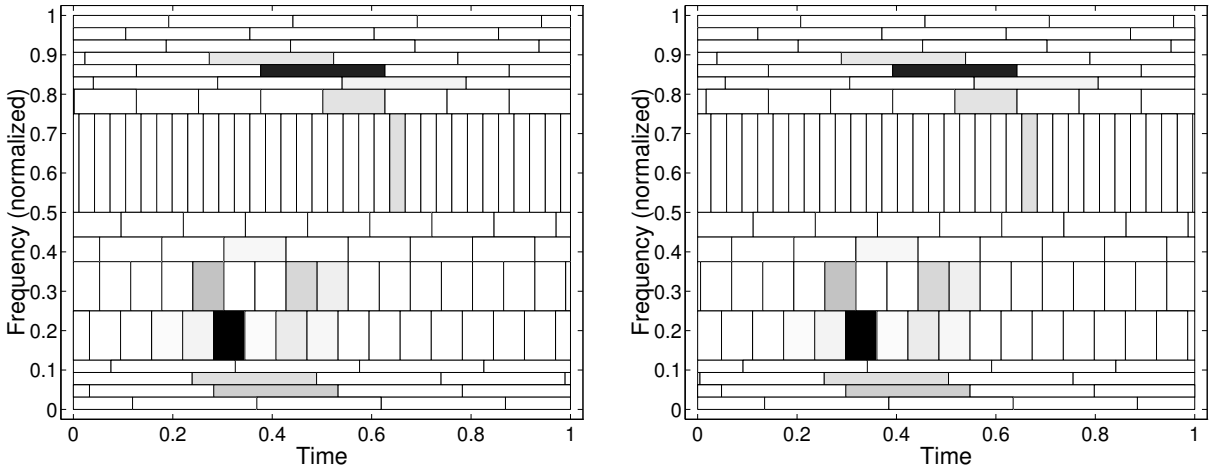
$$d_\ell = \begin{cases} d, & d - L \leq \ell \leq 0 \\ L + \ell, & \text{otherwise.} \end{cases} \tag{2.22}$$

For $d = 1$ or at the coarsest resolution level $\ell = -L$ we have $C_{\ell,n,m,d} = B_{\ell,n,m}$. For $\ell > -L$ and $d > 1$ $C_{\ell,n,m,d}$ is obtained recursively according to

$$C_{\ell,n,m,d} = \begin{cases} B_{\ell,n,m} \\ C_{\ell-1,2n,m,d-1} \oplus C_{\ell-1,2n+1,m,d-1} \\ C_{\ell-1,2n,m+2^{-\ell},d-1} \oplus C_{\ell-1,2n+1,m+2^{-\ell},d-1} \end{cases} \tag{2.23}$$

where $C_{\ell,n,m,d}$ takes on that value which minimizes the cost function $\mathcal{M}$.

The shift-invariance is retained for all $1 \leq d \leq L$. The cases $d = L$ and $d < L$ should be viewed as *optimal* and *sub-optimal* with respect to the prescribed information cost function $(\mathcal{M})$. The best-basis search algorithm of Coifman and Wickerhauser [45] corresponds to the special case $m_c = m$ for all nodes $(d \equiv 0)$. Quite expectedly, the non-adaptive selection yields representations that are not, in general, shift-invariant. Fig. 2.10 depicts the time-frequency representations of the signals $g(t)$ and $g(t - 2^{-6})$, using the sub-optimal SIWPD(d=1) with 8-tap Daubechies least asymmetric wavelet filters. The resultant entropy is higher than is obtained using the optimal SIWPD (Fig. 2.3). Yet, the valuable property of shift-invariance is provided with a significant reduction in the computational complexity.



(a) (b)

Figure 2.10: Time-frequency representation using the sub-optimal (d=1) SIWPD with 8-tap Daubechies least asymmetric wavelet filters: (a) The signal $g(t)$; Entropy= 2.32. (b) The signal $g(t - 2^{-6})$; Entropy= 2.32.

Since, at each level ℓ, the subtrees employed in estimating the shift indices are restricted to d_ℓ levels depth ($d_\ell \leq d$), the complexity is now $O[N2^d(L - d + 2)]$. More specifically, the algorithm requires $rN[2^d(L - d + 2) - 2]$ real multiplications, where r is the length of the filters. In the extreme case, $d = 1$, the complexity, $O(2NL)$, resembles that associated with WPD, and the representation merges with that proposed in [57]. As a rule, the larger d and L, the larger the complexity, however, the determined best basis is of a higher quality; namely, characterized by a lower information cost.

2.5.1 Example

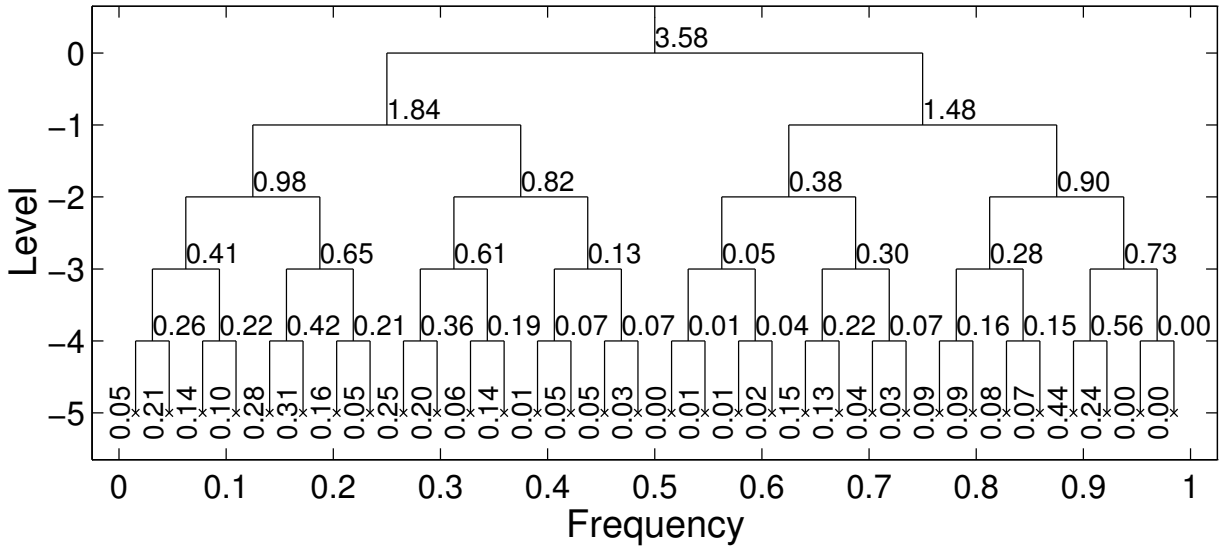
To demonstrate the trade-off between information-cost and complexity we refer to Figs. 2.11–2.13. These figures depict the expansion trees of the signal $g(t)$, either when the relative shifts are arbitrarily set to zero (the WPD algorithm), estimated using one-level-depth subtrees (sub-optimal SIWPD with d=1), or estimated using two-levels-depth subtrees (sub-

optimal SIWPD with d=2). The numbers associated with the nodes of the tree represent the entropies of g in the corresponding subspaces. For the best expansion trees, the numbers represent the minimum entropies obtained by the best-basis algorithms.

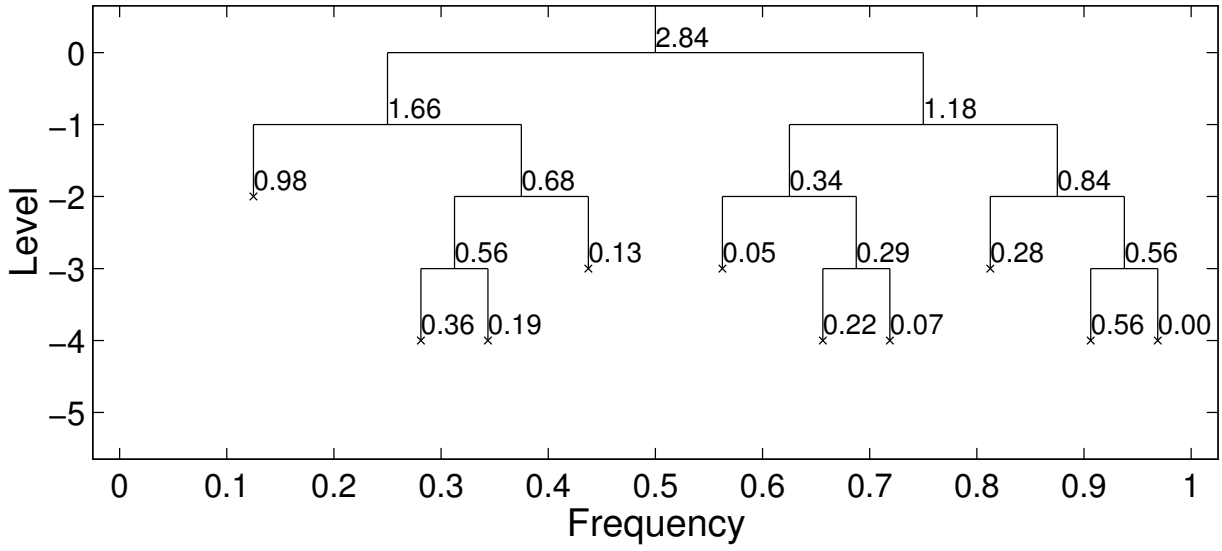
The initial entropy of the signal g is 3.58. The children-nodes of the root-node have lower entropy when we introduce a relative shift (regarding to Figs. 2.11(a) and 2.12(a): $1.85 + 1.41 < 1.84 + 1.48$). Hence the root-node decomposition in Fig. 2.12(a) is carried out with "heavy lines". Now, consider the expansion of the node specified by $(\ell, n, m) = (-1, 0, 1)$ (the left node at the level $\ell = -1$). If the relative shift is based on a one-level-depth subtree, then no relative shift is required (regarding to Figs. 2.12(a) and 2.13(a): $1.02 + 0.63 < 1.09 + 0.70$). However, a deeper subtree reveals that a relative shift is actually more desirable, and a lower entropy for the node $(-1, 0, 1)$ is attainable (regarding to Figs. 2.12(b) and 2.13(b): $1.23 < 1.49$). The eventual entropy of the signal g is 2.84 when implementing the WPD algorithm, 2.32 when using the sub-optimal SIWPD(d=1), and 1.92 when using the sub-optimal SIWPD(d=2). The number of real multiplications required by these algorithms are respectively $rNL = 5120$, $2rNL = 10240$ or $rN(4L - 2) = 18432$, where the length of the signal is $N = 128$, the number of decomposition levels is $L = 5$, and the filters' length is $r = 8$. In this particular example, larger d values do not yield a further reduction in the information cost, since $d = 2$ has already reached the optimal SIWPD (compare Figs. 2.13(b) and 2.3(a)).

2.5.2 Experiment

Normally, as was the case for the above example, the information cost decreases when the shift indices are evaluated based on deeper subtrees (larger d). Notwithstanding an assured reduction in information cost using the optimal SIWPD, sub-optimal SIWPD may anomalously induce an increase. We have performed an experiment on 50 acoustic transients, generated by explosive charges at various distances (these signals are detected

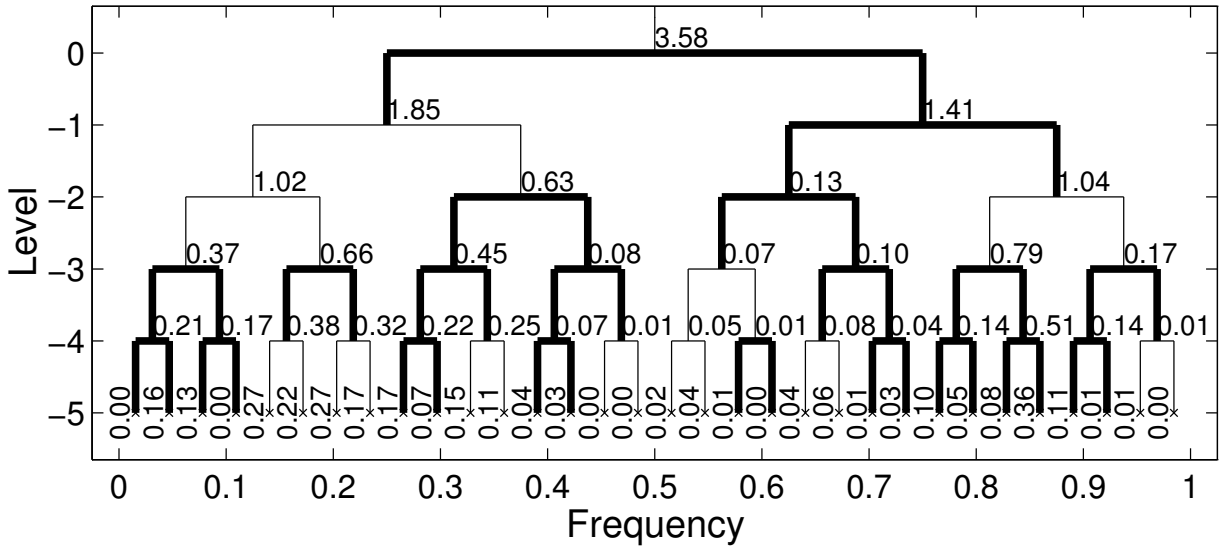


(a)

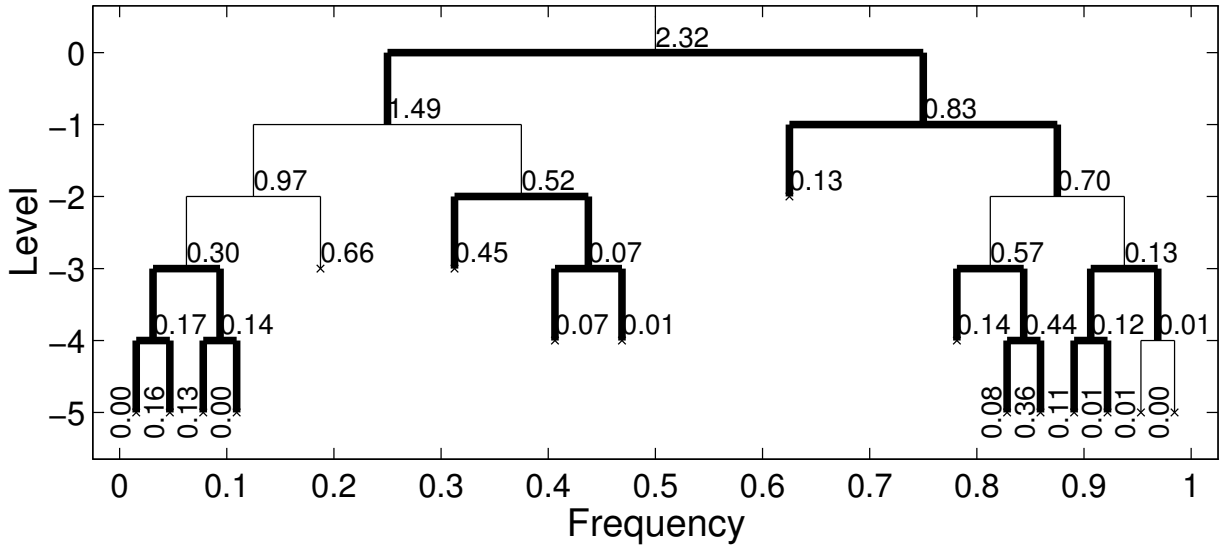


(b)

Figure 2.11: Wavelet packet library trees of the signal $g(t)$: (a) Five-level expansion tree; The numbers represent the entropies of g in the corresponding subspaces. (b) The best expansion tree; The numbers represent the minimum entropies obtained by the best-basis algorithm.

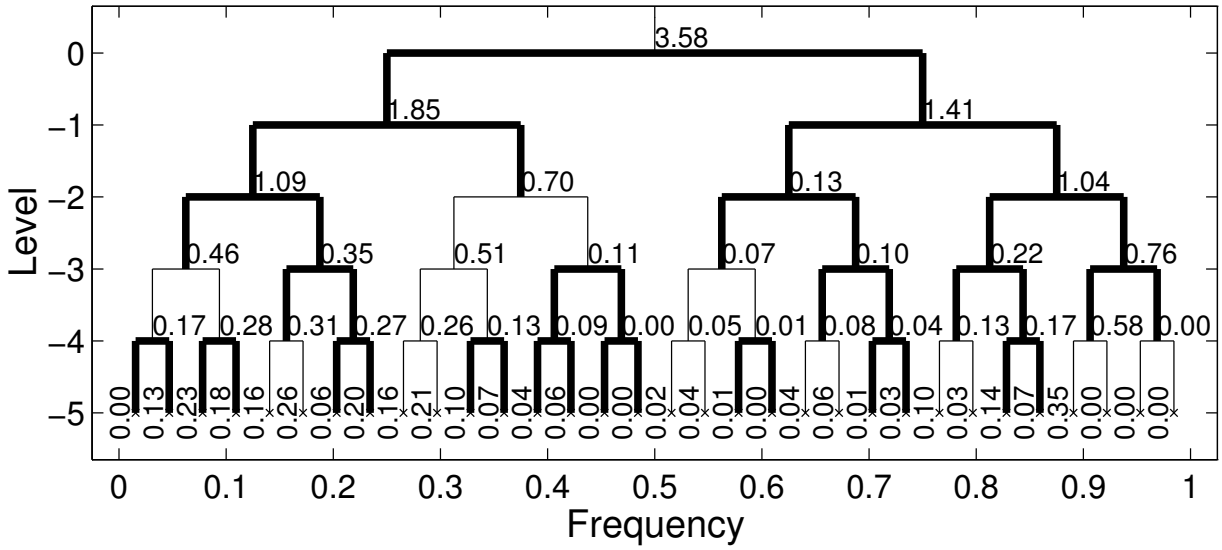


(a)

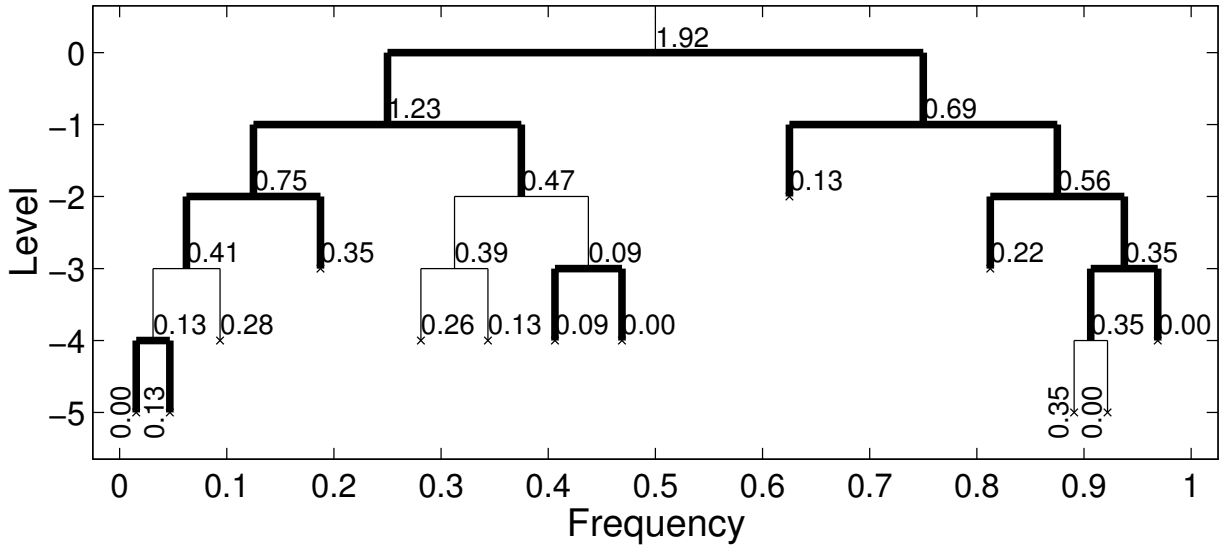


(b)

Figure 2.12: Shifted wavelet packet library trees of the signal $g(t)$: (a) Five-level expansion tree, where the relative shifts are estimated using one-level-depth subtrees (d=1); The numbers represent the entropies of g in the corresponding subspaces. (b) The best expansion tree; The numbers represent the minimum entropies obtained by the sub-optimal (d=1) best-basis algorithm.



(a)



(b)

Figure 2.13: Shifted wavelet packet library trees of the signal $g(t)$: (a) Five-level expansion tree, where the relative shifts are estimated using two-levels-depth subtrees (d=2); The numbers represent the entropies of g in the corresponding subspaces. (b) The best expansion tree; The numbers represent the minimum entropies obtained by the sub-optimal (d=2) best-basis algorithm.

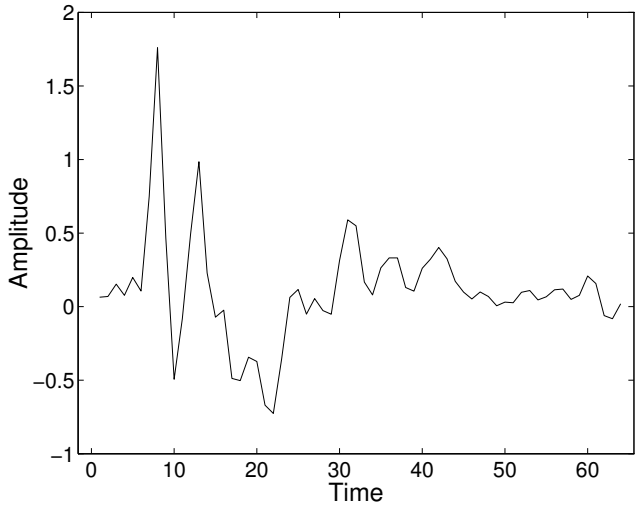


Figure 2.14: Typical acoustic pressure waveform in free air from explosive charges.

by an array of receivers and used to evaluate the location of explosive devices). Fig. 2.14 shows a typical acoustic pressure waveform containing 64 samples. We applied the WPD algorithm, the sub-optimal SIWPD with d=1 or d=2, and the optimal SIWPD to the compression of this data set. The decomposition was carried out to maximum level L=5 using 8-tap Daubechies minimum phase wavelet filters. The number of real multiplications required by these algorithms for expanding a given waveform in its best basis are respectively 2560, 5120, 9216 and 31744.

Table 2.2 lists the attained entropies by the best-basis algorithms for an arbitrary subset of ten waveforms. Clearly, the average entropy is lower when using the SIWPD. It decreases when d is larger, and a minimum value is reached using the optimal SIWPD ($d = L$). Moreover, the variations in the information cost, which indicate performance robustness across the data set, are also lower when using the SIWPD. Notice the irregularity pertaining to the eighth waveform. While its minimum entropy is expectedly obtained by implementing the optimal SIWPD, the sub-optimal SIWPD with d=1 fails to reduce the entropy in comparison with the conventional WPD.

waveform#	WPD	SIWPD		
	$L = 5$	$d = 1$	$d = 2$	$d = L = 5$
1	1.829	1.706	1.659	1.494
2	2.463	1.997	1.997	1.997
3	2.725	2.347	2.256	2.045
4	2.501	2.086	2.078	2.078
5	1.656	1.606	1.606	1.593
6	2.398	2.339	2.251	2.212
7	2.461	2.281	2.020	2.020
8	2.277	2.280	2.151	2.141
9	1.720	1.572	1.449	1.419
10	2.154	1.626	1.623	1.623
mean	2.218	1.984	1.909	1.862
variance	0.367	0.327	0.297	0.295

Table 2.2: Entropies attained by the conventional WPD, sub-optimal SIWPD (d <L) and optimal SIWPD (d=L) for acoustic pressure waveforms. The average entropy and the variance are lower when using the SIWPD, and they further decrease when d is larger.

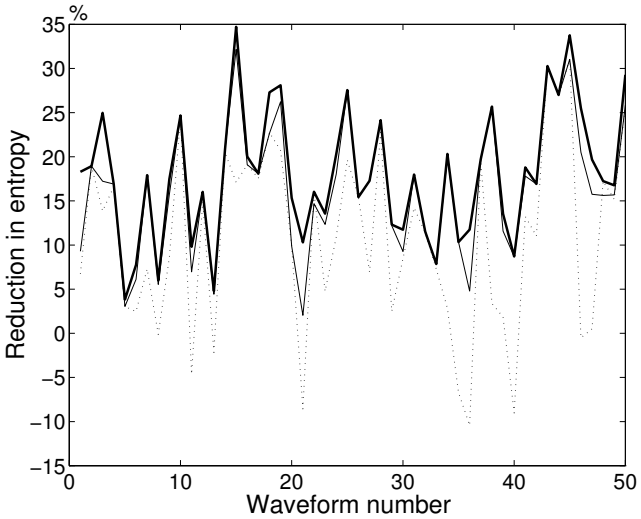


Figure 2.15: Percentage of reduction in entropy over the conventional WPD using the optimal SIWPD (heavy solid line), the sub-optimal SIWPD with d=2 (fine solid line) and the sub-optimal SIWPD with d=1 (dotted line).

To illustrate the improvement in information cost of the SIWPD with various d values over the conventional WPD, we plot in Fig. 2.15 the reduction in entropy relative to the entropy obtained using the WPD. We can see that for some signals the entropy is reduced by more than 30%. The average reduction is 10.8% by the sub-optimal SIWPD(d=1), 16.4% by the sub-optimal SIWPD(d=2), and 18.1% by the optimal SIWPD. Thus the average performance of SIWPD is increasingly improved as we deepen the subtrees used in estimating the shift indices.

2.6 Extension to 2D Wavelet Packets

Referring to Section 2.3, the best-basis representation of a signal is rendered shift-invariant by allowing a relative shift between a parent-node and its respective children-nodes in the expansion tree. The procedure remains essentially the same and leads to analogous results when applied to 2D signals [59, 89]. In this case, a shift with respect to the origin is a vector quantity $m = (m_x, .m_y)$. If we desire to generate a best-basis decomposition that remains invariant under shifts in the X–Y plane, we must permit a, now two-dimensional, parent-children relative shift, to be determined adaptively. Let m_p and m_c denote the parent and children shift with respect to the origin ($x = y = 0$). The relative shift $(m_c - m_p)$ may take on any one of four values

$$m_c - m_p = \left\{(0,0), (2^{-\ell}, 0), (0, 2^{-\ell}), (2^{-\ell}, 2^{-\ell})\right\} .$$

The value to be adapted is, once more, the one that minimizes the information cost. The proof follows along the lines charted in the one-dimensional case.

It should be stressed, however, that while the 2D expansion thus attained is shift-invariant in x and y, it is not invariant under rotation.

2.7 Summary

We have defined an extended library of wavelet packets that included all the shifted versions of wavelet packet bases, and presented efficient search algorithms for selecting the best basis. When compared with the conventional WPD algorithm [45], SIWPD is determined to be advantageous in three respects. First, it leads to a best basis expansion that is shift-invariant. Second, the resulting representation is characterized by a lower information cost. Third, the complexity is controlled at the expense of the information cost.

The stated advantages, namely the shift-invariance as well as the lower information cost, may prove crucial to signal compression, identification or classification applications. Furthermore, the shift-invariant nature of the information cost, renders this quantity a characteristic of the signal for a prescribed wavelet packet library. It should be possible now to quantify the relative efficiency of various libraries (i.e., various scaling function selections) with respect to a given cost function. Such a measure would be rather senseless for shift-variant decompositions.

The complexity associated with the SIWPD algorithm is $O(2^d N(L-d+2))$ (recall, N denotes the length of the signal, L is the number of tree decomposition levels and d limits through (2.22) the depth of the subtrees used to estimate the optimal children-nodes). One may exercise a substantial control over the complexity. The key to controlling the complexity is the built-in flexibility in the choice of d. Lower d implies lower complexity at the expense of a higher information cost. At its lower bound, $d = 1$, the attained level of complexity, $O(NL)$, resembles that of WPD while still guaranteeing shift-invariance.

The presented procedure is based on the general approach: extend the library of bases to include all their shifted versions, organize it in a tree structure and provide an efficient "best-basis" search algorithm. Clearly, it is not restricted to wavelet-packets and shift-invariance. In the next chapter we show that local trigonometric bases can be used as well,

and various extensions lead to enhanced representations.

Chapter 3

Shift-Invariant Local Trigonometric Decompositions

3.1 Introduction

Local trigonometric decompositions [2, 45, 102] can be considered as conjugates of the wavelet packet decompositions, where the partitioning of the frequency axis is replaced by smoothly partitioning the time axis. With this decomposition, a prescribed signal is first split into overlapping intervals. Then a folding operator [150] "folds" overlapping parts into the segments, and a standard sine or cosine transform is applied on each segment. In this case, the basis functions are sines or cosines multiplied by smooth window functions.

Similar to the wavelet packet bases, the local trigonometric bases construct a library of orthonormal bases, which is organized into a binary tree structure. The best basis which minimizes a certain information cost function is searched using the divide-and-conquer algorithm [45]. Unfortunately, the local trigonometric decomposition possesses the same drawback of the wavelet packet decomposition: it is sensitive to translation of the input signal. The expansion, as well as the information cost, are significantly influenced by the alignment of the input signal with respect to the basis functions.

In this chapter, the strategy for obtaining shift-invariance for the wavelet packet decomposition is applied to smooth local trigonometric bases. We extend the library to include all their shifted versions, organize it in a tree structure and provide efficient "best-basis" search algorithms. To further enhance the resultant representation, we introduce an adaptive-polarity folding operator which splits the prescribed signal and "folds" *adaptively* overlapping parts back into the segments. It is proved that the proposed algorithms lead to best-basis representations, which are shift-invariant, orthogonal and characterized by lower information cost.

We show that the shift-invariance stems from a relative shift between expansions in distinct resolution levels. Accordingly, at any resolution level ℓ it suffices to examine and select one of two relative shift options — a zero shift or a $2^{-\ell-1}$ shift. The choice between these two options, enabled by the extended library, is made in accordance with minimizing the information cost. Hence, the attained representation is not only shift-invariant, but also characterized by a lower information cost when compared to the local cosine decomposition (LCD) [45]. One of the proposed algorithms, namely the *shift-invariant adapted-polarity local trigonometric decomposition* (SIAP-LTD), further enhances the resultant representation by introducing an additional degree of freedom, which incorporates a periodic folding operator into the best decomposition tree. The polarity of the folding operation is locally adapted to the signal at the finest resolution level, and a recursive process is carried out towards the coarsest resolution level merging segments where beneficial. Each segment of the signal is then represented by a trigonometric basis which possesses the same parity properties at the end-points.

The computational complexity of the SIAP-LTD is $O[N(L+2^{\log_2 N-L+1})\log_2 N]$, where N denotes the length of the signal and $L+1$ is the number of resolution levels ($L \leq \log_2 N$). This complexity is comparable to that of the LCD ($O(NL\log_2 N)$) [45] with the benefits of shift-invariance and a higher quality (lower "information cost") "best-basis". To

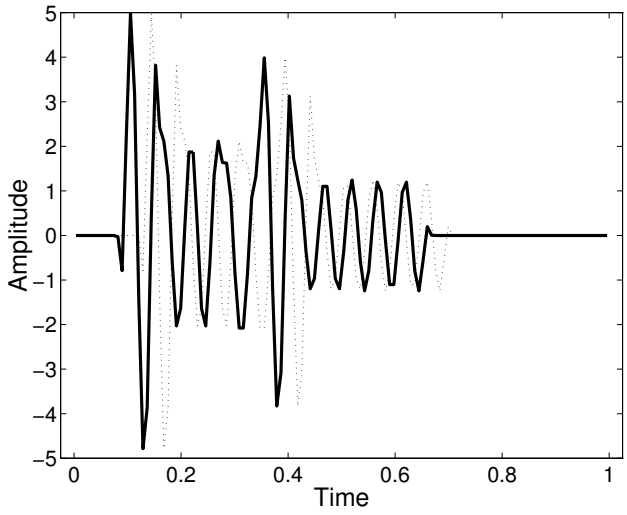
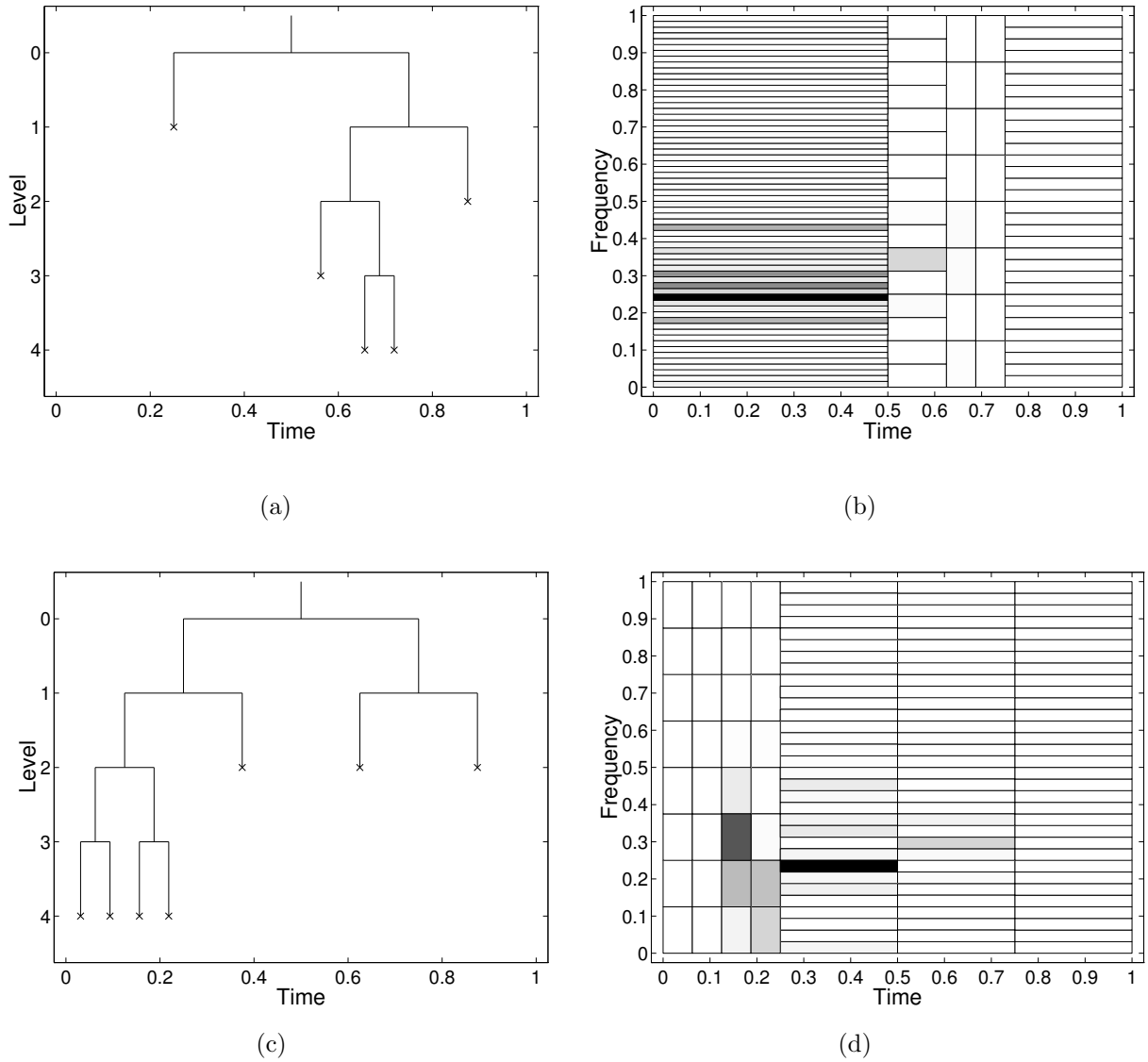


Figure 3.1: The signals $g(t)$ (solid) and $g(t-5\cdot 2^{-7})$ (dotted), sampled at 2^7 equally spaced points.

demonstrate the shift-invariant properties of SIAP-LTD, compared to LCD which lacks this feature, we refer to the expansions of the signals $g(t)$ and $g(t-5\cdot 2^{-7})$ (Fig. 3.1). These signals contain $2^7 = 128$ samples. For definiteness, we choose the Shannon entropy as the cost function. Figs. 3.2 and 3.3 depict the "best-basis" expansions under the LCD and the SIAP-LTD algorithms, respectively. A comparison of Figs. 3.2(b) and (d) readily reveals the sensitivity of LCD to temporal shifts while the "best-basis" SIAP-LTD representation is indeed shift-invariant and characterized by a lower entropy (Fig. 3.3).

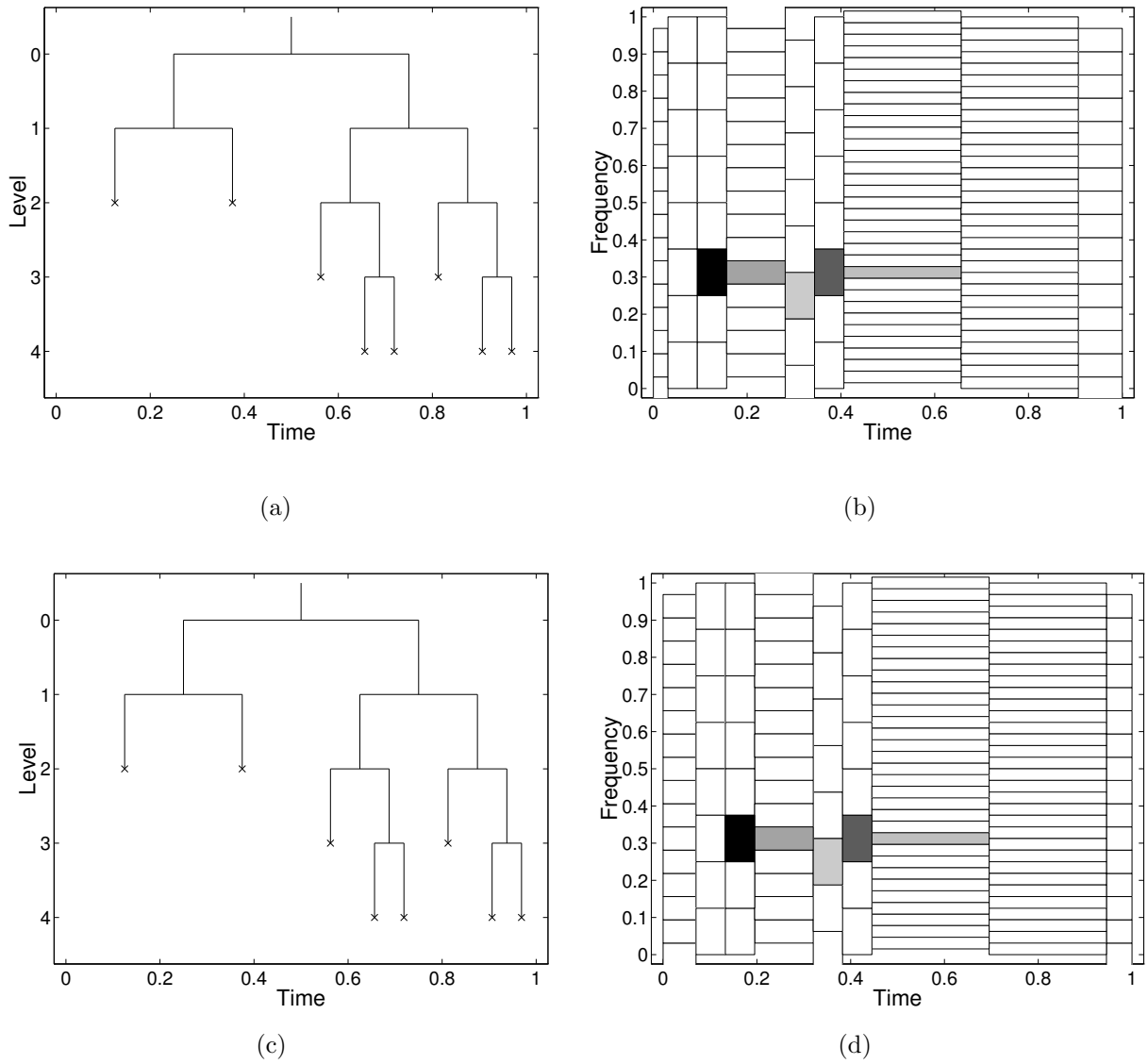
The organization of this chapter is as follows. In Section 3.2, we present the collection of basis functions, which is extended to include all the translations of basis-functions within the library. Section 3.3 introduces the periodic folding operator and formulates an efficient computation of the expansion coefficients. The construction of a tree-structured library of bases is described in Section 3.4. The best-basis search algorithm, namely the shift-invariant adapted-polarity local trigonometric decomposition, is presented in Section 3.5. Suboptimal variants, which entail a reduced complexity and higher information cost, are described in



(a) (b)

(c) (d)

Figure 3.2: Local Cosine Decomposition (LCD): (a) The best expansion tree of $g(t)$. (b) The time-frequency representation of $g(t)$ in its best-basis. Entropy=2.57. (c) The best expansion tree of $g(t - 5 \cdot 2^{-7})$. (d) The time-frequency representation of $g(t - 5 \cdot 2^{-7})$ in its best-basis. Entropy=2.39.



(a) (b)

(c) (d)

Figure 3.3: Shift-Invariant Adapted-Polarity Local Trigonometric Decomposition (SIAP-LTD): (a) The best expansion tree of $g(t)$. (b) The time-frequency representation of $g(t)$ in its best-basis. Entropy=1.44. (c) The best expansion tree of $g(t - 5 \cdot 2^{-7})$. (d) The time-frequency representation of $g(t - 5 \cdot 2^{-7})$ in its best-basis. Entropy=1.44.

Section 3.6.

3.2 Smooth Local Trigonometric Bases

Let $r = r(t)$ be a function in the class $C^s(\mathbb{R})$ for some $s \geq 0$ (class of s-times continuously differentiable functions), satisfying the following conditions:

$$|r(t)|^2 + |r(-t)|^2 = 1 \quad \text{for all } t \in \mathbb{R} \tag{3.1}$$

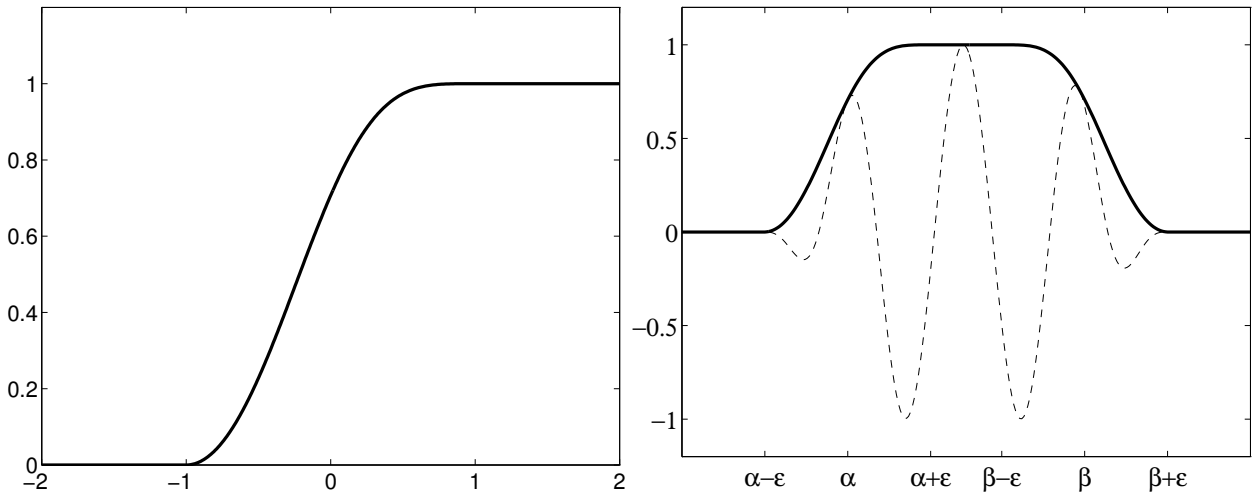
$$r(t) = \begin{cases} 0 & \text{if } t \leq -1\,, \\ 1 & \text{if } t > 1\,. \end{cases} \tag{3.2}$$

Then $r((t-\alpha)/\epsilon)r((\beta-t)/\epsilon)$ defines a *window* function, which is supported on the interval $[\alpha - \epsilon, \beta + \epsilon]$. The function r is called a *rising cutoff function* [150], since it rises from being identically zero to being identically one as t goes from $-\infty$ to $+\infty$. The role of $\epsilon > 0$ is to allow overlap of windows, and thus control the smoothness of the window function [98, 99, 100]. An example of a continuously differentiable real-valued rising cutoff function $r_1 \in C^1$ is given by

$$r_1(t) = \begin{cases} 0 & \text{if } t \leq -1\,, \\ \sin[\frac{\pi}{4}(1 + \sin\frac{\pi}{2}t)] & \text{if } -1 < t < 1\,, \\ 1 & \text{if } t \geq 1\,, \end{cases} \tag{3.3}$$

and depicted in Fig. 3.4, along with a corresponding window function on $[\alpha, \beta]$ for $\epsilon < (\beta - \alpha)/2$. By modulating a window function we obtain a smooth local trigonometric function that is supported on the same interval. Let us denote sets of modulating trigonometric functions by

$$C^{0,0}_{I,k}(t) = \sqrt{\frac{2}{\beta-\alpha}}\,\cos\left[\frac{\pi}{\beta-\alpha}(k+\frac{1}{2})(t-\alpha)\right]\,, \tag{3.4a}$$



(a) (b)

Figure 3.4: (a) An example of a rising cutoff function in C^1. (b) The corresponding window function on $[\alpha, \beta]$ for $\epsilon < (\beta - \alpha)/2$ (solid), and a modulated function (dashed).

$$C_{I,k}^{0,1}(t) = \begin{cases} 1/\sqrt{\beta-\alpha} & k = 0\,, \\ \sqrt{2/(\beta-\alpha)}\,\cos\left[\frac{\pi}{\beta-\alpha}k(t-\alpha)\right] & k \neq 0\,, \end{cases} \tag{3.4b}$$

$$C_{I,k}^{1,0}(t) = \sqrt{\frac{2}{\beta-\alpha}}\,\sin\left[\frac{\pi}{\beta-\alpha}(k+1)(t-\alpha)\right]\,, \tag{3.4c}$$

$$C_{I,k}^{1,1}(t) = \sqrt{\frac{2}{\beta-\alpha}}\,\sin\left[\frac{\pi}{\beta-\alpha}(k+\frac{1}{2})(t-\alpha)\right]\,, \tag{3.4d}$$

where $k \in \mathbb{Z}_+$. We define smooth local trigonometric functions on an interval $I = [\alpha, \beta]$ by

$$\phi_{I,k}^{\rho_0,\rho_1}(t) = \bar{r}(\frac{t-\alpha}{\epsilon})r(\frac{\beta-t}{\epsilon})C_{I,k}^{\rho_0,\rho_1}(t)\,, \quad k \in \mathbb{Z}_+, \quad \rho_0, \rho_1 \in \{0,1\}\,, \tag{3.5}$$

where $\bar{r}$ denotes the complex conjugate of r. The parities of the functions $\{C_{I,k}^{\rho_0,\rho_1} : k \in \mathbb{Z}_+,\ \rho_0, \rho_1 \in \{0,1\}\}$ at the end-points α and β are specified by ρ_0 and ρ_1, respectively, according to

$$C_{I,k}^{\rho_0,\rho_1}(\alpha + t) = (-1)^{\rho_0} C_{I,k}^{\rho_0,\rho_1}(\alpha - t) \tag{3.6a}$$

$$C_{I,k}^{\rho_0,\rho_1}(\beta+t) = -(-1)^{\rho_1} C_{I,k}^{\rho_0,\rho_1}(\beta-t) \tag{3.6b}$$

That is, even parity at the left end-point is specified by $\rho_0 = 0$ (respectively, odd parity by $\rho_0 = 1$), whereas even parity at the right end-point is specified by $\rho_1 = 1$ (respectively, odd parity by $\rho_1 = 0$). Each local trigonometric function $\phi_{I,k}^{\rho_0,\rho_1}$ is well localized in both time and frequency. In time, it is supported on $[\alpha - \epsilon, \beta + \epsilon]$ and thus has position uncertainty at most equal to the width of that compact interval. In frequency, $\hat{\phi}_{I,k}^{\rho_0,\rho_1}$ consists of two bumps centered at $\pm(2k + 1 + \rho_0 - \rho_1)/(\beta - \alpha)$, with uncertainty equal to that of the Fourier transform of the window function.

We can use the local trigonometric functions to produce smooth localized basis functions for various function spaces. For simplicity, we shall restrict ourselves to periodic functions with period 1, and use the ordinary inner product of $L_2[0,1]$. We designate this Hilbert space as $\hat{L}_2[0,1]$ to indicate the periodization. *i.e.*,

$$g(t+n) = g(t), \quad n \in \mathbb{Z}$$

$$\langle g, g \rangle = \int_0^1 |g(t)|^2 dt < \infty$$

for all $g \in \hat{L}_2[0,1]$.

Let $\mathcal{I} = \{I_{\ell,n,m}\}$ be a set of intervals of the form

$$I_{\ell,n,m} = [2^{-\ell}n + 2^{-J}m, 2^{-\ell}(n+1) + 2^{-J}m), \tag{3.7}$$

$0 \le \ell \le L \le J$, $0 \le n < 2^\ell$, $0 \le m < 2^{J-\ell}$, and consider the set of functions defined by

$$\left\{ \psi_{\ell,n,m,k}^{\rho_0,\rho_1}(t) \equiv \sum_{q \in \mathbb{Z}} \phi_{I_{\ell,n,m},k}^{\rho_0,\rho_1}(t+q) \ : \ I_{\ell,n,m} \in \mathcal{I},\ k \in \mathbb{Z},\ \rho_0, \rho_1 \in \{0,1\} \right\}. \tag{3.8}$$

We call ℓ the resolution-level index, n position index, m shift index, k frequency index ($k \in \mathbb{Z}_+$) and $\rho_0, \rho_1 \in \{0,1\}$ polarity indices. As proven in the sequel, the set of functions

defined in (3.8) is a redundant set that spans $\hat{L}_2[0,1]$. Our objective is to construct out of it a library of orthonormal bases, such that the best basis can be efficiently searched for a given signal and the resultant representation is shift-invariant.

3.3 The Periodic Folding Operator

The transform of a given function to an orthonormal basis involves computations of inner products with the basis functions. For the basis functions define in (3.8), an efficient computation is attainable by introducing a folding operator $F : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ and a periodic folding operator $Q : \hat{L}_2[0,1] \to \hat{L}_2[0,1]$ defined, respectively, by

$$F(\alpha,\rho)g(t) = \begin{cases} r(\frac{t-\alpha}{\epsilon})g(t) + (-1)^{\rho} r(\frac{\alpha-t}{\epsilon})g(2\alpha - t), & \text{if } \alpha < t < \alpha + \epsilon\,, \\ \bar{r}(\frac{\alpha-t}{\epsilon})g(t) - (-1)^{\rho} \bar{r}(\frac{t-\alpha}{\epsilon})g(2\alpha - t), & \text{if } \alpha - \epsilon < t < \alpha\,, \\ g(t), & \text{otherwise}\,, \end{cases} \tag{3.9}$$

and

$$Q(\alpha,\rho) = \prod_{q\in\mathbb{Z}} F(\alpha+q,\rho)\,. \tag{3.10}$$

The adjoint of F is given by

$$F^*(\alpha,\rho)g(t) = \begin{cases} \bar{r}(\frac{t-\alpha}{\epsilon})g(t) - (-1)^{\rho} r(\frac{\alpha-t}{\epsilon})g(2\alpha - t), & \text{if } \alpha < t < \alpha + \epsilon\,, \\ r(\frac{\alpha-t}{\epsilon})g(t) + (-1)^{\rho} \bar{r}(\frac{t-\alpha}{\epsilon})g(2\alpha - t), & \text{if } \alpha - \epsilon < t < \alpha\,, \\ g(t), & \text{otherwise}\,. \end{cases} \tag{3.11}$$

Observe that the *action region* of $F(\alpha,\rho)$ and $F^*(\alpha,\rho)$ is $(\alpha-\epsilon, \alpha+\epsilon)$, since outside this region they acts like the identity. Also, by the property (3.1) of the rising cutoff function, we have

$$F^*(\alpha,\rho)F(\alpha,\rho)g(t) = F(\alpha,\rho)F^*(\alpha,\rho)g(t) = \left(|r((t-\alpha)/\varepsilon)|^2 + |r(-(t-\alpha)/\varepsilon)|^2\right) g(t) = g(t)$$

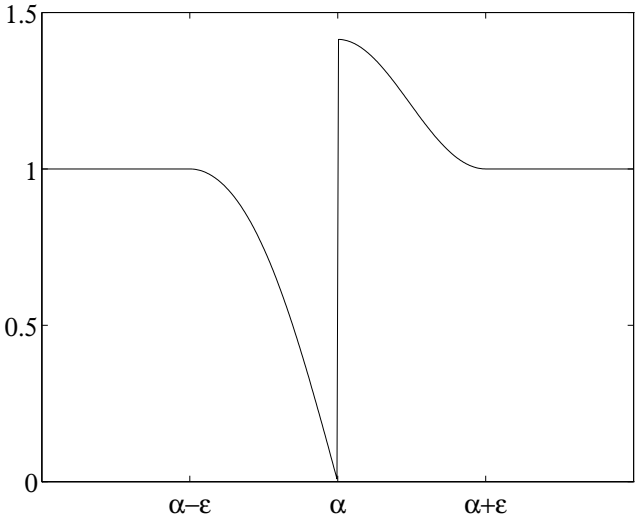


Figure 3.5: Action of $F(\alpha, 0)$ on the constant function $g(t) = 1$.

for all $t \neq 0$. Hence, F and F^* are unitary isomorphism of $L_2(\mathbb{R})$. Consequently, Q and Q^* are unitary isomorphism of $\hat{L}_2[0, 1]$.

The polarity of $F(\alpha, \rho)$ around $t = \alpha$ is odd-even for $\rho = 0$ and even-odd for $\rho = 1$. That is, if g is smooth, then folding it at α with polarity $\rho = 0$, for example, makes the left part, specifically[1] $\mathbf{1}_{(-\infty,\alpha]}F(\alpha, 0)g$, a function that is smooth when extended odd to the right, and makes the right part $(\mathbf{1}_{[\alpha,\infty)}F(\alpha, 0)g)$ a function that is smooth when extended even to the left. Fig. 3.5 shows the result of the action of $F(\alpha, 0)$ on the constant function $g(t) \equiv 1$, using the cutoff function defined by (3.3).

A pair of unfolding operators $F^*(\alpha, \rho_0)$ and $F^*(\beta, \rho_1)$ commute whenever $\epsilon < (\beta - \alpha)/2$. In this case, the actions of $F^*(\alpha, \rho_0)$ and $F^*(\beta, \rho_1)$ on a function g that is supported on an interval $I = [\alpha, \beta]$, simplify to multiplications by $\bar{r}(\frac{t-\alpha}{\epsilon})$ and $r(\frac{\beta-t}{\epsilon})$ respectively:

$$F^*(\alpha, \rho_0)F^*(\beta, \rho_1)g(t) = \bar{r}(\frac{t-\alpha}{\epsilon})r(\frac{\beta-t}{\epsilon})\tilde{g}(t) \tag{3.12}$$

[1] $\mathbf{1}_I$ denotes the indicator function for the interval I, *i.e.* the function that is 1 in I and 0 elsewhere.

where $\tilde{g}$ is the extension of g to the outside region of I, with the appropriate parities at the end-points:

$$\tilde{g} = \begin{cases} (-1)^{\rho_0} g(2\alpha - t), & \text{if } t < \alpha\,, \\ g(t), & \text{if } t \in [\alpha, \beta]\,, \\ -(-1)^{\rho_1} g(2\beta - t), & \text{if } t > \beta\,. \end{cases} \tag{3.13}$$

Denote by χ_I the periodic extension of the indicator function for the interval I, *i.e.*, $\chi_{[\alpha,\beta]} \equiv \mathbf{1}_{\{[\alpha+q,\beta+q]\,:\,q\in\mathbb{Z}\}}$. Accordingly, the local trigonometric functions defined in Eq. (3.5), and the basis functions defined in Eq. (3.8) satisfy

$$\phi_{I,k}^{\rho_0,\rho_1}(t) = F^*(\alpha,\rho_0)F^*(\beta,\rho_1)\mathbf{1}_I C_{I,k}^{\rho_0,\rho_1}(t) \tag{3.14}$$

$$\psi_{I,k}^{\rho_0,\rho_1}(t) = Q^*(\alpha,\rho_0)Q^*(\beta,\rho_1)\chi_I C_{I,k}^{\rho_0,\rho_1}(t) \tag{3.15}$$

whence

$$\left\langle \psi_{I,k}^{\rho_0,\rho_1}, g \right\rangle = \left\langle \chi_I C_{I,k}^{\rho_0,\rho_1}, Q(\alpha,\rho_0)Q(\beta,\rho_1)g \right\rangle = \left\langle C_{I,k}^{\rho_0,\rho_1}, \mathbf{1}_I F(\alpha,\rho_0)F(\beta,\rho_1)g \right\rangle\,. \tag{3.16}$$

Accordingly, we can compute the inner product of a given signal g with a basis function in two conventional stages: At the first stage the signal is preprocessed by folding it. Then, each segment is transformed by a trigonometric basis which has the appropriate parity properties at the end-points. In the discrete case, the trigonometric transform is DCT-II for even-even parity, DCT-IV for even-odd parity, DST-II for odd-odd parity and DST-IV for odd-even parity; all having fast implementation algorithms [13, 97, 119].

3.4 Tree-Structured Library of Bases

In this section, we organize the collection of the basis functions, defined in (3.8), in a binary tree structure, and by choosing various subsets from this collection we construct an orthonormal library of smooth local trigonometric bases.

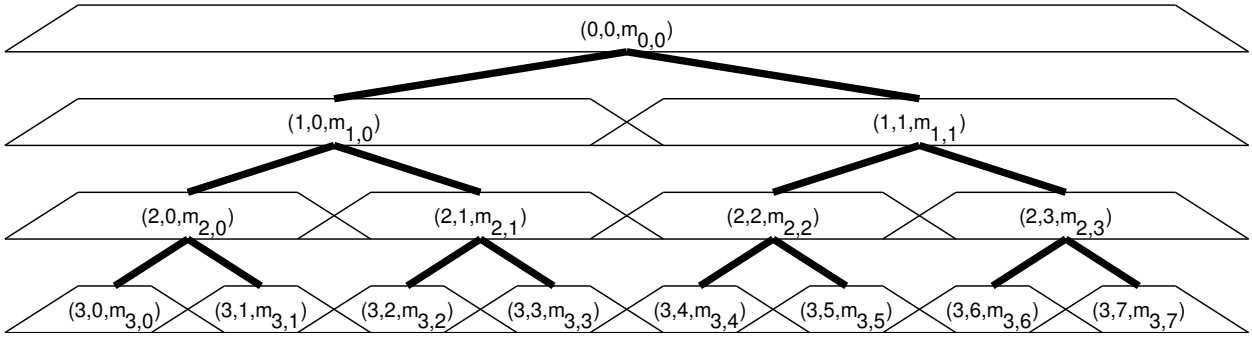


Figure 3.6: The smooth local trigonometric bases organized in a binary tree structure. Each node in the tree is indexed by the triplet (ℓ, n, m) and represents a subset of the basis functions.

The tree configuration is depicted in Fig. 3.6. Each node in the tree is indexed by the triplet (ℓ, n, m), and represents a subspace with different time-frequency localization characteristics:

$$B_{\ell,n,m}^{\rho_0,\rho_1} = \left\{ \psi_{\ell,n,m,k}^{\rho_0,\rho_1} \ : \ k \in \mathbb{Z}_+ \right\} \tag{3.17}$$

$$V_{\ell,n,m}^{\rho_0,\rho_1} = \operatorname{clos}_{L^2(\mathbb{R})} \left\{ B_{\ell,n,m}^{\rho_0,\rho_1} \right\} \tag{3.18}$$

Lemma 3.1 *The set $B_{\ell,n,m}^{\rho_0,\rho_1}$ is an orthonormal basis of the subspace $V_{\ell,n,m}^{\rho_0,\rho_1}$.*

Proof: We need to show that[2]

$$\left\langle \psi_{I,i}^{\rho_0,\rho_1}, \psi_{I,j}^{\rho_0,\rho_1} \right\rangle = \delta_{i,j} \tag{3.19}$$

for all $I \in \mathcal{I}$, $i, j \in \mathbb{Z}_+$ and $\rho_0, \rho_1 \in \{0, 1\}$, where $\delta_{i,j}$ denotes the Kronecker delta.

This can be proved by a straightforward computation of the inner product and using the properties of the rising cutoff function. Here we provide a simpler proof that is based on Equation (3.15). Define

$$\Lambda_{i,j} = \left\langle \psi_{I,i}^{\rho_0,\rho_1}, \psi_{I,j}^{\rho_0,\rho_1} \right\rangle ,$$

[2]To simplify notation, we sometimes replace the set of indices (ℓ, n, m) by their related interval $I = I_{\ell,n,m}$.

then by (3.15),

$$\Lambda_{i,j} = \left\langle Q^*(\alpha,\rho_0)Q^*(\beta,\rho_1)\chi_I C_{I,i}^{\rho_0,\rho_1}, Q^*(\alpha,\rho_0)Q^*(\beta,\rho_1)\chi_I C_{I,j}^{\rho_0,\rho_1} \right\rangle ,$$

where α and β are the end-points of the interval I. Since Q, the periodic folding operator, is unitary, it follows that

$$\Lambda_{i,j} = \left\langle \chi_I C_{I,i}^{\rho_0,\rho_1}, \chi_I C_{I,j}^{\rho_0,\rho_1} \right\rangle = \left\langle C_{I,i}^{\rho_0,\rho_1}, \mathbf{1}_I C_{I,j}^{\rho_0,\rho_1} \right\rangle .$$

Whence $\Lambda_{i,j} = \delta_{i,j}$ because the set $\left\{ \mathbf{1}_I C_{I,k}^{\rho_0,\rho_1} \ : \ k \in \mathbb{Z}_+ \right\}$ is an orthonormal basis for $L_2(I)$.

□

The window of the local trigonometric function defined in (3.5) has an ascending part which is supported on $[\alpha-\epsilon, \alpha+\epsilon]$ and a descending part which is supported on $[\beta-\epsilon, \beta+\epsilon]$. If the ascending parts of adjacent windows are disjoint, as well as their descending parts, then their associated intervals are called *compatible* [2]. That is, intervals $I' = [\alpha, \beta)$ and $I'' = [\beta, \gamma)$ are called compatible if $\alpha + \epsilon < \beta - \epsilon < \beta + \epsilon < \gamma - \epsilon$.

Lemma 3.2 *If I' and I'' are adjacent compatible intervals and the corresponding subspaces have the same connecting polarity index, then an orthogonal sum of the subspaces corresponds to the union of the intervals,* i.e.*, $V_{I'}^{\rho_0,\rho_1} \oplus V_{I''}^{\rho_1,\rho_2} = V_{I'\cup I''}^{\rho_0,\rho_2}$.*

Proof: Let $I' = [\alpha_0, \alpha_1)$ and $I'' = [\alpha_1, \alpha_2)$ be adjacent compatible intervals, and let $Q_j = Q(\alpha_j, \rho_j)$, $j = 0, 1, 2$ be the periodic folding operators at the end-points. First we show that the operators $P_{I'} = Q_0^* Q_1^* \chi_{I'} Q_1 Q_0$, $P_{I''} = Q_1^* Q_2^* \chi_{I''} Q_2 Q_1$ and $P_{I'\cup I''} = Q_0^* Q_2^* \chi_{I'\cup I''} Q_2 Q_0$ are orthogonal projections onto $V_{I'}^{\rho_0,\rho_1}$, $V_{I''}^{\rho_1,\rho_2}$ and $V_{I'\cup I''}^{\rho_0,\rho_2}$ respectively.

It follows from Equation (3.15) and the unitarity of Q_0 and Q_1 that

$$P_{I'}\psi_{I',k}^{\rho_0,\rho_1} = Q_0^* Q_1^* \chi_{I'} Q_1 Q_0 Q_0^* Q_1^* \chi_{I'} C_{I',k}^{\rho_0,\rho_1} = Q_0^* Q_1^* \chi_{I'} C_{I',k}^{\rho_0,\rho_1} = \psi_{I',k}^{\rho_0,\rho_1} . \tag{3.20}$$

Since the set $B_{I'}^{\rho_0,\rho_1} = \left\{\psi_{I',k}^{\rho_0,\rho_1} \,:\, k \in \mathbb{Z}_+\right\}$ is an orthonormal basis of $V_{I'}^{\rho_0,\rho_1}$ (Lemma 3.1), we have

$$P_{I'} v = v, \qquad \text{for all } v \in V_{I'}^{\rho_0,\rho_1}. \tag{3.21}$$

Now let $w \in \hat{L}_2[0,1]$ be in $(V_{I'}^{\rho_0,\rho_1})^{\perp}$. Then

$$\left\langle w, \psi_{I',k}^{\rho_0,\rho_1} \right\rangle = \left\langle w, Q_0^* Q_1^* \chi_{I'} C_{I',k}^{\rho_0,\rho_1} \right\rangle = \left\langle Q_1 Q_0 w, \mathbf{1}_{I'} C_{I',k}^{\rho_0,\rho_1} \right\rangle = 0 \tag{3.22}$$

for all $k \in \mathbb{Z}_+$. Thus $Q_1 Q_0 w$ is identically zero on I', because $\left\{\mathbf{1}_{I'} C_{I',k}^{\rho_0,\rho_1} \,:\, k \in \mathbb{Z}_+\right\}$ is an orthonormal basis for $L_2(I')$, and so

$$P_{I'} w = Q_0^* Q_1^* \chi_{I'} Q_1 Q_0 w = 0\,. \tag{3.23}$$

Consequently,

$$P_{I'}(v + w) = v, \qquad \text{for all } v \in V_{I'}^{\rho_0,\rho_1} \text{ and } w \perp V_{I'}^{\rho_0,\rho_1}\,. \tag{3.24}$$

In the same manner, $P_{I''} = Q_1^* Q_2^* \chi_{I''} Q_2 Q_1$ and $P_{I'\cup I''} = Q_0^* Q_2^* \chi_{I'\cup I''} Q_2 Q_0$ are orthogonal projections onto $V_{I''}^{\rho_1,\rho_2}$ and $V_{I'\cup I''}^{\rho_0,\rho_2}$ respectively.

The two sets of operators $\{Q_0, Q_1, Q_2\}$ and $\{Q_0^*, Q_1^*, Q_2^*\}$ form commuting families due to the compatibility condition of the intervals. Furthermore, Q_0 and Q_0^* commute with $\chi_{I''}$, Q_2 and Q_2^* commute with $\chi_{I'}$, and Q_1 and Q_1^* commute with $[\chi_{I'} + \chi_{I''}] = \chi_{I'\cup I''}$. Thus

$$\begin{aligned} P_{I'} + P_{I''} \quad &= Q_0^* Q_1^* \chi_{I'} Q_1 Q_0 + Q_1^* Q_2^* \chi_{I''} Q_2 Q_1 = Q_1^* \left[Q_0^* \chi_{I'} Q_0 + Q_2^* \chi_{I''} Q_2\right] Q_1 \\ &= Q_0^* Q_2^* Q_1^* \left[\chi_{I'} + \chi_{I''}\right] Q_1 Q_2 Q_0 = Q_0^* Q_2^* \chi_{I'\cup I''} Q_2 Q_0 = P_{I'\cup I''}\,, \end{aligned} \tag{3.25}$$

and

$$P_{I'} P_{I''} = Q_0^* Q_1^* \chi_{I'} Q_1 Q_0 Q_1^* Q_2^* \chi_{I''} Q_2 Q_1 = Q_0^* Q_1^* Q_2^* \chi_{I'} \chi_{I''} Q_0 Q_2 Q_1 = 0\,. \tag{3.26}$$

Hence $V_{I'}^{\rho_0,\rho_1} \oplus V_{I''}^{\rho_1,\rho_2} = V_{I'\cup I''}^{\rho_0,\rho_2}$. □

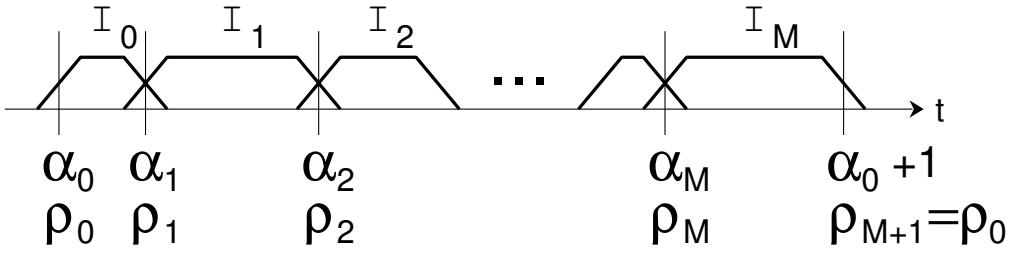


Figure 3.7: Indexing scheme of a compatible partition of a unit length interval, employed for generating smooth local trigonometric bases.

This lemma implies that a basis on the interval $I' \cup I''$ can be switched with a union of bases on I' and I''. Accordingly, an orthonormal basis of $\hat{L}_2[0,1]$ is constructed by taking the collection of basis functions that correspond to a disjoint compatible cover of $[0,1)$, or any other interval of a unit length. The indexing scheme of such a compatible cover is depicted in Fig. 3.7, and refers to the following proposition:

Proposition 3.1 *Let* $\{I_j\}_{j=0}^{M}$ *be a compatible partition of a unit-length interval by intervals* $I_j = [\alpha_j, \alpha_{j+1})$. *Let* $\{\rho_j\}_{j=0}^{M+1}$ *be a collection of* $\rho_j \in \{0,1\}$ *where* $\rho_{M+1} = \rho_0$. *Then* $\left\{B_{I_j}^{\rho_j,\rho_{j+1}} : 0 \le j \le M\right\}$ *forms an orthonormal basis of* $\hat{L}_2[0,1]$.

Proof: Lemmas 3.1 and 3.2 imply that $\left\{B_{I_j}^{\rho_j,\rho_{j+1}} : 0 \le j \le M\right\}$ is an orthonormal basis of $V_{[0,1)}^{\rho_0,\rho_0}$, so we shall show that $V_{[0,1)}^{\rho,\rho} \equiv \hat{L}_2[0,1]$ for $\rho \in \{0,1\}$.

Clearly, $V_{[0,1)}^{\rho,\rho} \subset \hat{L}_2[0,1]$ since $\left\{\psi_{[0,1),k}^{\rho,\rho} : k \in \mathbb{Z}_+\right\} \in \hat{L}_2[0,1]$ is an orthonormal basis of $V_{[0,1)}^{\rho,\rho}$. Now suppose that $g \perp V_{[0,1)}^{\rho,\rho}$, $g \in \hat{L}_2[0,1]$, then

$$\left\langle \psi_{[0,1),k}^{\rho,\rho}, g \right\rangle = \left\langle C_{[0,1),k}^{\rho,\rho}, \mathbf{1}_{[0,1)} F(0,\rho) F(1,\rho) g \right\rangle = \left\langle \mathbf{1}_{[0,1)} C_{[0,1),k}^{\rho,\rho}, Q(0,\rho) g \right\rangle = 0$$

for all $k \in \mathbb{Z}_+$. Hence g is identically zero, because $\left\{\mathbf{1}_{[0,1)} C_{[0,1),k}^{\rho,\rho} : k \in \mathbb{Z}_+\right\}$ is an orthonormal basis of $L_2[0,1]$, and Q is a unitary isomorphism of $\hat{L}_2[0,1]$. □

Recall that the collection of basis functions defined in (3.8) is structured in a tree whose nodes are associated with the intervals $I_{\ell,n,m}$. We can build out of this set a library of orthonormal bases by taking subsets which correspond to a compatible partition of a unit-length interval. The polarity-indices of the basis functions are practicably specified by a single integer P ($0 \leq P < 2^{2^L}$). Let $p(j)$ denote the polarity index at $t = \alpha_0 + j2^{-L}$, and let $P = [p(2^L - 1), ..., p(1), p(0)]_2$ be the binary representation of P. Then the polarity-indices of the basis functions on an interval $I = [\alpha, \beta)$, that belong to the disjoint cover of $[\alpha_0, \alpha_0 + 1)$, are given by

$$\rho(\alpha) = p[2^L(\alpha - \alpha_0)]\,, \tag{3.27a}$$

$$\rho(\beta) = p[2^L(\beta - \alpha_0)]\,. \tag{3.27b}$$

Notice that the length of an interval at the resolution level ℓ ($\ell \leq L$) is a multiple of 2^{-L}. Thus $(\alpha - \alpha_0)$ and $(\beta - \alpha_0)$ are also multiples of 2^{-L}, whenever $[\alpha, \beta)$ belongs to the partition of $[\alpha_0, \alpha_0 + 1)$. Derivable from Proposition 3.1, we have the following:

Proposition 3.2 *Let $E = \{(\ell, n, m)\}$ denote a collection of indices $0 \leq \ell \leq L$, $0 \leq n < 2^\ell$ and $0 \leq m < 2^{J-\ell}$ satisfying*

(i) The segments $\{I_{\ell,n,m} : (\ell, n, m) \in E\}$ are a disjoint compatible cover of $[\alpha_0, \alpha_0 + 1)$, for some $0 \leq \alpha_0 < 1$.

(ii) Nodes (ℓ, n_1, m_1), $(\ell, n_2, m_2) \in E$ at the same resolution level have identical shift index ($m_1 = m_2$).

Then for any polarity $0 \leq P < 2^{2^L}$ we have an orthonormal basis of $\hat{L}_2[0, 1]$, given by

$$\left\{ B^{\rho(\alpha),\rho(\beta)}_{[\alpha,\beta)} \;:\; [\alpha, \beta) = I_{\ell,n,m},\, (\ell, n, m) \in E \right\},$$

and the set of all (E, P) as specified above generates a library of orthonormal bases.

Condition (ii) precludes a relative shift between nodes within the same resolution level. This condition is actually unnecessary for the construction of a library of orthonormal bases. However, such a supplementary constraint limits the size of the library and thus controls the computational complexity of the best-basis search algorithm, while still retaining shift-invariance.

3.5 The Shift-Invariant Adaptive Polarity Local Trigonometric Decomposition

Let $\mathcal{B}$ represent the library of orthonormal bases of Proposition 3.2. Denote by $\mathcal{M}$ an additive information cost functional, and by $\mathcal{M}(Bg)$ the information cost of representing g on a basis B. The best basis for $g \in \hat{L}_2[0,1]$ in $\mathcal{B}$ relative to $\mathcal{M}$ is defined as that $B \in \mathcal{B}$ for which $\mathcal{M}(Bg)$ is minimal [45]. In this section we introduce an efficient search algorithm for the best basis, that relies on the tree structure of $\mathcal{B}$.

Denote by $A^{\rho_0,\rho_1}_{\ell,n,m}$ the best basis for g restricted to the subspace $V^{\rho_0,\rho_1}_{\ell,n,m}$. Since $B^{p,p}_{0,0,m}$ spans $\hat{L}_2[0,1]$ for any shift index m ($0 \leq m < 2^J$) and polarity index $p \in \{0,1\}$ (refer to Proposition 3.1), the best basis for g is $A^{p,p}_{0,0,m}$ combined with the best shift and polarity indices. These parameters, namely m and p, are determined recursively together with the best basis.

Let $m_0 = m$ and $p_0(0) = p$ designate respectively the shift and polarity at the coarsest resolution level ($\ell = 0$). Suppose that at the resolution level ℓ we have found m_ℓ, $\{p_\ell(i) \,|\, 0 \leq i < 2^\ell\}$ and $A^{p_\ell(n),p_\ell(n+1)}_{\ell,n,m_\ell}$ for all $0 \leq n < 2^\ell$, where we set $p_\ell(2^\ell + i) = p_\ell(i)$ owing to the periodicity of $\hat{L}_2[0,1]$. Then we will choose $m_{\ell-1}$, $\{p_{\ell-1}(i) \,|\, 0 \leq i < 2^{\ell-1}\}$ and $A^{p_{\ell-1}(n),p_{\ell-1}(n+1)}_{\ell-1,n,m_{\ell-1}}$ for $0 \leq n < 2^{\ell-1}$ so as to minimize the information cost.

It is shown in the sequel that shift-invariance is acquired by merely considering two

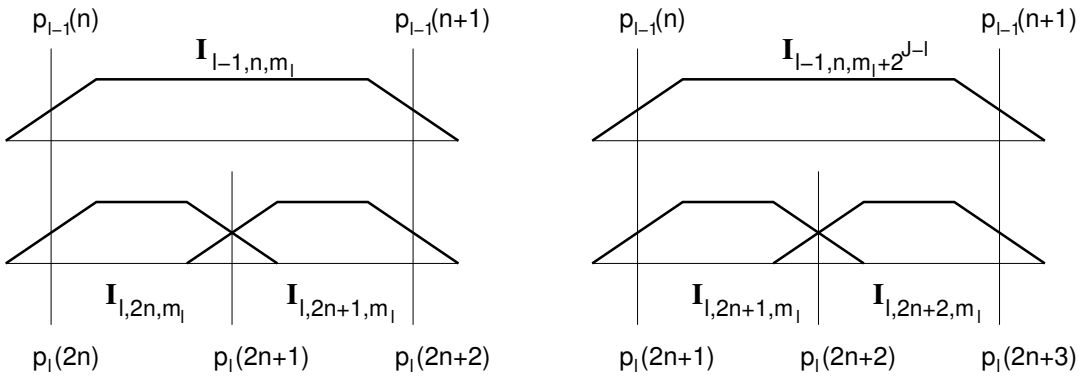


Figure 3.8: Joining up adjacent intervals at the resolution level ℓ into a parent interval at a coarser resolution level: (a) The levels have the same shift index. (b) The intervals at the level $\ell - 1$ are translated with respect to the intervals at the level ℓ.

optional values of $m_{\ell-1}$: m_ℓ and $m_\ell + 2^{J-\ell}$. These two options correspond to either no relative shift or $2^{-\ell}$ shift between the resolution levels $\ell - 1$ and ℓ. For each of the two choices we find the polarity indices and the best basis for g at the resolution level $\ell - 1$. Then we compare the information costs and select that value of $m_{\ell-1}$ which yields a cheaper representation.

Fig. 3.8(a) depicts the relation between intervals at the resolution levels $\ell - 1$ and ℓ, for the case where $m_{\ell-1} = m_\ell$. The interval $I_{\ell,2n+1,m_\ell}$ can be joined up with its *left* adjoining interval $I_{\ell,2n,m_\ell}$ into a parent interval $I_{\ell-1,n,m_\ell}$. Accordingly, the polarity indices at the resolution level $\ell - 1$ are simply $p_{\ell-1}(n) = p_\ell(2n)$, and by employing Lemma 3.2 we have

$$A_{\ell-1,n,m_\ell}^{p_{\ell-1}(n),p_{\ell-1}(n+1)} = \begin{cases} B_{\ell-1,n,m_\ell}^{p_{\ell-1}(n),p_{\ell-1}(n+1)}, \text{ if } \mathcal{M}'_B \le \mathcal{M}'_A\,, \\ A_{\ell,2n,m_\ell}^{p_\ell(2n),p_\ell(2n+1)} \cup A_{\ell,2n+1,m_\ell}^{p_\ell(2n+1),p_\ell(2n+2)}, \text{ otherwise} \end{cases} \tag{3.28}$$

where $\mathcal{M}'_A = \mathcal{M}(A_{\ell,2n,m_\ell}^{p_\ell(2n),p_\ell(2n+1)} g) + \mathcal{M}(A_{\ell,2n+1,m_\ell}^{p_\ell(2n+1),p_\ell(2n+2)} g)$ is the information cost of the children and $\mathcal{M}'_B = \mathcal{M}(B_{\ell-1,n,m_\ell}^{p_{\ell-1}(n),p_{\ell-1}(n+1)} g)$ is the information cost of the parent. In this

case, the information cost of g when expanded at the resolution level $\ell - 1$ is given by

$$\mathcal{M}'_{\ell-1} = \sum_{n=0}^{2^{\ell-1}-1} \mathcal{M}(A_{\ell-1,n,m_\ell}^{p_\ell(2n),p_\ell(2n+2)} g)\,. \tag{3.29}$$

For the other alternative of $m_{\ell-1}$ ($m_{\ell-1} = m_\ell + 2^{J-\ell}$), the relation between the intervals at the resolution levels $\ell-1$ and ℓ is depicted in Fig. 3.8(b). Now, the interval $I_{\ell,2n+1,m_\ell}$ can be joined up with its *right* adjoining interval $I_{\ell,2n+2,m_\ell}$ into a parent interval $I_{\ell-1,n,m_\ell+2^{J-\ell}}$. The polarity indices at the resolution level $\ell - 1$ are given by $p_{\ell-1}(n) = p_\ell(2n+1)$, and consequently

$$A_{\ell-1,n,m_\ell+2^{J-\ell}}^{p_{\ell-1}(n),p_{\ell-1}(n+1)} = \begin{cases} B_{\ell-1,n,m_\ell+2^{J-\ell}}^{p_{\ell-1}(n),p_{\ell-1}(n+1)}, \text{ if } \mathcal{M}''_B \le \mathcal{M}''_A\,, \\ A_{\ell,2n+1,m_\ell}^{p_\ell(2n+1),p_\ell(2n+2)} \cup A_{\ell,2n+2,m_\ell}^{p_\ell(2n+2),p_\ell(2n+3)}, \text{ otherwise} \end{cases} \tag{3.30}$$

where $\mathcal{M}''_A = \mathcal{M}(A_{\ell,2n+1,m_\ell}^{p_\ell(2n+1),p_\ell(2n+2)} g) + \mathcal{M}(A_{\ell,2n+2,m_\ell}^{p_\ell(2n+2),p_\ell(2n+3)} g)$ is the information cost of the children and $\mathcal{M}''_B = \mathcal{M}(B_{\ell-1,n,m_\ell+2^{J-\ell}}^{p_{\ell-1}(n),p_{\ell-1}(n+1)} g)$ is the information cost of the parent. In this case, the information cost of g when expanded at the resolution level $\ell - 1$ is given by

$$\mathcal{M}''_{\ell-1} = \sum_{n=0}^{2^{\ell-1}-1} \mathcal{M}(A_{\ell-1,n,m_\ell+2^{J-\ell}}^{p_\ell(2n+1),p_\ell(2n+3)} g)\,. \tag{3.31}$$

The value of $m_{\ell-1}$ is thus determined according to the lower information cost, *i.e.*,

$$m_{\ell-1} = \begin{cases} m_\ell, \text{ if } \mathcal{M}'_{\ell-1} \le \mathcal{M}''_{\ell-1}\,, \\ m_\ell + 2^{J-\ell}, \quad \text{otherwise.} \end{cases} \tag{3.32}$$

The corresponding best basis and polarity indices at the resolution level $\ell - 1$ are retained for the next stage of the procedure, which is carried out up to the level $\ell = 0$. The algorithm is initiated at the level $\ell = L$ ($L \le J$), specified by the shortest intervals that

are required for segmentation. At this level, we estimate the shift index m_L and polarity indices $\{p_L(n),\ 0 \le n < 2^L\}$, and impose

$$A_{L,n,m_L}^{p_L(n),p_L(n+1)} = B_{L,n,m_L}^{p_L(n),p_L(n+1)}, \quad 0 \le n < 2^L\,. \tag{3.33}$$

To simplify notation, the set of polarity indices at the resolution level L is organized into a single integer P_L ($0 \le P_L < 2^{2^L}$), using its binary representation $P_L = [p_L(2^L - 1), ..., p_L(1), p_L(0)]_2$. The optimal shift and polarity at the finest resolution level are given by

$$(m_L, P_L) = \arg \min_{\substack{0 \le m < 2^{J-L} \\ 0 \le P < 2^{2^L}}} \left\{ \sum_{n=0}^{2^L-1} \mathcal{M}(B_{L,n,m}^{p(n),p(n+1)} g) \right\}. \tag{3.34}$$

Definition 3.1 *$f, g \in \hat{L}_2[0,1]$ are said to be identical to within a resolution J time-shift ($J > 0$) if there exists $q \in \mathbb{Z}$, $0 \le q < 2^J$, such that $g(t) = f(t - 2^{-J}q)$ for all $t \in [0,1]$.*

Definition 3.2 *Bases $B_1, B_2 \in \mathcal{B}$ are said to be identical to within a resolution J time-shift ($J > 0$) if there exists $q \in \mathbb{Z}$, $0 \le q < 2^J$, such that $\psi(t - 2^{-J}q) \in B_2$ if and only if $\psi(t) \in B_1$.*

Definition 3.3 *A best-basis decomposition is said to be shift-invariant up to a resolution level J ($J > 0$) if for any $f, g \in \hat{L}_2[0,1]$ which are identical to within a resolution J time-shift, their respective best bases A_f and A_g are identical to within the* same *time-shift.*

It is evident that a best-basis decomposition, which is shift-invariant up to a resolution level J, is also shift-invariant up to a lower resolution level, because the translation is on a finer grid. In case of uniformly sampled discrete functions of length $N = 2^J$, an invariance to discrete translation is equivalent to shift-invariance up to a resolution level J.

Proposition 3.3 *The best basis expansion stemming from the previously described recursive algorithm is shift-invariant up to a resolution level J.*

Proof: Let $f, g \in \hat{L}_2[0,1]$ be identical to within a resolution J time-shift. Then there exists an integer $0 \leq q < 2^J$ such that $g(t) = f(t - q2^{-J})$. Denote the best bases for f and g by A_f and A_g, respectively. It is shown in Appendix A.2 that $B^{\rho_0,\rho_1}_{[\alpha,\beta)} \subset A_f$ implies $B^{\rho_0,\rho_1}_{[\alpha+q2^{-J},\beta+q2^{-J})} \subset A_g$ for all $[\alpha,\beta) \in \mathcal{I}$ and $\rho_0, \rho_1 \in \{0,1\}$. Consequently, if $\psi(t)$ is a basis-function in A_f, then $\psi(t - q2^{-J})$ is a basis-function in A_g. Thus A_f and A_g are identical to within a $q2^{-J}$ time-shift. □

3.6 Practical Variants of Suboptimal Foldings

3.6.1 Locally-Adapted Foldings

Normally, the influence of the polarity indices on the information cost is less significant than the influence of the shift index. Furthermore, an ill-adapted polarity-bit (a single polarity index specified at a certain end-point) is possibly eliminated at a coarser level by merging intervals on its both sides. Hence to maintain a manageable computational complexity for the minimization process of the information cost, we settle for suboptimal polarity indices which are locally adapted to the signal. Instead of pursuing a global minimum, as advised in (3.34), we estimate for each $0 \leq m < 2^{J-L}$ the locally adapted polarity indices, and choose that $m = m_L$ which leads to the lowest information cost.

For an additive information-cost functional, the orthogonal decomposition (Lemma 3.2) implies that any polarity-bit affects only the costs of its two adjoining segments. In particular, the value of the n-th polarity-bit $p_L(n)$, is completely subject to the values of his adjacent polarity-bits, namely $p_L(n-1)$ and $p_L(n+1)$. Denote by $\pi_m(n)$ the optimal

value of the n-th polarity-bit for a shift m. On the supposition that $\pi_m(n-1)$ and $\pi_m(n+1)$ correspond to the minimal local information cost about the n-th end-point, we have

$$\pi_m(n) = \begin{cases} 0, & \text{if } C_{m,n}(0) \leq C_{m,n}(1) \\ 1, & \text{otherwise.} \end{cases} \tag{3.35}$$

where

$$C_{m,n}(\rho) = \min_{\rho_0,\rho_1 \in \{0,1\}} \left\{ \mathcal{M}(B^{\rho_0,\rho}_{L,n,m} g) + \mathcal{M}(B^{\rho,\rho_1}_{L,n+1,m} g) \right\}, \quad \rho \in \{0,1\} \tag{3.36}$$

designates the local information cost about the n-th end-point for a shift m. If the assumption is true for all polarity indices and for all shifts, then the optimal shift and polarity at the finest resolution level are given by

$$m_L = \arg \min_{0 \leq m < 2^{J-L}} \left\{ \sum_{n=0}^{2^L-1} \mathcal{M}(B^{\pi_m(n),\pi_m(n+1)}_{L,n,m} g) \right\} \tag{3.37}$$

$$p_L(n) = \pi_{m_L}(n), \quad 0 \leq n < 2^L \,. \tag{3.38}$$

Clearly, the optimal shift and polarity, obtainable by (3.34), minimize the global information cost but not necessarily the local costs about each end-point. Hence the shift and locally-adapted polarity, computed by (3.37) and (3.38), are suboptimal and may result in a higher information cost. However, the representation is still shift-invariant due to the consistency in their computation. The following steps summarize the execution of SIAP-LTD:

Step 0 *Specify an information cost functional $\mathcal{M}$ and maximum depth of decomposition L.*

Step 1 *Use Eq. (3.16) and the trigonometric transforms DCT-II, DCT-IV, DST-II and DST-IV to expand g into the subsets $B^{\rho_0,\rho_1}_{L,n,m}$ for $0 \leq n < 2^L$, $0 \leq m < 2^{J-L}$ and $\rho_0, \rho_1 \in \{0,1\}$.*

Step 2 *Estimate the shift and polarity indices at the finest resolution level using Eqs. (3.37) and (3.38), and impose Eq. (3.33).*

Step 3 *For* $\ell = L, \ldots, 1$*:*

1. *Expand* g *into the subsets* $B_{\ell-1,n,m_\ell}^{p_{\ell-1}(n),p_{\ell-1}(n+1)}$ *and* $B_{\ell-1,n,m_\ell+2^{J-\ell}}^{p_{\ell-1}(n),p_{\ell-1}(n+1)}$ *for* $0 \le n < 2^{\ell-1}$.

2. *Let* $m_{\ell-1} = m_\ell$ *and compute the information cost of* g *at the resolution level* $\ell - 1$ *by Eq. (3.29).*

3. *Let* $m_{\ell-1} = m_\ell + 2^{J-\ell}$ *and compute the information cost of* g *at the resolution level* $\ell - 1$ *by Eq. (3.31).*

4. *Determine the value of* $m_{\ell-1}$ *according to (3.32) and keep the corresponding* $p_{\ell-1}(n)$ *and* $A_{\ell-1,n,m_{\ell-1}}^{p_{\ell-1}(n),p_{\ell-1}(n+1)}$ *for* $0 \le n < 2^{\ell-1}$.

The computational complexity of executing SIAP-LTD is $O[N(L + 2^{J-L+1}) \log_2 N]$, where N denotes the length of the signal. More specifically, Steps 1 and 2 take, respectively, $O(2^{J-L+1} N \log_2 N)$ and $O(2^{J-L+2} N)$ operations, and Step 3 requires twice as much operations as the conventional LCD [45], *i.e.*, $O(NL \log_2 N)$ operations. The complexity of SIAP-LTD is thus comparable to that of LCD with the benefits of shift-invariance and a higher quality (lower "information cost") "best-basis".

3.6.2 Fixed-Polarity Foldings

The LCD may be viewed as a degenerate form of SIAP-LTD characterized by a polarity $P_L = 0$ and shift $m_0 = 0$. In this case, no relative shift between resolution levels is allowed for (m_ℓ is non-adaptively set to zero for all $0 \le \ell \le L$), and the resultant representation is shift-variant. The SIAP-LTD provides two degrees of freedom that generate independently

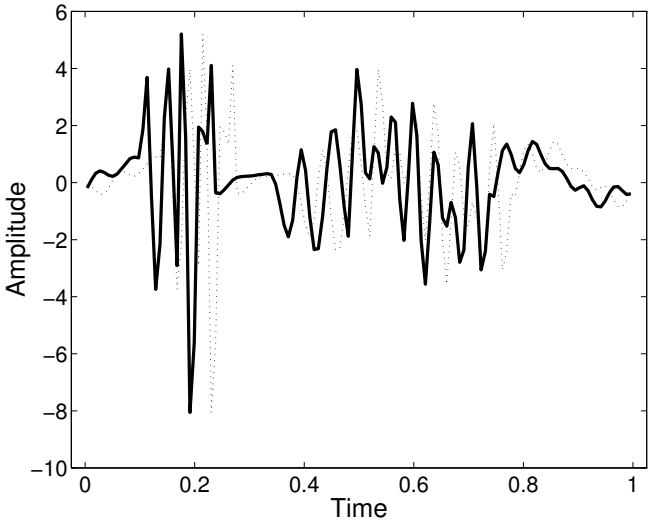
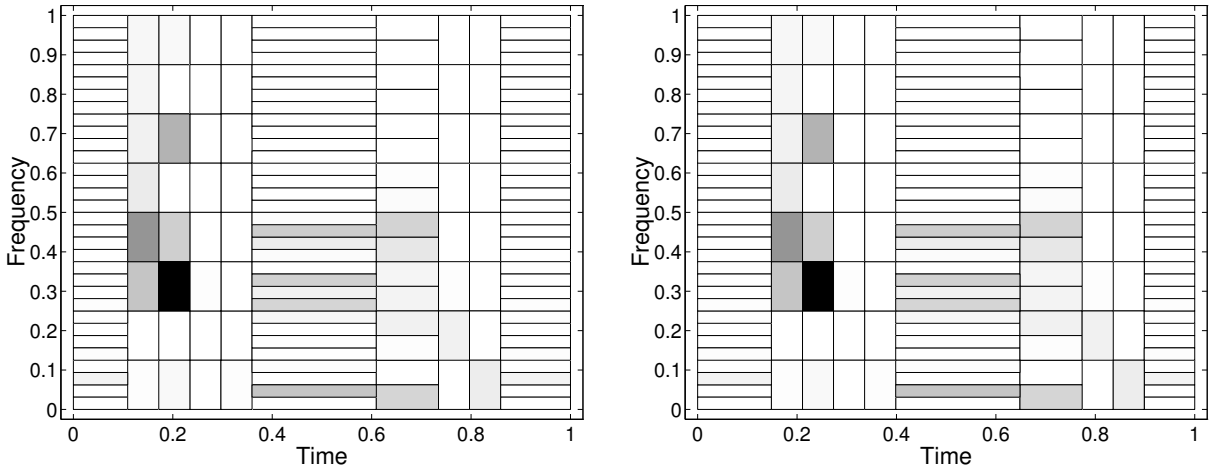


Figure 3.9: The signals $f(t)$ (solid) and $f(t-5\cdot 2^{-7})$ (dotted), sampled at 2^7 equally spaced points.

shift-invariance and adaptive-polarity foldings. While the relative shifts between resolution levels are required to obtain shift-invariance, the adaptation of the polarity indices at the finest resolution level is mainly intended to reduce the information cost and thus improve the time-frequency representation. This improvement is notable for signals that have dominant frequencies within each segment, such as the signal which is depicted in Fig. 3.1, or for signals that possess definite parity properties at the end-points of the segments. Otherwise, the polarity can be forced to a value whose bits are identical ($P_L = 0$ or $P_L = 2^{2^L} - 1$), without suppressing the shift-invariance. It follows from the definitions of the basis functions (Eq. (3.8)) and the modulating trigonometric functions (Eqs. (3.4a)–(3.4d)) that if the polarity-bits are restricted to zeros (respectively ones), then the library of bases consists of smooth local cosines (respectively sines). Accordingly, we call the best-basis search algorithms *Shift-Invariant Local Cosine Decomposition* (SI-LCD) when P_L is forced to zero, and *Shift-Invariant Local Sine Decomposition* (SI-LSD) when P_L is forced to $2^{2^L} - 1$.

As an example, we refer to the signals $f(t)$ and $f(t - 5 \cdot 2^{-7})$ depicted in Fig. 3.9.

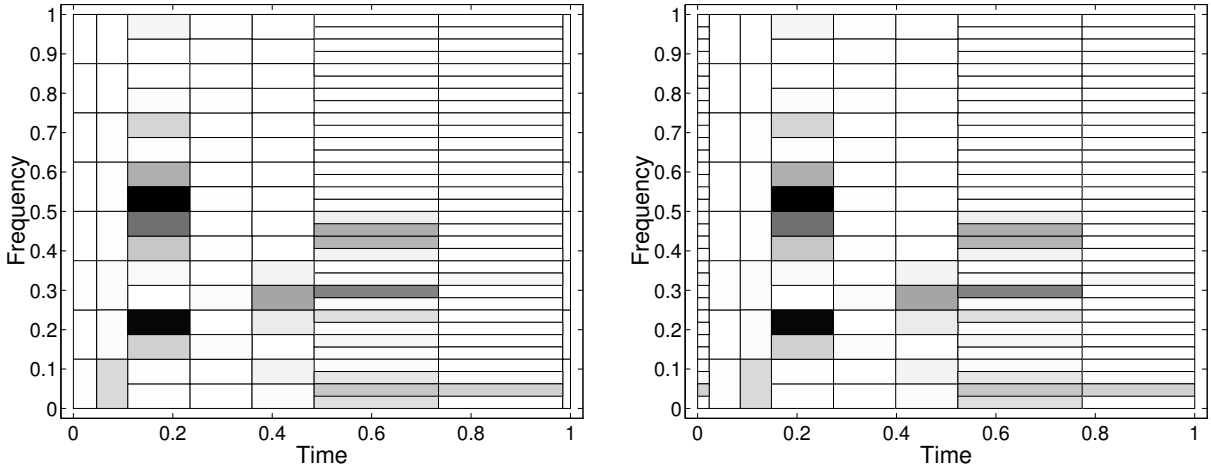


(a) (b)

Figure 3.10: Shift-Invariant Local Cosine Decomposition (SI-LCD): (a) The time-frequency representation of $f(t)$ in its best-basis. Entropy=3.01. (b) The time-frequency representation of $f(t - 5 \cdot 2^{-7})$ in its best-basis. Entropy=3.01.

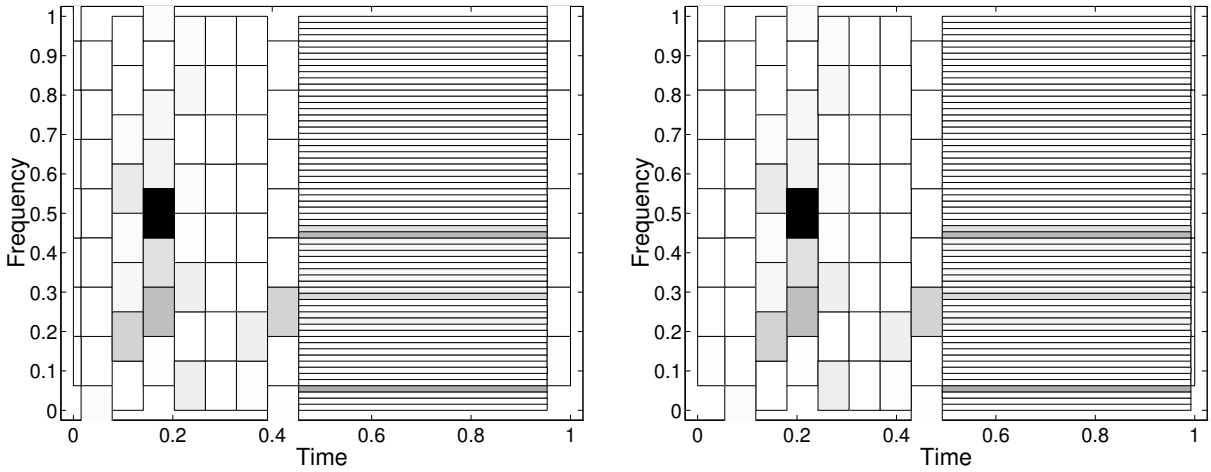
The time-frequency representations attained by SI-LCD (Fig. 3.10), SI-LSD (Fig. 3.11) and SIAP-LTD (Fig. 3.12) are all shift-invariant and have similar information costs. Whereas that obtained by LCD (Fig. 3.13) yields variations in the energy spread and leads to a higher shift-dependent information cost.

It is worth mentioning that while a fixed action-region was used for the folding operator (a fixed ϵ in (3.9)), it is possible to dilate it in coarser resolution levels, as long as the segments of the signal are compatible. That is, in each resolution level, if a parent-node has been chosen for the best expansion then the radii of the action-regions at its end-points are maximized, subject to the compatibility restriction. Such a *variable* action-region may lead to better time-frequency localization properties of basis functions, compared to fixed action-region [45] and multiple foldings [65].



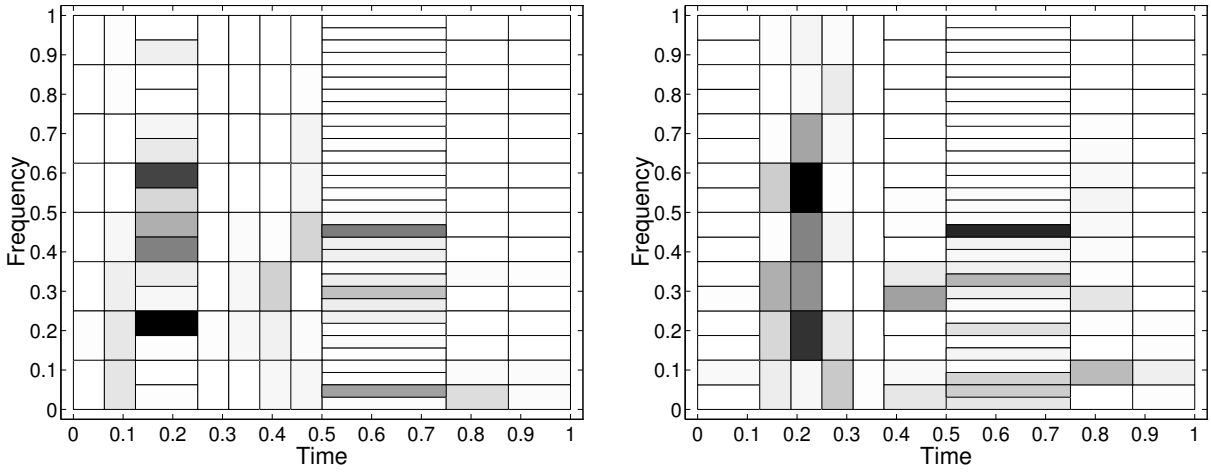
(a) (b)

Figure 3.11: Shift-Invariant Local Sine Decomposition (SI-LSD): (a) The time-frequency representation of $f(t)$ in its best-basis. Entropy=3.07. (b) The time-frequency representation of $f(t - 5 \cdot 2^{-7})$ in its best-basis. Entropy=3.07.



(a) (b)

Figure 3.12: Shift-Invariant Adapted-Polarity Local Trigonometric Decomposition (SIAP-LTD): (a) The time-frequency representation of $f(t)$ in its best-basis. Entropy=2.86. (b) The time-frequency representation of $f(t - 5 \cdot 2^{-7})$ in its best-basis. Entropy=2.86.



(a) (b)

Figure 3.13: Local Cosine Decomposition (LCD): (a) The time-frequency representation of $f(t)$ in its best-basis. Entropy=3.12. (b) The time-frequency representation of $f(t-5\cdot 2^{-7})$ in its best-basis. Entropy=3.27.

3.7 Summary

We have defined a tree-structured library of smooth local trigonometric bases, and described efficient search algorithms for selecting the best basis. The proposed algorithms are the *shift-invariant adapted-polarity local trigonometric decomposition* (SIAP-LTD), *Shift-Invariant Local Cosine Decomposition* (SI-LCD) and *Shift-Invariant Local Sine Decomposition* (SI-LSD). When compared with the local cosine decomposition (LCD) [45], these algorithms are determined to be advantageous in three respects. First, they leads to best-basis expansions that are shift-invariant. Second, the resulting representations are characterized by lower information costs. Third, the polarity of the folding operator may be adapted to the parity properties of the segmented signal at the end-points.

The shift-invariance stems from an adaptive relative shift of expansions in distinct resolution levels. We showed that at any resolution level ℓ it suffices to examine and select one of two relative shift options — a zero shift or a $2^{-\ell-1}$ shift. The choice between these

two options, enabled by the extended library, was made in accordance with minimizing the information cost. Thus, the attained representation was also characterized by a reduced information cost. The SIAP-LTD further enhanced the representation by applying a periodic folding operator, whose polarity was adapted to the parity properties of the signal.

Chapter 4

Adaptive Decompositions of Time-Frequency Distributions

4.1 Introduction

The Wigner distribution (WD) has long been of special interest, because it possesses a number of desirable mathematical properties [18, 23], including maximal autocomponent concentration in the time-frequency plane. In spite of its desirable properties, the practical application of the WD is often restricted due to the presence of interference terms. These terms render the WD of multicomponent signals extremely difficult or impossible to interpret.

Several methods, developed to reduce noise and cross-components at the expense of reduced time-frequency energy concentration, employ smoothing kernels or windowing techniques [15, 78, 79, 73]. Unfortunately, the specific choice of kernel dramatically affects the appearance and quality of the resulting time-frequency representation. Consequently, adaptive representations [79, 4, 50] often exhibit performance far surpassing that of fixed-kernel representations. However, such methods are either computationally expensive or have a very limited adaptation range. Another approach striving for cross-term suppression

with minimal resolution loss [144, 114] uses the Gabor expansion to decompose the WD. Interference terms are readily identified as cross WD of distinct basis functions. Here, a major drawback is the dependence of the performance on the choice of the Gabor window. An appropriate window selection depends on the data and may vary for different components of the same signal. Furthermore, distinct basis functions which are "close" in the time-frequency plane are often related to the same signal component. Accordingly, their cross-terms are not interpretable as interference terms, but rather may have a significant effect on the time-frequency resolution. Qian and Chen [115] proposed to decompose the WD into a series of Gabor expansions, where the order of the expansion is defined by the maximum degree of oscillation. They showed that such harmonic terms contribute minimally to the useful properties, but are directly responsible for the appearance of interference terms. In this case, the manipulation of cross-terms is equivalent to including cross-terms of Gabor functions whose *Manhattan distance* is smaller than a certain threshold. However, the order of the expansion has to be determined adaptively and generally depends on the local distribution of the signal. In [142], the signal is decomposed into frequency bands, and the Wigner distributions of all the subbands are superimposed. This attenuates interferences between subbands, but still suffers interferences within the subbands. Therefore it is merely suitable for signals that possess a single component in each subband. Moreover, the exclusion of beneficial cross-terms, which join neighboring basis-functions, invariably degrades the energy concentration and may artificially split a given signal component into several frequency-bands.

In this chapter, we present *adaptive* decompositions of the WD using *extended* libraries of orthonormal bases. Analogous to the approach in [144, 114, 115], a prescribed signal is expanded on a certain basis and subsequently transformed into the Wigner domain. Here, the basis is selectively constructed out of a redundant library of waveforms to best match the signal components, thereby concentrating its representation into a relatively small number

of significant expansion coefficients. The waveforms of the library are well localized in the time-frequency plane, and organized in a binary tree structure facilitating efficient search algorithms for the best orthonormal basis. It is demonstrated that the best *orthonormal* bases in *extended* libraries of bases are more advantageous to "optimal" expansions (*e.g.*, Matching Pursuit [96] and Basis Pursuit [14]) in *conventional* libraries. The extension of a library provides a fundamental flexibility in the expansion of a prescribed signal, while the restriction of the search procedure to orthonormal bases maintains a manageable computational complexity. In particular we focus on the shift-invariant decomposition in the extended library of wavelet packets, which was introduced in Chapter 2. The best-basis representation was proved preferable to the standard wavelet packet decomposition due to its desirable properties. Namely, shift-invariance, lower information cost and improved time-frequency resolution.

The interference terms in the Wigner domain are controlled by adaptively thresholding the cross WD of interactive basis functions according to their distance and amplitudes in the idealized time-frequency plane. When the distance-threshold is set to zero, the *modified* Wigner distribution precludes (MWD) any cross-terms, so essentially there is no interference terms but the energy concentration of the individual components is unacceptably low. When the amplitude-threshold is set to zero and the distance-threshold goes to infinity, the MWD converges to the conventional WD. By adjusting the distance and amplitude thresholds, one can effectively balance the cross-term interference, the useful properties of the distribution, and the computational complexity.

The distance measure in the idealized plane is related to a degree of adjacency by weighing the Euclidean time-frequency distance with the self distribution of the basis-functions. Since the basis-functions are adapted to the signal's local distribution, the thresholding of the cross-terms is also adapted to the local distribution of the signal. This dispenses with the need for local adjustments of the associated distance-threshold.

We note that the MWD constitutes an effective tool for resolving multicomponent signals. By defining equivalence classes in the time-frequency plane, we show that a prescribed component of a multicomponent signal can be determined as a partial sum of basis-functions. The signal components are well delineated in the time-frequency plane, and can be recovered from the energy distribution to within a constant phase factor.

The organization of this chapter is as follows. In Section 4.2, we review the Wigner distribution, the origin of interference terms and the relation to Cohen's class of distributions. Section 4.3 introduces the MWD. We present adaptive decompositions of the WD and show that the interference terms can be eliminated by thresholding the cross-terms according to a degree of adjacency in the idealized time-frequency plane. In particular, the superiority of the modified distribution is demonstrated by employing the shift-invariant wavelet packet decomposition. The general properties of the MWD are presented in Section 4.4. Inversion and uniqueness of the MWD are the subjects of Section 4.5.

4.2 The Wigner Distribution

Let $R_g(t,\tau)$ be the instantaneous auto-correlation of a complex signal $g(t)$, defined as

$$R_g(t,\tau) = g(t+\tau/2)g^*(t-\tau/2) \tag{4.1}$$

where g^* denotes the complex conjugate of g. The Wigner distribution of $g(t)$ is then defined as the Fourier transform (FT) of $R_g(t,\tau)$ with respect to the lag variable τ [151]:

$$W_g(t,\omega) = \int R_g(t,\tau)e^{-j\omega\tau}\,d\tau = \int g(t+\tau/2)g^*(t-\tau/2)e^{-j\omega\tau}\,d\tau\,, \tag{4.2}$$

or equivalently as

$$W_g(t,\omega) = \frac{1}{2\pi}\int G(\omega+\xi/2)G^*(\omega-\xi/2)e^{j\xi t}\,d\xi\,, \tag{4.3}$$

where $G(\omega)$ is the Fourier transform of $g(t)$ (the range of integrals is from $-\infty$ to $+\infty$ unless otherwise stated). Similarly, but with a different physical meaning, the symmetrical ambiguity function (AF) is defined as the inverese Fourier transform (IFT) of $R_g(t,\tau)$ with respect to the time variable t [153]:

$$A_g(\theta,\tau) = \frac{1}{2\pi}\int R_g(t,\tau)e^{j\theta t}\,dt\,. \tag{4.4}$$

Thus, the WD and AF are related by the 2-D FT:

$$W_g(t,\omega) = \iint A_g(\theta,\tau)e^{-j(\theta t+\omega\tau)}\,d\theta\,d\tau\,. \tag{4.5}$$

The WD satisfies a large number of desirable mathematical properties [18, 23]. In particular, the WD is always real-valued, it preserves time and frequency shifts and satisfies the marginal properties:

$$\tfrac{1}{2\pi}\textstyle\int W_g(t,\omega)\,d\omega = |g(t)|^2 \tag{4.6}$$

$$\textstyle\int W_g(t,\omega)\,dt = |G(\omega)|^2\,. \tag{4.7}$$

One major drawback of the WD is the interference terms between signal components. Suppose that a given signal consists of two components,

$$g(t) = g_1(t) + g_2(t) \tag{4.8}$$

Then, by substituting this into (4.2) we have

$$W_g(t,\omega) = W_{g_1}(t,\omega) + W_{g_2}(t,\omega) + 2\,\mathrm{Re}\{W_{g_1,g_2}(t,\omega)\} \tag{4.9}$$

where

$$W_{g_1,g_2}(t,\omega) = \int g_1(t+\tau/2)g_2^*(t-\tau/2)e^{-j\omega\tau}\,d\tau \tag{4.10}$$

is the cross WD of $g_1(t)$ and $g_2(t)$. This shows that the WD of the sum of two signals is not the sum of their respective WDs, but has the additional term $2\,\mathrm{Re}\{W_{g_1,g_2}(t,\omega)\}$. This term

is often called the interference term or cross term and it is often said to give rise to artifacts. However, one has to be cautious with the interpretations these words evoke, because any signal can be broken up into an arbitrary number of parts and the so-called cross terms are therefore not generally unique and do not characterize anything but our own division of a signal into parts [26]. There exists a natural decomposition where beneficial cross terms, which enhance the energy concentration, are distinguished from the undesirable interference terms, which obscure the time-frequency representation. This issue is addressed in the next section and in Section 4.5.

The WD, as well as the Choi-williams [15] and cone-kernel distributions [154] are members of a more general class of distributions, called *Cohen's class* [21]. Each member of this class is given by

$$C_g(t,\omega;\phi) = \frac{1}{2\pi}\iiint e^{j(-\theta t-\omega\tau+\theta u)}\phi(\theta,\tau)g(u+\tau/2)g^*(u-\tau/2)\,du\,d\theta\,d\tau \tag{4.11}$$

and related to the WD and ambiguity function by

$$C_g(t,\omega;\phi) = \iint W_g(u,\xi)\Phi(t-u,\omega-\xi)\,du\,d\xi \tag{4.12}$$

$$= \iint A_g(\theta,\tau)\phi(\theta,\tau)e^{-j(\theta t+\omega\tau)}\,d\theta\,d\tau \tag{4.13}$$

where $\phi(\theta,\tau)$ is the kernel of the distribution, and $\Phi(t,\omega)$ is the 2-D Fourier transform of $\phi(\theta,\tau)$. Different kernels produce different distributions obeying different properties. For example, $\phi(\theta,\tau) = 1$, $e^{j\theta|\tau|/2}$, $e^{-\theta^2\tau^2/\sigma}$ and $w(\tau)\,|\tau|\,\sin(\alpha\theta\tau)/\alpha\theta\tau$ correspond to the Wigner, Page, Choi-Williams and Cone-kernel distributions, respectively [73]. The spectrogram, the squared magnitude of the short-time Fourier transform, is also a member of Cohen's class, since it can be obtained as a 2-D convolution of the WD's of the signal and the window.

The interference terms associated with the WD are highly oscillatory, whereas the auto terms are relatively smooth. Therefore, the reduced-interference distributions are designed

to attenuate the interference terms by smoothing the WD with a low-pass kernel [78, 152]. Unfortunately, this procedure invariably entails a loss of time-frequency concentration. Accordingly, high energy concentration and effective suppression of interference terms cannot be achieved simultaneously by merely smoothing the Wigner distribution.

4.3 Adaptive Decomposition of the Wigner Distribution and Elimination of Interference Terms

In this section, we present adaptive decompositions of the WD using overcomplete libraries of orthonormal bases. The Wigner domain interference terms are controlled adaptively by thresholding the cross WD of interactive basis functions according to their degree of adjacency in the idealized time-frequency plane. In particular, we demonstrate the superiority of the modified distribution by employing the shift-invariant wavelet packet decomposition, which was introduced in Chapter 2.

Let $\mathcal{B}$ denote an overcomplete library of orthonormal bases, and let

$$g(t) = \sum_{\lambda \in \mathbf{N}} c_\lambda \varphi_\lambda(t)\,, \quad \{\varphi_\lambda\}_{\lambda \in \mathbf{N}} \in \mathcal{B} \tag{4.14}$$

be the best-basis expansion of the signal g. Then by inserting (4.14) into (4.2), the Wigner distribution of g can be written as

$$W_g(t,\omega) = \sum_{\lambda,\lambda' \in \mathbf{N}} c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega) \tag{4.15}$$

$$= \sum_{\lambda \in \mathbf{N}} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2 \sum_{\lambda > \lambda'} \mathrm{Re}\{c_\lambda c^*_{\lambda'} W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega)\}\,. \tag{4.16}$$

Equation (4.16) partitions the traditional WD into two subsets. The superposition of the auto WD of the basis-functions, represents the auto-terms. The second summation, comprising cross WD of basis-functions, represents the cross-terms. Cross terms associated with

the Wigner distribution, and other bilinear distributions, should not be always interpreted as interference terms, since they are not uniquely defined. Any signal can be sub-divided in an infinite number of ways, each generating different cross terms. Therefore, we need to distinguish between generally undesirable interference-terms and beneficial cross-terms that primarily enhance useful time-frequency features.

The cross WD of distinct basis-functions is oscillating and centered in the midway of the corresponding auto-terms [71, 72]. The oscillation rate is proportional to the distance between the auto-terms. On the other hand, useful properties such as the time marginal, frequency marginal, energy concentration and the instantaneous frequency property [23], are achieved by *averaging* the Wigner distribution. Therefore the overall contribution of each cross-term component is *inversely* proportional to the distance between the corresponding basis-functions in the time-frequency plane [115, 116].

A useful distance measure between pairs of basis-functions is obtainable in the idealized time-frequency plane. Recall that in the idealized plane, each basis-function is symbolized by a rectangular cell (tile) whose area is associated with Heisenberg's uncertainty principle, and its shade is proportional to the corresponding squared coefficient. We define the distance between a pair of basis-functions by

$$d(\varphi_\lambda, \varphi_{\lambda'}) = \left[\frac{(\bar{t}_\lambda - \bar{t}_{\lambda'})^2}{\Delta t_\lambda \Delta t_{\lambda'}} + \frac{(\bar{\omega}_\lambda - \bar{\omega}_{\lambda'})^2}{\Delta \omega_\lambda \Delta \omega_{\lambda'}} \right]^{1/2} \tag{4.17}$$

where $(\bar{t}_\lambda, \bar{\omega}_\lambda)$ is the position of the cell associated with φ_λ; Δt_λ and $\Delta \omega_\lambda$ denote the time and frequency widths (uncertainties), respectively. Similar notations apply to $\varphi_{\lambda'}$.

Since the best basis tends to represent the signal using a relatively small number of significant expansion coefficients, the summations in (4.16) can be restricted to basis-functions whose coefficients are above a prescribed cutoff, and to pairs that are "close" (sufficiently small values of $d(\varphi_\lambda, \varphi_{\lambda'})$). The modified Wigner distribution (MWD) is then

given by

$$T_g(t,\omega) = \sum_{\lambda\in\Lambda} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2 \sum_{\{\lambda,\lambda'\}\in\Gamma} \mathrm{Re}\{c_\lambda c_{\lambda'}^* W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega)\} \tag{4.18}$$

where

$$\Lambda = \{\lambda \,|\, |c_\lambda| \geq \varepsilon M\}, \quad M \equiv \max_\lambda\{|c_\lambda|\} \tag{4.19}$$

$$\Gamma = \{\{\lambda,\lambda'\} \,|\, 0 < d(\varphi_\lambda,\varphi_{\lambda'}) \leq D,\ |c_\lambda c_{\lambda'}| \geq \varepsilon^2 M^2\}\,. \tag{4.20}$$

ε and D denote thresholds of relative amplitude and time-frequency distance, respectively. When $D = 0$, the MWD precludes any cross-terms, so essentially there are no interference terms but the energy concentration of the individual components is generally low. As D goes to infinity and ε goes to zero, the MWD converges to the conventional WD. By adjusting the distance and amplitude thresholds, one can effectively balance the cross-term interference, the useful properties of the distribution, and the computational complexity.

Here, rather that the usual Euclidean distance ($\sqrt{(\bar{t}_\lambda - \bar{t}_{\lambda'})^2 + (\bar{\omega}_\lambda - \bar{\omega}_{\lambda'})^2}$) or the Manhattan distance ($|\bar{t}_\lambda - \bar{t}_{\lambda'}| + |\bar{\omega}_\lambda - \bar{\omega}_{\lambda'}|$) [115], we use the measure defined in (4.17), which weighs the time-frequency distance with the self distribution of the basis elements. Since the basis elements are selected to best match the signal's local distribution, such a distance measure implicitly characterizes the signal itself. Accordingly, the thresholding of the cross-terms is also adapted to the local distribution of the signal, dispensing with the need for local adjustments of the associated distance-threshold.

The extended library of wavelet packets (Chapter 2) includes basis-functions of the form

$$\psi_{\ell,n,m,k}(t) = 2^{\ell/2}\psi_n\left[2^\ell(t - 2^{-L}m) - k\right] \tag{4.21}$$

where ℓ is the resolution-level index ($0 \leq \ell \leq L$), n is the frequency index ($0 \leq n < 2^{L-\ell}$), m is the shift index ($0 \leq m < 2^{L-\ell}$) and k is the position index ($0 \leq k < 2^\ell$). Each

basis-function is symbolically associated with a rectangular tile in the time-frequency plane which is positioned about

$$\bar{t} = 2^{-\ell}k + 2^{-L}m + (2^{L-\ell} - 1)C_h + (C_h - C_g)R(n)\,, \tag{4.22}$$

$$\bar{f} = 2^{\ell-L}[GC^{-1}(n) + 0.5]\,, \tag{4.23}$$

where

$$C_h \triangleq \frac{1}{\|h\|^2}\sum_{k\in\mathbb{Z}} k|h_k|^2\,, \quad C_g \triangleq \frac{1}{\|g\|^2}\sum_{k\in\mathbb{Z}} k|g_k|^2\,, \tag{4.24}$$

are respectively the energy centers of the low-pass and high-pass quadrature filters [150, 70], $R(n)$ is an integer obtained by bit reversal of n in a $L-\ell$ bits binary representation, and GC^{-1} is the inverse Gray code permutation. The width and height of the tile are given by

$$\Delta t = 2^{-\ell}\,, \quad \Delta f = 2^{\ell-L}\,. \tag{4.25}$$

For a given signal, the SIWPD yields the best expansion in the extended library with respect to an additive cost function. It is demonstrated below that it would be advantageous to search for the best *orthonormal* basis using an *extended* library of wavelet packets, rather than using computationally expensive algorithms for searching optimal (not necessarily orthonormal) expansions in a *conventional* wavelet packet library. The extended library provides flexibility in expanding the signal, while the orthonormality contributes to a manageable complexity of the search procedure.

For example, Fig. 4.1 illustrates a test signal $g(t)$, which comprises a short pulse, a tone and a component with nonlinear frequency modulation. The corresponding Wigner distribution and spectrogram are displayed in Fig. 4.2. The spectrogram has no interference terms, at the expense of comparatively poor energy concentration. The optimal expansions of $g(t)$ obtained by the Method of Frames (minimum l^2 norm) [52], Matching Pursuit [96], Basis Pursuit (minimum l^1 norm) [14] and WPD are illustrated in Fig. 4.3. While these algorithms use the conventional library of wavelet packets and fail to represent the signal

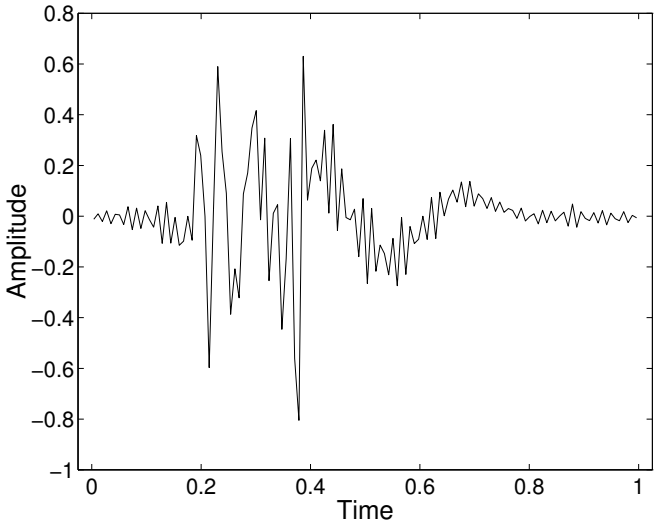
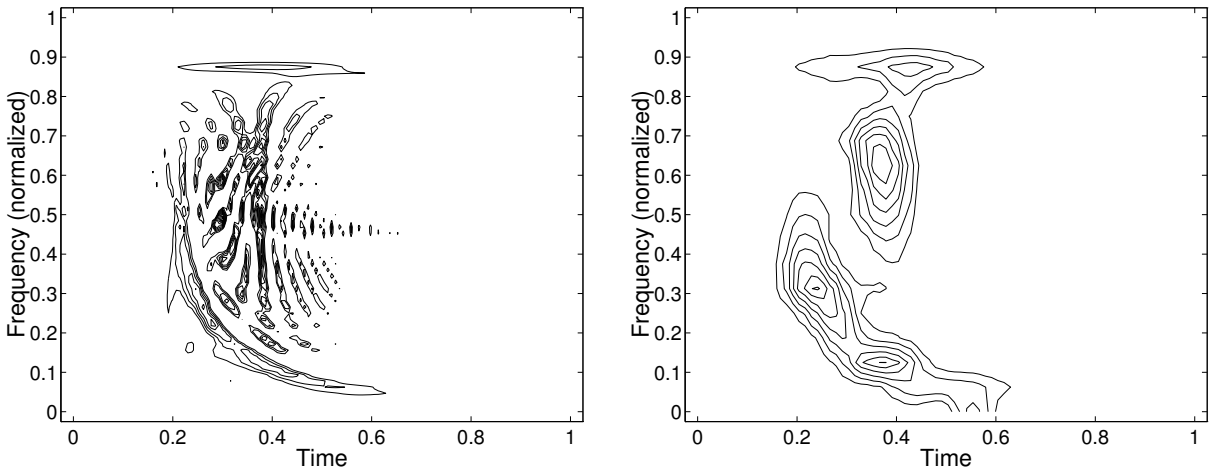


Figure 4.1: Test signal $g(t)$ consisting of a short pulse, a tone and a nonlinear chirp.



(a) (b)

Figure 4.2: Contour plots for the signal $g(t)$: (a) Wigner distribution; (b) Spectrogram. Compared with the WD, the spectrogram does not have undesirable interference terms but the energy concentration is poor.

efficiently, the SIWPD (Fig. 4.3(f)) facilitates an efficient representation by a small number of coefficients. Furthermore, its computational complexity ($\sim 3,580$ multiplications) is significantly lower than those associated with the Matching Pursuit ($\sim 44,800$ multiplications) and the Basis Pursuit ($\sim 331,500$ multiplications).

Fig. 4.4 illustrates the MWD for $g(t)$, using various distance-thresholds. When $D = 0$, there are no interference terms, but the energy concentration of individual components is insufficient. $D = 2$ leads to improved energy concentration, yet, no significant interference terms are present. As D gets larger, the interference between components becomes visible and the MWD converges to the conventional WD (*cf.* Fig. 4.2(a)). An acceptable compromise is usually found between $D = 1.5$ and $D = 2.5$.

Fig. 4.5(a) shows the MWD for $g(t)$, obtained via the SIWPD with thresholds $D = 2$ and $\varepsilon = 0.1$. Figs. 4.5(b), (c), (d), (e) and (f) describe, respectively, the WD, the Smoothed pseudo Wigner distribution, the Choi-Williams distribution, the cone-kernel distribution and the reduced interference distribution [73]. Clearly, the SIWPD based MWD achieves high time-frequency resolution, and is superior in eliminating interference terms.

The particular basis, selected for representing a prescribed signal, plays an important role in the MWD. As long as the "best" basis elements are localized in time-frequency and reasonably matched to the local distribution of the signal, each signal component is characteristically represented by a few significant elements. Thus, by restricting the cross-terms to neighboring basis-functions, we eliminate interference terms between distinct components, and even within components having a nonlinear frequency modulation. On the other hand, whenever the signal is arbitrarily decomposed into elements that have no relation to the actual signal distribution, the performance of the MWD may deteriorate. The SIWPD constitutes an efficient algorithm for selecting the most suitable basis. Similarly to standard WPD, the SIWPD library is generated by a single "mother-wavelet" (Chapter 2). Although the library is flexible and versatile enough to describe various local features

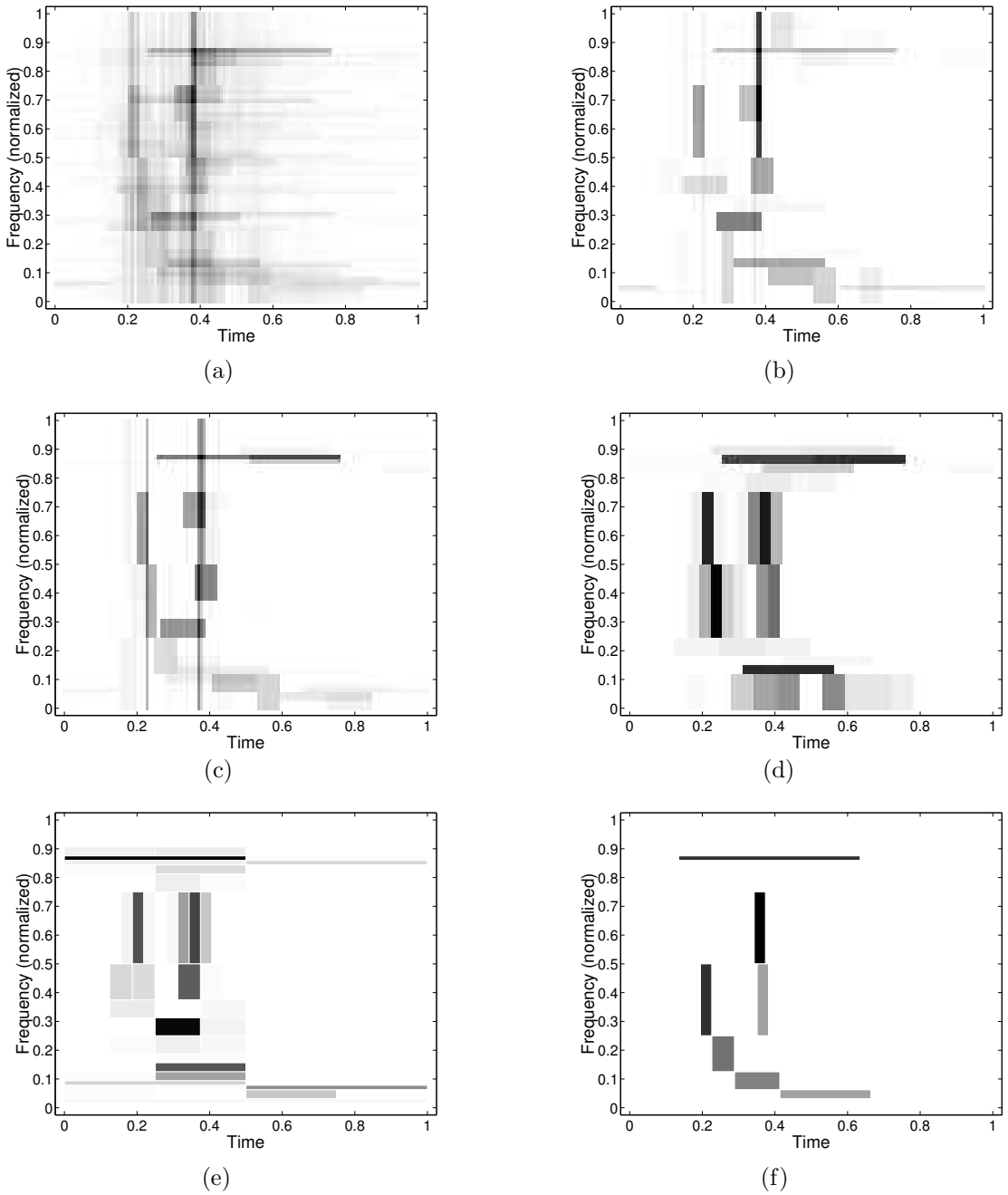


Figure 4.3: Time-frequency tilings for the signal $g(t)$, using the library of wavelet packet bases (generated by 12-tap coiflet filters) and various best-basis methods: (a) Method of Frames (minimum l^2 norm). (b) Matching Pursuit. (c) Basis Pursuit (minimum l^1 norm). (d) Wavelet Packet Decomposition (minimum l^1 norm). (e) Wavelet Packet Decomposition (minimum Shannon entropy). (f) Shift-Invariant Wavelet Packet Decomposition (minimum Shannon entropy).

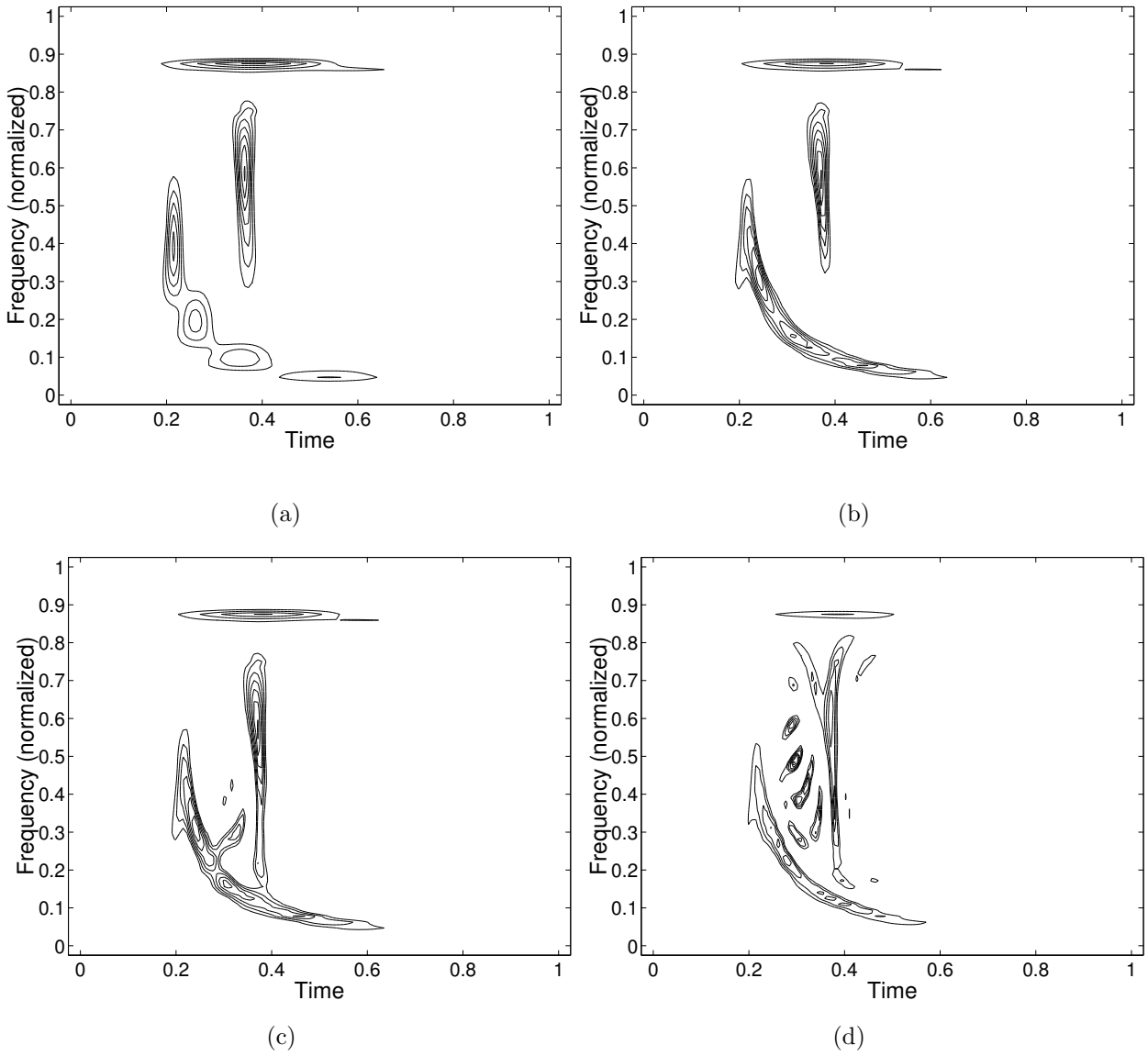


Figure 4.4: The modified Wigner distribution for the signal $g(t)$, combined with the SIWPD and various distance-thresholds: (a) $D = 0$; (b) $D = 2$; (c) $D = 3$; (d) $D = 5$. For $D = 0$, the energy concentration is not sufficient. For $D = 2$, the energy concentration is improved by cross-terms within components. As D gets larger, the interference between components becomes visible and the modified Wigner distribution converges to the conventional WD (*cf.* Fig. 4.2). A good compromise has been found for $1.5 \leq D \leq 2.5$.

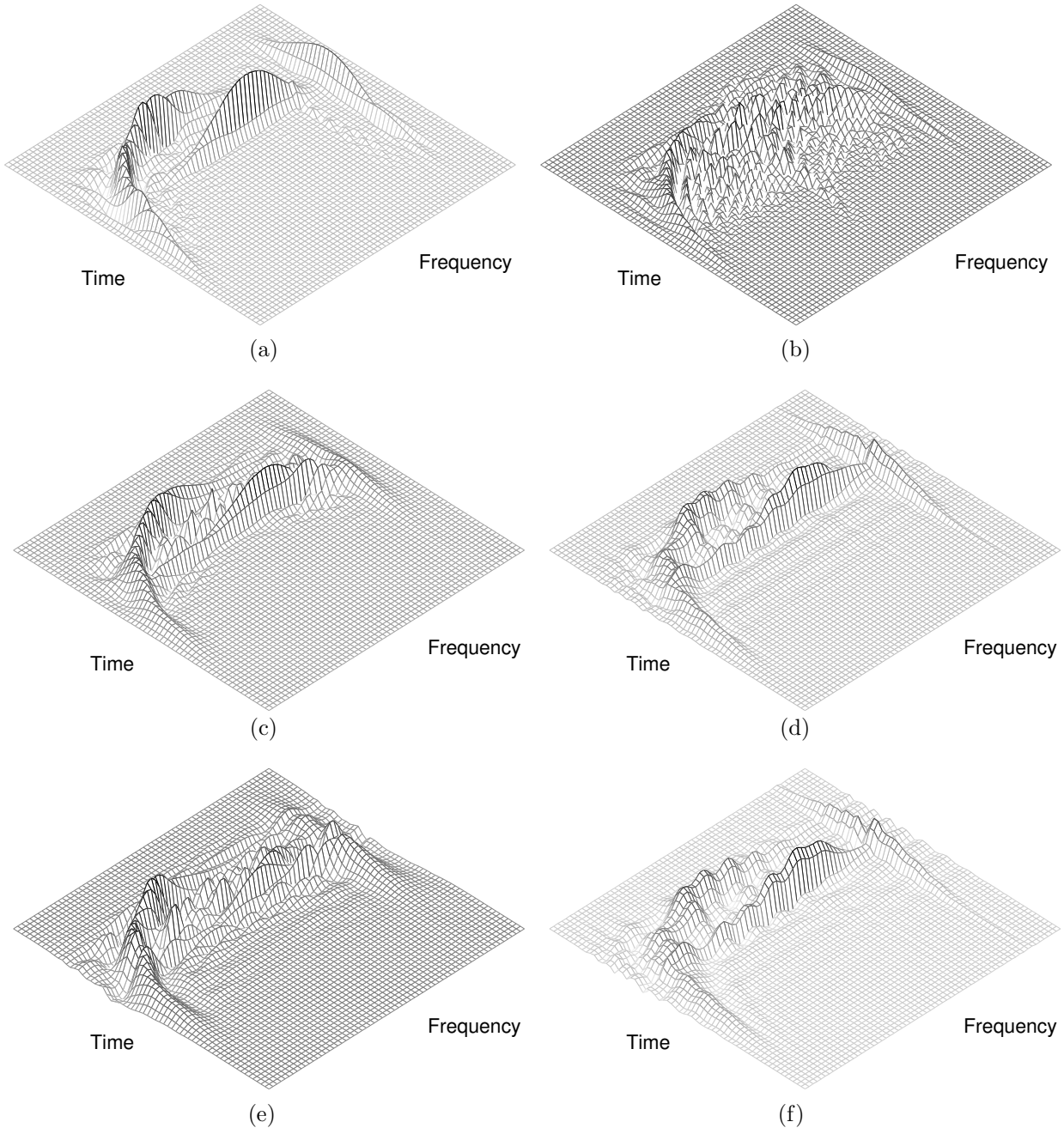
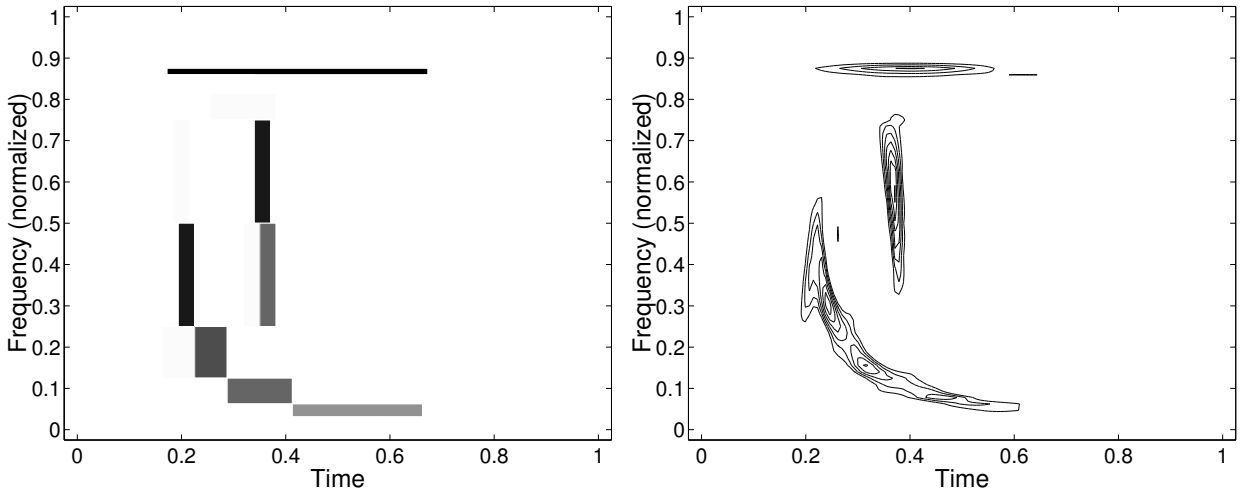


Figure 4.5: Mesh plots for the signal $g(t)$: (a) The modified Wigner distribution combined with the SIWPD and distance-threshold $D = 2$; (b) Wigner distribution; (c) Smoothed pseudo Wigner distribution; (d) Choi-Williams distribution; (e) Cone-kernel distribution; (f) Reduced interference distribution. The modified Wigner distribution yields an *adaptive* distribution where high resolution, high concentration, and suppressed interference terms are attainable.



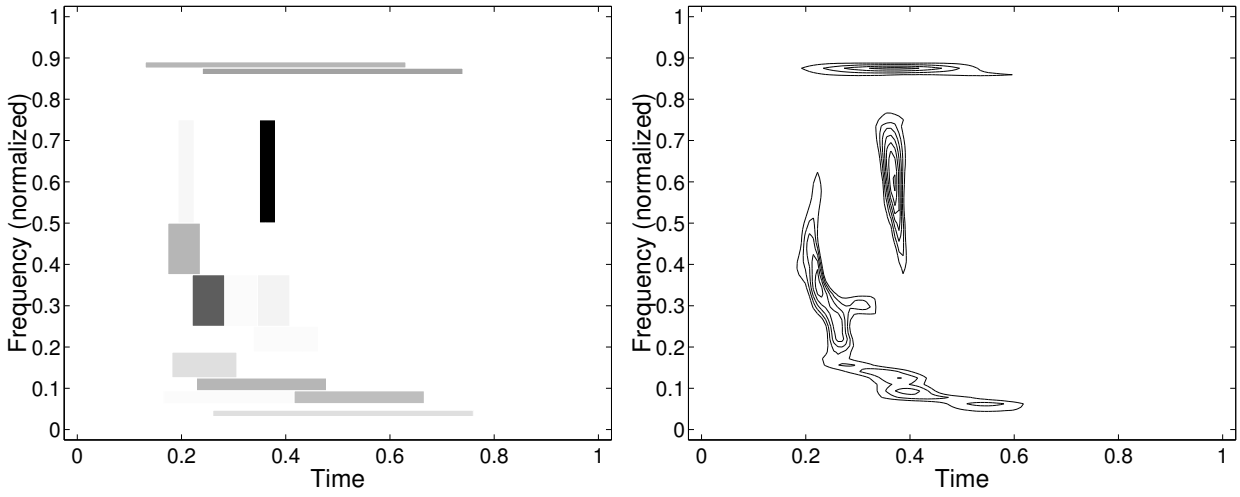
(a) (b)

Figure 4.6: Time-frequency representation for the signal $g(t)$, using the SIWPD with 6-tap Daubechies least asymmetric wavelet filters: (a) The best-basis tiling; entropy= 2.09. (b) The modified Wigner distribution ($D = 2$, $\varepsilon = 0.1$).

of the signal, the choice of the mother-wavelet may affect the eventual performance.

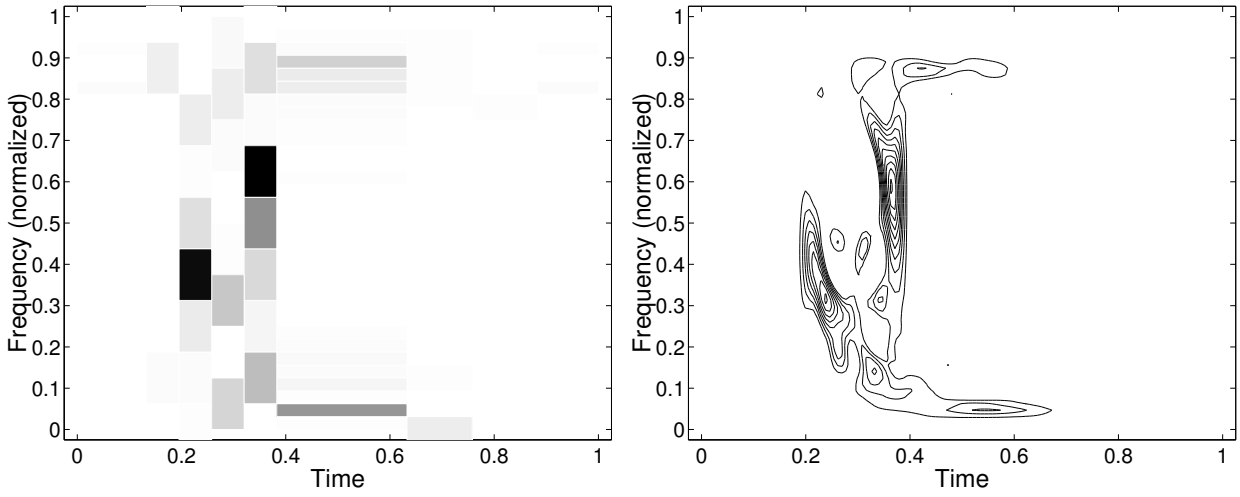
The signal $g(t)$, depicted in Fig. 4.1, can be represented by seven basis-function, belonging to the extended wavelet packet library with C_{12} as the mother-wavelet (*cf.* Fig. 4.3(f). C_{12} corresponds to 12-tap coiflet filters [53, p. 261]). If the SIWPD utilizes decomposition filters that correspond to a different mother-wavelet, then the entropy of the representation is expected to be higher and correspondingly the performance of the MWD will deteriorate. Figs. 4.6 and 4.7 illustrate best-basis expansions and MWDs for $g(t)$, obtained by the SIWPD with D_6 and S_9 as mother-wavelets (D_6 corresponds to 6-tap Daubechies least asymmetric wavelet filters, and S_9 corresponds to 9-tap Daubechies minimum phase wavelet filters [53, pp. 195,198]). A comparison with Figs. 4.3(f) and 4.4(b) shows that despite variations in the time-frequency tilings, the MWD managed to delineate the components of the signal and effectively eliminate the interference terms.

Fig. 4.8 illustrates the best-basis expansion and MWD for $g(t)$, obtained using the SIAP-



(a) (b)

Figure 4.7: Time-frequency representation for the signal $g(t)$, using the SIWPD with 9-tap Daubechies minimum phase wavelet filters: (a) The best-basis tiling; entropy= 2.32. (b) The modified Wigner distribution ($D = 2$, $\varepsilon = 0.1$).



(a) (b)

Figure 4.8: Time-frequency representation for the signal $g(t)$, using the SIAP-LTD: (a) The best-basis tiling; entropy= 2.81. (b) The modified Wigner distribution.

LTD. Here, the basis-functions fail to represent the signal efficiently. We may compare the entropy (= 2.81) with that obtained with the SIWPD (1.88 with C_{12}, 2.09 with D_6, and 2.32 with S_9). The reduced performance of the SIAP-LTD for this particular signal stems from the fact that short pulses, expanded on local trigonometric bases, require a large number of decomposition levels (Chapter 3). This entails a steeper rising cutoff function, and consequently basis-functions which are less localized in frequency [150]. Notice that the "visual quality" of the MWD is well correlated with the entropy attained by the best basis expansion. Lower entropy generally yields "better" (well delineated components, high resolution and concentration) MWD. It appears that "entropy" can serve as a reasonable measure for a quantitative comparison between MWDs.

4.4 General Properties

In this section we investigate the MWD in more detail.

Realness: The MWD is always real, even if the signal or the basis functions are complex.

$$T_g^* = T_g \tag{4.26}$$

This property is a direct consequence of the realness of the Wigner distribution.

Shift-Invariance: Shifting a signal by $\tau = k \cdot 2^{-J} \quad (k, J \in \mathbb{Z})$, where J is finest resolution level of the best-basis decomposition, entails an identical shift of the MWD, *i.e.*,

$$\text{if} \quad \tilde{g}(t) = g(t-\tau) \quad \text{then} \quad T_{\tilde{g}}(t,\omega) = T_g(t-\tau,\omega)\,. \tag{4.27}$$

This property follows from the shift-invariance property of the best-basis decomposition. To see this, let $g(t) = \sum_\lambda c_\lambda \varphi_\lambda(t)$ be the best-basis expansion of g, and let $\tilde{g}(t) = g(t-\tau)$, $\tau =$

$k \cdot 2^{-J}$. Then, using the shift-invariance of the best-basis decomposition, we have

$$\tilde{g}(t) = \sum_{\lambda} \tilde{c}_\lambda \tilde{\varphi}_\lambda(t) = \sum_{\lambda} c_\lambda \varphi_\lambda(t-\tau) \tag{4.28}$$

is the best-basis expansion of $\tilde{g}$, *i.e.*, the best-basis for $\tilde{g}$ is identical to within a time-shift τ to the best-basis for g, and the corresponding expansion coefficients are the same. The MWD of g and $\tilde{g}$ are given by

$$T_g(t,\omega) = \sum_{\lambda\in\Lambda} |c_\lambda|^2 W_{\varphi_\lambda}(t,\omega) + 2 \sum_{\{\lambda,\lambda'\}\in\Gamma} \mathrm{Re}\{c_\lambda c_{\lambda'}^* W_{\varphi_\lambda,\varphi_{\lambda'}}(t,\omega)\}\,, \tag{4.29}$$

$$T_{\tilde{g}}(t,\omega) = \sum_{\lambda\in\tilde{\Lambda}} |c_\lambda|^2 W_{\varphi_\lambda}(t-\tau,\omega) + 2 \sum_{\{\lambda,\lambda'\}\in\tilde{\Gamma}} \mathrm{Re}\{c_\lambda c_{\lambda'}^* W_{\varphi_\lambda,\varphi_{\lambda'}}(t-\tau,\omega)\}\,, \tag{4.30}$$

where we used the shift-invariance property of the Wigner distribution,

$$W_{\tilde{\varphi}_\lambda,\tilde{\varphi}_{\lambda'}}(t,\omega) = W_{\varphi_\lambda,\varphi_{\lambda'}}(t-\tau,\omega)\,, \qquad W_{\tilde{\varphi}_\lambda}(t,\omega) = W_{\varphi_\lambda}(t-\tau,\omega)\,. \tag{4.31}$$

Now, since the expansion coefficients of g and $\tilde{g}$ are identical ($\tilde{c}_\lambda = c_\lambda$), and the time-frequency distance between pairs of basis-functions remains unchanged ($d(\tilde{\varphi}_\lambda, \tilde{\varphi}_{\lambda'}) = d(\varphi_\lambda, \varphi_{\lambda'})$), the sets $\tilde{\Lambda}$ and $\tilde{\Gamma}$ are identical to Λ and Γ, respectively. It is therefore concluded that $T_{\tilde{g}}(t,\omega) = T_g(t-\tau,\omega)$.

Symmetry in Frequency: Real signals have symmetrical spectra. For symmetric spectra, the Wigner distribution is symmetric in the frequency domain,

$$W_g(t,-\omega) = W_g(t,\omega), \quad W_{g,s}(t,-\omega) = W_{s,g}(t,\omega)\,. \tag{4.32}$$

Thus, for real signals and real basis-functions, the MWD retains the same symmetries, *i.e.*,

$$T_g(t,-\omega) = T_g(t,\omega)\,. \tag{4.33}$$

Symmetry in Time: For symmetrical signals, the Wigner distribution is symmetrical in the time domain,

$$W_g(-t,\omega) = W_g(t,\omega), \quad W_{g,s}(-t,\omega) = W_{s,g}(t,\omega)\,. \tag{4.34}$$

However, the MWD is not necessarily symmetric, since the best-basis decomposition is generally asymmetric. Still, confining ourselves to symmetric basis-functions (entailing either biorthogonal or complex-valued basis-functions [53]) and restricting $\mathcal{B}$, the library of bases, to those bases satisfying

$$\{\varphi_\lambda\}_{\lambda\in\mathrm{N}} \in \mathcal{B} \quad \Longrightarrow \quad \{\varphi_\lambda(t)\}_{\lambda\in\mathrm{N}} = \{\varphi_\lambda(-t)\}_{\lambda\in\mathrm{N}} \,,$$

the best-basis decomposition becomes symmetric, rather than shift-invariant. In that case, the MWD is symmetric in time,

$$\begin{aligned} T_g(-t,\omega) &= \sum_{k\in\Lambda} |c_k|^2 W_{\varphi_k}(-t,\omega) + 2 \sum_{\{k,\ell\}\in\Gamma} \mathrm{Re}\{c_k c_\ell^* W_{\varphi_k,\varphi_\ell}(-t,\omega)\} \\ &= \sum_{k'\in\Lambda} |c_{k'}|^2 W_{\varphi_{k'}}(t,\omega) + 2 \sum_{\{k',\ell'\}\in\Gamma} \mathrm{Re}\{c_{k'} c_{\ell'}^* W_{\varphi_{k'},\varphi_{\ell'}}(t,\omega)\} \\ &= T_g(t,\omega)\,. \end{aligned}$$

Total Energy: Integrating the general form of the MWD with respect to time and frequency shows that the total energy is bounded by the energy of the signal:

$$\frac{1}{2\pi}\int dt \int d\omega \; T_g(t,w) = \sum_{\lambda\in\Lambda} |c_\lambda|^2 \le \sum_{\lambda} |c_\lambda|^2 = \|g\|^2 \tag{4.35}$$

where we have used

$$\begin{aligned} \tfrac{1}{2\pi}\textstyle\int dt \int d\omega \; W_{\varphi_k,\varphi_\ell}(t,w) &= \tfrac{1}{2\pi}\textstyle\int dt \int d\omega \int d\tau \; \varphi_k(t+\tfrac{\tau}{2})\varphi_\ell^*(t-\tfrac{\tau}{2}) e^{-j\omega\tau} \\ &= \textstyle\int dt \int d\tau \; \varphi_k(t+\tfrac{\tau}{2})\varphi_\ell^*(t-\tfrac{\tau}{2})\delta(\tau) = \langle \varphi_k, \varphi_\ell\rangle = \delta_{k,\ell}\,. \end{aligned}$$

Observe that the difference between the total energy and the energy of the signal essentially stems from the smallest expansion coefficients. In fact, if we set the amplitude threshold (ε) to zero, the set of indices Λ runs over all the basis-functions, and thus the total energy equals the energy of the signal.

Positivity: The interpretation of the conventional WD as a pointwise time-frequency energy density is generally restricted by the uncertainty principle and by the fact that the WD may locally assume negative values [106, 22, 75]. However, the nonnegativity and interference terms are closely related, and in many cases the suppression of interference terms accompanies reduction of negative values in magnitude [78]. Thus, reduction of the interference terms associated with the WD, entails comparable attenuation of its negative values.

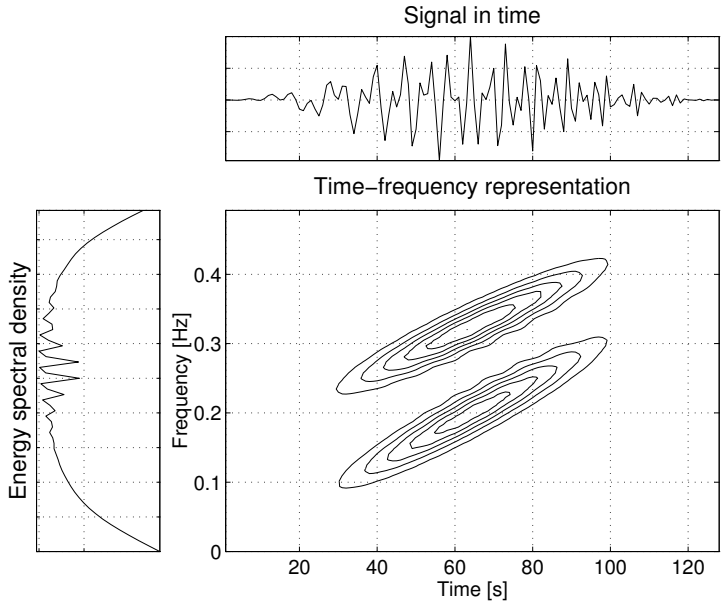
4.5 Inversion and Uniqueness

In this section we show that the components that comprise a given signal can be recovered from the MWD, to within an arbitrary constant phase factor and to within the errors caused by neglecting low weight basis constituents.

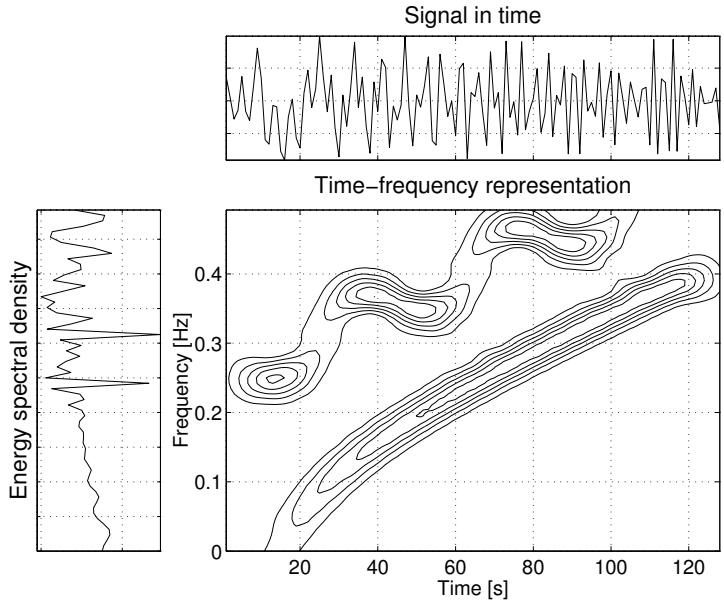
4.5.1 Equivalence Classes in the Time-Frequency Plane

A multicomponent signal is one that has well delineated regions in the time-frequency plane. Examples of multicomponent signals are illustrated in Fig. 4.9. One of the advantages of the MWD is its capability to resolve a multicomponent signal into disjoint time-frequency regions.

Definition 4.1 *Let* $X = \Lambda \cup \{\lambda \mid \{\lambda, \lambda'\} \in \Gamma$ *for some* $\lambda' \in \Lambda\}$ *be the indices set of the significant basis functions, i.e., the basis functions which contribute to the MWD. A pair of indices* $k, \ell \in X$ *are said to be equivalent, denoted by* $k \sim \ell$*, if* $k \equiv \ell$ *or alternatively there exists a finite series* $\{\lambda_i\}_{i=1}^{N}$ *such that* $\{\lambda_i, \lambda_{i+1}\} \in \Gamma$ *for* $i = 1, 2, \ldots, N-1$ *and* $\{k, \lambda_1\}, \{\ell, \lambda_N\} \in \Gamma$.



(a)



(b)

Figure 4.9: Examples of multicomponent signals: (a) Superposition of two linear chirps. (b) Superposition of two nonlinear chirps. Neither the time representation nor the energy spectral density indicate whether the signals are multicomponent. The joint time-frequency representations, however, show that the signals are well delineated into regions.

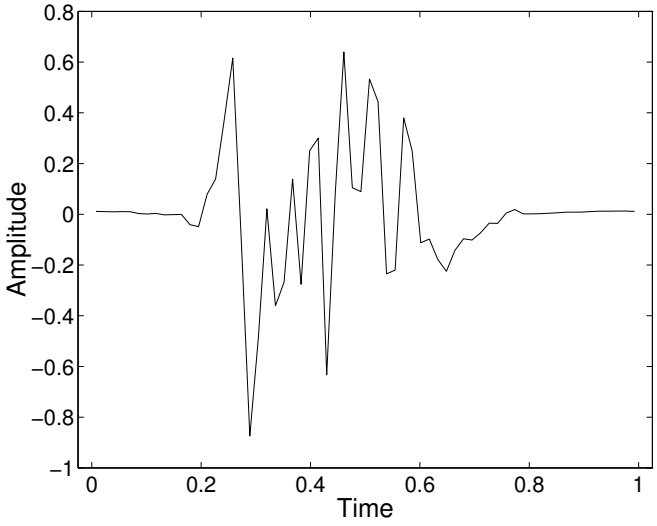


Figure 4.10: A multicomponent signal $s(t)$.

Clearly, $\sim$ is an equivalence relation on X, since it is reflexive ($k \sim k$ for all $k \in X$) symmetric ($k \sim \ell$ implies $\ell \sim k$) and transitive ($k \sim \ell$ and $\ell \sim m$ imply $k \sim m$). The equivalence relation means that the corresponding basis-functions are linked in the time-frequency plane by a series of consecutive adjacent basis-functions.

Denote by

$$\Lambda_k = \{\lambda \in X \mid \lambda \sim k\} \tag{4.36}$$

the equivalence class for $k \in X$. Then, for any $k, \ell \in X$ either $\Lambda_k = \Lambda_\ell$ or $\Lambda_k \cap \Lambda_\ell = \emptyset$. Hence, $\{\Lambda_k \mid k \in X\}$ forms a partition of X, and each equivalence class can be related to a single component of the signal. The number of components which comprise the signal g is determined by the number of distinct equivalence classes in X.

For example, refer to the multicomponent signal $s(t)$, depicted in Fig. 4.10. Its best-basis decomposition (Fig. 4.11) shows that it can be expressed as the sum of six basis-functions: $s(t) = \sum_{k=1}^{6} c_k \varphi_k$. In this case, with an appropriate distance-threshold ($D = 2$), we obtain

$$\Lambda = \{1, 2, 3, 4, 5, 6\} = X\,,$$

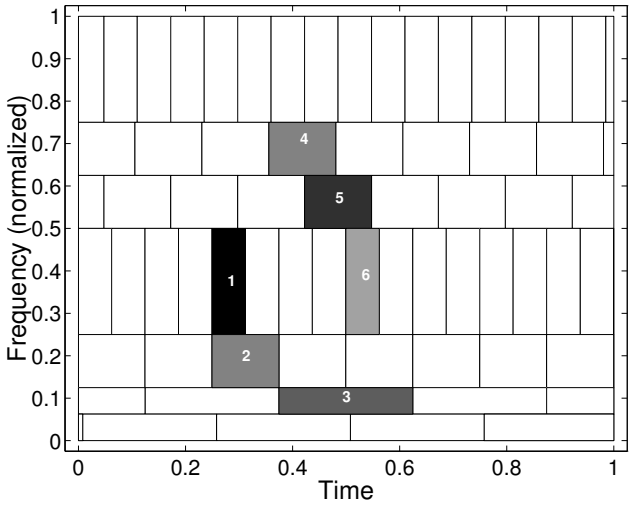


Figure 4.11: The best-basis decomposition of $s(t)$.

$$\Gamma = \{\{1,2\}\ \{2,3\}\ \{4,5\}\ \{5,6\}\}\ .$$

Thus there are two distinct equivalence classes on X,

$$\Lambda_1 = \Lambda_2 = \Lambda_3 = \{1,2,3\} \equiv \Lambda_I\,,$$
$$\Lambda_4 = \Lambda_5 = \Lambda_6 = \{4,5,6\} \equiv \Lambda_{II}\,.$$

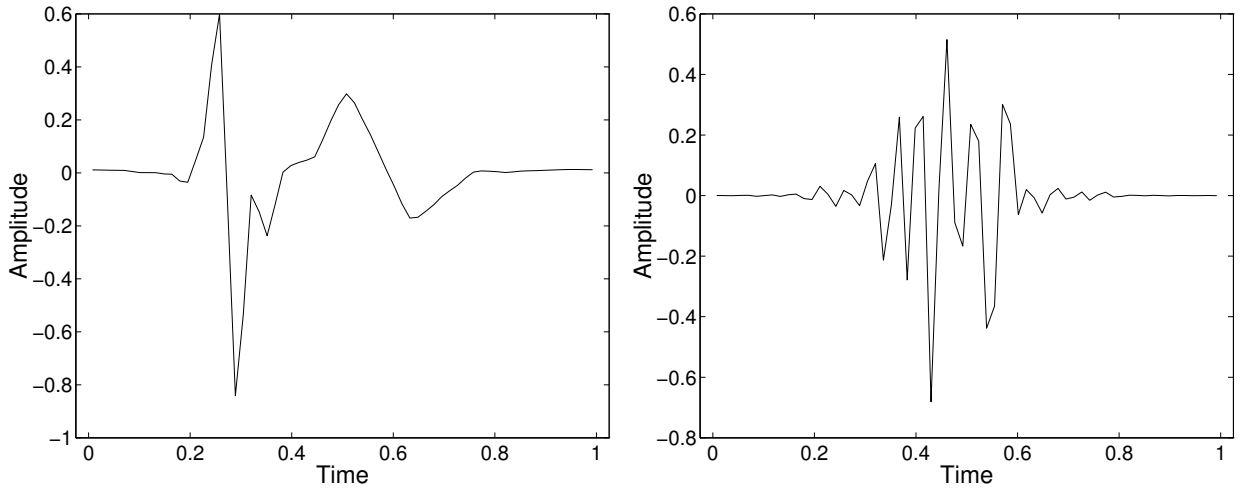
Accordingly, we presume that the signal consists of two components:

$$s = s_I + s_{II}$$

where

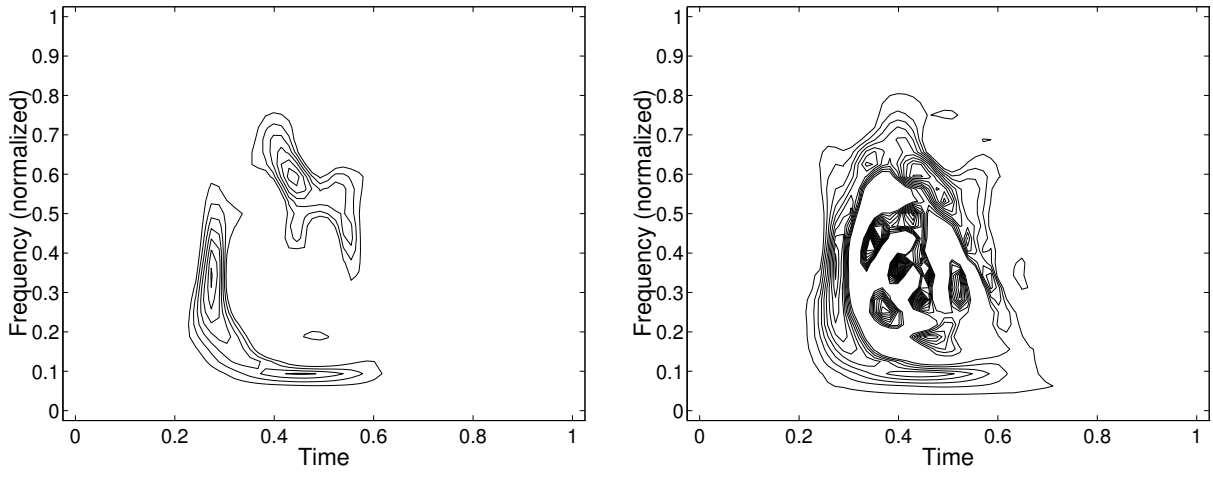
$$s_I = \sum_{k\in\Lambda_I} c_k\varphi_k\,, \quad s_{II} = \sum_{k\in\Lambda_{II}} c_k\varphi_k\,.$$

These components, depicted in Fig. 4.12, are associated with the two well delineated time-frequency regions in the MWD domain (Fig. 4.13(a)).



(a) (b)

Figure 4.12: The components of the signal s. (a) The component s_I associated with the equivalence class Λ_I. (b) The component s_{II} associated with the equivalence class Λ_{II}.



(a) (b)

Figure 4.13: Contour plots for the signal $s(t)$: (a) Modified Wigner distribution; (b) Wigner distribution.

4.5.2 Recovering the Components of a Multicomponent Signal

The components of a multicomponent signal are given by the partial sums of basis-functions with respect to equivalence classes. They can also be recovered from the MWD to within an arbitrary constant phase factor in each signal component, and to within errors generated by neglecting small basis constituents (small auto-terms, small cross-terms, as well as interference terms that correspond to distant basis functions).

Lemma 4.1 *Let $\{\varphi_k\}_{k\in\mathbb{N}}$ be the best basis for $g(t)$, and let $W_{k,\ell} \equiv W_{\varphi_k,\varphi_\ell}$ be the cross Wigner distribution of pairs of basis-functions. Then the set $\{W_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ is an orthonormal basis for $L_2(\mathbb{R}^2)$, and the expansion coefficients for the MWD are given by*

$$c_{k,\ell} = \langle T_g, W_{k,\ell}\rangle = \begin{cases} c_k c_\ell^*, & \text{if } \ k=\ell\in\Lambda \ \text{ or } \ \{k,\ell\}\in\Gamma\,, \\ 0, & \text{otherwise }\,, \end{cases} \tag{4.37}$$

where

$$\langle T_g, W_{k,\ell}\rangle \stackrel{\triangle}{=} \frac{1}{2\pi}\iint T_g(t,\omega) W_{k,\ell}^*(t,\omega)\,dt\,d\omega$$

Proof: We first need to show that the system $\{W_{k,\ell}\}_{k,\ell\in\mathbb{N}}$ is orthonormal and complete in $L_2(\mathbb{R}^2)$. Orthonormality is given by

$$\begin{aligned} \langle W_{k,\ell}, W_{m,n}\rangle &= \frac{1}{2\pi}\int dt\int d\omega\int d\tau\int d\tau'\ \varphi_k(t+\frac{\tau}{2})\varphi_\ell^*(t-\frac{\tau}{2})\varphi_m^*(t+\frac{\tau'}{2})\varphi_n(t-\frac{\tau'}{2})e^{-j\omega(\tau-\tau')} \\ &= \int dt\int d\tau\ \varphi_k(t+\frac{\tau}{2})\varphi_\ell^*(t-\frac{\tau}{2})\varphi_m^*(t+\frac{\tau}{2})\varphi_n(t-\frac{\tau}{2}) \\ &= \langle\varphi_k,\varphi_m\rangle\,\langle\varphi_n,\varphi_\ell\rangle = \delta_{k,m}\delta_{\ell,n}\,, \end{aligned}$$

and completeness is satisfied by

$$(2\pi)^{-1}\sum_{k,\ell\in\mathbb{N}} W_{k,\ell}(t,\omega)W_{k,\ell}^*(t',\omega')$$

$$= \frac{1}{2\pi} \sum_{k,\ell \in \mathbf{N}} \int d\tau \int d\tau' \; \varphi_k(t + \frac{\tau}{2}) \varphi_\ell^*(t - \frac{\tau}{2}) e^{-j\omega\tau} \varphi_k^*(t' + \frac{\tau'}{2}) \varphi_\ell(t' - \frac{\tau'}{2}) e^{+j\omega'\tau'}$$

$$= \frac{1}{2\pi} \int d\tau \int d\tau' \; \delta(t - t' - \frac{\tau}{2} + \frac{\tau'}{2}) \delta(t - t' + \frac{\tau}{2} - \frac{\tau'}{2}) e^{-j\omega\tau + j\omega'\tau'}$$

$$= \delta(t - t') \cdot \frac{1}{2\pi} \int e^{j\tau(\omega' - \omega)} d\tau = \delta(t - t') \delta(\omega - \omega') \, .$$

Now, the MWD can be expressed in the following form:

$$T_g = \sum_{k \in \Lambda} |c_k|^2 W_{k,k} + 2 \sum_{\{k,\ell\} \in \Gamma} \mathrm{Re}\{c_k c_\ell^* W_{k,\ell}\} = \sum_{k,\ell \in \mathbf{N}} c_{k,\ell} W_{k,\ell} \, . \tag{4.38}$$

Therefore, by the uniqueness of the expansion, the relation in Eq. (4.37) holds. □

Let $k \in \Lambda$, and let Λ_k be its equivalence class. Then for any $\ell \in \Lambda_k$ there exists a finite series $\{\lambda_i\}_{i=1}^N$ such that $\{\lambda_i, \lambda_{i+1}\} \in \Gamma$ for $i = 1, \ldots, N-1$ and $\{k, \lambda_1\}, \{\ell, \lambda_N\} \in \Gamma$. By Eq. (4.37) we have

$$|c_k|^2 = \langle T_g, W_{k,k} \rangle \; , \tag{4.39}$$

$$c_k c_{\lambda_1}^* = \langle T_g, W_{k,\lambda_1} \rangle \; , \tag{4.40}$$

$$c_{\lambda_i} c_{\lambda_{i+1}}^* = \langle T_g, W_{\lambda_i,\lambda_{i+1}} \rangle \, , \qquad i = 1, \ldots, N-1 \, , \tag{4.41}$$

$$c_{\lambda_N} c_\ell^* = \langle T_g, W_{\lambda_N,\ell} \rangle \; , \tag{4.42}$$

which shows that c_ℓ has a recursive relation to c_k, and c_k can be recovered from the MWD up to a phase factor. Accordingly, each component of the signal can also be recovered up to an arbitrary constant phase factor by

$$s_k = \sum_{\ell \in \Lambda_k} c_\ell \varphi_\ell \, . \tag{4.43}$$

The constant phase factor in each component of the signal clearly drops out when we calculate the MWD (as it does for the WD). Therefore, it cannot be recovered. Summation

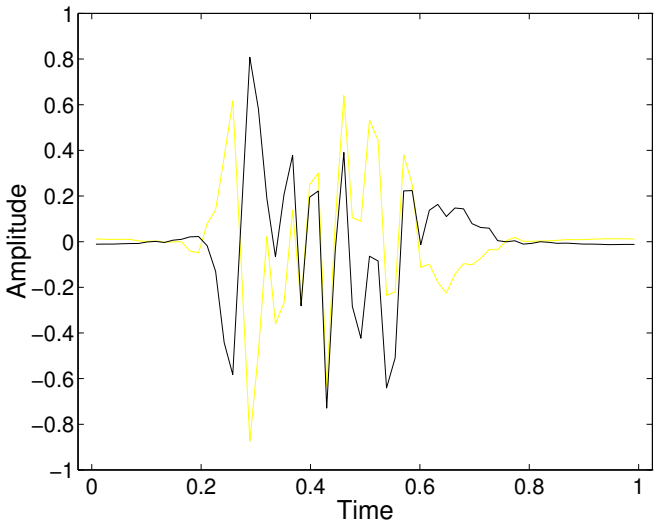


Figure 4.14: The signals $\tilde{s} = -s_I + s_{II}$ (bold line) and $s = s_I + s_{II}$ (light line) are different. However, since they consist of the same components, they have the same modified Wigner distribution.

of distinct signal components generally yields a different signal that has the same MWD. For example, we observed that the signal s in Fig. 4.10 consists of two components, $s = s_I + s_{II}$. The difference of these components, generates another signal $\tilde{s} = s_{II} - s_I$ (*cf.* Fig. 4.14), which has the same MWD as s. In some applications, such as pattern recognition, it is actually desirable that signals consisting of the same components will be identified, irrespective of their relative phase. The MWD provides an efficient technique for doing so.

4.6 Summary

The main issue investigated in this chapter is that of adaptive decompositions of the Wigner distribution and suppression of interference terms, leading to a newly defined modified Wigner space. A prescribed signal is expanded on its best basis, and subsequently transformed into the Wigner domain. The resulting distribution is modified by restricting

the auto-terms and cross-terms to basis-functions whose normalized coefficients are larger in magnitude than a certain amplitude-threshold ε, and to pairs whose time-frequency distance is smaller than a specified critical distance D. We have shown that the distance and amplitude thresholds control the cross-term interference, the useful properties of the distribution, and the computational complexity. A smaller distance-threshold better eliminates the interference terms, but tends to lower the energy concentration. A larger distance-threshold improves the time-frequency resolution at the expense of retaining additional interference terms. When the amplitude-threshold is set to zero and the distance-threshold goes to infinity, the MWD converges to the conventional WD. Appropriate threshold values ($D \approx 2$, $\varepsilon \approx 0.1$) combine high resolution, high concentration and suppressed cross-term interference at a manageable computational complexity.

We have compared alternative selections of libraries, showing that interference terms between distinct components, and even within components having a nonlinear frequency modulation, can be efficiently eliminated, as long as the localization properties of basis elements aptly resemble that of the signal. The visual quality of the MWD is well correlated with the entropy attained by the best basis expansion (improved distributions ensued from lower entropies), facilitating a quantitative comparison between energy distributions. The MWD is also effective for resolving multicomponent signals. The signal components are determined as partial sums of basis-functions over certain equivalence classes in the time-frequency plane. They are well delineated in the time-frequency plane, and can be recovered from the energy distribution to within a constant phase factor and to within the errors caused by neglecting low weight basis constituents.

The proposed methodology is extendable to other distributions (*e.g.*, the Cohen class) and other "best-basis" decompositions. However, the properties of the resulting modified forms clearly depend on the particular distribution, library of bases and best-basis search algorithm which are employed.

Chapter 5

Translation-Invariant Denoising Using the Minimum Description Length Criterion

5.1 Introduction

Recently, the use of wavelet bases for estimation of signals embedded in noise has been the object of considerable research. While traditional methods often remove noise by low-pass filtering, thus blurring the sharp features in the signal, wavelet-based methods show good performance for a wide diversity of signals, including those with jumps, spikes and other nonsmooth features [62, 47, 49]. Originally, the *wavelet shrinkage* method, developed by Donoho and Johnston [64], used a fixed wavelet basis in the following scheme: transformation of the noisy data into the wavelet-domain, soft or hard thresholding of the resulting coefficients, transformation back into the original space. It was recognized that the success of such denoising scheme is determined by the extent to which the transform compresses the unknown signal into few significant coefficients [63]. Since the compression of a given signal depends on the selected basis, adaptive transforms into libraries of bases, such as the wavelet packet and local trigonometric decompositions, have become more acceptable

[61, 84, 130].

Given a library of bases and a noisy measurement, researchers proposed several different approaches to choosing the "best" basis and threshold value, leading to the best signal estimate. In [61, 84], the selection of the adapted basis and the choice of the threshold are based on a criterion of minimum mean-squared error. That threshold is also used in [14] together with the minimum complexity-penalized argument to find the "best" representation of the underlying signal. Saito [130] proposed to use an information-theoretic criterion, called a *Minimum Description Length* (MDL) principle [125], for the noise removal. He suggested that the MDL criterion gives the best compromise between the estimation fidelity (noise suppression) and the efficiency of representation (signal compression). However, the cost function according to his method is not additive. Thus, he employed the Shannon entropy as the primary cost function to determine the best basis in each library of bases, and the MDL principle merely as a secondary criterion. In [85, 104], the MDL principle is further investigated to derive efficient procedures for the selection of the basis as well as the selection of the threshold. They show that it is possible to define an additive "denoising" criterion such that the conventional WPD is applicable.

Coifman and collaborators [49, 7, 130] observed that denoising with the conventional wavelet transform and WPD may exhibit visual artifacts, such as pseudo-Gibbs phenomena in the neighborhood of discontinuities and artificial symmetries across segmentation points in the frequency domain. These artifacts are related to the lack of *shift-invariance*, and therefore can be reduced by averaging the translation dependence: applying a range of shifts to the noisy data, denoising the shifted versions with the wavelet transform, then unshifting and averaging the denoised data. This procedure, termed *Cycle-Spinning*, generally yields better visual performance on smooth parts of the signal. However, transitory features may be significantly attenuated [145]. Furthermore, the MDL principle and related information-theoretic arguments cannot be applied.

Another approach to attaining shift-invariance is to optimize the time localization of the signal, so that its features are well-aligned with the basis-functions. In the case of WPD, Pesquet *et al.* [111, 112] suggested to adapt the shift of the signal as follows: (*i*) To each node of the expansion tree assign an information-cost by averaging the Shannon entropy over all translations. (*ii*) Determine the best expansion tree using the conventional WPD algorithm of Coifman and Wickerhauser. (*iii*) Compare the entropy of the 2^κ orthonormal representations resulting from 2^κ different shift-options, where κ is the number of nodes in the best expansion tree, and choose that representation (shift-option) which minimizes the entropy. This procedure is sub-optimal compared with the SIWPD introduced in Chapter 2, since the expansion tree is determined by the *averaged* entropy. Additionally, the shift-options in step (*iii*) are examined one by one, whereas the SIWPD not only provides a *recursive* selection method for the optimal shift, but also offers an inherent trade-off between the computational complexity and the information cost.

In this chapter, we present a translation-invariant signal estimator, which is based on the SIWPD, the SIAP-LTD and the MDL criterion. We define a collection of signal models and show that the description length of the noisy observed data can be minimized through optimizing the expansion-tree associated with the SIWPD or SIAP-LTD. This yields an optimal representation and optimal signal estimate by applying hard-thresholding to the resulting coefficients. The proposed estimator is not influenced by the alignment of the observation with respect to the basis functions. Furthermore, the advantages of the SIWPD and SIAP-LTD over the conventional WPD and LTD supply the estimator with beneficial properties which make it superior to other methods.

The proposed algorithms for signal estimation are also useful for estimating the time-frequency distributions of noisy signals. Since the Wigner distribution is very sensitive to noise, it is often necessary to employ some kind of smoothing to reduce the noise effects [23, 108]. However, smoothing suppresses noise at the expense of considerable smearing of

the signal components. Consequently, we propose to combine the signal estimator with the adaptive time-frequency distribution which was described in Chapter 4. Illustrations show that the resultant distribution is robust to noise and characterized by high resolution, high concentration and suppressed interference-terms.

The organization of this chapter is as follows. In Section 5.2, we formulate our problem. The estimation of the signal is described as a problem of choosing the best model from a given collection pertaining to an extended library of orthonormal bases. In Section 5.3, the MDL principle is applied to determine the description length of the data. We show that minimum description length is attainable by optimizing the expansion-tree. In Section 5.4, we present a corresponding algorithm for the optimal tree design and signal estimation. We also propose an MDL-based estimator for the time-frequency distribution and a few modifications to the algorithm to optimize either the library or the filter banks used at each node of the expansion-tree. Examples illustrating the execution and performance of the proposed algorithms are presented in Section 5.5. Finally, we discuss the connection of these algorithms with other approaches in Section 5.6.

5.2 Problem Formulation

We assume the following model for signal estimation:

$$y(t) = f(t) + z(t) \tag{5.1}$$

where $y(t)$ represents the noisy observed data, $f(t)$ is the unknown signal to be estimated, and $z(t)$ is a white Gaussian noise (WGN) with zero mean and a presumingly known power spectral density (PSD) σ^2. We assume that $f(t)$ is real-valued and belongs to V_0, where

$$V_0 = \mathrm{clos}_{L^2(\mathbb{R})} \{\psi_0(t-k) \,:\, k \in \mathbb{Z}\} \,, \tag{5.2}$$

so that Eq. (5.1) can be projected onto V_0 (this assumption amounts to some weak regularity condition on $f(t)$ [92]). Furthermore, $f(t)$ is assumed to have a compact support, so that there exists a finite integer N such that

$$\langle f\ ,\ \psi_{\ell,n,m,k}\rangle = 0 \quad \text{if } k < 0 \text{ or } k \geq N2^{\ell} \tag{5.3}$$

where

$$\psi_{\ell,n,m,k}(t) \equiv 2^{\ell/2}\psi_n\left(2^{\ell}(t-m)-k\right)\ , \tag{5.4}$$

$-log_2 N \leq -L \leq \ell \leq 0,\ 0 \leq n, m < 2^{-\ell}$. The integer N designates the number of wavelet packet coefficients retained at the finest resolution level $\ell = 0$.

To estimate $f(t)$ from the noisy observation $y(t)$, we use the SWP library $\mathcal{B}$, which was introduced in Chapter 2, defined as the collection of all the orthonormal bases for $U_{0,0,0}$ which are subsets of

$$\left\{B_{\ell,n,m}\ :\ -L \leq \ell \leq 0,\ 0 \leq n, m < 2^{-\ell}\right\}\ , \tag{5.5}$$

where

$$B_{\ell,n,m} = \left\{\psi_{\ell,n,m,k}\ :\ 0 \leq k < N2^{\ell}\right\}\ , \tag{5.6}$$

$$U_{\ell,n,m} = \mathrm{clos}_{L^2(\mathbb{R})}\left\{B_{\ell,n,m}\right\}\ . \tag{5.7}$$

Definition 5.1 *A collection of indices* $E = \{(\ell,n,m)\ :\ -L \leq \ell \leq 0, \quad 0 \leq n, m < 2^{-\ell}\}$ *is called a tree-set if it satisfies*

(i) The segments $I_{\ell,n} = [2^{\ell}n,\ 2^{\ell}(n+1))$ *are a disjoint cover of* $[0,\ 1)$.

(ii) The shift indices of a pair of nodes (ℓ_1, n_1, m_1), $(\ell_2, n_2, m_2) \in E$ *are related by*

$$m_1 \mod 2^{-\hat{\ell}+1} = m_2 \mod 2^{-\hat{\ell}+1} \tag{5.8}$$

where $\hat{\ell}$ *is the level index of a dyadic interval* $I_{\hat{\ell},\hat{n}}$ *that contains both* I_{ℓ_1,n_1} *and* I_{ℓ_2,n_2}.

By Proposition 2.1, a tree-set E corresponds to the terminal-nodes of a SIWPD tree, *i.e.*, $\{B_{\ell,n,m} : (\ell,n,m) \in E\}$ is an orthonormal basis for $U_{0,0,0}$, and the collection of all tree-sets E as specified above generates a SWP library (Section 2.2). Eq. (5.3) implies that $f(t)$ belongs to $U_{0,0,0} \subset V_0$. Consequently, $f(t)$ can be estimated from

$$\left\{\langle y\,,\,\psi_{\ell,n,m,k}\rangle \,:\, (\ell,n,m) \in E,\, 0 \le k < N2^{\ell}\right\}.$$

Since the bases in the SWP library compress signals very well and the tree-set E is adapted to the signal, it is reasonable to assume that $f(t)$ is adequately represented by a small number $K < N$ of orthogonal directions. Accordingly, we consider a signal estimate of the form

$$\hat{f}(t) = \sum_{k=1}^{K} f_k \phi_k(t) \tag{5.9}$$

where

$$\phi_k \in \{B_{\ell,n,m} \,:\, (\ell,n,m) \in E\}\,. \tag{5.10}$$

The problem is to find the best tree-set E and the best number of terms K (best model) such that the estimate (5.9) is optimal according to the MDL principle.

5.3 The Minimum Description Length Principle

The MDL principle [123, 124, 125] asserts that given a data set and a collection of competing models, the best model is the one that yields the minimal description length of the data. The description length of the data is counted for each model in the collection as the codelength (in bits) of encoding the data using that model, and the codelength needed to specify the model itself. The rationale is that a good model is judged by its ability to "explain" the data, hence the shorter the description length, the better the model.

In order to apply the MDL principle to our problem, we compute the codelength required to encode the data $y(t)$ using the following model

$$y(t) = \sum_{k=1}^{N} y_k \phi_k(t)\,, \tag{5.11}$$

$$f(t) = \sum_{k=1}^{N} f_k \phi_k(t)\,, \qquad f_k \neq 0 \quad \text{iff } k \in \{k_n\}_{1\leq n\leq K}\,, \tag{5.12}$$

$$\{\phi_k \,:\, 1 \leq k \leq N\} = \{B_{\ell,n,m} \,:\, (\ell,n,m) \in E\}\,, \tag{5.13}$$

$$y_k = f_k + z_k\,, \qquad 1 \leq k \leq N \tag{5.14}$$

where $y_k = \langle y, \phi_k \rangle$ and $f_k = \langle f, \phi_k \rangle$ are, respectively, expansion coefficients of the observed data and the unknown signal, and $z_k = \langle z, \phi_k \rangle$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ by the orthonormality of the transform. The encoding, and hence the computation of the codelength, is carried out in three steps: (i) encoding the observed data assuming E, K and $\{k_n\}_{1\leq n\leq K}$ are given; (ii) encoding the number of signal terms K and their locations $\{k_n\}_{1\leq n\leq K}$ assuming that E is given; and (iii) encoding the tree-set E. Accordingly, the total description length of the data is given by

$$\mathcal{L}(y) = \mathcal{L}\left(y \;\middle|\; E, K, \{k_n\}_{1\leq n\leq K}\right) + \mathcal{L}\left(K, \{k_n\}_{1\leq n\leq K} \;\middle|\; E\right) + \mathcal{L}(E)\,. \tag{5.15}$$

We start with the encoding of the observed data assuming E, K and $\{k_n\}_{1\leq n\leq K}$ are given. It was established by Rissanen [125, pp. 56, 87] that the shortest codelength for encoding the data set $\{y_k\}_{1\leq k\leq N}$ using the probabilistic model $P(\{y_k\}_{1\leq k\leq N} \mid \mu)$, where μ is an unknown parameter vector, is asymptotically given by

$$\mathcal{L}(\{y_k\}_{1\leq k\leq N}) = -\log_2 P(\{y_k\}_{1\leq k\leq N} \mid \hat{\mu}) + \frac{q}{2}\log_2 N \tag{5.16}$$

where $\hat{\mu}$ is the maximum likelihood estimator of μ:

$$\hat{\mu} = \arg\max_{\mu} P(\{y_k\}_{1\leq k\leq N} \mid \mu) \tag{5.17}$$

and q is the number of free real parameters in the vector μ.

Recalling that the expansion coefficients of the noise $\{z_k\}_{1\le k\le N}$ are i.i.d. $\mathcal{N}(0,\sigma^2)$, it follows from Eq. (5.14) that the probability of observing the data given all model parameters is,

$$P\left(y \mid \mu\right) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{n=1}^{K}(y_{k_n}-f_{k_n})^2 + \sum_{n=K+1}^{N} y_{k_n}^2\right)\right) \tag{5.18}$$

where

$$\mu = (E, K, \{k_n\}_{1\le n\le K}, \{f_{k_n}\}_{1\le n\le K}) \tag{5.19}$$

is the parameter vector, and

$$\{k_n\}_{K+1\le n\le N} = \{1,\ldots,N\}\backslash\{k_n\}_{1\le n\le K}\,. \tag{5.20}$$

Thus, from Eq. (5.16), the codelength required to encode the observed data, assuming E, K and $\{k_n\}_{1\le n\le K}$ are given, is

$$\begin{aligned}\mathcal{L}\left(y \mid E, K, \{k_n\}_{1\le n\le K}\right) &= -\log_2 P\left(y \mid E, K, \{k_n\}_{1\le n\le K}, \{\hat{f}_{k_n}\}_{1\le n\le K}\right) + \frac{K}{2}\log_2 N \\ &= \frac{1}{2\sigma^2\ln 2}\sum_{n=K+1}^{N} y_{k_n}^2 + \frac{N}{2}\log_2(2\pi\sigma^2) + \frac{K}{2}\log_2 N \end{aligned} \tag{5.21}$$

where

$$\hat{f}_{k_n} = y_{k_n}\,, \quad 1\le n\le K \tag{5.22}$$

are the maximum likelihood estimates of $\{f_{k_n}\}_{1\le n\le K}$.

Next, we encode the number of signal terms K and their locations $\{k_n\}_{1\le n\le K}$ assuming that E is given. The integer K ($1\le K\le N$) requires $\log_2 N$ bits (clearly, if the probability density function for K, $P_K(k)$, is known, then $\mathcal{L}(K) = -\sum_{k=1}^{N} P_K(k)\log_2 P_K(k) \le \log_2 N$). The indices $\{k_n\}_{1\le n\le K}$ can be specified by a binary string of length N containing exactly

K 1s. Since there are $\binom{N}{K}$ such possible strings, the codelength is given by

$$\mathcal{L}\,(K, \{k_n\}_{1\le n\le K} \mid E) = \log_2 N + \log_2 \binom{N}{K} = \log_2 \frac{N \cdot N!}{K!(N-K)!} \tag{5.23}$$

By applying Stirling's formula[1] to the factorials we have

$$\mathcal{L}\,(K, \{k_n\}_{1\le n\le K} \mid E) = Nh(K/N) - \frac{1}{2}\log_2[K(N-K)] - \frac{1}{12\ln 2}\left(\frac{\theta_1}{K} + \frac{\theta_2}{N-K}\right) + c \tag{5.24}$$

where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function and θ_1, θ_2 and c are constants independent of K ($0 < \theta_1, \theta_2 < 1$). For $N \gg K$, ignoring constant terms which are independent of K, the codelength can be approximated by

$$\mathcal{L}\,(K, \{k_n\}_{1\le n\le K} \mid E) \approx K \log_2 N\,. \tag{5.25}$$

Since our goal is to obtain the shortest codelength, the optimal number of signal terms K^* and their optimal locations $\{k_n^*\}_{1\le n\le K}$ are obtained by minimizing the sum of codelengths given by Eqs. (5.21) and (5.25):

$$\begin{aligned} \mathcal{L}\,(y \mid E) &= \frac{1}{2\sigma^2 \ln 2} \sum_{n=K+1}^{N} y_{k_n}^2 + \frac{3K}{2}\log_2 N \\ &= \frac{1}{2\sigma^2 \ln 2}\left[\sum_{n=K+1}^{N} y_{k_n}^2 + \sum_{n=1}^{K}(3\sigma^2 \ln N)\right] \end{aligned} \tag{5.26}$$

where the constant terms are discarded. Clearly,

$$\sum_{n=1}^{N} \min\left(y_n^2\,,\, 3\sigma^2 \ln N\right) \le \sum_{n=K+1}^{N} y_{k_n}^2 + \sum_{n=1}^{K}(3\sigma^2 \ln N) \tag{5.27}$$

for all $1 \le K \le N$ and $\{k_n\}_{1\le n\le K} \subset \{1, \ldots, N\}$. Equality in (5.27) holds for the optimal values given by

$$K^* = \#\left\{y_n^2 > 3\sigma^2 \ln N \mid 1 \le n \le N\right\} \tag{5.28}$$

[1] $x! = \sqrt{2\pi}\, x^{x+1/2} \exp(-x + \frac{\theta}{12x}) \quad (x > 0,\, 0 < \theta < 1)$

and

$$\{k_n^*\}_{1\le n\le K^*} = \left\{ n \mid y_n^2 > 3\sigma^2 \ln N \,,\, 1 \le n \le N \right\} . \tag{5.29}$$

Specifically, given E we compute the expansion coefficients of the observed data, and then K^* is the number of coefficients exceeding the threshold $\sigma\sqrt{3\ln N}$ in absolute value, and $\{k_n^*\}_{1\le n\le K^*}$ are their locations (notice that $K^* = 0$ implies $\hat{f} \equiv 0$). Thus the codelength in Eq. (5.26) reduces to

$$\mathcal{L}\left(y \mid E\right) = \frac{1}{2\sigma^2 \ln 2} \sum_{n=1}^{N} \min\left(y_n^2 \,,\, 3\sigma^2 \ln N\right) . \tag{5.30}$$

To encode the tree-set E, we associate a 3-ary string with the SIWPD tree as follows: For each node (ℓ, n, m), use 0 if its shift-index m is identical to the shift-index of its child-nodes; use 1 if its child-nodes, $(\ell - 1, 2n, m_c)$ and $(\ell - 1, 2n + 1, m_c)$, have a different shift-index $(m_c \neq m)$; and use 2 if it is a terminal-node $((\ell, n, m) \in E)$. Now, traverse the tree from node to node, top-down from left to right, starting at the root at the top. The string for the example shown in Fig. 5.1 is 0210222.

A SIWPD tree includes $|E|$ terminal nodes and $|E| - 1$ internal nodes, where $|E|$ is the cardinality of E. Since the tree always ends with a terminal node, the last 2 in the string can be discarded, and thus we need to encode a sequence containing $|E| - 1$ 2s and $|E| - 1$ symbols from $\{0, 1\}$. The description length of such sequence is

$$\mathcal{L}\left(E\right) = \log_2 \binom{2|E| - 2}{|E| - 1} + (|E| - 1) + \log_2 |E| \,, \tag{5.31}$$

where the first term is required to specify the locations of 2s in the sequence, the second term to discriminate between 0s and 1s, and the third term to encode the number of terminal terms. Applying Stirling's formula to the factorials, the description length of the tree is

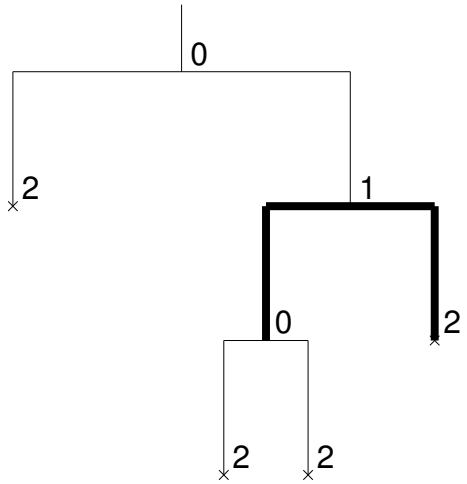


Figure 5.1: Exemplifying the description of SIWPD trees by 3-ary strings. Terminal nodes are represented by 2s, and internal nodes by either 0s or 1s, depending on their expansion mode. In the present example, the string is 0210222.

given by

$$\mathcal{L}(E) = 3|E| + \log_2 \frac{|E|}{\sqrt{|E|-1}} + \frac{\alpha_1 - 4\alpha_2}{24(|E|-1)\ln 2} + c' \tag{5.32}$$

where α_1, α_2 and c' are constants independent of E ($0 < \alpha_1, \alpha_2 < 1$). For $|E| \gg 1$, the codelength can be approximated by

$$\mathcal{L}(E) \approx 3|E| \tag{5.33}$$

where the constant terms are ignored. Adding the codelength $\mathcal{L}(y \mid E)$ (Eq. (5.30)), the total description length of the observed data is given by

$$\mathcal{L}(y) = \mathcal{L}(E) + \mathcal{L}(y \mid E) = 3|E| + \frac{1}{2\sigma^2 \ln 2} \sum_{n=1}^{N} \min\left(y_n^2,\, 3\sigma^2 \ln N\right) . \tag{5.34}$$

Observe that the dependence of $\mathcal{L}(y)$ on the tree-set E is introduced through the number of terminal nodes and the values of the expansion coefficients $\{y_n\}_{1 \le n \le N}$. Since the total

energy of the coefficients $\sum_{n=1}^{N} y_n^2 = \|y\|^2$ is independent of E, we want that the relative energy contained in the coefficients exceeding $\sigma\sqrt{3\ln N}$ in magnitude will be as large as possible. At the same time, we want to minimize the complexity of the expansion tree (the number of terminal nodes). In the next section we propose an efficient algorithm to search for the best tree-set E such that $\mathcal{L}(y)$ is minimized.

5.4 The Optimal Tree Design and Signal Estimation

Let $\mathcal{B}$ represent the SWP library of orthonormal bases. Since each basis B in the library is related to a tree-set E by

$$B = \{B_{\ell,n,m} \,:\, (\ell,n,m) \in E\}\ , \tag{5.35}$$

the search for the optimal E is equivalent to the search for the optimal basis in $\mathcal{B}$. Denote by $\mathcal{L}(By)$ the description length of y represented on a basis B. Then, by Eq. (5.34)

$$\mathcal{L}(By) = \sum_{(\ell,n,m)\in E} \mathcal{L}(B_{\ell,n,m}y) \tag{5.36}$$

where

$$\mathcal{L}(B_{\ell,n,m}y) = 3 + \frac{1}{2\sigma^2\ln 2}\sum_{k=1}^{2^\ell N} \min\left\{C_{\ell,n,m,k}^2(y)\,,\ 3\sigma^2\ln N\right\} \tag{5.37}$$

is the codelength for the terminal node $(\ell,n,m) \in E$, and

$$B_{\ell,n,m}y = \left\{C_{\ell,n,m,k}(y) = \langle y\,,\ \psi_{\ell,n,m,k}\rangle \,:\, 1 \le k \le 2^\ell N\right\} \tag{5.38}$$

are the expansion coefficients of the observed data.

Definition 5.2 *The optimal basis for y in $\mathcal{B}$ with respect to the MDL principle is $B \in \mathcal{B}$ for which $\mathcal{L}(By)$ is minimal.*

The codelength in Eq. (5.36) is an additive cost function, which directly results from the expressions and approximations derived in the previous section. Accordingly, we can apply the SIWPD on the observed data y, as described in Section 2.3, in order to find its optimal basis.

Denote by $A_{\ell,n,m}$ the optimal basis for y restricted to the subspace $U_{\ell,n,m}$. Then, the optimal basis $A \equiv A_{0,0,0}$ can be determined recursively by setting

$$A_{\ell,n,m} = \begin{cases} B_{\ell,n,m} & \text{if } \mathcal{L}(B_{\ell,n,m}y) \leq \mathcal{L}(A_{\ell-1,2n,m_c}y) + \mathcal{L}(A_{\ell-1,2n+1,m_c}y)\,, \\ A_{\ell-1,2n,m_c} \oplus A_{\ell-1,2n+1,m_c}, & \text{otherwise}\,, \end{cases} \tag{5.39}$$

where the shift indices of the respective children-nodes are obtained by

$$m_c = \begin{cases} m, & \text{if } \sum_{i=0}^{1} \mathcal{L}(A_{\ell-1,2n+i,m}y) \leq \sum_{i=0}^{1} \mathcal{L}(A_{\ell-1,2n+i,m+2^{-\ell}}y) \\ m + 2^{-\ell}, & \text{otherwise}\,. \end{cases} \tag{5.40}$$

At the coarsest resolution level $\ell = -L$ the subspaces $U_{-L,n,m}$ are not further decomposed, *i.e.*,

$$A_{-L,n,m} = B_{-L,n,m} \tag{5.41}$$

for $0 \leq n, m < 2^L$.

The optimal basis A minimizes the description length of the observed data. Thus, from Eqs. (5.22), (5.28) and (5.29), the optimal estimate of $f(t)$ is obtained by expanding the observed data $y(t)$ on the optimal basis $A = \left\{\hat{\phi}_k\right\}_{1\leq k\leq N}$ and *hard-thresholding* the coefficients by $\tau \equiv \sigma\sqrt{3\ln N}$. Specifically,

$$\hat{f}(t) = \sum_{k=1}^{N} \eta_\tau(y_k)\hat{\phi}_k(t) \tag{5.42}$$

where $y_k = \left\langle y, \hat{\phi}_k \right\rangle$, and $\eta_\tau(c) \stackrel{\triangle}{=} c\mathbf{1}_{\{|c|>\tau\}}$ is the *hard-threshold* function.

Form Proposition 2.2 it turns out that the signal estimation by the above process is shift-invariant. That is, if the observed data $y(t)$ is translated in time by $q \in \mathbb{Z}$, then the signal estimate $\hat{f}(t)$ is also translated by q. Observe that the restriction of the translations to integers stems from the fact that the initial (finest) resolution level of representing the observed signal is $\ell = 0$, as the unknown signal $f(t)$ is assumed to be in V_0. If we use a finer resolution level $J > 0$ for the initial discrete representation, the shift-invariance is satisfied for finer translations of the form $2^{-J}q$, where $q \in \mathbb{Z}$. However, the resolution levels $0 < \ell \leq J$ add no information to estimating the signal, and consequently the execution of SIWPD over the resolution levels $\ell > 0$ merely increases the computational complexity without improving the performance of the estimator.

The following steps summarize the execution of optimal signal estimation by the MDL principle:

Step 0 *Choose an extended library of wavelet packet bases $\mathcal{B}$ (i.e, specify a mother wavelet for the SWP library) and specify the maximum depth of decomposition L ($L \leq \log_2 N$).*

Step 1 *Expand the data y into the library $\mathcal{B}$. i.e., obtain the coefficients $B_{\ell,n,m}y = \{C_{\ell,n,m,k}(y)\}_{1\leq k\leq 2^\ell N}$ for $-L \leq \ell \leq 0$, $0 \leq n, m < 2^{-\ell}$.*

Step 2 *Use Eq. (5.37) to determine $\mathcal{L}(B_{\ell,n,m}y)$ for $-L \leq \ell \leq 0$, $0 \leq n, m < 2^{-\ell}$, and set $A_{-L,n,m} = B_{-L,n,m}$ for $0 \leq n, m < 2^L$.*

Step 3 *Determine the optimal basis $A \equiv A_{0,0,0}$ and the minimum description length $\mathcal{L}(Ay)$ using Eqs. (5.39)–(5.40).*

Step 4 *Threshold the expansion coefficients in the selectred basis by $\tau = \sigma\sqrt{3\ln N}$ and reconstruct the signal estimate, as expressed by (5.42).*

The computational complexity of executing an optimal SIWPD best-basis expansion is $O(N2^{L+1})$. Yet, as demonstrated in Section 2.5, one may resort to a *sub-optimal* SIWPD procedure entailing a reduced complexity, and higher description length (*i.e.*, information cost) while still retaining the desirable shift-invariance property. In that case, the depth of a subtree, used at a given parent-node to determine its shift index, is restricted to d resolution levels ($1 \leq d \leq L$), and the computational complexity reduces to $O[2^d(L-d+2)N]$. In the extreme case $d = 1$, the complexity, $O(NL)$, is similar to that associated with the conventional WPD. The larger d and L, the larger the complexity, however, the determined optimal basis generally yields a shorter description length.

Similar to the algorithm described in [130], our algorithm can also be extended to find the optimal basis in more than one library. Given a collection of libraries $\{\mathcal{B}_i\}_{1\leq i\leq P}$ including a few SWP libraries and extended libraries of local trigonometric bases, we can find the optimal basis that minimizes the description length as follows: For each library $\mathcal{B}_i$ ($1 \leq i \leq P$), find the optimal basis $A_i \in \mathcal{B}_i$ and the description length $\mathcal{L}(A_i y)$ as described above. Then, choose the optimal basis A such that $\mathcal{L}(Ay) = \min\{\mathcal{L}(A_i y) : 1 \leq i \leq P\}$. In the case of an extended library of local trigonometric bases, the codelength associated with a terminal node is also approximated by Eq. (5.37). Each node in a SIAP-LTD tree has only two expansion alternatives, for it is either decomposed or selected as a terminal node (in contrast to the SIWPD tree, where each node has three expansion alternatives). However, another bit is required for each terminal node to specify its polarity (Section 3.5). Therefore, the description lengths of SIAP-LTD and SIWPD trees are approximately the same.

Finding the optimal basis $A = \left\{\hat{\phi}_k\right\}_{1\leq k\leq N}$, the signal estimate is once again obtained by Eq. (5.42). Alternatively, the decomposition filters can be adapted to the statistics of the signal in each node [104]. Joint adaptation of filter banks and tree structures has been utilized in image coding applications [56, 105], and a fast algorithm for maximizing energy

compaction was introduced in [103]. In our case, to compute the description length of the observed data, the codelength of an internal node should include the specification of the filters applied to expand it. Since the number of internal nodes is relative to the number of terminal nodes (there are $|E| - 1$ internal nodes and $|E|$ terminal nodes), the MDL can be obtained by adding to $\mathcal{L}(B_{\ell,n,m}y)$ (expression (5.37)) the codelength required to specify the filter banks. Specifically, the codelength of a terminal node is given by

$$\mathcal{L}(B_{\ell,n,m}y) = \log_2 M + 3 + \frac{1}{2\sigma^2 \ln 2} \sum_{k=1}^{2^\ell N} \min\left\{ C^2_{\ell,n,m,k}(y)\,,\; 3\sigma^2 \ln N \right\}\,, \tag{5.43}$$

where M is the number of different decomposition filters being examined at each *internal* node.

The proposed algorithm for signal estimation is also useful for estimating the time-frequency distributions of noisy signals. While the conventional Wigner distribution is very sensitive to noise and smoothing is usually applied to reduce noise at the expense of considerable smearing of the signal components [23, 108], the above signal estimate, combined with the modified Wigner distribution that was described in Chapter 4, yields robust time-frequency representations. Denote by $\hat{T}_y$ the time-frequency distribution estimate of y. Then, from Eqs. (5.42) and (4.18)–(4.20),

$$\hat{T}_y(t,\omega) = \sum_{k\in\Lambda} |y_k|^2 W_{\hat{\phi}_k}(t,\omega) + 2 \sum_{\{k,k'\}\in\Gamma} \mathrm{Re}\{y_k y^*_{k'} W_{\hat{\phi}_k,\hat{\phi}_{k'}}(t,\omega)\} \tag{5.44}$$

where

$$\Lambda = \left\{ k \;:\; |y_k| > \sigma\sqrt{3\ln N},\, 1 \le k \le N \right\}, \tag{5.45}$$

$$\Gamma = \left\{ \{k,k'\} \;:\; k,k' \in \Lambda,\, 0 < d(\hat{\phi}_k, \hat{\phi}_{k'}) \le D \right\}. \tag{5.46}$$

In the next section we show by examples that the above estimate of the time-frequency distribution is robust to noise and possesses the all useful properties of the modified Wigner

distribution, which were extensively discussed in the previous chapter (high energy concentration, well delineated components, low interference-terms, *etc*).

5.5 Examples

In this section, we give two examples for demonstrating the execution and performance of the proposed estimators.

Example 5.1 *Synthetic signal.*

We created a synthetic signal $f_1(t)$ by a linear superposition of a few shifted wavelet packets, generated by the C_{12} scaling function (C_{12} corresponds to 12−tap coiflet filters [53, page 261] [54]). The signal contains $N = 2^7$ samples and is depicted in Fig. 5.2(a). Its SIWPD is illustrated in Fig. 5.2(b), where the Shannon entropy is used as the cost function. The noisy observation $y_1(t)$ (Fig. 5.2(c)) was created by adding WGN to $f_1(t)$ with signal-to-noise ratio SNR= 7dB. The optimal SIWPD of $y_1(t)$ using the MDL criterion is shown in Fig. 5.2(d). Notice the remarkable resemblance between the optimal representation of the noisy signal using the MDL principle and the ordinary SIWPD of the original signal using the Shannon entropy. This resemblance stems from fact that according to the MDL principle, the relative energy, contained in the coefficients exceeding $\sigma\sqrt{3\ln N}$ in magnitude, should be as large as possible (refer to Eq. (5.34)). While by the Shannon entropy, the expansion coefficients in the best-basis should decrease as rapidly as possible, when rearranged in a decreasing magnitude order. Therefore, the Shannon entropy applied to the original signal and the MDL criterion applied to the noisy signal generally produce similar SIWPD, as long as the threshold level (noise) is lower than the expansion coefficients of the original signal in the best-basis.

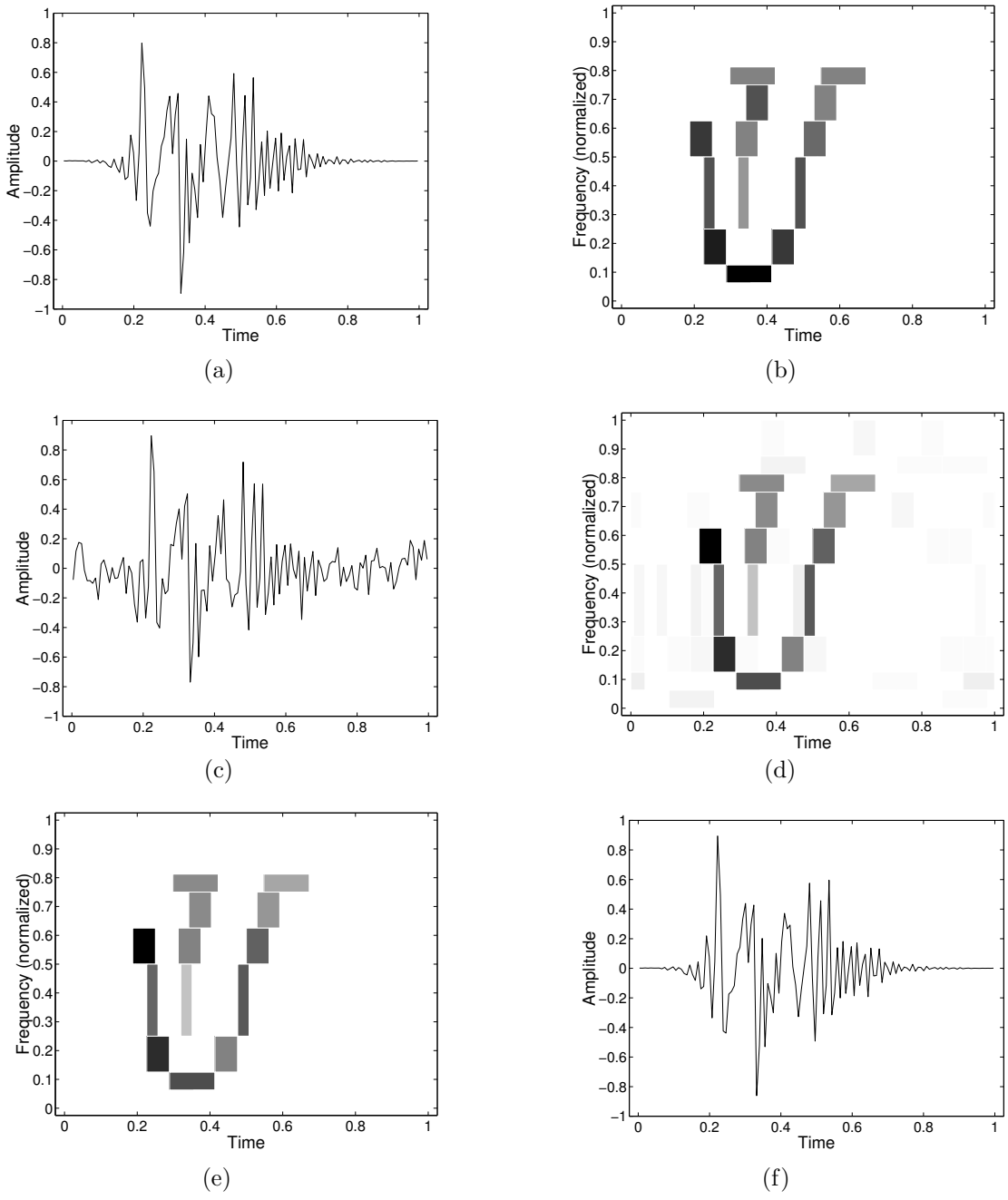


Figure 5.2: Signal estimation by SIWPD and MDL principle: (a) Synthetic signal $f_1(t)$. (b) SIWPD of $f_1(t)$ using the Shannon entropy. (c) Noisy measurement $y_1(t)$; SNR= 7dB. (d) SIWPD of $y_1(t)$ using the MDL principle. (e) The expansion coefficients of $y_1(t)$ after hard-thresholding. (f) The signal estimate $\hat{f}_1(t)$; SNR= 19dB.

Pursuing the estimation procedure with the MDL criterion, the expansion coefficients of $y_1(t)$ in the optimal basis are thresholded by $\sigma\sqrt{3\ln N}$ and transformed back into the signal domain. Figs. 5.2(e) and (f) show, respectively, the retained coefficients and the signal estimate $\hat{f}_1(t)$. Compared to the noisy measurement $y_1(t)$, the signal estimate is enhanced to SNR= 19dB.

Fig. 5.3 illustrates the usefulness of our algorithm for estimating the time-frequency distribution of the noisy data. While the WD of the original signal is corrupted by interference terms and even worsens by the noise (Figs. 5.3(a) and (b)), the Smoothed pseudo Wigner distributions are more readable and less sensitive to noise (Figs. 5.3(c) and (d)). However, the energy concentration of the signal components is poor. The estimate of the modified Wigner distribution, given by Eq. (5.44), is not only robust to noise (compare Figs. 5.3(e) and (f)), but also characterized by high resolution, high concentration and suppressed interference-terms.

Example 5.2 *Evolution of electromagnetic pulse in a relativistic magnetron.*

Fig. 5.4(a) shows a noisy measurement of an electromagnetic pulse ($\approx$ 100 nanoseconds long) generated by high power ($\approx$ 100 MegaWatts) relativistic magnetron. The measurement involves heterodyning at 2.6GHz, filtering at 500kHz and sampling at 1GHz [132]. The Wigner distribution, depicted in Fig. 5.4(b), is clearly ineffective as a time-frequency analysis tool, for its high noise sensitivity. Yet, the estimates of the signal and the modified Wigner distribution, as shown in Figs. 5.4(c) and (d), are potentially valuable when analyzing the measurements and studying the non-stationary phenomena, such as mode build-up and competition and pulse shortening [5], which are common in such high power microwave tubes.

In this example, we employed the SIAP-LTD, since it yielded a shorter description length than the SIWPD (probably because the energy of the pulse is concentrated in the cavity-

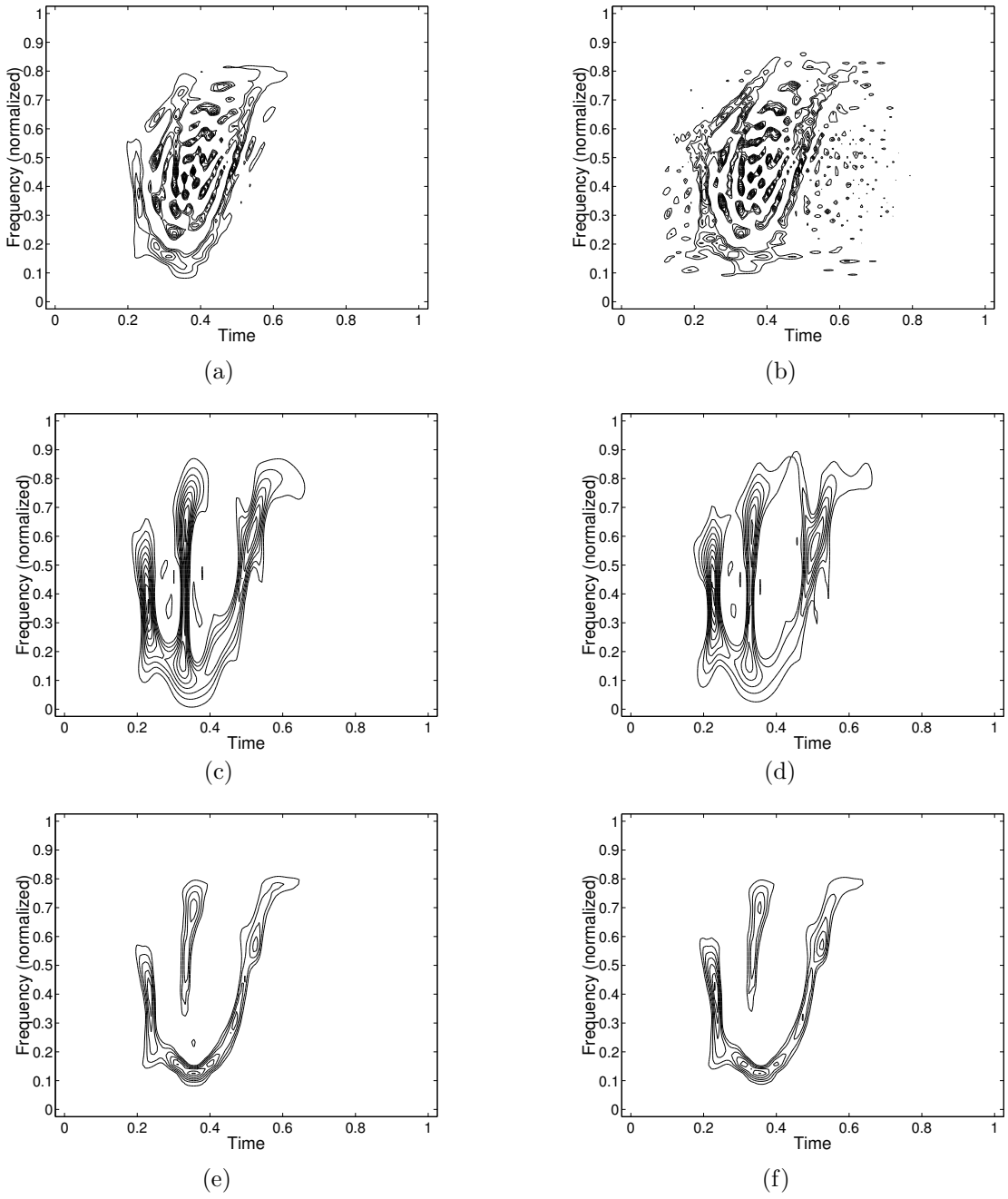


Figure 5.3: Contour plots of time-frequency distributions: (a) Wigner distribution for the original signal $f_1(t)$. (b) Wigner distribution for the noisy measurement $y_1(t)$. (c) Smoothed pseudo Wigner distribution for $f_1(t)$. (d) Smoothed pseudo Wigner distribution for $y_1(t)$. (e) The modified Wigner distribution for $f_1(t)$. (f) The estimate of the modified Wigner distribution for $y_1(t)$ by the MDL principle.

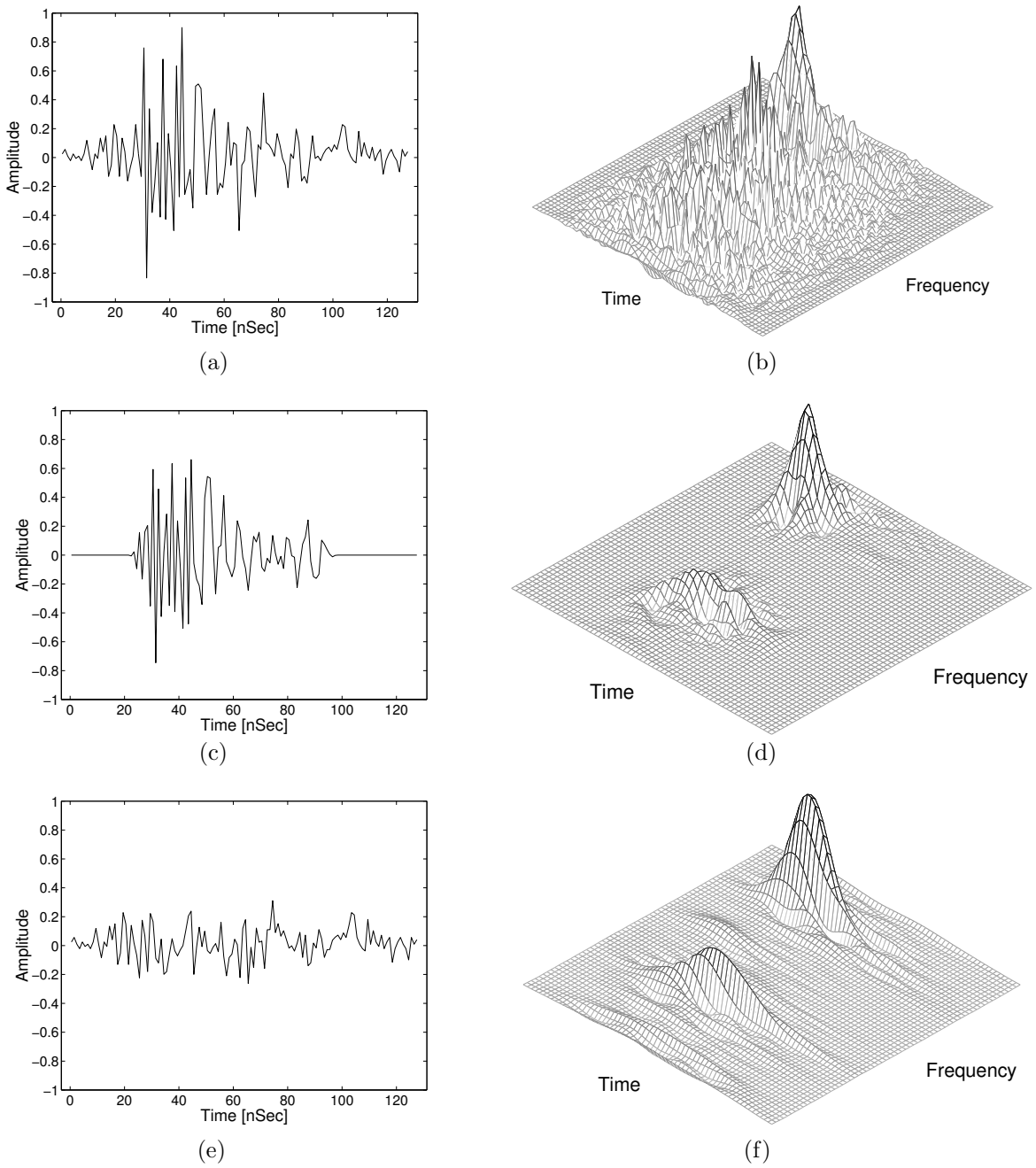


Figure 5.4: Electromagnetic pulse in a relativistic magnetron (heterodyne detection; local oscillator= 2.6GHz): (a) Noisy measurement $y_2(t)$. (b) Wigner distribution for $y_2(t)$. (c) The signal estimate $\hat{f}_2(t)$ by the MDL principle. (d) The estimate of the modified Wigner distribution for $y_2(t)$. (e) Residual between $y_2(t)$ and $\hat{f}_2(t)$. (f) Smoothed pseudo Wigner distribution for $y_2(t)$.

modes of the magnetron, and local trigonometric bases are more appropriate for describing oscillations). The residual between the noisy measurement and the signal estimate is depicted in Fig. 5.4(e). To ascertain that this residual is actually the noise component, we compare the estimate of the modified Wigner distribution with the smoothed pseudo Wigner distribution of the noisy measurement (Fig. 5.4(f)). Since these two distributions are similar, in view of the fact that smoothing in the Wigner domain reduces the noise at the expense of smearing the signal components, it is reasonable to assume that the signal estimate contains all the signal components and the residual is mostly noise.

5.6 Relation to Other Work

Our algorithm has a close relationship with the "simultaneous noise suppression and signal compression" algorithm developed by Saito [130]. For a given collection of orthonormal bases $\{B_p\}_{1\le p\le P}$ consisting of standard wavelet-packet and local trigonometric bases, his algorithm first selects the optimal basis $A \equiv B_{p^*}$ and the optimal number of retained coefficients $K^* < N$ by the MDL principle:

$$\{p^*, K^*\} = \arg \min_{\substack{1 \le p \le P \\ 0 \le K < N}} \left\{ \mathcal{L}(B_p y) = \frac{3}{2} K \log N + \frac{N}{2} \log(\textstyle\sum_{k=K+1}^{N} C_{p,k}^2(y)) \right\} \qquad (5.47)$$

where $\{C_{p,k}(y) \equiv \langle y, \phi_{p,k} \rangle\}_{1\le k\le N}$ are the expansion coefficients of y represented in the basis $B_p = \{\phi_{p,k}(t)\}_{1\le k\le N}$, sorted in order of decreasing magnitude. Then, the signal estimate is reconstructed from the K^* largest expansion coefficients in the optimal basis:

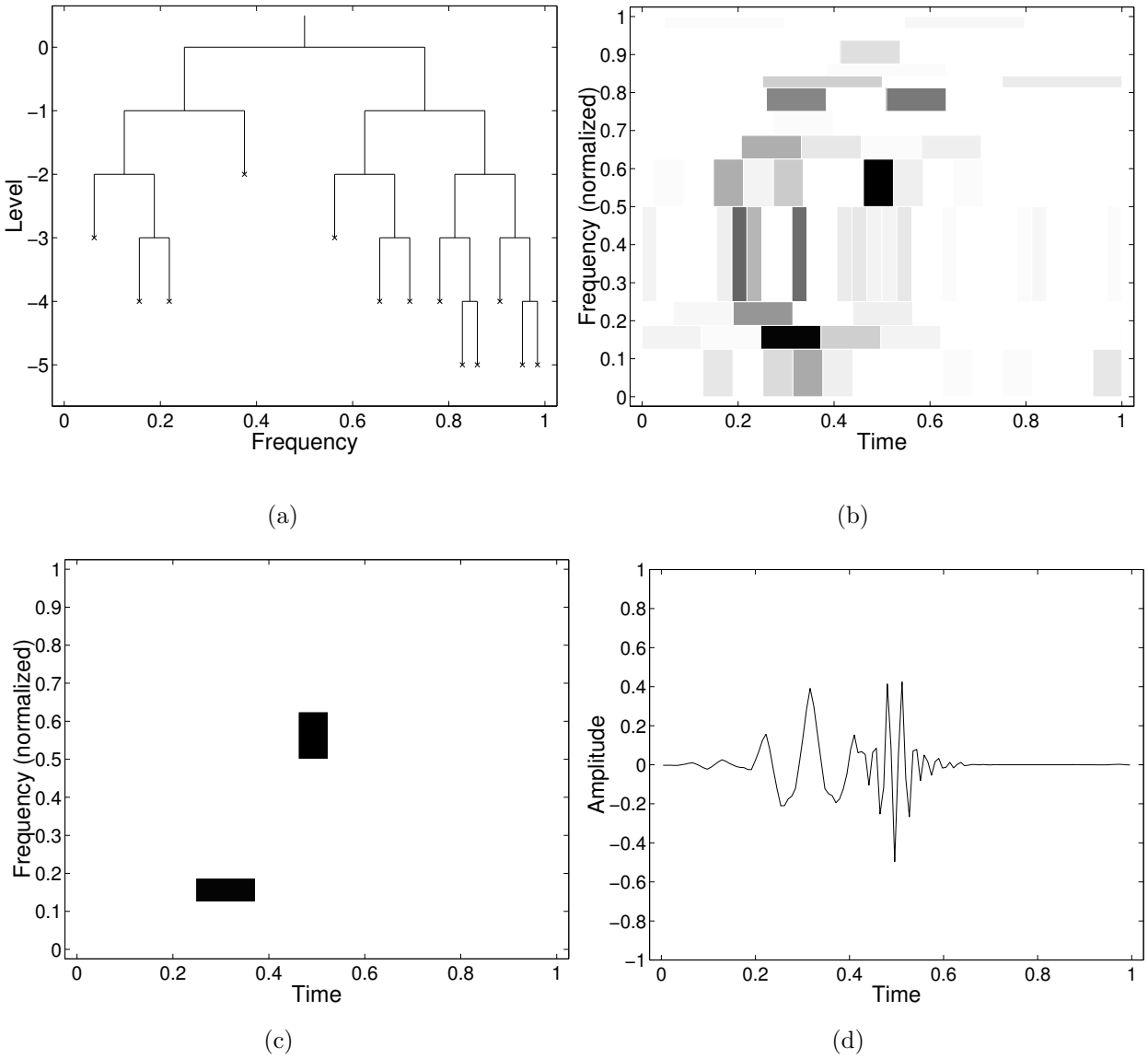
$$\hat{f}(t) = \sum_{k=1}^{K^*} C_{p^*,k}(y) \phi_{p^*,k}(t) \qquad (5.48)$$

(compare Eqs. (5.47) and (5.48) with (5.28) and (5.42)). To maintain a manageable computational complexity, when considering *libraries* of bases only *one* basis out of each

library is being examined, by taking that basis which minimizes the Shannon entropy of the observed data. The main differences between our algorithm and that of Saito are:

- Our method selects the optimal basis by the MDL principle whereas his method first minimizes the Shannon entropy to determine the "best-basis" in each library and only then applies the MDL principle to select the optimal basis among the "best-bases".
- His method ignores the codelength required to specify the best-basis in its library, and thus complex expansion trees are not penalized. On the other hand, our method imposes a significant penalty (up to $3 \cdot 2^L$ bits) for complex trees.
- Our method assumes that the PSD of the noise (σ^2) is known whereas his method estimates it from the $N - K$ smallest coefficients by $\frac{1}{N}\sum_{k=K+1}^{N} C_{p,k}^2(y)$ (maximum-likelihood estimate). In our algorithm we can use different measurements or more advanced methods to estimate the noise, whereas the above estimate of σ^2 heavily relies on the assumption that $f(t)$ is orthogonal to $\{\phi_{p^*,k}(t)\}_{K^*+1\le k\le N}$.
- Our method translates the MDL criterion into an additive information cost function and thus best-basis search algorithms are applicable, whereas his method computes the description length in each basis one at a time.

Figs. 5.5–5.7 demonstrate the comparison between our algorithm and that of Saito, using the synthetic signal analyzed in Example 5.1. Suppose that the library of bases includes the wavelet packet bases generated by the C_{12} scaling function (recall that the synthetic signal $f_1(t)$ was formed using this library), then according to Saito, the best basis is obtained by a conventional WPD with the Shannon entropy employed as the cost function. The resultant expansion-tree and coefficients of the noisy observation $y_1(t)$ are illustrated in Figs. 5.5(a) and (b), respectively. Since the compression of the signal by the WPD is insufficient, some of the coefficients containing signal energy are regarded as noise and set to zero. The



(a) (b)

(c) (d)

Figure 5.5: Signal estimation by the Saito method using the WPD: (a) The best expansion tree of $y_1(t)$ (the signal is depicted in Fig. 5.2(c)). (b) The expansion coefficients of $y_1(t)$. (c) The retained coefficients. (d) The signal estimate; SNR= 1.1dB.

retained coefficients are shown in Fig. 5.5(c). The signal estimate, reconstructed from these coefficients, is depicted in Fig. 5.5(d). Observe that the SNR for the signal estimate got worse than for the noisy measurement (1.1dB< 7dB).

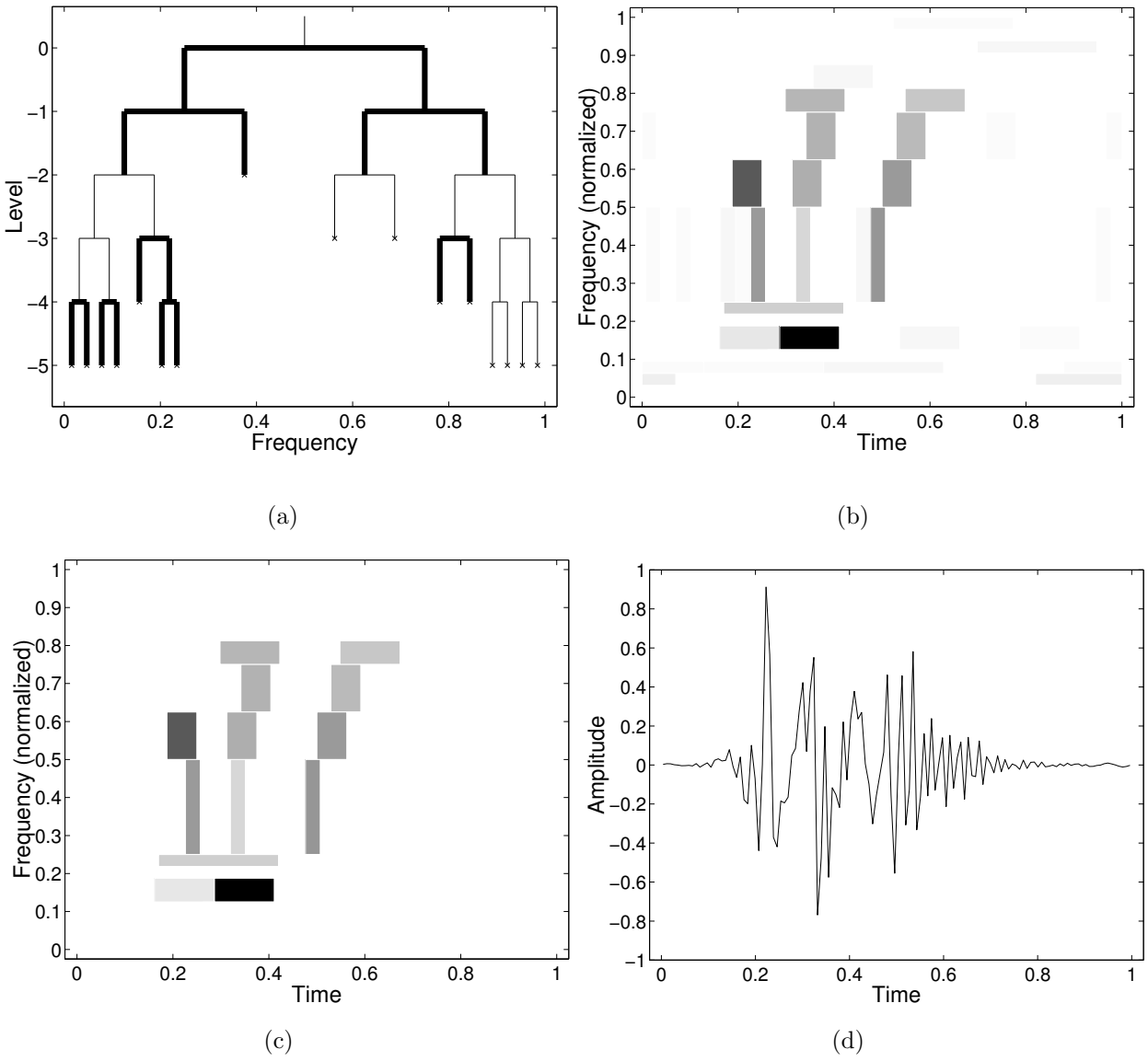
The WPD is a special case of the SIWPD (Chapter 2). Therefore, the SIWPD yields sparser representations and better estimates than the WPD, even using the Saito method (compare Figs. 5.6 and 5.5). Still, the selection of the best-basis by the Shannon entropy criterion, as discussed above, is not optimal with regard to the MDL principle. The results obtained using our method are depicted in Fig. 5.7. The expansion of the signal estimate by the MDL principle (Fig. 5.7(c)) is similar to the expansion of the original signal (Fig. 5.2(b)). The SNR for the signal estimate is significantly higher than for the noisy measurement (19dB> 7dB).

Our algorithm is also intimately connected to the denoising algorithm of Krim and Pesquet [85]. Their algorithm first applies the WPD to the observed data using the information cost

$$\mathcal{M}(\{y_n\}) = \sum_n \min\left(y_n^2\,,\, 2\sigma^2 \log_2 N\right)\,, \tag{5.49}$$

and then reconstructs the signal estimate from the coefficients that are larger than $\sigma\sqrt{2\log_2 N}$ in magnitude. Their method, however, disregards the description length of the expansion tree (compare Eqs. (5.49) and (5.34)). Furthermore, while our method attains shift-invariance by utilizing the SIWPD and SIAP-LTD, their method, restricted by the WPD, admits of signal estimates and performances which are significantly influenced by the alignment of the observation with respect to the basis functions.

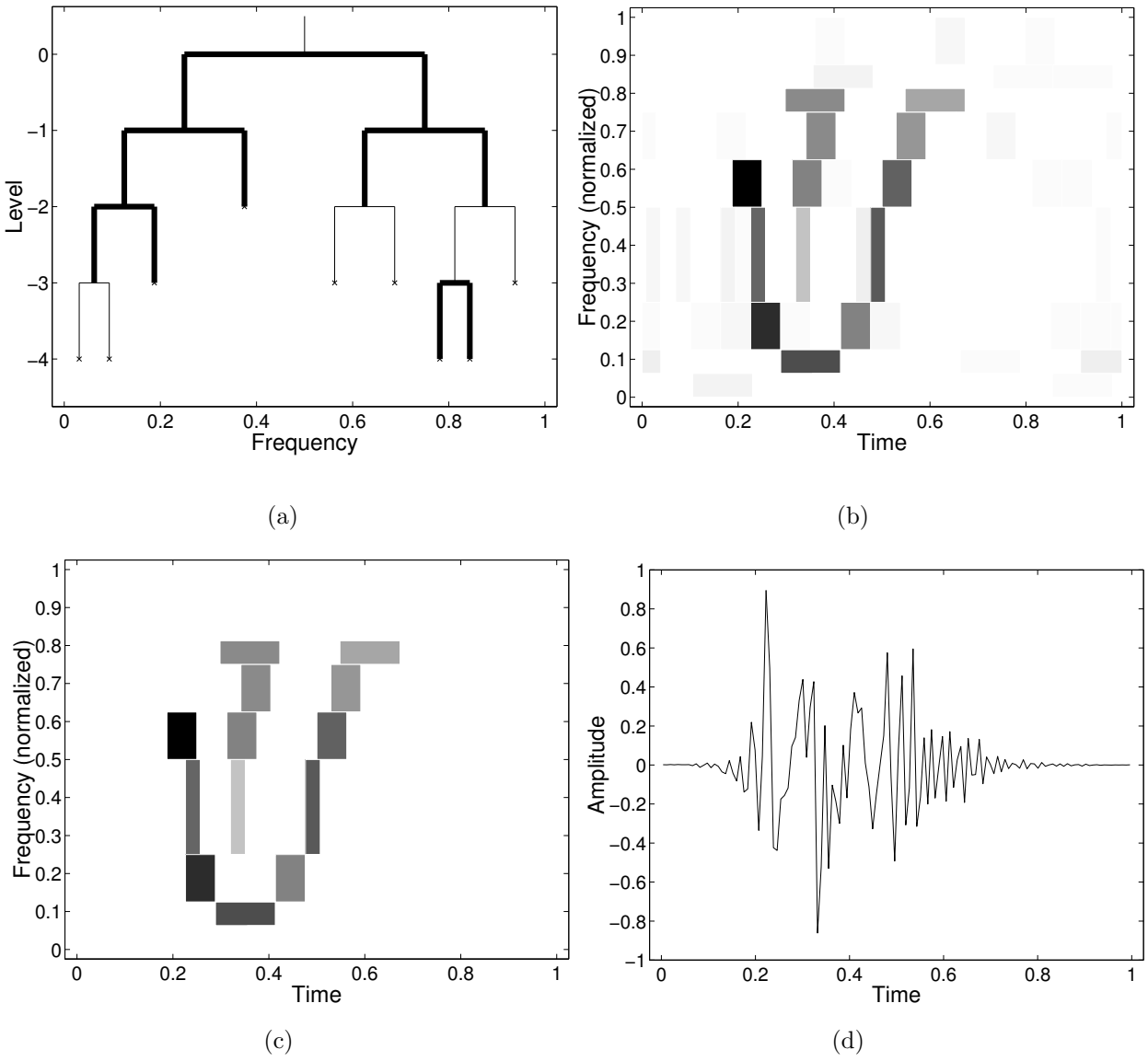
Donoho and Johnstone [61] used a different approach to select from a library of bases the "ideal basis" for the signal estimator. Rather than the MDL principle, their criterion was the mean-squared error. They showed that from this point of view, the best-basis for



(a) (b)

(c) (d)

Figure 5.6: Signal estimation by the Saito method using the SIWPD: (a) The best expansion tree of $y_1(t)$. (b) The expansion coefficients of $y_1(t)$. (c) The retained coefficients. (d) The signal estimate; SNR= 12.8dB.



(a) (b)

(c) (d)

Figure 5.7: Signal estimation by the proposed method: (a) The optimal expansion tree of $y_1(t)$. (b) The expansion coefficients of $y_1(t)$. (c) The retained coefficients. (d) The signal estimate; SNR= 19dB.

denoising is one minimizing

$$\mathcal{M}(\{y_n\}) = \sum_n \min\left(y_n^2\,,\,\zeta^2\right)\,, \tag{5.50}$$

where $\zeta = \nu\sigma(1+\sqrt{2\ln M_N})$, M_N is the number of distinct basis-functions contained in the library (for WPD, $M_N = N\log_2 N$) and $\nu > 8$. The signal is then reconstructed in the best-basis from the coefficients which are larger than ζ in magnitude. The threshold ζ is larger than $\tau = \sigma\sqrt{3\ln N}$, obtained by the MDL principle (see Eq. (5.42)), by at least a factor of $8\sqrt{2/3}$. Thus, the criterion (5.50) imposes a larger penalty on nonzero coefficients, but nothing for the complexity of the expansion-tree (compare with Eq. (5.34)).

The methods mentioned above try to recover the signal from a few basis-functions that belong to one of the bases in a library. Alternatively, one could gather all the basis-functions which comprise the library into a *dictionary* of functions, and then search for the "best" reconstruction (not necessarily orthogonal) of the signal estimate according to a specified criterion. Let $\mathcal{D}$ denote an overcomplete dictionary of waveforms, and let

$$\hat{f}(t) = \sum_{k=1}^{N} \hat{f}_k\phi_k(t)\,, \qquad \{\phi_k\}_{1\le k\le N} \subset \mathcal{D} \tag{5.51}$$

be the signal estimate model. Chen and Donoho [14] proposed to choose the optimal set of elements $\{\phi_k\}_{1\le k\le N}$ and optimal set of coefficients $\{\hat{f}_k\}_{1\le k\le N}$ by solving the penalized problem

$$\min_{\hat{f}}\left\{\frac{1}{2}\left\|y-\hat{f}\right\|_2^2 + \sigma\xi\cdot\sum_{k=1}^{N}\left|\hat{f}_k\right|\right\} \tag{5.52}$$

where $\xi = \sqrt{2\ln M_N}$, and M_N is the cardinality of the dictionary. They showed that the solution to this problem can be obtained by linear programming, and compared it by examples to: (*i*) the Donoho-Johnstone estimator described above; (*ii*) the Method-of-

Frames denoising (MOFDN), which refers to the solution of

$$\min_{\hat{f}} \left\{ \left\| y - \hat{f} \right\|_2^2 + \xi \cdot \sum_{k=1}^{N} \left| \hat{f}_k \right|^2 \right\} ; \tag{5.53}$$

and (*iii*) the Matching-Pursuit denoising (MPDN), which runs Matching-Pursuit [96] until the coefficient associated with the selected waveform gets below the threshold ξ. The solution to (5.52), which was named *Basis-Pursuit* denoising (BPDN), generally results in fewer significant coefficients than the MOFDN, more stable than the MPDN, and outperforms the Donoho-Johnstone estimator when the true signal has a moderate number of nonorthogonal components. However, the BPDN is computationally much more expensive than the other methods.

It is interesting to recognize that part of the criterion in our method, which is based on the MDL principle, is similar to expressions (5.52) and (5.53). Inserting Eqs. (5.12) and (5.22) into (5.26), we have that $\mathcal{L}(y \mid E)$, the description length of the noisy data given the expansion-tree, can be written as

$$\mathcal{L}(y \mid E) = \frac{1}{2\sigma^2 \ln 2} \left\{ \left\| y - \hat{f} \right\|_2^2 + \sigma^2 (3 \ln N) \cdot \sum_{n=1}^{K} \left| \hat{f}_{k_n} \right|^0 \right\} . \tag{5.54}$$

Here, the penalty term includes an ℓ^0 norm of the coefficients, whereas BPDN and MOFDN use ℓ^1 and ℓ^2 norms, respectively. Considering again the estimation problem described in Example 5.1, Fig. 5.8 shows the signal estimates of the synthetic signal obtained by the Donoho-Johnstone method, MOFDN, BPDN and MPDN. The dictionary of basis-elements employed in these algorithms is derived from the WPD with the C_{12} scaling function. Compared to the signal estimate in our method (Fig. 5.2(f)), the above estimates have very low signal-to-noise ratios (Table 5.1). The deficient recovery of the original signal results from the restricted compression capability of the WPD-dictionary. While the SIWPD optimizes the representation of the signal by incorporating translations of wavelet-packets

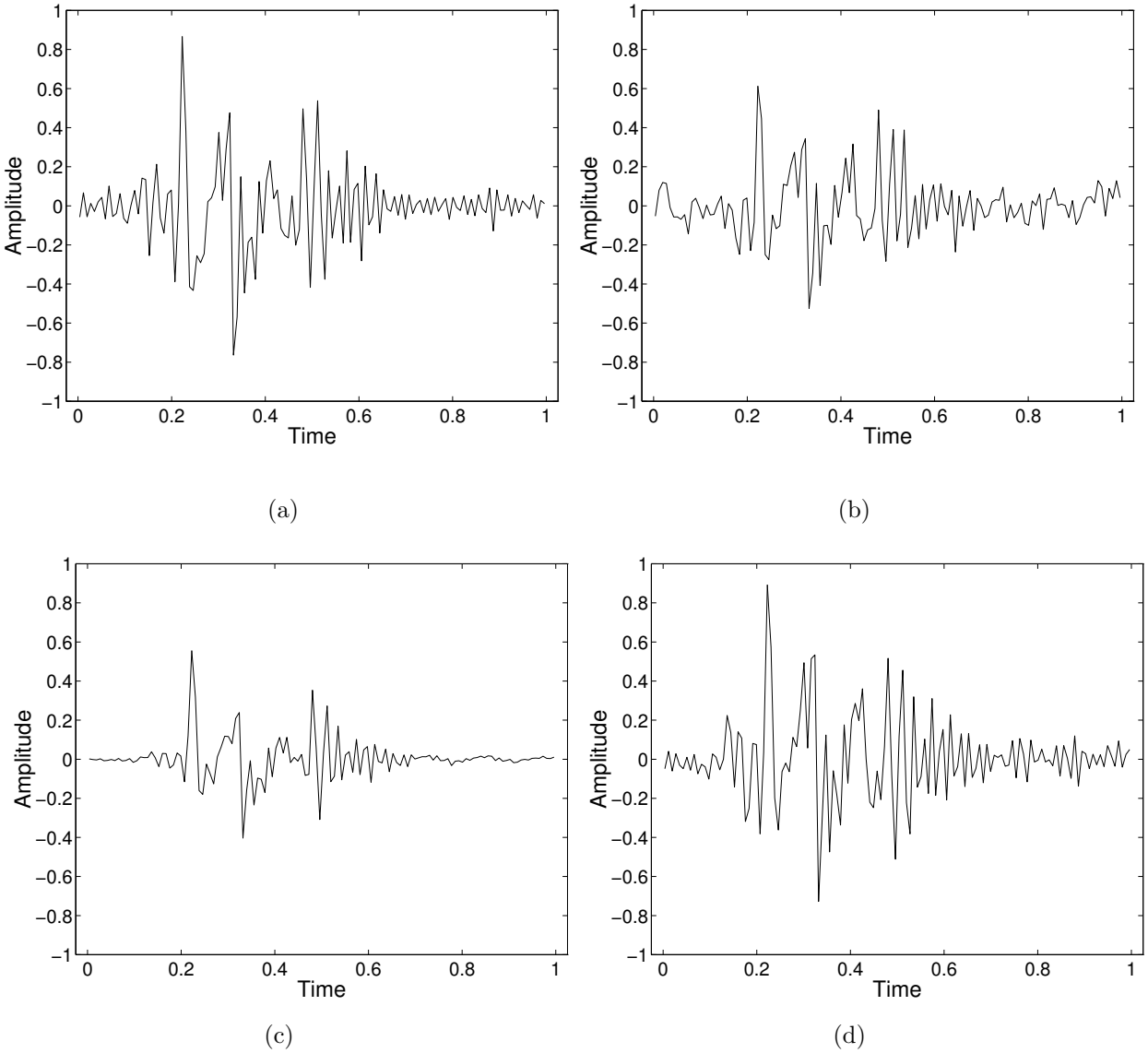


Figure 5.8: Signal estimates of the synthetic signal using the library of wavelet packets (12-tap coiflet filters): (a) The Donoho-Johnstone method; SNR= 6.4dB. (b) The Method-of-Frames denoising (MOFDN); SNR= 7.1dB. (c) The Basis-Pursuit denoising (BPDN); SNR= 4.3dB. (d) The Matching-Pursuit denoising (MPDN); SNR= 7.5dB.

Denoising Method	SNR (dB)
Saito + WPD	1.1
Basis-Pursuit	4.3
Donoho-Johnstone	6.4
Method-of-Frames	7.1
Matching-Pursuit	7.5
Saito + SIWPD	12.8
The proposed method	19.1

Table 5.1: Signal-to-noise ratios for the signal estimates of the synthetic signal using the library of wavelet packets (12-tap coiflet filters) and various denoising methods. The SNR obtained by the proposed *MDL-based Translation-Invariant Denoising* method is significantly higher than those obtained with alternative methods.

into the dictionary, the WPD-dictionary is inadequate for signal components that are not aligned with the basis elements. Thus, combing the extended libraries of orthonormal bases with the fast best-basis search algorithms (*e.g.*, the SIWPD and SIAP-LTD), the proposed method facilitates shift-invariant estimators at a manageable computational complexity, which are based on the MDL criterion.

5.7 Summary

In this chapter, we have described a translation-invariant denoising method, which uses the MDL criterion and tree-structured best-basis algorithms. We have defined a collection of signal models based on an extended library of bases, and applied the MDL principle to determine the description length of the noisy observed data. We showed that the optimal model is obtainable through optimizing the expansion-tree associated with the SIWPD or SIAP-LTD. The signal estimator was combined with the modified Wigner distribution and thus generated robust time-frequency representations. Both synthetic and real data examples have shown the usefulness of our algorithms compared to other methods.

Chapter 6

Conclusion

6.1 Summary

We have derived a general approach for achieving shift-invariant best-basis decompositions of signals and time-frequency distributions using libraries of wavelet packets and local trigonometric bases. These decompositions are used for developing MDL-based methods for translation-invariant denoising and robust time-frequency signal analysis.

The strategy for obtaining shift-invariance is based on a three-step procedure. First, we extend the conventional libraries of wavelet packets and local trigonometric bases to include all their shifted versions. Second, the extended libraries are organized in tree structures, where each node represents a subset of basis functions with different time-frequency localization properties. Third, efficient algorithms compare different combinations of nodes that construct orthonormal bases, and choose that combination (basis) leading to the minimal information cost.

The extension of a given library admits into the library translated versions of bases, so that a time-shift of a prescribed signal results in a similar translation of the best-basis elements. Accordingly, the expansion coefficients of a signal represented on its best-basis do

not vary with the changing time-position of the signal. To facilitate the comparison between alternative bases in the extended libraries, we configure the basis elements in a tree structure such that the following two conditions are satisfied: 1) Each node is associated with an orthonormal subset of the basis elements. 2) The subspace connected with a given parent-node (the closure of linear span of the basis elements) is an orthogonal sum of subspaces connected with the corresponding children-nodes. Consequently, whenever the cost function is of an additive nature, the implementation of the best-basis selection procedure is strictly local, *i.e.*, the representation of the signal at each node of the expansion-tree is optimized independently of other nodes at the same resolution level.

In the case of wavelet packets, we have defined an extended library (the SWP library) that is larger than the ordinary WP library by a square power (the size of the SWP library is $O(2.5^N)$, and that of the WP library is $O(1.5^N)$). Still, the tree configuration of the basis elements retains the complexity level of the best-basis search algorithm significantly below the squared WPD complexity (the SIWPD complexity is $O[2^d(L-d+2)N]$; the squared WPD complexity is $O(N^2L^2)$; L is the depth of the expansion-tree and d is the depth of a subtree used at a given parent-node to determine its shift index). Moreover, we showed that the computational complexity may be further reduced at the expense of the attained information cost down to $O(NL)$. The key to controlling the complexity is the built-in flexibility in the choice of d. While the shift-invariance is satisfied for all $1 \leq d \leq L$, larger d values generally yield higher quality best-bases, *i.e.*, bases characterized by a lower and more stable information cost, at the expense of a higher computational complexity. An experiment on real data has shown a significant improvement in the performance when using the SIWPD rather than the WPD. The reduction in the entropy was more than 30% for some signals, and more than 18% on the average. The variations in the entropy across the data set were lessened on the average by approximately 20%.

The SIWPD is implemented by adapting the down-sampling following the high-pass

and low-pass filters in the expansion tree. That is, upon expanding a prescribed node with minimization of the information cost in mind, we test as to whether or not the information cost indeed decreases and whether to choose even or odd down-sampling. The special case where, at any resolution level, only low frequency nodes are further expanded is interpreted as shift-invariant *wavelet* transform.

The additional degree of freedom in the expansion-tree yields relative shifts between parent-nodes and their respective children-nodes (or more precisely, between the relevant basis-functions). The even and odd down-sampling correspond to a zero shift and a $2^{-\ell}$ shift, respectively, where ℓ denotes the resolution level of the parent-node. The generated branches at the expansion-tree are designated by either fine or heavy lines, depending on the adaptive selection of the odd or the even down-sampling. We proved that the SIWPD expansion leads to trees configurations that are independent of the time-origin. Specifically, the best expansion trees for translated signals are identical to within a time-shift (fine and heavy lines at each resolution level may exchange positions).

In the case of local trigonometric bases, we have extended the library of bases to include smooth local trigonometric functions with different parity properties, as well as different time positions and durations. The extended library is constructed into a tree structure whose nodes are associated with intervals, having four optional polarity values. We have shown that adjacent compatible intervals with the same connecting polarity can be merged, thus replacing the bases on the intervals with the basis on their union. In this scheme, each basis in the library corresponds to a compatible partition of a unit-length interval, with certain shift and polarity indices.

The resultant algorithm, the *shift-invariant adapted-polarity local trigonometric decomposition* (SIAP-LTD), provides two degrees of freedom that generate independently shift-invariance and adaptive-polarity foldings. The shift-invariance stems from a relative shift between expansions in distinct resolution levels. At any resolution level ℓ we examine and

select one of two relative shift options — a zero shift or a $2^{-\ell-1}$ shift. The additional degree of freedom, which incorporates an adaptive folding operator into the best decomposition tree, is mainly intended to reduce the information cost and thus improve the time-frequency representation. This improvement is notable for signals that have dominant frequencies within each segment or for signals that possess definite parity properties at the end-points of the segments. Otherwise, the polarities at the finest resolution level can be forced to any fixed values, while still retaining the shift-invariance property. The special cases where the polarities at the finest resolution level are restricted to zeros, respectively ones, correspond to Shift-Invariant Local *Cosine*, respectively *Sine*, Decompositions.

The implementation of SIAP-LTD for a given signal involves an efficient computation of inner products with the basis functions, selection of the best relative shifts between resolution levels and determination of the folding polarities at the finest resolution level. The computation of the expansion coefficients is carried out in two conventional stages. At the first stage, the segments of the signal are preprocessed by folding overlapping parts back into the segments. Then, each segment is transformed by a trigonometric basis which has the appropriate parity properties at the end-points (in the discrete case, the trigonometric transform is DCT-II for even-even parity, DCT-IV for even-odd parity, DST-II for odd-odd parity and DST-IV for odd-even parity). The polarities of the folding operators are determined at the finest resolution level. However, an ill-adapted polarity at a certain point may be eliminated at a coarser resolution level by merging intervals on its both sides. Thus, instead of pursuing the polarities which minimize the total information cost at the finest resolution level, we tolerate polarities which are *locally* adapted to the signal, *i.e.*, minimize the *local* information cost. The entailed computational complexity is significantly reduced, while still retaining the shift-invariance property.

When compared with the conventional WPD and LCD algorithms, the SIWPD and SIAP-LTD are determined to be advantageous in the following respects:

1. Shift-invariance.

2. Lower information cost.

3. Improved time-frequency resolution.

4. More stable information cost across a prescribed data set.

5. The computational complexity is controlled at the expense of the information cost.

These advantages may prove crucial to signal compression, identification or classification applications. Furthermore, the shift-invariant nature of the information cost, renders this quantity a characteristic of the signal for a prescribed wavelet packet library. It should be possible now to quantify the relative efficiency of various libraries (i.e., various scaling function selections) with respect to a given cost function. Such a measure would be rather senseless for shift-variant decompositions.

Another issue investigated in this work, closely related to the problem of shift-invariance, is that of adaptive decompositions of time-frequency distributions and suppression of interference terms associated with bilinear distributions. We have shown that utilizing the SIWPD and SIAP-LTD, various useful properties relevant to time-frequency analysis, including high energy concentration and suppressed interference terms, can be achieved simultaneously with simple operations in the Wigner domain. Instead of smoothing, which broadens the energy distribution of signal components, we proposed best-basis decompositions and cross-term manipulations that are adapted to the local distribution of the signal via a certain time-frequency distance measure.

A prescribed signal is expanded on its best basis using the SIWPD or SIAP-LTD, and subsequently transformed into the Wigner domain. The distribution of the signal is partitioned into two subsets: One subset, representing the auto-terms, includes the auto WD of the basis elements. The other subset, representing the cross-terms, comprises the

cross WD of distinct pairs of basis-functions. It follows that the contribution of a cross term to the desirable time-frequency properties (*e.g.*, time marginal, frequency marginal, energy concentration and instantaneous frequency property) is inversely proportional to the distance between the corresponding basis-functions. Therefore, we modify the WD by restricting the cross terms to *neighboring* pairs of basis-functions, *i.e.*, basis-functions whose time-frequency distance is smaller than a certain distance-threshold D. The auto terms, as well as the cross terms, are restricted to basis elements whose normalized coefficients are larger in magnitude than a certain amplitude-threshold ε, to further reduce the computational complexity. We have shown that the distance and amplitude thresholds can balance the cross-term interference, the useful properties of the distribution, and the computational complexity. A smaller distance-threshold better eliminates the interference terms, but generally renders the energy concentration low. A larger distance-threshold improves the time-frequency resolution at the cost of including more interference terms. When the amplitude-threshold is set to zero and the distance-threshold goes to infinity, the modified Wigner distribution converges to the conventional WD. Appropriate threshold values ($D \approx 2$, $\varepsilon \approx 0.1$) yield enhanced time-frequency decompositions, which achieve high resolution, high concentration and suppressed cross-term interference at a manageable computational complexity.

The distance measure for a pair of basis-functions is defined in the idealized time-frequency plane by weighing their Euclidean distance with their time and frequency uncertainties (the widths and heights of the corresponding tiles). Since these uncertainties are adapted to the signal's local distribution, the thresholding of cross-terms is also adapted to the local distribution of the signal, which dispenses with the need for local adjustments of the associated distance-threshold. We compared alternative selections of libraries, showing that as long as the localization properties of the basis elements are adapted to the signal, the interference terms between distinct components of a given signal, and even within

nonlinear components, can be efficiently eliminated. Furthermore, the visual quality of the resultant modified Wigner distribution conforms with the entropy of the best basis expansion (improved distributions ensued from lower entropies). This property is indeed valuable as it facilitates a quantitative comparison between energy distributions.

Another advantage of the modified Wigner distribution is the potential for resolving multicomponent signals. By defining equivalence classes in the time-frequency plane, we showed that the components of a multicomponent signal can be determined as partial sums of basis-functions. The components are well delineated in the time-frequency plane, and can also be recovered from the energy distribution up to a constant phase factor and up to the contribution of insignificant basis functions.

The final problem we addressed is that of translation-invariant denoising, based on the Minimum Description Length criterion. We have defined collections of signal models based on extended libraries of wavelet packet and local trigonometric bases, and applied the MDL principle to derive approximate additive cost functions. The description length of the noisy observed data was then minimized by utilizing the SIWPD and SIAP-LTD, thus optimizing the expansion-trees associated with the best-basis algorithms, and thresholding the resulting coefficients. We have combined the signal estimator with the *modified* Wigner distribution, and introduced robust time-frequency representations. The proposed method was compared to alternative existing methods, and its superiority was demonstrated by synthetic and real data examples.

6.2 Future Research

The methodology we have established in this thesis opens a number of interesting topics for future study:

1. Relative translations of signal components.

   In this work, as well as in other approaches, the shift-invariance is defined as insensitivity to time-domain translations of the analyzed signal. A more profound issue, which still deserves a careful consideration, concerns independent translations of signal components. This problem was too difficult to cope with, since signal components are generally not uniquely defined and the conventional shift-invariance was still under research. Resolving multicomponent signals, as proposed in Chapter 4, and applying shift-invariant best-basis decompositions on individual components, is possibly an efficient way to handle this problem.

2. Statistical analysis of the MDL-based translation-invariant denoising.

   The translation-invariant denoising method, proposed in Chapter 5, selects the best basis which minimizes the description length of the noisy observed data and then applies hard-thresholding to the resulting coefficients. The criterion for the best basis selection is a random variable. Accordingly, statistical analysis, numerical evaluation and statistical tests for improving the robustness of the search for the best basis require further study.

3. Designing "mother-wavelets" for particular applications and signal classes.

   As demonstrated, the shift-invariant nature of the information cost and its stability across a prescribed data set, renders this quantity a characteristic of the signal class for a prescribed wavelet packet library. Research on the relative efficiency of various libraries (i.e., various mother-wavelet selections) with respect to given applications and cost functions may provide us with efficient methods for designing mother-wavelets.

4. Merging libraries of wavelet packet and local trigonometric bases.

   Restricting the depth of an expansion tree, the computational complexity of the associated best-basis search algorithm is significantly reduced. However, the representation of tones, respectively transients, with wavelet packets, respectively with local trigonometric

bases, becomes inefficient in the sense that it spreads over more basis functions. In that case, signals containing transitory, as well as tonal, features should be expanded with basis functions from both libraries by iteratively "pealing off" the features. First extract the major features (either tones or transients) by some shift-invariant best-basis decomposition (SIWPD or SIAP-LTD), then get the secondary features by applying the other algorithm to the residual signal, and repeat the process (this procedure has a close relationship to the "matching pursuit" algorithm [96])

5. Classification, identification and local discriminant bases.

   Saito and Coifman [129] have described a best-basis method for signal classification problems, which is based on the conventional WPD and LTD. Their algorithm uses a discrimination measure, such as relative entropy, as a basis selection criterion, picks out a few most important basis functions that serve as feature extractors, and then utilizes an ordinary classifier. The shift sensitivity of the expansion coefficients is a serious drawback of such classification method. Saito [128] proposed to somewhat reduce the sensitivity by creating from each training signal a few circularly-shifted versions. This increases the computational complexity, while the feature extractors are not optimized over translations and the expansion coefficients of a given signal still depend on its time shift. The proposed shift-invariant decompositions may extend the ideas proposed in [128] and make them robust.

Appendix A

Proofs

A.1 Proof of Proposition 2.2

Let $f, g \in V_j$ be identical to within a time-shift, and let A_f and A_g denote their respective best bases. Hence there exists $q \in \mathbb{Z}$ such that

$$g(x) = f(x - q2^{-j})\,. \tag{A.1}$$

We show by induction that

$$B_{\ell,n,m} \subset A_f \tag{A.2}$$

implies

$$B_{\ell,n,\tilde{m}} \subset A_g, \qquad \tilde{m} = (m+q) \bmod (2^{-\ell}) \tag{A.3}$$

for all $m, n \in \mathbb{Z}_+$ and $\ell \in \mathbb{Z}_-$.

First we validate the claim for the coarsest resolution level $\ell = -L$. Suppose that

$$B_{-L,n,m_0} \subset A_f, \qquad 0 \le n < 2^L\,. \tag{A.4}$$

That is, $m = m_0$ minimizes the information cost for representing f in the subspace $U_{-L,n,m}$, *i.e.*

$$Arg \min_{0 \le m < 2^L} \{\mathcal{M}(B_{-L,n,m} f)\} = m_0\,. \tag{A.5}$$

It stems from (A.1) that

$$\langle g(x), \psi_n[2^\ell(2^j x - m) - k]\rangle = \langle f(x), \psi_n[2^\ell(2^j x - m + q) - k]\rangle, \quad l, n, j, k, m \in \mathbb{Z} \tag{A.6}$$

and accordingly

$$\mathcal{M}(B_{\ell,n,m}g) = \mathcal{M}(B_{\ell,n,m-q}f)\,. \tag{A.7}$$

Hence the information cost for representing g in the subspace $U_{-L,n,m}$ is minimized for $m = m_0 + q$, *i.e.*

$$Arg \min_{0\le m<2^L}\{\mathcal{M}(B_{-L,n,m}g)\} = (m_0 + q) \bmod (2^L) \tag{A.8}$$

and

$$B_{-L,n,\tilde{m}_0} \subset A_g, \qquad \tilde{m}_0 = (m_0 + q) \bmod (2^L)\,. \tag{A.9}$$

Now, suppose that the claim is true for all levels coarser than ℓ_0 ($\ell_0 > -L$), and assume that (A.2) exists for $\ell = \ell_0$. Then by (2.11)

$$\mathcal{M}(B_{\ell_0,n,m}f) \le \mathcal{M}(A_{\ell_0-1,2n,m_c}f) + \mathcal{M}(A_{\ell_0-1,2n+1,m_c}f), \quad m_c \in \{m, m + 2^{-\ell_0}\}\,. \tag{A.10}$$

The inductive hypothesis together with equation (A.7) lead to

$$\mathcal{M}(A_{\ell_0-1,2n+\epsilon,m_c}f) = \mathcal{M}(A_{\ell_0-1,2n+\epsilon,m_c+q}g), \quad \epsilon \in \{0, 1\}\,. \tag{A.11}$$

Consequently,

$$\mathcal{M}(B_{\ell_0,n,m+q}g) \le \mathcal{M}(A_{\ell_0-1,2n,m_c+q}g) + \mathcal{M}(A_{\ell_0-1,2n+1,m_c+q}g), \quad m_c \in \{m, m + 2^{-\ell_0}\} \tag{A.12}$$

and again by (2.11) we have

$$B_{\ell_0,n,\tilde{m}} \subset A_g, \qquad \tilde{m} = (m + q) \bmod (2^{-\ell_0}) \tag{A.13}$$

proving as well the validity of the claim for ℓ_0. Thus, A_f and A_g are *identical to within a time-shift.*

A.2 Proof of Proposition 3.3

Let $f, g \in \hat{L}_2[0,1]$ be identical to within a resolution J time-shift, and let A_f and A_g denote their respective best bases. Then there exists an integer $0 \le q < 2^J$ such that

$$g(t) = f(t - q2^{-J})\,. \tag{A.14}$$

We show by induction that

$$B^{\rho_0,\rho_1}_{[\alpha,\beta)} \subset A_f \tag{A.15}$$

implies

$$B^{\rho_0,\rho_1}_{[\alpha+q2^{-J},\beta+q2^{-J})} \subset A_g \tag{A.16}$$

for all $I = [\alpha, \beta) \in \mathcal{I}$ and $\rho_0, \rho_1 \in \{0, 1\}$. Or equivalently,

$$B^{\rho_0,\rho_1}_{\ell,n,m} \subset A_f \tag{A.17}$$

implies

$$B^{\rho_0,\rho_1}_{\ell,\tilde{n},\tilde{m}} \subset A_g, \qquad \tilde{n} \equiv n + (m+q) \text{ div } 2^{J-\ell},\ \tilde{m} \equiv (m+q) \bmod\ 2^{J-\ell} \tag{A.18}$$

for all $0 \le \ell \le L$, $0 \le n < 2^\ell$, $0 \le m < 2^{J-\ell}$ and $\rho_0, \rho_1 \in \{0, 1\}$.

First we validate the claim for the finest resolution level $\ell = L$. Suppose that $B^{\rho_0,\rho_1}_{L,n_0,m_L} \subset A_f$. Then the information cost for representing f at the finest resolution level is minimized for shift m_L and polarity P_L where $p_L(n_0) = \rho_0,\ p_L(n_0+1) = \rho_1$. That is,

$$(m_L, P_L) = \arg \min_{\substack{0 \le m < 2^{J-L} \\ 0 \le P < 2^{2^L}}} \left\{ \sum_{n=0}^{2^L-1} \mathcal{M}(B^{p(n),p(n+1)}_{L,n,m} f) \right\}, \tag{A.19}$$

$$P_L = [p_L(2^L - 1), ..., p_L(1), p_L(0)]_2\,, \quad p_L(2^L) = p_L(0)\,, \tag{A.20}$$

$$p_L(n_0) = \rho_0\,, \quad p_L(n_0+1) = \rho_1\,. \tag{A.21}$$

It stems from (A.14) and definition of $\psi^{\rho_2,\rho_3}_{\ell,n,m,k}$ that

$$\left\langle f, \psi^{\rho_2,\rho_3}_{\ell,n,m,k} \right\rangle = \left\langle g, \psi^{\rho_2,\rho_3}_{\ell,\tilde{n},\tilde{m},k} \right\rangle , \qquad \tilde{n} = n+(m+q) \text{ div } 2^{J-\ell}, \tilde{m} = (m+q) \text{ mod } 2^{J-\ell} \quad \text{(A.22)}$$

and accordingly,

$$\mathcal{M}(B^{\rho_2,\rho_3}_{\ell,n,m} f) = \mathcal{M}(B^{\rho_2,\rho_3}_{\ell,\tilde{n},\tilde{m}} g) \quad \text{(A.23)}$$

for all $0 \le \ell \le L$, $0 \le n < 2^\ell$, $0 \le m < 2^{J-\ell}$, $k \in \mathbb{Z}_+$ and $\rho_2, \rho_3 \in \{0,1\}$. Hence the information cost for representing g at the finest resolution level is minimized for shift $\tilde{m}_L = (m_L + q) \text{ mod } 2^{J-L}$ and polarity $\tilde{P}_L$ where $\tilde{p}_L(\tilde{n}) = p_L(n)$ and $\tilde{n} = n + (m_L + q) \text{ div } 2^{J-L}$. That is,

$$(\tilde{m}_L, \tilde{P}_L) = \arg \min_{\substack{0 \le m < 2^{J-L} \\ 0 \le P < 2^{2^L}}} \left\{ \sum_{n=0}^{2^L-1} \mathcal{M}(B^{p(n),p(n+1)}_{L,n,m} g) \right\} , \quad \text{(A.24)}$$

$$\tilde{P}_L = [\tilde{p}_L(2^L - 1), ..., \tilde{p}_L(1), \tilde{p}_L(0)]_2 , \quad \tilde{p}_L(2^L) = \tilde{p}_L(0) , \quad \text{(A.25)}$$

$$\tilde{p}_L(\tilde{n}) = p_L(n) . \quad \text{(A.26)}$$

Consequently,

$$B^{\rho_0,\rho_1}_{L,\tilde{n}_0,\tilde{m}_L} = B^{\tilde{p}_L(\tilde{n}_0),\tilde{p}_L(\tilde{n}_0+1)}_{L,\tilde{n}_0,\tilde{m}_L} \subset A_g . \quad \text{(A.27)}$$

Now, suppose that the claim is true for all levels finer than l ($l < L$), and assume that (A.17) exists for $\ell = l$. Then by (3.28) and (3.30)

$$\mathcal{M}(B^{\rho_0,\rho_1}_{l,n,m_l} f) \le \mathcal{M}(A^{\rho_0,\rho_2}_{l+1,2n+\gamma,m_{l+1}} f) + \mathcal{M}(A^{\rho_2,\rho_1}_{l+1,2n+1+\gamma,m_{l+1}} f) \quad \text{(A.28)}$$

where $m_{l+1} \equiv m_l - \gamma \cdot 2^{J-l-1}$, $\rho_2 \equiv p_{l+1}(2n+1+\gamma)$ and $\gamma \in \{0,1\}$ such that $m_{l+1} \in [0, 2^{J-l-1})$. Notice that γ is an indication of a relative shift between the resolution levels l and $l+1$.

The inductive hypothesis together with equation (A.23) lead to the identities

$$\mathcal{M}(A^{\rho_0,\rho_2}_{l+1,2n+\xi+\gamma,m_{l+1}}f) = \mathcal{M}(A^{\rho_0,\rho_2}_{l+1,2\tilde{n}+\xi+\tilde{\gamma},\tilde{m}_{l+1}}g), \quad \xi \in \{0,1\}\,, \tag{A.29}$$

$$\mathcal{M}(A^{\rho_2,\rho_1}_{l+1,2n+\xi+\gamma,m_{l+1}}f) = \mathcal{M}(A^{\rho_2,\rho_1}_{l+1,2\tilde{n}+\xi+\tilde{\gamma},\tilde{m}_{l+1}}g), \quad \xi \in \{0,1\}\,, \tag{A.30}$$

where $\tilde{m}_{l+1} \equiv \tilde{m}_l - \tilde{\gamma}\cdot 2^{J-l-1}$ and $\tilde{\gamma} \in \{0,1\}$ such that $\tilde{m}_{l+1} \in [0, 2^{J-l-1})$. Consequently, using again (A.23) we have

$$\mathcal{M}(B^{\rho_0,\rho_1}_{l,\tilde{n},\tilde{m}_l}g) \le \mathcal{M}(A^{\rho_0,\rho_2}_{l+1,2\tilde{n}+\tilde{\gamma},\tilde{m}_{l+1}}g) + \mathcal{M}(A^{\rho_2,\rho_1}_{l+1,2\tilde{n}+1+\tilde{\gamma},\tilde{m}_{l+1}}g)\,, \tag{A.31}$$

$$\tilde{p}_l(\tilde{n}) = p_l(n)\,, \quad \tilde{n} = n + (m+q) \text{ div } 2^{J-l}\,. \tag{A.32}$$

So by (3.28) and (3.30) we conclude that

$$B^{\rho_0,\rho_1}_{l,\tilde{n},\tilde{m}_l} = B^{\tilde{p}_l(\tilde{n}),\tilde{p}_l(\tilde{n}+1)}_{l,\tilde{n},\tilde{m}_l} \subset A_g\,, \tag{A.33}$$

proving as well the validity of the claim for $\ell = l$. Thus, A_f and A_g are identical to within a $q2^{-J}$ time-shift.

Bibliography

[1] G. Aharoni, A. Averbuch, R. Coifman and M. Israeli, "Local cosine transform — A method for the reduction of the blocking effect in JPEG", J. Math. Imag. Vision, Vol. 3, 1993, pp. 7–38.

[2] P. Auscher, G. Weiss and M. V. Wickerhauser, "Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets", in: C. K. Chui, ed., *Wavelets — A Tutorial in Theory and Applications*, Academic Press, Inc., San Diego, 1992, pp. 237–256.

[3] F. Bao and N. Erdol, "The optimal wavelet transform and translation invariance", Proc. of the 19th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-94, Adelaide, Australia, April 19–22 1994, pp. III.13–III.16.

[4] R. G. Baraniuk and D. J. Jones, "A signal-dependent time-frequency representation: Fast algorithm for optimal kernel design", IEEE Trans. on Signal Processing, Vol. 42, No. 1, Jan. 1994, pp. 134–146.

[5] J. Benford and J. Swegle, *High Power Microwaves*, Artech House, Norwood, 1992.

[6] S. A. Benno and J. M. F. Moura, "Nearly shiftable scaling functions", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1097–1100.

[7] J. Berger, R. R. Coifman and M. Goldberg, "Removing noise from music using local trigonometric bases and wavelet packets", J. Audio Eng. Soc., Vol. 42, Dec. 1994, pp. 808–818.

[8] Z. Berman and J. S. Baras, "Properties of the multiscale maxima and zero-crossings representations", IEEE Trans. on Signal Processing, Vol. 41, No. 12, Dec. 1993, pp. 3216–3231.

[9] G. Beylkin, "On the representation of operators in bases of compactly supported wavelets", SIAM J. Numer. Anal., Vol. 6, No. 6, Dec. 1992, pp. 1716–1740.

[10] P. J. Burt, "Fast Filter transforms for image processing", Comput. Graphics and Image Proc., Vol. 16, 1981, pp. 20–51.

[11] A. E. Cetin and R. Ansari, "Signal recovery from wavelet transform maxima", IEEE Trans. on Signal Processing, Vol. 42, No. 1, Jan. 1994, pp. 194–196.

[12] T. Chang and C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform", IEEE Trans. Image Processing, Vol. 2, No. 4, Oct. 1993, pp. 429–441.

[13] W. Chen, C. H. Smith and S. Fralic, "A fast computational algorithm for the discrete cosine transform", IEEE Trans. on Communication, Vol. 25, Sept. 1977, pp. 1004–1009.

[14] S. Chen and D. L. Donoho, "Atomic decomposition by basis pursuit", Technical Report, Dept.of Statistics, Stanford Univ., Feb. 1996 (http://playfair.stanford.edu /reports/chen_s).

[15] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multi-component signals using exponential kernels", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 37, No. 6, June 1989, pp. 862–871.

[16] C. K. Chui, *An Introduction to Wavelets*, Academic Press, Inc., San Diego, 1992.

[17] C. K. Chui, ed., *Wavelets — A Tutorial in Theory and Applications*, Academic Press, Inc., San Diego, 1992.

[18] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution — a tool for time-frequency signal analysis. Part I. Continuous-time signals", Philips J. Res., Vol. 35, 1980, pp. 217–250.

[19] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution — a tool for time-frequency signal analysis. Part II. Discrete-time signals", Philips J. Res., Vol. 35, 1980, pp. 276–300.

[20] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution — a tool for time-frequency signal analysis. Part III. Relations with other time-frequency signal transformations", Philips J. Res., Vol. 35, 1980, pp. 372–389.

[21] L. Cohen, "Generalized phase-space distribution functions", J. Math. Phys., Vol. 7, 1966, pp. 781–786.

[22] L. Cohen and T. Posch, "Positive time-frequency distribution functions", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 33, 1985, pp. 31–38.

[23] L. Cohen, "Time-frequency distributions — a review", Proc. IEEE, Vol. 77, No. 7, July 1989, pp. 941–981.

[24] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall Inc., 1995.

[25] A. Cohen and E. Sere, "Time-frequency localization with non stationary wavelet packets", Cahiers de Mathematiques de la Decision, No. 9401, Jan. 1994.

[26] A. Cohen and J. Kovačević, "Wavelets: The mathematical background", Proc. IEEE, Vol. 84, No. 4, 1996, pp. 514–522.

[27] I. Cohen, S. Raz and D. Malah, "Shift invariant wavelet packet bases", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1081–1084.

[28] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant wavelet packet decomposition and representation", Signal Processing, Vol. 57, No. 3, Mar. 1997, pp. 251–270. (also EE PUB No. 953, Technion - Israel Institute of Technology, Haifa, Israel, Jan. 1995).

[29] I. Cohen, S. Raz and D. Malah, "Shift-invariant adaptive local trigonometric decomposition", Proc. of the 4th European Conference on Speech, Communication and Technology, EUROSPEECH'95, Madrid, Spain, 18–21 Sep. 1995, pp. 247–250.

[30] I. Cohen, S. Raz and D. Malah, "Shift-invariant adaptive representations in libraries of bases", Technical Report, CC PUB No. 128, Technion - Israel Institute of Technology, Haifa, Israel, Dec. 1995.

[31] I. Cohen, S. Raz, D. Malah and I. Schnitzer, "Best-basis algorithm for orthonormal shift-invariant trigonometric decomposition", Proc. of the 7th IEEE Digital Signal Processing Workshop, DSPWS'96, Loen, Norway, 1–4 Sep. 1996, 1–4 Sep. 1996, pp. 401–404.

[32] I. Cohen, S. Raz and D. Malah, "Orthonormal shift-invariant adaptive local trigonometric decomposition", Signal Processing, Vol. 57, No. 1, Feb. 1997, pp. 43–64.

[33] I. Cohen, S. Raz and D. Malah, "Eliminating interference terms in the Wigner distribution using extended libraries of bases", Proc. of the 22th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-97, Munich, Germany, 20–24 Apr. 1997, pp. 2133–2136.

[34] I. Cohen, S. Raz and D. Malah, "Adaptive suppression of Wigner interference-terms using shift-invariant wavelet packet decompositions", Technical Report, CC PUB

No. 245, Technion - Israel Institute of Technology, Haifa, Israel, June 1998 (submitted to Signal Processing).

[35] I. Cohen, S. Raz and D. Malah, "Translation-invariant denoising using the minimum description length criterion", Technical Report, CC PUB No. 246, Technion - Israel Institute of Technology, Haifa, Israel, June 1998 (submitted to Signal Processing).

[36] I. Cohen, S. Raz and D. Malah, "Time-frequency analysis and noise suppression with shift-invariant wavelet packets", Proc. of the 11th Int. Conf. on High-Power Electromagnetics, EUROEM'98, Tel-Aviv, Israel, 14–19 June 1998.

[37] I. Cohen, S. Raz and D. Malah, "Adaptive time-frequency distributions via the shift-invariant wavelet packet decomposition", Proc. of the 4th IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Pittsburgh, Pennsylvania, 6–9 Oct. 1998.

[38] I. Cohen, S. Raz and D. Malah, "MDL-based translation-invariant denoising and robust time-frequency representations", Proc. of the 4th IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Pittsburgh, Pennsylvania, 6–9 Oct. 1998.

[39] R. R. Coifman and Y. Meyer, "Orthonormal wave packet bases", Dept. of Mathematics, Yale Univ., New Haven, Aug. 1989 (In: ftp pascal.math.yale.edu /pub/wavelets/wavepkt.tex).

[40] R.R. Coifman, Y. Meyer, S. Quake and M.V. Wickerhauser, "Signal processing and compression with wave packets", Proc. Conf. on Wavelets, Marseilles, Spring 1989.

[41] R. R. Coifman and M. V. Wickerhauser, "Best-adapted wavelet packet bases", Yale Univ., New Haven, Feb. 1990 (In: ftp pascal.math.yale.edu /pub/wavelets/ bestbase.tex)

[42] R. R. Coifman and Y. Meyer, "Remarques sur l'analyse de Fourier à fenêtre", Comptes Rendus de l'Académie des Sciences, Vol. 312, pp. 259–261, 1991.

[43] R. R. Coifman, Y. Meyer and M. V. Wickerhauser, "Wavelet analysis and signal processing", in: M. B. Ruskai et al., ed., *Wavelets and Their Applications*, Jones and Bartlett, Boston, 1992, pp. 153–178.

[44] R. R. Coifman, Y. Meyer and M. V. Wickerhauser, "Size properties of wavelet packets", in: M. B. Ruskai et al., ed., *Wavelets and Their Applications*, Jones and Bartlett, Boston, 1992, pp. 453–470.

[45] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection", IEEE Trans. Inform. Theory, Vol. 38, No. 2, Mar. 1992, pp. 713–718.

[46] R. R. Coifman, Y. Meyer, S. Quake and M. V. Wickerhauser, "Signal processing and compression with wavelet packets", in: Y. Meyer and S. Roques, ed., *Progress in Wavelet Analysis and Applications*, Proc. Int. Conf. "Wavelet and Applications", Toulouse, France, 8–13 June 1992, pp. 77–93.

[47] R. R. Coifman and F. Majid, "Adapted waveform analysis and denoising", in: Y. Meyer and S. Roques, eds., *Progress in Wavelet Analysis and Applications*, Editions Frontieres, France, 1993, pp. 63–76.

[48] R. R. Coifman and M. V. Wickerhauser, "Adapted waveform analysis as a tool for modeling, feature extraction, and denoising", Optical Engineering, Vol. 33, No. 7, July 1994, pp. 2170–2174.

[49] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 125–150.

[50] R. N. Czerwinski and D. J. Jones, "Adaptive cone-kernel time-frequency analysis", IEEE Trans. on Signal Processing, Vol. 43, No. 7, July 1995, pp. 1715–1719.

[51] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Commun. Pure Appl. Math., Vol. 41, 1988, pp. 909-996.

[52] I. Daubechies, "Time-frequency localization operators: a geometric phase space approach", IEEE Trans. Inform. Theory, Vol. 34, 1988, pp. 605-612.

[53] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM Press,Philadelphia, Pennsylvania, 1992

[54] I. Daubechies, "Orthonormal bases of compactly supported wavelets, II. Variations on a theme", SIAM J. Math. Anal., Vol. 24, No. 2, 1993, pp. 499–519.

[55] G. Davis, S. Mallat and Z. Zhang, "Adaptive time-frequency decompositions", Optical Engineering, Vol. 33, No. 7, July 1994, pp. 2183–2191.

[56] P. Delsarte, B. Macq and D. T.M. Slock, "Signal adapted multiresolution transform for image coding", IEEE Trans. Inform. Theory, Vol. 38, No. 2, 1992, pp. 897–904.

[57] S. Del Marco, J. Weiss and K. Jagler, "Wavepacket-based transient signal detector using a translation invariant wavelet transform", Proc. SPIE, Vol. 2242, 1994, pp. 792–802.

[58] S. Del Marco and J. Weiss, "M-band wavepacket-based transient signal detector using a translation-invariant wavelet transform", Optical Engineering, Vol. 33, No. 7, July 1994, pp. 2175–2182.

[59] S. Del Marco P. Heller and J. Weiss, "An M-band 2-dimensional translation-invariant wavelet transform and applications", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1077–1080.

[60] D. N. Dononho, "Nonlinear wavelet methods for recovery of signals, images and densities from noisy and incomplete data", in: I. Daubechies, ed., *Different Perspectives on Wavelets*, Providence, RI:American Mathematical Society, 1993, pp. 173–205.

[61] D. L. Donoho and I. M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases", Comptes Rendus Acad. Sci., Ser. I, Vol. 319, 1994, pp. 1317–1322.

[62] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", Biometrica, Vol. 81, 1994, pp. 425–455.

[63] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation", Applied and Computational Harmonic Analysis, Vol. 1, 1994, pp. 100–115.

[64] D. L. Donoho, "De-noising by soft thresholding", IEEE Trans. Inform. Theory, Vol. 41, May 1995, pp. 613–627.

[65] X. Fang and E. Séré, "Adapted multiple folding local trigonometric transforms and wavelet packets", Applied and Computational Harmonic Analysis, Vol. 1, 1994, pp. 169–179.

[66] R. A. Haddad, A. N. Akansu and A. Benyassine, "Time-frequency localization in transforms, subband, and wavelets: a critical review", Optical Engineering, Vol. 32, No. 7, Jul. 1993, pp. 1411–1429.

[67] C. Herley, J. Kovačević, K. Ramchandran and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms", IEEE Trans. on Signal Processing, Vol. 41, No. 12, Dec. 1993, pp. 3341–3359.

[68] C. Herley, Z. Xiong, K. Ramchandran and M. T. Orchard, "An efficient algorithm to find a jointly optimal time-frequency segmentation using time-varying filter banks",

Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1516–1519.

[69] N. Hess-Nielsen, "Control of frequency spreading of wavelet packets", Applied and Computational Harmonic Analysis, Vol. 1, 1994, pp. 157–168.

[70] N. Hess-Nielsen and M. V. Wickerhauser, "Wavelets and time-frequency analysis", Proc. IEEE, Vol. 84, No. 4, 1996, pp. 523–540.

[71] F. Hlawatsch, "Interference terms in the Wigner distribution", Proc. Int. Conf. on Digital Signal Processing, Florence, Italy, Sept. 5–8,1984, pp. 363–367.

[72] F. Hlawatsch and P. Flandrin, "The interference structure of the Wigner distribution and related time-frequency signal representations", in: W. Mecklenbrāuker, ed., *The Wigner Distribution - Theory and Applications in Signal Processing*, North Holland, Elsevier Science, 1992.

[73] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations", IEEE SP Magazine, Apr. 1992, pp. 21–67.

[74] R. Hummel and R. Moniot, "Reconstructions from zero crossings in scale space", IEEE Trans. Acoust. Speech and Signal Processing, Vol. 37, No. 12, Dec. 1989, pp. 2111–2130.

[75] A. J. E. M. Janssen and T. A. C. M. Claasen, "On positivity of time-frequency distributions", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 33, No. 4, Aug. 1985, pp. 1029–1032.

[76] B. Jawerth, Y. Liu and W. Sweldens, "Signal compression with smooth local trigonometric bases", Optical Eng., Vol. 33, No. 7, July 1994, pp. 2125–2135.

[77] B. Jawerth and W. Sweldens, "Biorthogonal smooth local trigonometric bases", Research Report 1994:05, Univ. of South Carolina, 1994 (To appear in J. of Fourier Analysis and Applications)

[78] J. Jeong and W. J. Williams, "Kernel design for reduced interference distributions", IEEE Trans. on Signal Processing, Vol. 40, No. 2, Feb. 1992, pp. 402–412.

[79] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 38, No. 12, Dec. 1990, pp. 2127–2135.

[80] D. L. Jones and R. G. Baraniuk, "A simple scheme for adapting time-frequency representations", IEEE Trans. on Signal Processing, Vol. 42, No. 12, Dec. 1994, pp. 3530–3535.

[81] S. Kadambe R. S. Orr and M. J. Lyall, "Cross-term deleted wigner representation (CDWR) based signal detection methodologies", Proc. of the 21th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-96, Atlanta, Georgia, 7–10 May 1996, Vol. 5, pp. 2583–2586.

[82] S. Kadambe and R. S. Orr, "Instantaneous frequency estimation using the cross-term deleted wigner representation (CDWR)", Proc. of the 3rd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18–21 June 1996, pp. 289–292.

[83] H. Krim, J.-C. Pesquet and A. S. Willsky, "Robust multiscale representation of processes and optimal signal reconstruction", Proc. of the 2nd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Philadelphia, PA, 25–28 Oct. 1994, pp. 1–4.

[84] H. Krim, S. Mallat, D. Donoho and A. S. Willsky, "Best basis algorithm for signal enhancement", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1561–1564.

[85] H. Krim, and J.-C. Pesquet, "On the statistics of best bases criteria", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 193–207.

[86] R. Kronland-Martinet, J. Morlet and A. Grossman, "Analysis of sound patterns through wavelet transforms", Int. J. Patt. Rec. Art. Intell., Vol. 1, No. 2, 1987, pp. 273–301.

[87] A. Laine and J. Fan, "Texture classification by wavelet packet signatures", IEEE Trans. PAMI, Vol. 15, No. 11, Nov. 1993, pp. 1186–1191.

[88] J. Liang and T. W. Parks, "A translation invariant wavelet compression algorithm and its applications", Proc. Conf. on Information Sciences and Systems, Princeton, NJ, pp. 1097–1102, March 1994.

[89] J. Liang and T. W. Parks, "A two-dimensional translation invariant wavelet representation and its applications", Proc. Int. Conf. on Image Processing, Austin, TX, 13–16 Nov. 1994, pp. 66–70.

[90] J. Liang and T. W. Parks, "A translation invariant wavelet representation algorithm with applications", IEEE Trans. on Signal Processing, Vo. 44, No. 2, Feb. 1996, pp. 225–232.

[91] W. R. Madych, "Some elementary properties of multiresolution analyses of $L^2(\mathbf{R}^n)$, in: C. K. Chui, ed., *Wavelets — A Tutorial in Theory and Applications*, Academic Press, Inc., San Diego, 1992, pp. 259–294.

[92] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet decomposition", IEEE Trans. PAMI, Vol. 11, No. 7, July 1989, pp. 674–693.

[93] S. Mallat, "Zero crossings of a wavelet transform", IEEE Trans. Inf. Theory, Vol. 37, No. 4, July 1991, pp. 1019–1033.

[94] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets", IEEE Trans. Inf. Theory, Vol. 38, No. 2, Mar. 1992, pp. 617–643.

[95] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges", IEEE Trans. PAMI, Vol. 14, No. 7, July 1992, pp. 710–732.

[96] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries", IEEE Trans. on Signal Processing, Vol. 41, No. 12, Dec. 1993, pp. 3397–3415.

[97] H. S. Malvar, "Fast computations of the discrete cosine transform through fast Hartley transform", Electron Let., Vol. 22, March 1986, pp. 352–353.

[98] H. S. Malvar and D. H. Staelin, "The LOT: Transform coding without blocking effects", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 37, April 1989, pp. 553–559.

[99] H. S. Malvar, "Lapped transforms for efficient transform/subband coding", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 38, No. 6, June 1990, pp. 969–978.

[100] H. S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Inc., 1992.

[101] D. Marr and E. Hildreth, "Theory of edge detection", Proc. Roy. Soc. London, Vol. 207, 1980, pp. 187–217.

[102] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.

[103] P. Moulin, "A new look at signal-adapted QMF bank design", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1312–1315.

[104] P. Moulin, "Signal estimation using adapted tree-structured bases and the MDL principle", Proc. of the 3rd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18–21 June 1996, pp. 141–143.

[105] P. Moulin, K. Ramchandran and V. Pavlovic, "Transform image coding based on joint adaptation of filter banks and tree structures", Proc. Int. Conf. on Image Processing, ICIP'96, Lausanne, Switzerland, Sep. 1996.

[106] M. Mugur-Schachter, "A study of Wigner's theorem on joint probabilities", Found. Phys., Vol. 9, 1979, pp. 389–404.

[107] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 281–300.

[108] A. H. Nuttall, "Wigner distribution function: Relation to short-term spectral estimation, smoothing, and performance in noise", Naval Underwater Systems Center, Technical Report, No. 8225, 1988.

[109] R. S. Orr, J. M. Morris and S. Qian, "Use of the Gabor representation for Wigner distribution crossterm suppression", Proc. of the 17th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-92, San Francisco, CA, 23–26 March 1992, Vol. 5, pp. 29–31.

[110] M. Pasquier, P. Gonçalvès and R. Baraniuk, "Hybrid linear/bilinear time-scale analysis", Proc. of the 3rd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18–21 June 1996, pp. 513–516.

[111] J.-C. Pesquet, H. Krim, H. Carfantan and J. G. Proakis, "Estimation of noisy signals using time-invariant wavelet packets", Proc. of Asilomar Conference, Monterey, CA, USA, Vol. 1, Nov. 1993, pp. 31–34.

[112] J.-C. Pesquet, H. Krim and H. Carfantan, "Time-invariant orthonormal wavelet representations", IEEE Trans. on Signal Processing, Vol. 44, No. 8, Aug. 1996, pp. 1964–1996.

[113] F. Peyrin and R. Prost, "A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distribution", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. ASSP-36, No. 4, 1988, pp. 1681–1684.

[114] S. Qian and J. M. Morris, "Wigner distribution decomposition and cross-terms deleted representation", Signal Processing, Vol. 27, No. 2, May 1992, pp. 125–144.

[115] S. Qian and D. Chen, "Decomposition of the Wigner-Ville distribution and time-frequency distribution series", IEEE Trans. on Signal Processing, Vol. 42, No. 10, Oct. 1994, pp. 2836–2842.

[116] S. Qian and D. Chen, *Joint Time-Frequency Analysis: Methods and Applications*, Prentice-Hall Inc., 1996.

[117] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense", IEEE Trans. Image Processing, Vol. 2, No. 2, Apr. 1993, pp. 160–175.

[118] K. Ramchandran, M. Vetterli and C. Herley, "Wavelets, subband coding, and best bases", Proc. IEEE, Vol. 84, No. 4, 1996, pp. 541–560.

[119] K. R. Rao and P. Yip, *Discrete Cosine Transform*, Academic Press Inc., 1990.

[120] S. Raz, "Synthesis of signals from Wigner distribution: Representation on biorthogonal bases", Signal Processing, Vol. 20, No. 4, Aug. 90, pp. 303–314.

[121] O. Rioul and M. Vetterli, "Wavelets and signal processing", IEEE Signal Processing Magazine, Vol. 8, No. 4, Oct. 1991, pp. 14–38.

[122] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms", IEEE Trans. Inf. Theory, Vol. 38, No. 2, Mar. 1992, pp. 569–586.

[123] J. Rissanen, "Modeling by shortest data description", Automatica, Vol. 14, 1978, pp. 465–471.

[124] J. Rissanen, "Universal coding, information, prediction, and estimation", IEEE Trans. Inform. Theory, Vol. 30, No. 4, July 1984, pp. 629–636.

[125] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.

[126] M. B. Ruskai et al., ed., *Wavelets and Their Applications*, Jones and Bartlett, Boston, 1992.

[127] N. Saito and G. Beylkin, "Multiresolution representations using the auto-correlation functions of compactly supported wavelets", IEEE Trans. on Signal Processing, Vol. 41, No. 12, Dec. 1993, pp. 3584–3590.

[128] N. Saito, *Local Feature Extraction and Its Applications Using a Library of Bases*, Ph.D. Dissertation, Yale Univ., New Haven, Dec. 1994.

[129] N. Saito and R. R. Coifman, "Local discriminant bases", in: A. F. Laine and A. M. Unser, ed., *Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, Proc. SPIE, Vol. 2303, Jul. 1994.

[130] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion", in: E. Foufoula and P. Kumar, eds., *Wavelets in Geophysics*, Academic Press, 1994, pp. 299–324.

[131] N. Saito and R. R. Coifman, "On local orthonormal bases for classification and regression", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 1529–1532.

[132] I. Schnitzer, A. Rosenberg, C. Leibovitch, M. Botton, I. Cohen and J. Leopold, "Evolution of spectral power density in grounded cathode relativistic magnetron", Proc. SPIE, Intense Microwave Pulses IV, Vol. 2843, Aug. 1996.

[133] Y. Sheng, D. Roberge, H. Szu and T. Lu, "Optical wavelet matched filters for shift-invariant pattern recognition", Optics Letters, Vol. 18, No. 4, Feb. 1993, pp. 299–301.

[134] E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D. J. Heeger, "Shiftable multiscale transforms", IEEE Trans. on Information Theory, Vol. 38, No. 2, Mar. 1992, pp. 587–607.

[135] G. Strang, "Wavelets and dilation equations: a brief introduction", SIAM Rev., Vol. 31, No. 4, Dec. 1989, pp. 614–627.

[136] C. Taswell, "Near-best basis selection algorithms with non-additive information cost functions", Proc. of the 2nd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Philadelphia, PA, 25–28 Oct. 1994, pp. 13–16.

[137] C. Taswell, "Top-down and bottom-up tree search algorithms for selecting bases in wavelet packet transforms", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 345–360.

[138] C. Taswell, "WavBox 4: A software toolbox for wavelet transforms and adaptive wavelet packet decompositions", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 361–376.

[139] M. Vetterli, "Wavelet and filter banks for discrete-time signal processing", in: M. B. Ruskai et al., ed., *Wavelets and Their Applications*, Jones and Bartlett, Boston, 1992, pp. 17–52.

[140] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design", IEEE Trans. on Signal Processing, Vol. 40, No. 9, Sep. 1992, pp. 2207–2232.

[141] G. G. Walter, *Wavelets and Other Orthogonal Systems With Applications*, CRC Press, Inc., Boca Raton, Florida, 1994.

[142] M. Wang, A. K. Chan and C. K. Chui, "Wigner-Ville distribution decomposition via wavelet packet transform", Proc. of the 3rd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18–21 June 1996, pp. 413–416.

[143] E. Wesfreid and M. V. Wickerhauser, "Adapted local trigonometric transforms and speech processing", IEEE Trans. SP, Vol. 41, No. 12, Dec. 1993, pp. 3596–3600.

[144] J. Wexler and S. Raz, "On minimizing the cross-terms of the Wigner distribution", Technical Report, EE PUB No. 809, Technion - Israel Institute of Technology, Haifa, Israel, Nov. 1991.

[145] N. A. Whitmal, J. C. Rutledge and J. Cohen, "Wavelet-based noise reduction", Proc. of the 20th IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-95, Detroit, Michigan, 8–12 May 1995, pp. 3003–3006.

[146] N. A. Whitmal, J. C. Rutledge and J. Cohen, "Reduction of autoregresive noise with shift-invariant wavelet-packets", Proc. of the 3rd IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18–21 June 1996, pp. 137–140.

[147] M. V. Wickerhauser, "Fast approximate Karhunen-Loeve expansions", Yale Univ., New Haven, May 1990 (In: ftp pascal.math.yale.edu /pub/wavelets/ fakle.tex)

[148] M. V. Wickerhauser, "Picture compression by best-basis sub-band coding", Yale Univ., New Haven, Jan. 1990 (In: ftp pascal.math.yale.edu /pub/wavelets/ pic.tar)

[149] M. V. Wickerhauser, "Acoustic signal compression with wavelet packets", in: C. K. Chui, ed., *Wavelets — A Tutorial in Theory and Applications*, Academic Press, Inc., San Diego, 1992, pp. 679–700.

[150] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, AK Peters, Ltd, Wellesley, Massachusetts, 1994.

[151] E. Wigner, "On the quantum correction for thermodynamic equilibrium", Phys. Rev., Vol. 40, 1932, pp. 749–759.

[152] W. J. Williams, "Reduced interference distributions: biological applications and interpretations", Proc. IEEE, Vol. 84, No. 9, Sep. 1996, pp. 1264–1280.

[153] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, Pergamon, London, 1953.

[154] Y. Zhao, L. E. Atlas and R. J. Marks, "The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 38, No. 7, July 1990, pp. 1084–1091.

ייצוגי WAVELET מסתגלים משמרי-העתקה וייישומיהם

ישראל כהן

ייצוגי WAVELET מסתגלים משמרי-העתקה ויישומיהם

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר
דוקטור למדעים

ישראל כהן

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
אייר תשנ"ח חיפה מאי 1998

המחקר נעשה בהנחיית פרופ׳ שלום רז ופרופ׳ דוד מלאך בפקולטה להנדסת חשמל.

ברצוני להביע את תודתי והערכתי העמוקה לפרופ׳ שלום רז ופרופ׳ דוד מלאך על ההנחיה המסורה, היחס החם, הדיונים הנלהבים, ההכוונה והביקורת הבונה במהלך כל שלבי המחקר. אני שמח על הכבוד הרב שנפל בחלקי לעבוד בהנחייתם, ומאוד מעריך את הקשר ההדוק שנשמר גם בשנה בה יצאו לשבתון.

אני מודה לד״ר משה פורת על העזרה בתקופת השבתון של פרופ׳ רז ופרופ׳ מלאך.

תודה לעמיתי לעבודה ברפא״ל, תחום פיסיקה שימושית, על החברות והעידוד בתקופת לימודי. בפרט, תודה לד״ר מוטי בוטון, ד״ר יוחנן לאופולד, אלישע ברדוגו, ד״ר יצחק שניצר וד״ר בני מנדלבאום.

לבסוף, תודה מיוחדת לאשתי תמי, שתמיכתה ואהבתה עזרו לי רבות במהלך העבודה.

לאהבת חיי, תמי

ולזכרו של אבי, אהרן בן-ציון

תוכן העניינים

תוכן העניינים (המשך)

תוכן העניינים (המשך)

תקציר

בשנים האחרונות הולך וגובר השימוש בבסיסים המותאמים לאותות המיוצגים על-ידם על-פי מטרת ההתמרה. ייצוג אות בבסיס מתאים יכול לאפשר את דחיסתו וניתוח מאפייניו, ויכול לפשט יישומים כגון הפחתת רעשים, שיערוך פרמטרים, סיווג, גילוי וזיהוי [150, 102, 128]. מטרת ההתמרה קובעת את מידת ההתאמה של הבסיס לאות (או משפחה של אותות). באופן כללי, דרוש בסיס כך שרק מספר קטן של מקדמים (מקדמי פירוק כאשר האות מתואר כסכום ליניארי של פונקציות בסיס) הם בלתי זניחים, והתרומה הכוללת של כל המקדמים הזניחים גם היא זניחה. עבור דחיסה למשל, הבסיס נבחר כך שאמפליטודת מקדמי הפירוק, כאשר הם מסודרים לפי סדר יורד של האמפליטודה, דועכת מהר ככל האפשר. אם לעומת זאת מטרת ההתמרה היא סיווג האות לקטגוריות נתונות, דרוש בסיס שיאפשר הפרדה מרבית בין תחומים של אותות מסוגים שונים (ההפרדה היא במרחב N-ממדי, כאשר N הוא מספר הדגמים באות). במקרה זה הבסיס המתאים מקטין את ממד הבעיה ומדגיש את ההבדלים בין הקטגוריות השונות.

בחירת בסיס לייצוג אות חייבת להיעשות מתוך ספרייה מוגבלת של בסיסים. אחרת, סיבוכיות החישוב של הבסיס המתאים ביותר עבור אות נתון גבוהה מכדי שהיישומים המצוינים לעיל יהיו מעשיים. התכונות הרצויות עבור ספרייה של בסיסים הם:

1) ניתן לארגן את הספרייה בצורה שתאפשר חיפוש יעיל אחר הבסיס המתאים ביותר.

2) חישוב מקדמי הפירוק של אות בבסיס הוא יעיל ומהיר.

3) קיום אלגוריתמים מהירים לחישוב התמרה הופכית (שיחזור האות מתוך מקדמי הפירוק).

4) פונקציות הבסיס הן בעלות תומך-זמן סופי, או לפחות דועכות מהר בזמן.

5) פונקציות הבסיס הן חלקות (דועכות מהר בתדר).

שתי התכונות האחרונות תלויות כמובן אחת בשניה לפי אי-שוויון Heisenberg ומאפשרות ייצוג וניתוח של אות במרחב זמן-תדר. לוקליות בזמן של פונקציות הבסיס דרושה על-מנת להבחין בתופעות מעבר קצרות. הלוקליות בתדר דרושה על-מנת להבחין בנוכחות של תדירויות מאפיינות.

Coifman and Meyer [102, 39] הם הראשונים שהציגו ספריות בסיסים אורתונורמליים אשר מורכבות מאלמנטים (פונקציות בסיס) ממוקמים במרחב זמן-תדר ומסודרות בצורת עץ המאפשר חיפוש יעיל אחר הבסיס הטוב ביותר בספרייה. עבור הספרייה האחת, ספריית בסיסים טריגונומטריים ממוקמים (localized trigonometric bases), פונקציות הבסיס הן סינוסים או קוסינוסים מוכפלים בפונקציות חלון חלקות הנתמכות בקטעי זמן סופיים. הלוקליות בתדר של פונקציות הבסיס תלויה במידת החלקות של פונקציות החלון [65]. עבור הספרייה האחרת, ספריית בסיסי wavelet packets, פונקציות הבסיס הן העתקות ומתיחות של wavelet packets, ותכונות הלוקליות בזמן-תדר של ה-"mother wavelet" (הפונקציה היוצרת של משפחת wavelet

packets) קובעות תכונות אלו של כל פונקציות הבסיס בספרייה [69, 44]. בשני המקרים הספריות מסודרות במבנה של עץ בינארי בו מכל צומת מתפצלים שני ענפים (ארבעה ענפים במקרה הדו-ממדי) [45]. צומת בעץ של ספריית בסיסים טריגונומטריים או ספריית בסיסי wavelet packets מייצג תת-מרחב של האות המכסה קטע זמן או פס תדר מסוים, בהתאמה. בזכות התכונה שכל צומת ניתן לפיצול אורתוגונלי לשני צמתי-בן, ניתן לבחור בין פונקציות הבסיס של צומת-האב לבין פונקציות הבסיס של צמתי-הבן. גמישות זו בבחירת פונקציות הבסיס של כל תת-מרחב מאפשרת התאמת הבסיס לאות ולמטרת הייצוג. נניח שנתון אות ונתונה פונקצית מחיר לאינפורמציה המשמשת מדד לטיב הייצוג. אזי בחירת הבסיס של כל תת-מרחב תעשה באופן רקורסיבי ע"י השוואת מחיר האינפורמציה לייצוג האות בצומת-האב לסכום מחירי האינפורמציה בצמתי-הבן. הבסיס המתקבל עבור מרחב האות, כלומר עבור צומת שורש העץ, הוא הטוב ביותר לייצוג האות מבין כל הבסיסים בספרייה במובן שמחיר האינפורמציה של האות בבסיס זה הוא הנמוך ביותר.

קיים קשר הדוק בין פונקצית מחיר האינפורמציה ובין היישום הרצוי לייצוג האות בבסיס הטוב ביותר [136, 150]. למשל, אנטרופיה יכולה לשמש כפונקצית מחיר המובילה לאפיון קומפקטי של אות במרחב ההתמרה [48, 76, 143]. ניתוח סטטיסטי של מקדמי הפירוק האנרגטיים ביותר בבסיס הטוב ביותר ייתן חתימה אופיינית של האות במרחב זמן-תדר [12, 87], ויאפשר לכאורה יישומים של גילוי וזיהוי. הבעיה העיקרית בגישה זו היא חוסר אינווריאנטיות להעתקה. הייצוג במרחב ההתמרה רגיש מאוד למיקום הזמני של האות, ולכן ניתוח סטטיסטי של מקדמי הפירוק מורח את האנרגיה על פני מספר גדול של מקדמים וה"חתימה" המתקבלת אינה מאפיינת את האות די הצורך.

Herley et al. [67] הציעו ספרייה כללית יותר של בסיסים אורתונורמליים, ואלגוריתם מתאים (time-varying wavelet packet decomposition) לחיפוש אחר הבסיס הטוב ביותר בספרייה. על-פי אלגוריתם זה הבסיס הטוב ביותר משלב חלוקה זמנית טובה ביותר של האות עם הייצוג הטוב ביותר של כל קטע על-ידי wavelet packets. אולם למרות החלוקה הזמנית "הטובה ביותר", הייצוג בבסיס הטוב ביותר אינו אינווריאנטי להעתקה, מפני שהחלוקה הזמנית מוגבלת לסריג קבוע.

מגרעת נוספת של ייצוגים מסתגלים בספריות בסיסי wavelet, wavelet packets ובסיסים טריגונומטריים ממוקמים, היא חוסר יציבות. בפרט כאשר מגדילים את היקף החיפוש אחר הבסיס הטוב ביותר למספר ספריות במטרה לשפר את התאמתו, בסיסים טובים ביותר עבור אותות דומים עלולים להיות שונים במידה מרובה (מודגם במבוא של פרק 2).

חוסר האינווריאנטיות להעתקה של התמרות wavelet גורמת אף לפגיעה בביצועים במספר רב של יישומים, ביניהם - הסרת רעשים מאותות (denoising) באמצעות בסיסי wavelet [64, 61]. Coifman ואחרים [49, 7, 130] הבחינו שהסרת רעשים באמצעות התמרת wavelet או wavelet packet decomposition (WPD) עלולה לגרום לתופעות לוואי, כגון תופעת "Gibbs" בסביבה של אי-רציפויות ו"סימטרייה מלאכותית" בציר התדר [49]. תופעות אלו נובעות מחוסר

האינווריאנטיות להעתקה, ולכן הם הציעו למנוע את התלות בהעתקה ע"י הסרת הרעש מגרסאות מוזזות של האות ומיצוע של אותות "נקיים" מרעש (Cycle-Spinning). בתהליך זה אמנם ניתן לקבל ביצועים משופרים בחלקים "איטיים" של האות, אך תופעות המעבר מונחתות בדרך-כלל בצורה ניכרת [145]. נוסף על כך, עקרון "אורך קוד-תיאור מינימלי" (MDL – Minimum Description Length) [125], שהוכח כקריטריון יעיל לבחירת הבסיס המתאים להסרת הרעש [130, 146, 104, 84], לא ניתן ליישום בשיטה זו.

נושא אחר שקשור לבעיית חוסר האינווריאנטיות להעתקה, הוא פירוקים מסתגלים של פילוגי זמן-תדר והסרת רכיבי הפרעה (interference terms) בפילוגים ביליניאריים. פילוג Wigner למשל, מקיים מספר רב של תכונות הרלוונטיות לניתוח אותות במרחב משולב זמן-תדר [18, 23], אולם רכיבי ההפרעה מגבילים מאוד את יישומו. רכיבים אלה מקשים מאוד את הפענוח והניתוח של פילוגי Wigner. ניתן אמנם להקטין את רכיבי ההפרעה במידת-מה ע"י סינון פילוג Wigner במסננת מעבירה נמוכים (החלקה של הפילוג) [152, 78]. אולם, הסינון מקטין את ריכוזיות האנרגיה של רכיבי האות ומשפיע בצורה ניכרת על התמונה המתקבלת במרחב זמן-תדר.

המטרה המרכזית של המחקר היא לנסח פתרון כללי לבעיות שתוארו לעיל, ובד בבד לשפר ולפשט יישומים שהוגבלו בעטיין. בפרט, מטרות המחקר הן:

- לפתור את בעיית חוסר אינווריאנטיות להעתקה של ייצוגים מסתגלים בספריות בסיסים, ולפתח אלגוריתמים מהירים עבור ספריית בסיסי wavelet packets וספריית בסיסים טריגונומטריים ממוקמים, כך שהייצוג בבסיס הטוב ביותר אינווריאנטי להעתקה, אורתוגונלי ומאופיין במחיר אינפורמציה נמוך. (התמרת wavelet אינווריאנטית להעתקה תתקבל כמקרה פרטי ע"י הגבלת ספריית בסיסי wavelet packets לספריית בסיסי wavelets).

- לבחון את הקשר בין סיבוכיות החישוב ובין מחיר האינפורמציה כאשר האינווריאנטיות להעתקה נשמרת.

- לאפשר פילוגי זמן-תדר מסתגלים המקיימים תכונות שימושיות רלוונטיות לניתוח אותות במרחב המשולב. התכונות יכללו ריכוזיות גבוהה של האנרגיה, רכיבי-הפרעה נמוכים ואינווריאנטיות להעתקה.

- לפתח שיטה להסרת רעש מאותות ופילוגי זמן-תדר, שאינה תלויה בהעתקה זמנית ומבוססת על קריטריון אורך קוד-תיאור מינימלי.

הנושאים הראשונים של התזה הם פירוקי wavelet packets ופירוקים טריגונומטריים מסתגלים משמרי-העתקה. אנו מגדירים ספריות מורחבות של בסיסים המתאפיינות בכך שכל העתקה זמנית של בסיס בספרייה גם-כן כלול בספרייה, ומפתחים אלגוריתמים מהירים למציאת הבסיס הטוב ביותר לייצוג אות נתון. באופן זה, אם בסיס מסוים נבחר כטוב ביותר לייצוג האות הנתון, אזי

עבור העתקה זמנית של האות יבחר בסיס שהוא העתקה זמנית זהה של הבסיס הקודם ומקדמי הפירוק לא ישתנו.

הספרייה המורחבת של בסיסי wavelet packets כוללת את הספרייה המקורית שהוגדרה ע"י Coifman and Meyer [45], והיא גדולה ממנה ביותר מפקטור ריבועי. לפיכך, עבור אות נתון הבסיס הטוב ביותר הוא בהכרח מתאים יותר (מתאפיין במחיר אינפורמציה נמוך יותר). יחד עם זאת, התוספת לסיבוכיות החישוב של האלגוריתם לחיפוש אחר הבסיס הטוב ביותר בספרייה המורחבת היא קטנה משמעותית יחסית לגודל הספרייה. יעילות האלגוריתם היא פועל יוצא של ארגון הספרייה במבנה של עץ. צומת בעץ מייצג תת-מרחב של האות המכסה פס תדרים ברזולוציה מסוימת, ועץ הפירוק הטוב ביותר מתאים לבסיס הטוב ביותר. האלגוריתם Shift-Invariant Wavelet Packet Decomposition (SIWPD) בדומה לאלגוריתם WPD בודק אילו צמתים כדאי לפצל, ובנוסף מאפשר העתקה של צמתי-בן ביחס לצומת-אב. בפרק 2 מוכח שעבור כל צומת-אב מספיק לבדוק שתי אפשרויות להעתקה (העתקה אפס או $2^{-\ell}$, כאשר ℓ רמת הרזולוציה, $\ell \le 0$), והייצוג המתקבל בבסיס הטוב ביותר הוא אורתוגונלי ואינווריאנטי להעתקה. במקרה הפרטי שפיצול צמתים נעשה בכל רמת רזולוציה רק עבור ערוץ התדרים הנמוכים, מקבלים התמרת wavelet אינווריאנטית להעתקה.

דרך אחרת להסביר את האלגוריתם היא באמצעות filter banks [140, 139, 121]. פיצול של צומת מתקבל ע"י סינון דרך מסננות מעבירה-גבוהים ומעבירה-נמוכים ולאחר מכן דצימציה 2:1. בעוד שעבור WPD הדצימציה מתקבלת ע"י השארת הדגימות באינדקס זוגי (או אי-זוגי) והתעלמות משאר הדגימות, עבור SIWPD הדצימציה אדפטיבית לאות; בוחרים להשאיר את הדגימות באינדקס זוגי או אי-זוגי לפי מחיר האינפורמציה הנמוך יותר. בזכות האורתוגונליות של הפירוק והאדיטיביות של פונקצית מחיר האינפורמציה, ההחלטה לגבי אופן הדגימה בכל צומת היא בלתי תלויה בצמתים אחרים אשר באותה רמת רזולוציה.

יתרון נוסף של SIWPD לעומת WPD הוא התאפשרות השוואה כמותית של טיב ספריות והתאמתן ליישום המבוקש. ניתן להשוות סטטיסטית את מחיר האינפורמציה לייצוג אות (או משפחה של אותות) עבור ספריות שונות של wavelet packets (ספרייה כזו תלויה בבחירה הפונקציה היוצרת, "scaling function") ולבחור את הספרייה הטובה (מתאימה) ביותר. עבור WPD, בחירת הספרייה תלויה במיקום הזמני של האות ולכן ניתוח סטטיסטי כזה עלול להיות חסר משמעות (דוגמה במבוא של פרק 2).

במקרה של ספריית בסיסים טריגונומטריים, האינווריאנטיות להעתקה של "הייצוג בבסיס הטריגונומטרי הטוב ביותר" מתקבלת ע"י אפשור העתקה בין רמות רזולוציה. מוכח בפרק 3 שבכל רמת רזולוציה ℓ מספיק לבחור בין שתי אופציות להעתקה זמנית יחסית לרמת רזולוציה גבוהה יותר - העתקה אפס או $2^{-\ell-1}$. הבחירה תעשה על-פי מחיר האינפורמציה הנמוך ביותר לייצוג. לכן, בנוסף לאינווריאנטיות להעתקה, הייצוג מתאפיין במחיר אינפורמציה נמוך יותר בהשוואה לזה

המתקבל באמצעות האלגוריתם LTD. שיפור נוסף במחיר האינפורמציה מושג בעזרת אופרטור קיפול (folding operator) אדפטיבי. האופרטור מפריד את האות לקטעים ומקפל חלקים חופפים בקוטביות המתאימה באופן מקומי לאות. לאחר מכן הקטעים "המקופלים" של האות מותמרים למרחב התדר ע"י ייצוגם בבסיסים טריגונומטריים בעלי אותן תכונות זוגיות כשל הקטעים. השיפור במחיר האינפורמציה הוא בפרט משמעותי עבור אותות המתאפיינים בכל קטע בתדירות אחת עיקרית (רוב האנרגיה של האות בכל קטע היא בסביבה של תדירות בודדה). במקרה זה אופרטור הקיפול מנצל את תכונות הזוגיות של האות בנקודות הפרדה בין קטעים ומאפשר ייצוג יותר קומפקטי.

הנושא השני שנחקר בתזה הוא פירוקים מסתגלים של פילוגי זמן-תדר והסרת רכיבי הפרעה בפילוגים ביליניאריים. אנו מראים שבאמצעות ייצוגים משמרי-העתקה ופעולות מסוימות במרחב Wigner, ניתן להשיג מספר תכונות החשובות מאוד לניתוח אותות במרחב משולב זמן-תדר. התכונות כוללות ריכוזיות גבוהה של האנרגיה, רכיבי הפרעה נמוכים ואינווריאנטיות להעתקה. במקום פעולת החלקה על הפילוג, שגורמת למריחת האנרגיה של מרכיבי האות, אנו מציעים פירוק של הפילוג בבסיס "הטוב ביותר" ושליטה על רכיבי-הצלבה (cross-terms) של פונקציות בסיס באמצעות פונקצית מרחק מסוימת.

תחילה פורשים את האות הנתון בבסיס הטוב ביותר (אנטרופיה מינימלית), ומתמירים אותו למרחב Wigner. לאחר-מכן, מסירים את רכיבי-ההפרעה ע"י אי-הכללה של רכיבי-הצלבה מסוימים, שנקבעים לפי האמפליטודה של פונקציות הבסיס והמרחק ביניהם במרחב המשולב זמן-תדר. פונקצית המרחק משקללת את המרחק האוקלידי בין פונקציות בסיס בפילוג האנרגיה המקומי של האות, על-מנת לאפשר סף-מרחק שלא תלוי בפילוג המקומי של האות. סף האמפליטודה וסף המרחק נבחרים כך שקיים איזון אופטימלי בין רכיבי ההפרעה, התכונות המועילות של הפילוג והסיבוכיות החישובית.

התכונות של הפילוג המתקבל, פילוג modified Wigner, נחקרות, והביצועים המשופרים שלו מושווים לאלו המתקבלים בשיטות אחרות. אנו מדגימים יישום להפרדת רכיבים מאות מורכב, ובוחנים בחירות שונות של ספריות מורחבות. ניתן להסיק שהאיכות הויזואלית של פילוג modified Wigner היא בדרך-כלל תואמת את מחיר האינפורמציה בבסיס הטוב ביותר, ולפיכך ניתנת באופן זה לכימות.

הנושא האחרון בעבודה עוסק בשיטה משמרת-העתקה להסרת רעש מאותות ופילוגי זמן-תדר, כאשר קריטריון הטיב הוא "אורך קוד-תיאור מינימלי". אנו מגדירים עבור האות הרצוי תבנית, המבוססת על ספרייה מורחבת של בסיסים (wavelet packets או בסיסים טריגונומטריים), ומיישמים את עקרון "אורך קוד-תיאור מינימלי" על-מנת לקבוע פונקצית מחיר אדיטיבית מקורבת. אורך קוד-התיאור של האות הנמדד (האות הרועש) מקבל ערך מינימום ע"י אופטימיזציה של עץ הפירוק ואיפוס מקדמי הפירוק הנמוכים. אנו מראים שניתן לשלב בצורה יעילה את המשערך המתואר עם פילוג modified Wigner, ובאופן זה לקבל ייצוגי זמן-תדר מסתגלים חסינים

לרעש ובעלי תכונות משופרות. השיטות המוצעות בעבודה זו מושוות לשיטות אחרות הקיימות בספרות, והעדיפות של אלו הראשונות מוסברת ומתוארת באמצעות דוגמאות על אותות מלאכותיים ואותות אמיתיים.