AD-A256 172

PAGE

Form Approved
OMB No. 0704-0188

Public reporting b
gathering and ma
collection of infor
Davis Highway, Su

per response, including the time for reviewing instructions, searching existing data sources
of information. Send comment regarding this burden estimate or any other aspect of this
Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson
and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	28 August 1992	Final 1 Oct 91 to 15 Sep 92

4. TITLE AND SUBTITLE
AFIT/AFOSR Workshop on The Role of Wavelets in Signal Processing Applications

5. FUNDING NUMBERS
AFOSR-616-92-0019

(1)

6. AUTHOR(S)
Bruce W. Suter, Mark E. Oxley and Gregory T. Warhola

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Electrical and Computer Engineering
and
Department of Mathematics and Statistics

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/EN-TR-92-3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
Bolling AFB, DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
SEP 30 1992
S A D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

The Air Force Institute of Technology (AFIT) hosted a workshop on the Role of Wavelets in Signal Processing Applications on 12-13 March 1992 at the AFIT Science and Research Center, Wright-Patterson AFB, OH. The workshop was sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling AFB, DC. The major technical themes were: (1) time/scale analysis versus frequency analysis of signals, and (2) filter bank structures versus discrete wavelet expansions. Professors Alexander Grossmann, Patrick Flandrin and Albert Cohen were sponsored by the AFOSR Window on Science program.

14. SUBJECT TERMS
Wavelets, Signal Processing, Time-Frequency

15. NUMBER OF PAGES
469

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

AFIT/EN-TR-92-3

AFIT/AFOSR WORKSHOP on the ROLE of WAVELETS in SIGNAL PROCESSING APPLICATIONS

held

12-13 March 1992

at the

Air Force Institute of Technology
Wright-Patterson Air Force Base, Ohio

FINAL REPORT

Final report for the period 1 October 1991 to 15 September 1992

Bruce W. Suter, Ph.D.
Department of Electrical and Computer Engineering (AFIT/ENG)

Mark E. Oxley, Ph.D.
Department of Mathematics and Statistics (AFIT/ENC)

Gregory T. Warhola, Ph.D.
Major, USAF
Department of Mathematics and Statistics (AFIT/ENC)

28 August 1992

Approved for public release; distibution unlimited

92 9 29 047

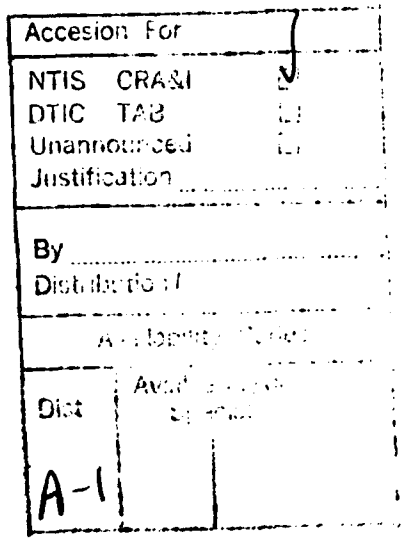
DTIC 01225
92-26158
4748

WORKSHOP ORGANIZING COMMITTEE

Brucw W. Suter, Chairman

Mark E. Oxley

Gregory T. Warhola



DTIC QUALITY INSPECTED 3

TABLE OF CONTENTS

Introduction
Bruce W. Suter 1

Summary
Jon Sjogren 2

Photograph of Participants 6

Participants' Address List 7

Speaker Abstracts 10

Papers
Thomas P. Barnwell III 15
A Time Domain View of Filter Banks and Wavelets

John J. Benedetto 21
Irregular Sampling and the Theory of Frames, I

Gregory Beylkin 45
Multiresolution Representations using the Auto-Correlation
Functions of Compactly Supported Wavelets

Albert Cohen and Ingrid Daubechies 48
Non-Separable Bidimensional Wavelet Bases

Leon Cohen 118
Scale and Inverse Frequency Representations

Ingrid Daubechies 130
Tiling time-frequency in an adaptive way

Patrick Flandrin 137
On the Time-Scale Analysis of Self-Similar Processes

Stephane Mallat 151
Structural Decomposition of Signals

Bruce W. Suter and Mark E. Oxley 166
On Variable Length Windows and Weighted Orthonormal Functions

Robert R. Tenney and Alan S. Willsky 179
Multi-Resolution Estimation for Image Processing and Fusion

P.P. Vaidyanathan 201
Paraunitary and Orthonormal Convolvers

Martin Vetterli 243
Best Wavelet Packet Bases in a Rate-Distortion Sense

Martin Vetterli 281
Wavelets and Recursive Filter Banks

Martin Vetterli 319
Multiresolution Broadcast for Digital HDTV Using Joint Source-Channel Coding

Alan Willsky 360
Modeling and Estimation of Multiresolution Stochastic Processes

Gregory W. Wornell and Alan V. Oppenheim 421
Wavelet-Bases Representations for a Class of Self-Similar Signals
with Application to Fractal Modulation

Introduction

Bruce W. Suter, AFIT

The Air Force Institute of Technology (AFIT) was pleased to host the international workshop on "The Role of Wavelets in Signal Processing Applications". This workshop, held 12-13 March 1992 at Wright-Patterson Air Force Base, Ohio, brought together leading researchers from both the mathematics and signal processing communities. As such, the workshop provided a forum for the interchange of ideas, practical experiences, and recent advances. This invitation-only workshop was intentionally kept small in size in order to encourage active participation by all attendees.

The purpose of this workshop was to gain a perspective of the role of wavelets in signal processing, and to form a vision of where we should look as a research community. With this in mind, the workshop sought to address the following objectives:

1. to highlight major accomplishments and limitations in the use of wavelets in applied mathematics and signal processing,

2. to define the current status of wavelets research in these fields,

3. to present the challenges for future wavelets research in mathematics and in signal processing.

In order to develop the workshop with these objectives and to stimulate discussions, brief presentations were given by several of the attendees. The sponsor of the workshop, the Air Force Office of Scientific Research (AFOSR), requested that an official document be generated to commemorate this workshop. As a result, attendees provided a paper to be included in these proceedings, since these papers were not refereed by the organizers and, it is acceptable for these papers to be submitted to other journals for publication.

In order to prepare the participants for meaningful discussions, the organizers requested that each one answer the following questions, prior to their arrival, regarding an objective comparison of wavelets against techniques in applied mathematical analysis and signal processing:

1. For what problems have wavelets been shown to be clearly superior to all other known techniques?

2. For what problems are the use of wavelets clearly inferior to all other known techniques?

3. In what problems have wavelets shown promise, but, to date the wavelets-based research results obtained are not superior to other known techniques?

While the participants were promised that their responses would not be published, their contributions to these proceedings certainly reflect them.

Summary

Jon Sjogren, AFOSR

The AFIT/AFOSR Workshop on "The Role of Wavelets in Signal Processing", sometimes known "Wavelets Workshop II" or "WW II", was unique in several respects. To a great extent it was shaped by the course of the previous year's meeting, "Symposium on Applications of Wavelets to Signal Processing" (WW I). Last year's meeting, also held at Wright-Patterson AFB, was an order of magnitude larger, having grown considerably from the modest plans and expectations of its organizers.

The first Workshop was kicked off with a superlative series of tutorial talks, delivered mainly by Maj Greg Warhola, USAF. The tutorials were followed by invited presentations, which served to acquaint the diverse audience with specific wavelet methodologies and their applications, from spread-spectrum communications, electronic warfare, modeling of noise processes, feature identification in EEG and PET scans, to speech and image compression. A banquet and panel discussion brought out some fundmental issues on the meaning and interpretation of signals. A focal point was the question of how important wavelets "ultimately" would turn out to be, say in 10 years, as a tool and technique in Signal Processing. On a scale of 0 to 10, responses ranged from "1" to "10+". Electrical engineering researchers tended to hold the more skeptical position, while mathematicians (mainly "wavelet" mathematicians) took a more exuberantly positive stand. Other observations included the importance of Conjugate Quadrature Filters as a precursor to analysis via wavelets. Alan Willsky of MIT, though technically not part of the panel, forcefully succeeded in raising the consciousness of the participants regarding the dynamics of scale as a variable of value equal to time. The festivities in 1991 were brought to a close as no one seemed to have a satisfactory answer to an annoying but persistent question: "what means frequency?"

A significant achievement of both WW I and WW II was that the cultural gap between theoreticians and practitioners was closed. This was evidenced by the large variance of the "important" assessments. Evidently a mutual learning process is accelerating.

A few major "technical" themes stand out from the talks and discussions of Wavelet Workshop II. I mention two: (1) time/scale analysis versus frequency analysis of signals, and (2) filter-bank structures versus discrete wavelet expansions to name a few. As Leon Cohen pointed out in the lead talk, scale and "reciprocal frequency" are related in a subtle way and cannot be taken as identical for all purposes – the operator algebras that they generate are different. On the second point, it is by now generally understood that discrete wavelet analysis can be completely carried out with banks of subband filters, down- and up-sampling, etc. Clearly there are many such filter bank configurations that may be useful in engineering systems, and employ adaptive features, which take into account varied distributions of particular signal components and so forth, that will not likely be of much interest to a mathematical theory of wavelets. But it is also coming to be acknowledged that studies made by the mathematical-physicist pioneers in wavelets (such as involving regularity properties)

are proving to be a significant consideration and guide in filter design. The lunchtime talks and discussions brought out both of these points and other issues as well.

It has been forcefully expressed that all these remarks, and especially those of Prof. Grossmann on the "History of Wavelets" at the Thursday dinner, be collected and included in this volume. This is eminently desirable, but the use of recording devices was ruled out as too inhibiting. In any case only a faint reflection of the mood, the spontaneity, the depth of experience that welled up on those occasions, could possibly be captured in print. We are consoled with the thought that future work, publications and talks will owe something to the rare conditions of insight and mutuality that prevailed for a short time.

Martin Vetterli applies multiresolution decomposition in channel coding to commercial digital broadcast. Transmitting a coarse signal separately from its details will allow a "gracefully degrading" received signal in this future broadcast environment.

Patrick Flandrin and Greg Wornell are both interested in signals with self- similar qualities at different scales ("fractal"). Noise with a $1/f^{\alpha}$ spectral characteristic can fall into this category. This self-similarity can be either stochastic ("fractional Brownian motion") or deterministic. The wavelet transform is close to the ideal decorrelating transform for fractional stochastic processes ("Karhunen-Loeve"). This gives a way to estimate the "alpha" parameter among others. Wornell shows how a scheme of encoding a message at several scales in a deterministically self- similar carrier can provide robustness in "difficult" environments where conventional modulation falls short.

The talks of Alan Willsky and Robert Tenney were related and scheduled separated only by the first lunch and discussion. Alan's talk introduced the concept, difficult to this writer, of stochastic processes defined on certain graphs, where correlations can be inferred across levels, which represent different magnitudes of scale. Bob Tenney demonstrated an interactive environment that has been built around this theoretical framework: rapid estimation based on image data for background/terrain that possesses self- similar characteristics. Such techniques seem especially suitable for reconstructing "ground truth" from diverse data sources ("sensor fusion").

Stephane Mallat sees a need for adaptive, possibly non-linear transforms to complement existing methods in time-frequency analysis. His "dictionaries" and "structure books" allow more efficient representation of a signal that has disparate types of component (sinusoidal, pulse, fractal) than would, say, a wavelet-packet basis on its own.

Ingrid Daubechies and Albert Cohen also approached the issue of making wavelet-packets even more flexible and adaptive. A significant problem here has been to define wavelet-bases for a finite interval. This is very recent mathematics, with substantial ramifications for functional analysis. In addition, this may well amount to a conceptual breakthrough for applications on the order of wavelet-packets themselves (which were brand-new only two years ago!)

Alexander Grossmann's evening historical retrospective was inspiring and set the mood for eloquent reminiscences by Thomas Barnwell and P.P. Vaidyanathan. Prof. Grossmann recounted the early ideas of Morlet in seismic analysis and how collaboration led to connections with unitary operators and their representations.

Gregory Beylkin pointed out some of the unesthetic features of compactly supported wavelets, such as lack of symmetry or shift invariance. He remedies this situation by considering shift-classes (bringing in the shift invariance), and using autocorrelation functions of wavelets as a basis, replacing the unsymmetrical wavelets.

In addition to his exposition of the history of filter banks for subband coding, Thomas Barnwell explained how a general setting using Lapped Orthogonal Transforms can lead to analysis/synthesis systems with higher frequency resolution than available with dyadic wavelets.

P.P. Vaidyanathan, concentrating in his historical account on the origin of the concept of paraunitary matrices (which are key to a 'polyphase' analysis of filtering/sampling) gave a succinct theorem for filter bank convolution, emulating the classical Fourier convolution theorem. Characteristically, P.P. immediately applies this to an improvement in the coding gain that results from a subband coding/vector quantization scheme, which uses band-by-band convolution in the synthesis step.

John Benedetto is a the master of sampling theory; the theory of frames in a Hilbert space, motivated by Wigner-Ville (Gabor) wavelet methods, is seen to be the right tool for expressing "interpolations" more generally than given by the Whittaker-Nyquist formula, especially for *irregular* sampling.

Bruce Suter talked about his work with AFIT colleague and Workshop co- organizer Mark Oxley on variable length windows and *weighted* orthonormal functions. This approach puts windowed ("short-time") Fourier analysis, which ostensibly is a lot of *ad hoc* engineering, into a consistent and logical context.

The participants seems unanimous that this was a rare meeting. Months later its special circumstances are recounted from Toulouse to Il Ciocco, from Berkeley to Oberwohlfach. It is not possible to reproduce the lively discussions that took place, but the participants will all remember a bit of what was said, and by whom, and can get in contact for elaboration. We expect to see some of the contentions and "resolutions" that arose propagate to magazine articles and learned forums. The sine function figures in the solution of Maxwell's equations, though it is not the only solution – is it thereby God-given? (Kronecker's assertion about the natural numbers notwithstanding.) More to the point, *light* does exhibit sinusoidal wave characteristics (before we knew all about E&M). Thus "spectral analysis" lets you discover the element Helium. Legendre gave us his polynomials (or God gave us Legendre) so that we can measure the shape of the earth. Classical physics is old, and such breakthroughs will be rarer. Wavelets did not leap to prominence as eigenfunctions of a differential operator (is there a Bessel wavelet?) but they have a claim to mathematical natural-ness (bases of Calderon), physical importance ("coherent states"), and are in tune with the Age of Computation (sparsifying Toeplitz operators). Speaking from the point of view of combinatorics/algebra, workers in this field as well would have stumbled sooner rather than later onto Perfect Reconstruction Filter Banks.

There is surely no better proponent of the mathematics of wavelets than R. Raphael Coifman, and no better proponent of an enlightened scrutiny of their potential in signal processing than Alan V. Oppenheim. In Prof. Coifman's after-lunch

remarks, he showed himself keenly aware of the complexities and trade-offs inherent in applying a great new idea, and came close to declaring himself a born-again engineer. Prof. Oppenheim, *mutatis mutandis*, acknowledged a renewed appreciation for the applied mathematics community's contribution just over the past year.

Some practical conclusions seemed to be in consensus. Wavelet-based methods of themselves cannot compete in speech compression, with the algorithms derived over decades by the expert speech modelers. For certain particular compression tasks (fingerprint images?), nearly "raw" wavelet methods are running very strongly. One must bear in mind that the various image and speech communities have not had time to incorporate the best of wavelet methods into their tradecraft. Preliminary results indicate that it is precisely a combination of time-scale (wavelet) and frequency methods that approaches optimality for a problem of importance such as estimating frequency-hopping parameters in a "covert communications" scheme.

If another Wavelet Workshop is held in a few years (with a different name?), we can anticipate some startling developments. On the one hand, multiresolution science will be more firmly entrenched in Control Theory, Time-Series modeling, and Dynamics. On the other hand, revelations of mathematical appropriateness will emerge. Recently it was seen how the Naparst equations for Range-Doppler imaging have a wavelet transform interpretation. This leads to optimal solutions for an emitted radar waveform in this imaging: they are elements of a wavelet basis! Another significant theoretical observation is the utility of locally transforming in the Fourier domain. This holds immediate promise for speeded-up and less expensive MRI scanning. These exciting topics are among those that would be treated in a future meeting.

The Dayton Workshop was indeed a watershed in terms of consolidating the achievements of the wavelet community. We feel that the Workshop succeeded if it has made it easier for those interested to tap into the corps of tools, techniques and understanding that transcends the collection of individual practitioners.



AFIT/AFOSR Wavelets Workshop Participants

Left to right: Greg Wornell, Stephane Mallat, Alexander Grossmann, Alan Oppenheim, Patrick Flandrin, Thomas Barnwell III, Leon Cohen, Gregory Beylkin, Jon Sjogren, Albert Cohen, Ingrid Daubechies, Martin Vetterli, P.P. Vaidyanathan, Robert Tenney, Ronald Coifman, Alan Willsky, Robert Ryan, Bruce Suter, Mark Oxley, John Bendetto, Greg Warhola

PARTICIPANTS LIST

Thomas Barnwell III
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

email: tom@eedsp.gatech.edu
FAX : (404)894-8363
Phone: (404)894-2914

John Benedetto
Department of Mathematics
University of Maryland
College Park, MD 20742

email: jjb@math.umd.edu
FAX : (301)405-9377
Phone: (301)405-5161

Gregory Beylkin
Program in Applied Mathematics
University of Colorado
Boulder, CO 80309-0526

email: beylkin@boulder.colorado.edu
FAX : (303)492-4066
Phone: (303)492-6935

Albert Cohen
Ceremade Universite Paris IX-Dauphine
Place du Marechal ce Lattre de Tassigny
75775 Paris Cedex 16
FRANCE

email: chavent@frulm63.bitnet
FAX : 011(33)1.47.55.4857
Phone: 011(33)1.47.27.7554

Leon Cohen
4465 Douglas Avenue
Apt 3G
Bronx, NY 10471

email: N/A
FAX : (212)772-5390
Phone: (212)548-4525
(908)932-0548 (at Rutgers)

Ronald Coifman
Mathematics Department
Yale University
10 Hillhouse Avenue
New Haven, CT 06520

email: coifman@lom1.math.yale.edu
FAX : (203)432-7316
Phone: (203)432-4175

Ingrid Daubechies
Mathematics Department
Busch Campus
Rutgers University
New Brunswick, NJ 08903

email: ingrid@research.att.com
ingrid@math.rutgers.edu
FAX : (908)932-5530
Phone: (908)932-3097 or 2393

Patrick Flandrin
Ecole Normale Superieure de Lyon
46 Allie D'Italie
69364 Lyon Cedex 07
FRANCE

email: flandrin@frensl61.bitnet
flandrin@ens-lyon.fr
FAX : 011(33)72.72.8080
Phone: 011(33)72.72.8160

Alexander Grossmann
Centre National de la Recherche Scientifique
Centre De Physique Theorique
Luminy Case 907F
13288 Marseille Cedex 9
FRANCE

email: grossman@frmopli.bitnet
FAX : 011(33)91.26.9553
Phone: 011(33)91.26.9526

Stephane Mallat
Computer Science Department
Courant Institute
New York University
New York, NY 10012

email: mallat@regard.nyu.edu
FAX : (212)995-4122
Phone: (212)998-3466

Alan Oppenheim
Room 36-615
Massachusetts Institute of Technology
Cambridge, MA 02139

email: avo@allegro.mit.edu
FAX : (617)258-7864
Phone: (617)253-4177
(508)548-1400x2640 (at Woods Hole)

Mark Oxley
Department of Mathematics and Statistics
Air Force Institute of Technology
AFIT/ENC
WPAFB, OH 45433-6583

email: moxley@afit.af.mil
FAX : (513)476-4055
Phone: (513)255-3098

Robert Ryan
Office of Naval Research
ONREUR
PSC 802 BOX 39
FPO AE 09499-0700

email: rryan@onreur-gw.navy.mil
FAX : 011(44)71.723.1873
Phone: 011(44)71.409.4539

Jon Sjogren
Air Force Office of Scientific Research
AFOSR/NM
Bolling AFB, DC 20332-6448

email: jas@src.umd.edu
FAX : (202)767-0466
Phone: (202)767-4940

Bruce Suter
Dept of Electrical and Computer Engineering
Air Force Institute of Technology
AFIT/ENG
WPAFB, OH 45433-6583

email: bsuter@afit.af.mil
FAX : (513)476-4055
Phone: (513)255-7210 or 3576

Robert Tenney
Alphatech Inc
Executive Place III
50 Mall Road
Burlington, MA 01803

email: N/A
FAX : (617)273-9345
Phone: (617)273-3388 x227

P.P. Vaidyanathan
Dept of Electrical Engineering (116-81)
California Institute of Technology
Pasadena, CA 91125

email: N/A
FAX : (818)564-9307
Phone: (818)356-4681

Martin Vetterli
Dept of Electrical Engineering
Room 1342 SW Mudd Bldg
Columbia University
500 W 120th Street
New York, NY 10027-6699

email: martin@ctr.columbia.edu
FAX : (212)923-9421
Phone: (212)854-3109

Alan Willsky
Room 35-437
Massachusetts Institute of Technology
Cambridge, MA 02139

email: willsky@lids.mit.edu
FAX : (617)258-8553
Phone: (617)253-2356
(617)273-3388 x232 (at Alphatech)

Gregory Warhola
Department of Mathematics and Statistics
Air Force Institute of Technology
AFIT/ENC
WPAFB, OH 45433-6583

email: gwarhola@afit.af.mil
FAX : (513)476-4055
Phone: (513)255-3098

Gregory Wornell
Room 36-615
Massachusetts Institute of Technology
Cambridge, MA 02139

email: gww@allegro.mit.edu
FAX : (617)258-7864
Phone: (617)253-3513

ABSTRACTS OF THE WORKSHOP

A FILTER BANK PERSPECTIVE ON DISCRETE WAVELET TRANSFORM

Thomas Barnwell
Georgia Institute of Technology

Discrete Wavelet Transform (DWT) can be modeled as special case of a general lapped transform based on filter banks. Design methodologies for such systems are quite mature. A fundamental issue involves the performance degredation implicit with the DWT as compared with other similar but unconstrained transforms.

MULTIRESOLUTION REPRESENTATIONS USING THE AUTO-CORRELATION FUNCTIONS OF COMPACTLY SUPPORTED WAVELETS

Gregory Beylkin
University of Colorado at Boulder

In my talk I will describe a multiresolution representation of signals using dilations and translations of the auto-correlation functions of compactly supported wavelets. This representation was developed together with Naoki Saito of SchlumbergerDoll Research. Although the set of dilations and translations of the auto-correlation functions does not form an orthonormal basis, a number of properties of these functions makes them useful for signal and image analysis. Unlike wavelet-based orthonormal representations, this representation has (1) symmetric analyzing functions, (2) shift-invariance, (3) natural and simple iterative interpolation schemes, (4) a simple algorithm for finding the locations of the multiscale edges as zero-crossings. It also leads to a non-iterative method for reconstructing signals from their zero-crossings and slopes at these zero-crossings.

WAVELET BASES ADAPTED TO AN INTERVAL

Albert Cohen
Ceremade University of Paris IX - Dauphine

Orthonormal and biorthogonal wavelet bases have found many interesting applications in signal and image processing (compression) as well as in fast numerical analysis. In these applications, one always deals with a signal or a function which has a finite expansion in space and/or time. The problem arises then of adapting these bases, usually constructed for the analysis on the whole real line, to a finite interval, [0,1] for example.

In this talk, we shall describe and discuss the different possibilities to solve this problem and the related algorithm. A construction obtained in a joint work with Ingrid Daubechies and Pierre Vial that provides with both a sharp analysis at the borders and a simple algorithmic structure will be explained in more details.

A SIMPLE APPROACH TO JOINT SCALE REPRESENTATIONS

Leon Cohen
Hunter College and Graduate Center of CUNY

Various authors have proposed representations involving scale and time or scale and frequency. The distributions they obtained are very different from each other. We will show that there is a simple conceptual principle for obtaining joint distributions and show that there have been two different notions of scale used. The approach presented clarifies this distinction. We obtain the distributions previously given in a simple direct manner. The concept of instantaneous frequencies is generalized to instantaneous scale and the uncertainty principle for scale is obtained.

ADAPTED WAVEFORM ANALYSIS

Ronald Coifman
Yale University

Local variable length libraries of windowed trigonometric bases are dual versions of wavelet and wavelet packet algorithms. Efficient parameter extractions and compressions for sounds and images can be obtained by selecting best bases out of these libraries, in either frequency or time domain.

NONSEPARABLE TWO-DIMENSIONAL WAVELETS

Ingrid Daubechies
Rutgers University

Many two-dimensional wavelet applications use a tensor product multiresolution analysis. One can also construct "genuine" (non tensor product) two-dimensional multiresoluition analyses, possibly corresponding to matrix dilations. A special case is given by "quincunx" subsampling. The talk shows how orthonormal and biorthogonal bases can be constructed for this case, and discusses their regularity

TIME-SCALE ANALYSIS OF SELF-SIMILAR SIGNALS

P. Flandrin
Ecole Normale Superieure de Lyon

In a number of different physical situations (1/f noises, turbulence, texture analysis, ...), we are faced with fractal or multifractal signals for which it would be desirable to have at hand methods which would allow one to estimate efficiently the corresponding scaling laws or the underlying self-similarity structures. The recently introduced techniques of time-scale analysis (wavelet transforms and generalizations) offer such a possibility, especially in the case of locally self-similar signals, i.e. those for which scaling laws are time-dependent. Starting from the idealized case of fractional Brownian motions, we will show which advantages can be gained from such approaches, either for obtaining almost Karhunen-Loeve (doubly orthogonal) representations via orthogonal wavelet bases, or for defining general classes of estimators (aimed at scaling exponents) via bilinear time-scale representations which generalize the usual Wigner-Ville distribution.

COMPOSITE WAVELETS

Alexander Grossmann
Centre National de la Recherche Scientifique

In some applications of continuous wavelet transforms, it is important to choose an analyzing wavelet with very peaked analyzing kernel. There exist inequalities showing that, in a certain sense, arbitrarily high peaking is not possible. However, wavelet reproducing kernels are precisely broad-band ambiguity functions, and there exists in the radar literature an extensive body of information which shows the way around these constraints. This information is applied to the construction of custommade analyzing wavelets. Both analytic and numerical aspects of the subject are discussed.

ADAPTIVE TIME/FREQUENCY SIGNAL REPRESENTATION

Stephane Mallat
Courant Institute

We proved that detecting the wavelet transform local maxima allows us to locate and characterize singularities. A close approximation of the signal can be recovered from these local maxima. Such an adaptive sampling of the wavelet transform has applications in pattern recognition, compact image coding and noise removal. We are extending this technique to a larger class of transforms that are local in the time/frequency plane. The transform is adapted in order to obtain a compact time/frequency signal representation.

ON VARIABLE LENGTH WINDOWS AND WEIGHTED ORTHONORMAL FUNCTIONS

Bruce Suter and Mark Oxley
Air Force Institute of Technology

A new formulation is presented for the analysis and synthesis of signals. This formulation is composed of a variable length window and a linear combination of weighted orthonormal functions. Tradeoffs in the specification of windows are considered. A sinusoidal example is considered and a fast algorithm is provided for its evaluation.

ESTIMATION ON MULTISCALE NETWORK MODELS MODELS OF RANDOM FIELDS

Robert Tenney
ALPHATECH, INC.

Multiscale treestructured models of twodimensional random processes are known to lead to exceedingly efficient estimation algorithms. However, realizations of the processes defined by these models contain artifacts atypical of most imagery, such as discontinuities along quadrant boundaries. Augmenting these tree models with one-dimensional multiscale models along those boundaries removes the artifacts in samples of the process. However, it also transforms the tree into a network. This talk presents samples of this class of stochastic process, along with a sketch of a general estimation theory which shows that (1) the complexity of the statistics to be maintained by an estimator is independent of the depth of the model, and (2) update of these statistics from a point measurement can be accomplished in time strictly proportional to the number of pixels at the finest scale.

PARAUNITARY CONVOLVER

P.P. Vaidyanathan
California Institute of Technology

The maximally decimated filter bank (perhaps with nonuniform decimation) can be regarded as a transformation from time to time-frequency. Examples of special cases include the DFT and the short time Fourier transform. The filter bank transformer has also been regarded as the discrete-time wavelet transformation by some researchers in the community. Now, for the case of the traditional Fourier transformation, the convolution theorem is well-known. That is, convolution in time is equivalent to multiplication in the transform domain. What is the corresponding theorem for the case of the filter bank transformer? The answer turns out to be particularly simple for the case of orthonormal (paraunitary) filter banks, and in fact offers some practical advantages (coding gain) in finite-precision implementations. This talk will address these issues.

WAVELETS, FILTER BANKS, AND APPLICATIONS

Martin Vetterli
Columbia University

Recent results on the connection between wavelets and filter banks will be reviewed. These include FIR/IIR constructions, as well as multidimensional ones.

The question of arbitrary linear tilings of the time/frequency or phase space will be addressed, showing that short-time Fourier and wavelet decompositions are two special cases.

Finally, applications in compression will be discussed. It will be indicated that subband coding schemes, which are essentially identical to wavelet methods, have been well studied over the last 15 years and achieve interesting compression results, but no spectacular improvements.

MULTIRESOLUTION STOCHASTIC MODELS AND FRACTAL REGULARIZATION

Alan Willsky
Massachusetts Institute of Technology

In this talk we describe our continuing effort to develop a framework for modeling stochastic processes at multiple resolutions and for developing efficient signal and image processing algorithms based on these models. We also illustrate the potential of this approach in one context, namely as the basis for the "fractal regularization" of ill-posed image processing and computer vision problems. Other potential areas of application will also be touched upon.

WAVELETS, SELF-SIMILAR SIGNALS. AND FRACTAL MODULATION

Gregory W. Wornell and Alan Oppenheim
Massachusetts Institute of Technology

Orthonormal wavelet bases provide highly efficient representations for several classes of self-similar signals. One such collection of self-similar signals we refer to as dy-homogeneous signals because they generalize the well-known homogeneous functions. These signals, which are characterized in terms of a deterministic scale-invariance relations, can be categorized into two classes; energy-dominated and power-dominated. We present wavelet-based constructions of orthonormal self-similar bases for the representation of such signals, as well as efficient discrete-time algorithms for their manipulation.

Synthesis of dy-homogeneous signals is potentially important in a range of engineering applications, including remote-sensing and communications. In the communications context, we demonstrate that orthonormal self-similar bases lead to an efficient strategy for embedding information into a self-similar signal on multiple time scales. The resulting "fractal modulation" strategy constitutes an interesting paradigm for communication that is naturally suited for use with noisy channels of unknown duration and bandwidth.

©1991 IEEE. Reprinted, with permission, from *Proceedings of the 1991 Asilomar Conference on Circuits and Systems* Pacific Grove, CA, Nov. 4-6, 1991 (invited)

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commerical advantage, the IEEE copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Institute of Electrical and Electronics Engineers. To copy otherwise, or to republish, requires a fee and specific permission.

A Time Domain View of Filter Banks and Wavelets

Kambiz Nayebi, Thomas P. Barnwell III, and Mark J. T. Smith

Digital Signal Processing Laboratory
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract

In this paper, we consider a time domain approach to the reconstruction problem for filter bank and wavelet decompositions. The development of this time domain reconstruction theory has resulted in a unified design approach for uniform, nonuniform, low delay, and efficient analysis-synthesis systems. The theory is based on FIR filters and can also be applied to the design of perfect reconstruction systems based on wavelets for multi-resolution analysis-synthesis systems.

In this paper, the procedure for the design of general decomposition and reconstruction systems, called Generalized Lapped Transforms (*GLT*), is discussed. GLT's include many classical transforms and the discrete-time wavelet transform (DTWT) as special cases. The new design procedure is used to design wavelets and DTWT systems. Because of the generality of the framework, regularity and phase conditions can easily be imposed on the wavelet. Also, because the design procedure can be used to design nonuniform band systems, systems with higher frequency resolution than dyadic wavelet-based systems can be designed and realized. A number of design examples are included in the paper.

1 Introduction

Signal analysis and reconstruction based on filter banks has been a popular field of research over the last decade. Many different time-frequency representations can be interpreted and implemented as filter bank structures. For example, the discrete short-time Fourier transform (DSTFT), which has long been used to generate spectrograms, can be implemented in a filter bank structure [1]. The discrete wavelet transform is also typically implemented using a tree-structured filter bank system [2]. From a more general point of view, filter bank systems are time-frequency block transforms in which the transformation matrix is $M \times N$ and $M \leq N$ (where N is the length of the longest system filter). From this viewpoint, filter bank systems include classical block transforms (DSTFT, Discrete Cosine Transform, etc.) and Lapped Orthogonal Transforms (LOT's) [3] as special cases.

Recent research on the design of perfect reconstruction analysis-synthesis systems has produced some significant results [3–8]. The invention of Conjugate Quadrature Filters (CQF's), perfect reconstruction filter banks with $N = 2M$, lossless analysis-synthesis systems, perfect reconstructing modulated analysis-synthesis filter banks, and the design of nonuniform filter banks with rational sampling rate changes are among the significant achievements within the last decade.

The time domain formulation introduced in [6,9,10] provides a unified framework for the design of a wide variety of analysis-synthesis systems which includes all known structures based on FIR filter banks. In the past, almost all decomposition-reconstruction systems based on filter banks where *designed by analysis*. This means that the design was accomplished by a thorough and complete analysis of the system of interest. In the new time domain design methodology, a complete analysis of the system is not required in order to accomplish the design. In fact, the only information that is fundamentally necessary in the design process is the desired system structure. Obviously, for the design to be possible, the system structure must be consistent with the desired frequency resolution. This design approach provides an elegant and powerful procedure for designing systems which are not completely understood analytically. For example, it was not known that it was possible to design low and minimum delay systems with perfect reconstruction until they were designed using the time domain design technique [11].

From a signal analysis viewpoint, analysis-synthesis systems are used to project the input signal onto different signal subspaces each of which may have different time and/or frequency characteristics. Enough information is preserved in signal subspaces so that the signal can be reconstructed. It should be clear that the class of analysis-synthesis systems based on filter banks, which is perhaps better called the class of fixed overlapping block transforms, is very large and includes many of the well known transforms as special cases. It is also obvious that there are infinitely

[1] This paper was published in the proceedings of the 25th Asilomar Conference on Signals, Systems and Computers, November 1991.

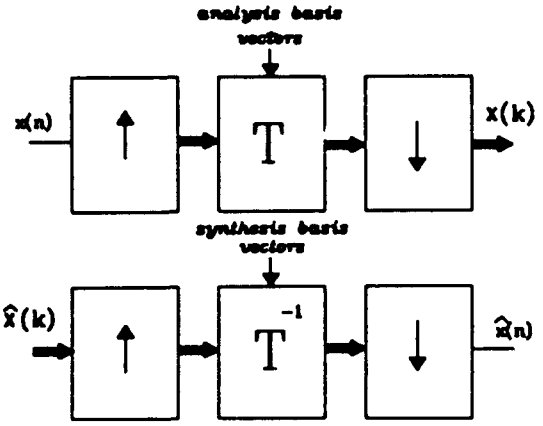


Figure 1: *A simple block diagram of the generalized lapped transform.*

many unique time-frequency decompositions that can be achieved using nonuniform band systems. These transforms can be considered to be an invertible mapping of the signal onto the time-frequency plane. The resolution in the time-frequency plane depends of the transformation characteristics.

The ability of the time domain approach to design systems with arbitrary nonuniform decompositions makes it possible to create a very large family of overlapped block transforms with any desired time and frequency resolution (within the constraints of the uncertainty principle). Because of its ability to design fixed lapped transform systems, and because the DTWT is a fixed lapped transform, it can be used to design wavelets with constrained regularity and phase properties. The relationship between compactly supported wavelets and filter banks is described in recent literature including [12,2]. In this next section, a general view of analysis-synthesis systems as lapped transforms is presented. It is shown that the STFT and the DTWT are special cases of this general transform. In the remainder of the paper, a summary of the time domain framework and its applications to the design of wavelets and DTWT systems are presented.

2 Generalized Lapped Transforms

In the DSTFT, the signal is typically viewed through a lowpass analysis window $h(n)$. To analyze the entire signal, the window is shifted in time and the DFT of the windowed signal is computed. In the resulting representation, the signal is mapped onto time-frequency plane in which the time and frequency resolution is fixed. The resolution in the time-frequency plane is determined by the amount of the window shift, the shape of the window function and the length of the DFT. From a filter bank point of view, the DSTFT is the decimated output of a bank of complex modulated filters whose prototype filter impulse response is the analysis window $h(n)$. From a signal decomposition point of view, the DSTFT may be interpreted as the decomposition of the signal into different sub-signals using a set of basis vectors and their time-shifted versions. In the DSTFT, these basis vectors are defined by the analysis window and its frequency modulated versions.

The DTWT offers an alternative time-frequency (known as time-scale) representation in which the resolution in the time-frequency plane is not uniform. As opposed to the STFT in which the signal is viewed from a window of fixed size and shape, the DTWT is based on a series of different windows which are related to a continuous wavelet function $w(t)$.

Each window is of different length and corresponds to a bandpass filter with a center frequency at $\frac{3\pi}{2^{j+2}}$ for $j = 0, 1, \ldots$ In the analysis of the signal, the signal is decomposed into different sub-signals at different resolutions. From a sliding window perspective, the sliding rate of different windows depends on their length. Thus, shorter windows are moved more frequently than longer ones. This results in a time-frequency representation that has high temporal resolution in high frequencies and low temporal resolution (high frequency resolution) in low frequencies. From a filter bank perspective, the DTWT is an octave-band tree-structure in which the low frequency band is further divided into high and low bands and downsampled.

Both the DSTFT and the DTWT can be considered to be special cases of a more general time-frequency transformation which we will refer to as the Generalized Lapped Transform (GLT). The GLT includes DSTFT's, uniform band filter banks, DSTCTs, LOT's, and DTWT's as special cases. These are all examples of fixed lapped transforms, where the basis vectors do not change with time. The GLT also can include systems in which the basis vectors change with time. The time domain design procedure can be used to design any GLT.

In the generalized lapped transform, the characteristics of analysis windows and the resolution in the time-frequency plane are only restricted by the reconstruction requirement of the input signal (equivalently, the invertibility of the transformation). As in the DSTFT and the DTWT, the GLT can be implemented in a filter bank structure whose general form is presented in Figure 1. In such a transform, the input signal is decomposed into a set of sub-signals with different resolutions. To achieve such a multi-resolution time-frequency representation of the signal, the signal is passed through a set of upsamplers before the transformation is applied. The redundancy of the representation is reduced by downsampling the sub-signals to appropriate rates. The basis vectors are designed to have the required time and frequency characteristics.

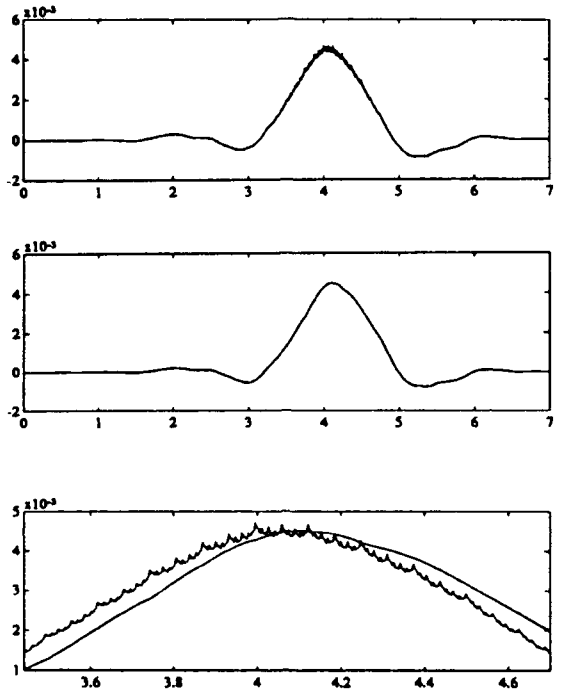


Figure 2: *(a) Scaling function, ϕ_0, generated from CQF-8, (b) Scaling function ϕ_1, (c) A closer view of ϕ_0 and ϕ_1 in the range [3.5,4.7].*

3 Time Domain Framework

The basic idea behind the time domain formulation of analysis-synthesis structures is to determine the set of necessary and sufficient conditions for exact reconstruction of the input at the system output. These conditions are expressed in a proper matrix product form which is used in the design procedure. The conditions can be expressed as

$$\mathbf{A}_i\mathbf{s}_i = \mathbf{b}_i \qquad i = 0, 1, \ldots, T-1, \tag{1}$$

where T is the shift-invariance period of the system. The elements of matrix $\mathbf{A}_i$ are related to the analysis basis vectors and elements of $\mathbf{s}_i$ are related to the synthesis basis vectors [13]. The vector $\mathbf{b}_i$ is constant.

After expressing the reconstruction conditions of the system (i.e. the invertibility conditions of the transform) in a matrix form, a cost function is defined as

$$\epsilon = \sum_{i=0}^{T-1} \|\mathbf{A}_i\mathbf{s}_i - \mathbf{b}_i\|. \tag{2}$$

This cost function is minimized (and brought to zero for perfect reconstruction, if possible) using an optimization routine. Other system constraints which can not be directly incorporated into the formulation (such as frequency responses) can also be added to the cost function for minimization. In the next section, the design approach is applied to the design of compactly supported wavelets and DTWT systems.

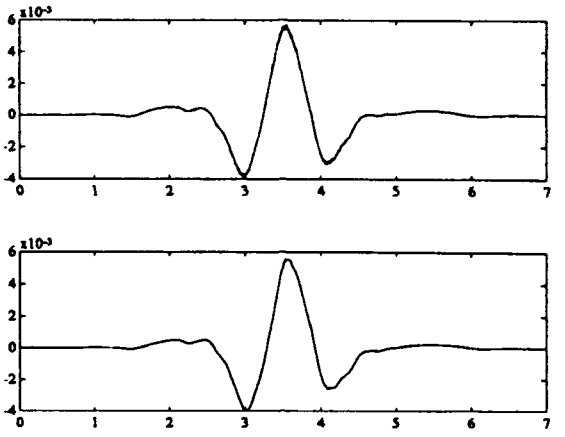


Figure 3: *(a) Wavelet, ψ_0, generated from CQF-8, (b) Wavelet ψ_1.*

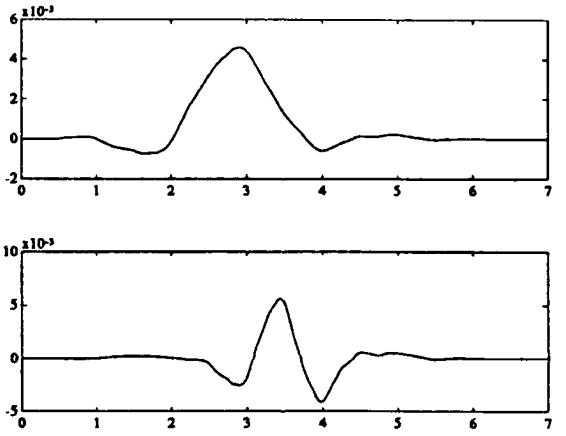


Figure 4: *(a) Synthesis scaling function, (b) Synthesis wavelet, which corresponds to ϕ_1 and ψ_1.*

4 Wavelet Design

In this section, the flexibility of the time domain design procedure in designing compactly supported wavelets is illustrated. Examples are designed to exhibit the flexibility of the design approach in imposing regularity condition on the wavelet, low delay or linear phase property on the wavelet, and their effects on the wavelet. Both orthogonal and biorthogonal wavelets are designed and presented. To design orthogonal wavelets, the synthesis filter coefficients are chosen to be the time-reversed version of the analysis filters. In designing biorthogonal wavelets, the synthesis filters are chosen to be

$$g_0(n) = h_1(n)(-1)^n \tag{3}$$
$$g_1(n) = -h_0(n)(-1)^n. \tag{4}$$

First, we consider the imposition of regularity constraints on the wavelet. Figure 2a shows the generated scaling function from the 8-tap Conjugate Quadrature

Filter (CQF-8). This scaling function is referred to as ϕ_0 in this paper. The relative irregularity of this function and its corresponding wavelet (ψ_0) shown in Figure 3a is obvious. Scaling functions are generated by iterating the system lowpass filter eight times. To impose regularity on the wavelet, a zero at $z = -1$ is imposed on the lowpass analysis filter, $h_0(n)$. To impose similar regularity on the synthesis side, a zero at $z = 1$ is imposed on the highpass filter, $h_1(n)$. Figure 2b shows the scaling function of this system, referred to as ϕ_1, which is much smoother than ϕ_0. Figure 3b shows the corresponding relatively regular wavelet, ψ_1. Figure 2c shows a closer view of both the scaling functions in the range [3.5,4.7]. Since the synthesis wavelet is different from the analysis wavelet, Figure 4 show the synthesis scaling function and wavelet. To make the wavelet even more regular, two zeros are located at $z = -1$ for $h_0(n)$ and two zeros at $z = 1$ for $h_1(n)$. The resulting scaling function and wavelet are referred to as ϕ_2 and ψ_2, respectively.

To compare the regularity properties of the three wavelets, the spectra of the three scaling functions, $\Phi_0(\omega)$, $\Phi_1(\omega)$, and $\Phi_2(\omega)$, are compared. In Figure 5, Φ_0 and Φ_1 are shown side-by-side for comparison. As seen from this figure, the high frequency components of Φ_1 are significantly smaller than those of Φ_0 which results in a smoother wavelet function. Figure 6 compares Φ_1 and Φ_2. As seen in this figure, since Φ_2 is generated from a lowpass filter with two zeros at $z = -1$, it has a higher decay rate than Φ_1.

To understand the effects of imposing zeros at $z = -1$ to achieve regularity, Figure 7 shows the log magnitude response of lowpass filters used to generate ϕ_0 and ϕ_1. Having the same passband and transition characteristics, a few dB loss in the stopband attenuation is the result of imposing one of the zeros to be at -1.

To exhibit the flexibility of the design approach, three more design examples are shown. Figure 8 shows the analysis and the synthesis wavelets with linear phase. These correspond to a two-band system with 16-tap system filters and two zeros at $z = -1$ for both analysis and synthesis sections. Orthogonal wavelets can also be easily designed. Figure 9a shows the scaling function and Figure 9b shows the corresponding orthogonal wavelet function. The regularity in this case in achieved by imposing two zeros at $z = -1$ for $h_0(n)$. In the last example, the design approach is applied to design low delay wavelet decomposition-reconstruction systems. In these systems, both analysis and synthesis wavelets have low group delay which results in a lower total system delay. Figure 10 shows the analysis and synthesis wavelets of such a system. These wavelets correspond to a two-band system with 16-tap filters and only 10 samples of system delay. This delay is normally 15 samples for CQF and QMF based systems. In this example, two zeros are imposed at $z = -1$ for $h_0(n)$ and at $z = 1$ for $h_1(n)$.

5 Conclusion

In this paper, the generalized lapped transform in its general form is introduced. It is shown that the many well known transforms, including the wavelet transform, are special cases of the GLT. The time domain formulation of the general analysis-synthesis systems based on FIR filters is used to design wavelets with different constraints. Constraints included the regularity, a linear phase, a low delay, and the orthogonality. All these can be imposed using the same design procedure.

References

[1] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time fourier analysis," *IEEE Transactions ASSP*, pp. 55–69, February 1980.

[2] M. Vetterli and C. Herley, "Wavelets and Filter Banks: Relationships and New Results," *Proceedings ICASSP*, pp. 1723 - 1726, April 1990.

[3] H. S. Malvar and D. H. Staelin, "The LOT: Transform Coding without Blocking Effects," *IEEE Transactions on ASSP*, pp. 553–559, April 1989.

[4] M. J. T. Smith and T. P. Barnwell, "Exact reconstrcution techniques for tree-structured subband coders," *IEEE Transaction ASSP*, pp. 434–411, June 1986.

[5] J. P. Princen and A. B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," *IEEE Transactions ASSP*, vol. ASSP-34, October 1986.

[6] K. Nayebi, T. P. Barnwell, and M. J. T. Smith, "The Time Domain Analysis and Design of Exactly Reconstructing FIR Analysis/Synthesis Filter Banks," *Proceedings ICASSP*, pp. 1735–1738, April 1990.

[7] P. P. Vaidyanathan, "Theory and design of M channel maximally decimated QMF with arbitrary M, having perfect reconstruction property," *IEEE Transactions on ASSP*, April 1987.

[8] M. Vetterli and D. L. Gall, "Perfect Reconstruction FIR Filter Banks: Lapped Transforms, Pseudo QMF's and Paraunitary Matrices," *Proceedings ISCAS*, pp. 2249–2253, 1988.

[9] K. Nayebi, T. P. Barnwell, and M. J. T. Smith, "General Time Domain Analysis and Design Framework for Exacly Reconstructing FIR Analysis/Synthesis Filter Banks," *Proceedings ISCAS*, pp. 2022–2025, May 1990.

[10] K. Nayebi, T. P. Barnwell, and M. J. T. Smith, "Time domain filter bank analysis: A new design theory," *To be published Trans. on ASSP.*, June 1992.

[11] K. Nayebi, T. P. Barnwell, and M. J. T. Smith, "Design of low delay FIR analysis-synthesis filter bank systems," *Proc. Conf. on Information Sciences and Systems*, 1991.

[12] I. Duabechies, "Orthonormal bases of compactly supported wavelets," *Comm. in Pure and Applied Math.*, vol. 41, pp. 909–996, 1988.

[13] K. Nayebi, T. P. Barnwell, and M. J. T. Smith, "The design of perfect reconstruction nonuniform band filter banks," *Proceedings ICASSP*, 1991.

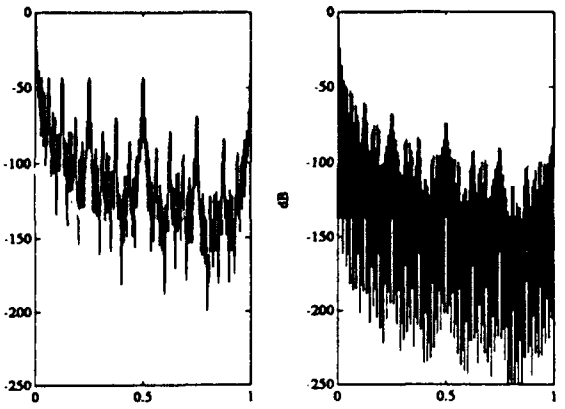


Figure 5: *Magnitude of $\Phi_0(\omega)$ (left) and $\Phi_1(\omega)$ (right) in dB.*

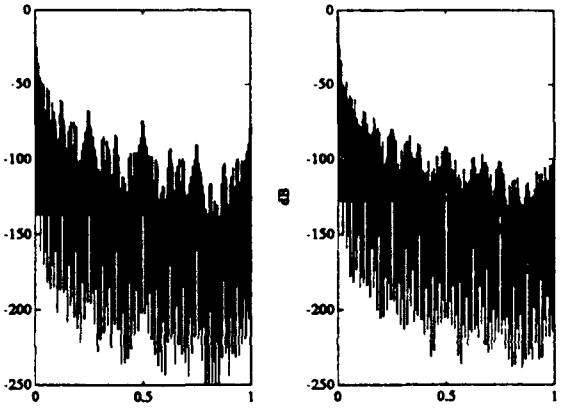


Figure 6: *Magnitude of $\Phi_1(\omega)$ (left) and $\Phi_2(\omega)$ (right) in dB.*

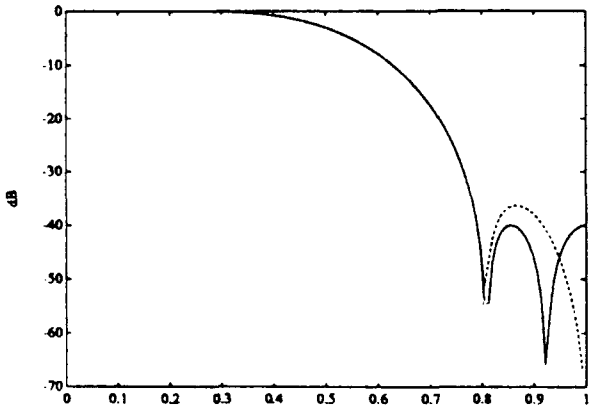


Figure 7: *Frequency response of the lowpass analysis filter with no zeros at $z = -1$ (full line) and one zero at $z = -1$ (dashed line).*

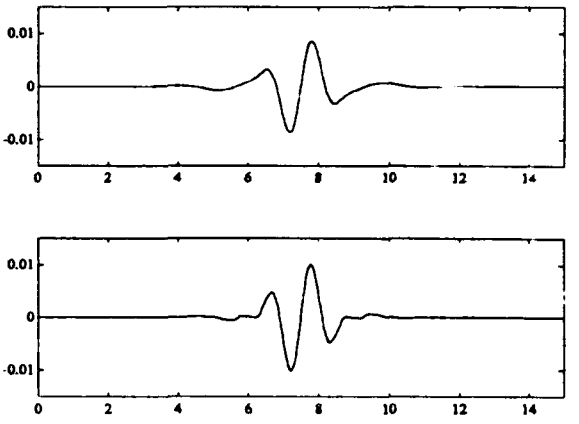


Figure 8: *(a) Analysis wavelet, (b) Synthesis wavelet with linear phase generated from 16-tap filters with two zeros at $z = -1$.*

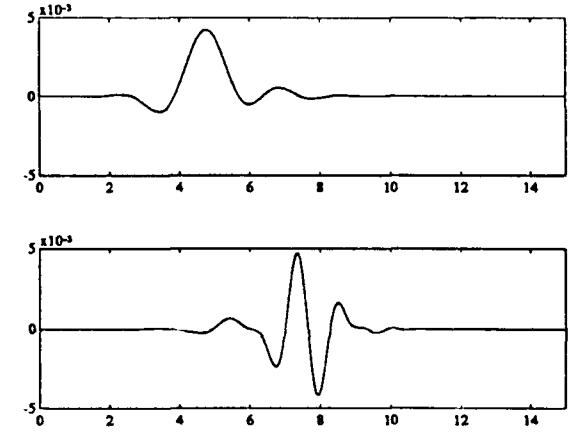


Figure 9: *(a) Scaling function, (b) Orthonormal wavelet generated from 16-tap orthogonal filters with two zeros at $z = -1$.*

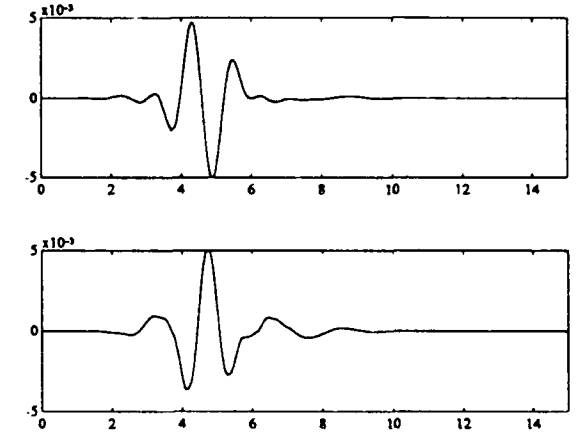


Figure 10: *(a) Analysis low delay wavelet, (b) Synthesis low delay wavelet generated from 16-tap filters of a 10 sample delay system with two zeros at $z = -1$.*

Irregular Sampling and the Theory of Frames, I

BY

JOHN J. BENEDETTO[1] AND WILLIAM HELLER

Prometheus Inc.
21 Arnold Ave.
Newport, RI 02840

Dept. of Math.
Univ. of Maryland
College Park, MD 20742

Invited Paper[2]

Dedicated to the Memory of

Professor Gottfried Köthe

Abstract. Irregular sampling expansions are proved in an elementary way by an analysis of the inverse frame operator. The expansions are of two dual types: in the first, the sampled values at irregularly spaced points are the coefficients; in the second, the sequence of sampling functions are irregularly spaced translates of a single sampling function. The results include regular sampling theory as well as the irregular sampling theory of Paley-Wiener, Levinson, Beutler, and Yao-Thomas. The use of frames also gives rise to a new interpretation of aliasing,

cf., J. Benedetto, The theory of frames and irregular sampling, A Tutorial in Wavelet Theory, C. Chui, ed., Academic Press 1992,

for relations with Beurling density and quadrature mirror filters.

[1]The first named author is also Professor of Mathematics at the University of Maryland, College Park, MD 20742. His work was supported in part by DARPA Contract DAAH01-90-C-0667.

[2]To appear in a special volume of **Note Mat.** dedicated to the memory of Professor Köthe. (1990)

Irregular Sampling and the Theory of Frames, I

BY

JOHN J. BENEDETTO[1] AND WILLIAM HELLER

Prometheus Inc.
21 Arnold Ave.
Newport, RI 02840

Dept. of Math.
Univ. of Maryland
College Park, MD 20742

Dedicated to the memory of Professor Gottfried Köthe

Abstract. Irregular sampling expansions are proved in an elementary way by an analysis of the inverse frame operator. The expansions are of two dual types: in the first, the sampled values at irregularly spaced points are the coefficients; in the second, the sequence of sampling functions are irregularly spaced translates of a single sampling function. The results include regular sampling theory as well as the irregular sampling theory of Paley-Wiener, Levinson, Beutler, and Yao-Thomas. The use of frames also gives rise to a new interpretation of aliasing.

1 Introduction.

The subject of sampling, whether as method, point of view, or theory, weaves its fundamental ideas through a panorama of engineering, mathematical, and scientific disciplines. Sampling is so pervasive that excellent expositions and surveys abound; [BSS] and [Hi2] are two such papers that are particularly appropriate for our perspective. Alas, our contributions focus on an important result by Köthe [K] (1936), on a new look at Duffin and Schaeffer's theory of frames [DS] (1952) in light of the emergence of wavelet theory, and on effective, elementary, and unifying methods for irregular sampling in terms of frames.

Köthe was the first to prove that all bounded unconditional bases are equivalent in a given separable Hilbert space. An explanation of this result and its relationship with the theory of frames are the content of **Theorem 2.5**. **Section 2** presents a crisp compendium of frame theory with **Theorem 2.5** as its focal point. To titillate the reader during this

[1]The first named author is also Professor of Mathematics at the University of Maryland, College Park, MD 20742. His work was supported in part by DARPA Contract DAAH01-90-C-0667.

dry compilation, we've pointed out yet another "first" by Vitali [V] in **Remark 2.3**. The technical device we extricate from **Section 2** is the inverse frame operator S^{-1} for weighted Fourier frames associated with the lattice $\{(na, mb)\}$, e.g., **Definition 2.6** and **Theorems 2.7** and **2.8**; and this operator is our basic tool in proving regular sampling theorems.

Section 3 is devoted to classical regular sampling expansions of the form,

$$f(t) = \sum_{n=-\infty}^{\infty} f(nT)\, s(t - nT), \tag{1}$$

where $T > 0$ is the sampling rate, $\{f(nT)\}$ is tne set of regularly sampled values of the signal f, and s is the sampling function. The point of **Section 3** is to prove (1) quickly in terms of frame decompositions. Regular sampling involves orthonormal bases of exponentials, and S^{-1} is used as a multiplier, e.g., **Theorem 3.1** and **Theorem 3.3**. An important consequence of this line of thinking is a new interpretation and explanation of aliasing in terms of frames, e.g., **Section 3.2**.

Since our main goal is to prove irregular sampling expansions analogous to (1), we develop the theory of weighted Fourier frames associated with irregular "lattices" $\{(a_n, b_m)\}$ in **Section 4**.

Our results on irregular sampling are the subject of **Sections 5** and **6**. In **Section 5**, irregularly sampled values of f are used in the expansions analogous to (1). In **Section 6**, irregular translates of a single sampling function are used in the expansions analogous to (1). These two expansions are dual in the context of frame theory in a way that is explained in the text. The results in **Section 5** use special frames associated with Köthe's work, and include the completeness and sampling theory of Paley–Wiener, Levinson, Beutler, and Yao–Thomas. The results in **Section 6** use ordinary frames, and lead to an algorithm providing insight into the role of irregularly sampled values for the expansions of this section. The irregular sampling of **Section 5** involves bounded unconditional bases, and S^{-1} is used in terms of biorthonormality, cf., our remark above, about **Section 3**, on the role of S^{-1}.

We indicated at the outset that sampling ideas have diverse theoretical foundations and catholic applicability. As such, the sequel to this paper has two components. First, there is

a critical comparison in **Part II** of other approaches to irregular sampling, cf., the analysis by Feichtinger and Gröchenig [FG]. Second, as regards applicability, **Part II** contains results dealing with aliasing, the algorithm, stability, and higher dimensions, all in the context of our frame theoretic approach. We have already indicated our technical direction for aliasing and the algorithm in **Sections 3 and 6**, respectively. In **Part II**, the aliasing method is fully developed for the irregular sampling case, and an error analysis is conducted on the algorithm for various truncations of the inverse frame operator. Our approach to stability builds on the ideas of Yao and Thomas [YT], and ties in with the results of Beurling and Malliavin [BM] and Landau [La]. Our approach to higher dimensions is direct.

Besides the usual notation in analysis as found in the books by Hörmander [Hö], Schwartz [S], and Stein and Weiss [SW], we shall use the conventions and notation described at the end of the paper.

Finally, in this paper we have only proved convergence in the L^2 norm. All of our results have been proved for other modes of convergence, and details are found in [H]. Also, we have dealt exclusively with bandlimited sampling functions.

2 Riesz bases and frames.

2.1 DEFINITION. a. A sequence $\{g_n\} \subseteq H$, a separable Hilbert space, is a **frame** if there exist $A, B > 0$ such that

$$\forall\, f \in H, \quad A\|f\|^2 \leq \sum |\langle f, g_n\rangle|^2 \leq B\|f\|^2,$$

where $\langle\ ,\ \rangle$ is the inner product on H and the norm of $f \in H$ is $\|f\| = \langle f, f\rangle^{1/2}$. A and B are the **frame bounds**, and a frame $\{g_n\}$ is **tight** if $A = B$. A frame $\{g_n\}$ is **exact** if it is no longer a frame when any one of its elements is removed. Clearly, if $\{g_n\}$ is an orthonormal basis of H then it is a tight exact frame with $A = B = 1$.

b. The **frame operator** of the frame $\{g_n\}$ is the function $S : H \to H$ defined as $Sf = \sum \langle f, g_n\rangle g_n$.

The theory of frames is due to Duffin and Schaeffer [DS] in 1952. Expositions include [Y] and [HW], the former presented in the context of non-harmonic Fourier series and the latter in the setting of wavelet theory.

2.2 THEOREM. *Let $\{g_n\} \subseteq H$ be a frame with frame bounds A and B.*

a. S is a topological isomorphism with inverse $S^{-1} : H \to H$. $\{S^{-1}g_n\}$ is a frame with frame bounds B^{-1} and A^{-1}, and

$$\forall\, f \in H, \quad f = \sum \langle f, S^{-1}g_n \rangle\, g_n = \sum \langle f, g_n \rangle\, S^{-1}g_n.$$

The first expansion is the **frame expansion** *and the second is the* **dual frame expansion**.

b. If $\{g_n\}$ is tight, $\|g_n\| = 1$ for all n, and $A = B = 1$, then $\{g_n\}$ is an orthonormal basis of H.

c. If $\{g_n\}$ is exact, then $\{g_n\}$ and $\{S^{-1}g_n\}$ are biorthonormal, i.e.

$$\forall\, m, n\ , \quad \langle g_m, S^{-1}g_n \rangle = \delta_{mn}.$$

2.3 REMARK. We comment on part **b** because it is surprisingly useful and because of a stronger result by Vitali (1921) [V].

To prove **b** we first use tightness and $A = 1$ to write,

$$\|g_m\|^2 = \|g_m\|^4 + \sum_{n \neq m} |\langle g_m, g_n \rangle|^2;$$

and obtain that $\{g_n\}$ is orthonormal since each $\|g_n\| = 1$. To conclude the proof we then invoke the well-known result: if $\{g_n\} \subseteq H$ is orthonormal then it is an orthonormal basis of H if and only if

$$\forall\, f \in H, \quad \|f\|^2 = \sum |\langle f, g_n \rangle|^2.$$

In 1921, Vitali proved that an orthonormal sequence $\{g_n\} \subseteq L^2[a, b]$ is complete, and so $\{g_n\}$ is an orthonormal basis, if and only if

$$\forall\, t \in [a, b], \quad \sum \Big| \int_a^t g_n(u)\, du \Big|^2 = t - a. \tag{2.1}$$

For the case $H = L^2[a, b]$, Vitali's result is stronger than part **b** since (2.1) is tightness with $A = 1$ for functions $f = \mathbf{1}_{[a,t]}$.

Other remarkable and important contributions by Vitali are highlighted in [B].

2.4 DEFINITION. Let H be a separable Hilbert space. A sequence $\{g_n\} \subseteq H$ is a **Schauder basis** or **basis** of H if each $f \in H$ has a unique decomposition $f = \sum c_n(f)\, g_n$. A basis $\{g_n\}$ is an **unconditional basis** if

$$\exists\, C \text{ such that } \forall\, F \subseteq \mathbf{Z}, \text{ where card } F < \infty, \text{ and}$$
$$\forall\, b_n, c_n \in \mathbf{C}, \text{ where } n \in F \text{ and } |b_n| \leq |c_n|,$$

$$\| \sum_{n \in F} b_n g_n \| \leq C \| \sum_{n \in F} c_n g_n \|.$$

An unconditional basis $\{g_n\}$ is **bounded** if

$$\exists\, A, B > 0 \text{ such that } \forall\, n,\ A \leq \|g_n\| \leq B.$$

Separable Hilbert spaces have orthonormal bases, and orthonormal bases are bounded unconditional bases.

Köthe's result mentioned in **Section 1** is the implication, **b** implies **c**, of the following theorem. The implication, **c** implies **b**, is straightforward; and the equivalence of a and c is found in [Y, pp. 188-189].

2.5 THEOREM. *Let H be a separable Hilbert space and let $\{g_n\} \subseteq H$ be a given sequence. The following are equivalent:*

a. *$\{g_n\}$ is an exact frame for H;*

b. *$\{g_n\}$ is a bounded unconditional basis of H;*

c. *$\{g_n\}$ is a* **Riesz basis**, *i.e., there is an orthonormal basis $\{u_n\}$ and a topological isomorphism $T : H \to H$ such that $Tg_n = u_n$ for each n.*

2.6 DEFINITION/REMARK. a. Given $g \in L^2(\mathbf{R})$ and sequences $\{a_n\}, \{b_m\} \subseteq \mathbf{R}$. Define $(T_{a_n} g)(t) = g(t - a_n)$ and $E_{b_m}(t) = e^{2\pi i t b_m}$. If $\{E_{b_m} T_{a_n} g\}$ is a frame for $L^2(\mathbf{R})$ it is called a **weighted Fourier frame** with weight g.

b. **Fourier frames** $\{E_{b_m}\}$ were defined in [DS] for $L^2[-T, T]$. Gabor's seminal paper [G] deals with "regularly latticed" systems $\{E_{mb} T_{na} g\}$, where g is the Gaussian; and it turns out that the Heisenberg group is fundamental in analyzing the structure of modulations

and translations. As such, the names "Gabor" and "Weyl-Heisenberg" have also been associated with these systems in the case of regular lattices.

c. $\{E_{b_m}T_{a_n}g\}$ is a frame for $L^2(\mathbf{R})$ if and only if $\{T_{a_n}(E_{b_m}g)\}$ is a frame for $L^2(\mathbf{R})$.

Also, our weighted Fourier frames will often be defined for $L^2(\hat{\mathbf{R}})$. As such we note that

$$(E_{a_n}T_{b_m}\hat{g})^\vee = e^{2\pi i a_n b_m}E_{b_m}T_{-a_n}g.$$

2.7 THEOREM. *Given $g \in L^2(\mathbf{R})$ and $a, b > 0$. Define*

$$G(t) = \sum |g(t - na)|^2.$$

Assume that there exist $A, B > 0$ such that

$$0 < A \le G(t) \le B < \infty \quad a.e. \ on \ \mathbf{R}, \tag{2.2}$$

and that $\operatorname{supp} g \subseteq I$ where I is an interval of length $1/b$. Then $\{E_{mb}T_{na}g\}$ is a frame for $L^2(\mathbf{R})$, with frame bounds $b^{-1}A$ and $b^{-1}B$, and

$$\forall\, f \in L^2(\mathbf{R}), \quad S^{-1}f = \frac{bf}{G}. \tag{2.3}$$

2.8 THEOREM. *Given $g \in L^2(\mathbf{R})$ and $a, b > 0$. Assume $\{E_{na}T_{mb}\hat{g}\}$ is a frame for $L^2(\hat{\mathbf{R}})$. Then*

$$S^{-1}(E_{na}T_{mb}\hat{g}) = E_{na}T_{mb}S^{-1}\hat{g}. \tag{2.4}$$

2.9 EXAMPLE. a. Given $g \in L^2(\mathbf{R})$ and $a, b > 0$ for which $ab = 1$. *If $\{E_{mb}T_{na}g\}$ is a frame then it is an exact frame.* This remarkable fact (for $ab = 1$) can be proved using properties of the Zak transform which we now define.

b. The **Zak transform** of $f \in L^2(\mathbf{R})$ is

$$Zf(x,\omega) = a^{1/2}\sum f(xa + ka)e^{2\pi i k\omega}$$

for $(x,\omega) \in \mathbf{R} \times \hat{\mathbf{R}}$ and $a > 0$. It turns out that the Zak transform is a unitary map of $L^2(\mathbf{R})$ onto $L^2(Q)$, $Q = [0,1) \times [0,1)$.

c. If $\{E_{mb}T_{na}g\}$ is a frame for $ab = 1$, it is a bounded unconditional basis (part a and **Theorem 2.5**); and, in particular, the frame decomposition

$$\forall\, f \in L^2(\mathbf{R}), \quad f = \sum c_{m,n}\, E_{mb}T_{na}g$$

(**Theorem 2.2a**) is unique. We shall verify that

$$\begin{aligned} c_{m,n} &= \langle f, S^{-1}E_{mb}T_{na}g\rangle \\ &= \int_0^1 \int_0^1 \frac{Zf(x,\omega)}{Zg(x,\omega)} e^{-2\pi imx} e^{-2\pi in\omega}\, dx\, d\omega. \end{aligned} \tag{2.5}$$

First, with the hypotheses that $\{E_{mb}T_{na}g\}$ is a frame for $L^2(\mathbf{R})$ and $ab = 1$, we compute

$$\forall\, F \in L^2(Q), \quad S_Z F = F|Zg|^2,$$

where $S_Z : L^2(Q) \to L^2(Q)$ is the frame operator for the frame $\{Z(E_{mb}T_{na}g)\}$. Thus,

$$\forall\, f \in L^2(\mathbf{R}), \quad S_Z^{-1}(Zf) = \frac{Zf}{|Zg|^2}. \tag{2.6}$$

Next, using (2.6), we compute

$$\begin{aligned} \forall\, f \in L^2(\mathbf{R}), \quad Zf &= S_Z S_Z^{-1}(Zf) \\ &= \sum \langle \frac{Zf}{Zg}, E_{m,n}\rangle E_{m,n} Zg, \end{aligned}$$

where $E_{m,n}(x,\omega) = e^{2\pi imx}e^{2\pi in\omega}$. Consequently,

$$\begin{aligned} \forall\, f \in L^2(\mathbf{R}), \quad f &= \sum \langle \frac{Zf}{Zg}, E_{m,n}\rangle Z^{-1}(E_{m,n}Zg) \\ &= \sum \langle \frac{Zf}{Zg}, E_{m,n}\rangle E_{mb}T_{na}g, \end{aligned}$$

so that (2.5) is obtained by the uniqueness of the representation.

3. Regular sampling and weighted Fourier frames.

The theme of this section is to prove classical sampling results by frame methods in the case that the inverse frame operator S^{-1} is a multiplier.

The **Paley–Wiener space**, PW_Ω, is the subset of $L^2(\mathbf{R})$ whose elements are Ω–bandlimited, i.e.

$$PW_\Omega = \{f \in L^2(\mathbf{R}) : supp\, \hat{f} \subseteq [-\Omega, \Omega]\}.$$

Clearly the elements of PW_Ω are entire functions.

3.1 THEOREM. *Given $T, \Omega > 0$ for which $0 < T \leq \frac{1}{2\Omega}$. Then*

$$\forall\, f \in PW_\Omega, \quad f = T \sum f(nT)\, T_{nT}\, d_{2\pi\Omega} \quad \text{in } L^2(\mathbf{R}), \tag{3.1}$$

where $d_{2\pi\Omega}$ is the $2\pi\Omega$ dilation of the Dirichlet function

$$d(t) = \frac{\sin t}{\pi t},$$

where $f(nT)$ is the value of f at $nT \in \mathbf{R}$, and where $T_{nT}\, d_{2\pi\Omega}$ is the translation

$$T_{nT}\, d_{2\pi\Omega}(t) = d_{2\pi\Omega}(t - nT) = \frac{\sin 2\pi\Omega\,(t - nT)}{\pi(t - nT)}.$$

PROOF: Let $g = \frac{1}{(2\Omega)^{1/2}}\, d_{2\pi\Omega}$ so that $\hat{g} = \frac{1}{(2\Omega)^{1/2}}\, \mathbf{1}_{(\Omega)}$ and $\|g\|_2 = 1$. Set $a = T$ and $b = 2\Omega$ so that $ab = 2T\Omega \leq 1$. Note that

$$\sum |\hat{g}(\gamma - mb)|^2 = \frac{1}{2\Omega} \quad a.e.,$$

$supp\, \hat{g} \subseteq [-\Omega, \Omega]$, and $|[-\Omega, \Omega]| \leq 1/a$. Thus, by **Theorem 2.7**, $\{E_{na} T_{mb} \hat{g}\}$ is a frame. Consequently, by **Theorem 2.2a** and **Theorem 2.8**,

$$\forall\, f \in L^2(\mathbf{R}), \quad \hat{f} = \sum \langle \hat{f}, E_{na} T_{mb} S^{-1} \hat{g} \rangle\, E_{na} T_{mb} \hat{g} \quad \text{in } L^2(\hat{\mathbf{R}}). \tag{3.2}$$

Since $supp\, \hat{g}$ is compact, we have

$$\forall\, h \in L^2(\hat{\mathbf{R}}), \quad S^{-1} \hat{h} = 2T\Omega\, \hat{h}$$

by **Theorem 2.7**; and, hence, (3.2) becomes

$$(3.3) \qquad \forall f \in L^2(\mathbf{R}), \quad f = 2T\Omega \sum \langle \hat{f}, E_{na} T_{mb} \hat{g} \rangle \, T_{-na} E_{mb} g \quad \text{in } L^2(\mathbf{R}).$$

If $f \in PW_\Omega$ then

$$(3.4) \qquad \langle \hat{f}, E_{na} T_{mb} \hat{g} \rangle = \begin{cases} \dfrac{1}{(2\Omega)^{1/2}} f(-nT), & \text{for } m = 0 \\ 0, & \text{for } m \neq 0. \end{cases}$$

The sampling formula, (3.1), follows from (3.3) and (3.4). ∎

The hypothesis, that $f \in PW_\Omega$, was essential in both parts of (3.4); and the above proof shows that only "t–information" (i.e., $m = 0$) is required in this case. When f is not Ω-bandlimited so that aliasing occurs, phase information contributed by $m \neq 0$ is required in the frame decomposition of a signal. To quantify this remark, we define the **aliasing pseudomeasure**, $\alpha_{t,\Omega}$, on $\mathbf{R}$ as the distributional Fourier transform, $\alpha_{t,\Omega} = A^{\vee}_{t,\Omega}$, where each t is fixed and

$$A_{t,\Omega} \equiv \sum (e^{2\pi i (2m\Omega t)} - 1)(T_{2m\Omega} \mathbf{1}_{(\Omega)}) \in L^\infty(\hat{\mathbf{R}}).$$

3.2 CALCULATION/DEFINITION. Let $f \in L^2(\mathbf{R})$ and assume $2T\Omega = 1$. Writing (3.3) as a sum, $\sum_{m=0,n} + \sum_{m\neq 0,n}$, we compute

$$(3.5) \qquad f(t) = T \sum f(nT) \, T_{nT} \, d_{2\pi\Omega}(t) + T \sum (f * \alpha_{t,\Omega})(nT) \, T_{nT} \, d_{2\pi\Omega}(t).$$

The **aliasing error** of f at t for the low pass filter $d_{2\pi\Omega}$ is

$$ae(f,t) = T \sum (f * \alpha_{t,\Omega})(nT) \, T_{nT} \, d_{2\pi\Omega}(t).$$

Formally, standard calculations give

$$(3.6) \qquad \|ae(f,\cdot)\|_\infty \leq 2 \int_{|\gamma| \geq \Omega} |\hat{f}(\gamma)| \, d\gamma.$$

In the following result we use sampling kernels s with more rapid decay than $d_{2\pi\Omega}$. The goal is better computational efficiency for low pass filters; the price to be paid is more sampling.

3.3 THEOREM. *Given $T, \Omega > 0$, for which $0 < T < \frac{1}{2\Omega}$, and $g \in \mathcal{S}(\mathbf{R})$ with the properties that $supp\, \hat{g} \subseteq [\frac{-1}{2T}, \frac{1}{2T}]$, $\hat{g} = 1$ on $[-\Omega, \Omega]$, and $\hat{g} > 0$ on $(\frac{-1}{2T}, -\Omega] \cup [\Omega, \frac{1}{2T})$. Set*

$$G(\gamma) = \sum |\hat{g}(\gamma - mb)|^2 \quad \text{and} \quad s(t) = (\frac{\hat{g}}{G})^{\vee}(t),$$

where $\Omega + \frac{1}{2T} \leq b < \frac{1}{T}$. Then $0 < A \leq G(\gamma) \leq B < \infty$, $s \in \mathcal{S}(\mathbf{R})$, $supp\, \hat{s} = supp\, \hat{g}$, $\hat{s} = \frac{1}{G}$ on $[-\Omega, \Omega]$, and

$$\forall\, f \in PW_{\Omega}, \quad f = T \sum f(nT)\, T_{nT}\, s \ \text{ in } L^2(\mathbf{R}). \tag{3.7}$$

PROOF: The assertions about G and s follow from our choice of b.

Set $a = T$ so that $|supp\, \hat{g}\,| = 1/a$. Thus, using the fact, $A \leq G(\gamma) \leq B$, and **Theorem 2.7**, we see that $\{E_{na} T_{mb} \hat{g}\}$ is a frame. Since $supp\, \hat{g}$ is compact, we have

$$\forall\, h \in L^2(\mathbf{R}), \quad S^{-1}\hat{h} = T\frac{\hat{h}}{G}$$

by **Theorem 2.7**; and, hence, we have the frame decomposition

$$\forall\, f \in L^2(\mathbf{R}), \quad \hat{f} = T \sum_{m,n} \langle \hat{f}, E_{na} T_{mb} \hat{g} \rangle E_{na} T_{mb}\, \hat{s}, \tag{3.8}$$

where we have used the fact that $S^{-1}(E_{na} T_{mb} \hat{g}) = E_{na} T_{mb} S^{-1} \hat{g}$ (**Theorem 2.8**).

If $f \in PW_{\Omega}$, then (3.4) is again valid since $\hat{g} = 1$ on $[-\Omega, \Omega]$. The sampling formula (3.7) follows from (3.4) and (3.8). ∎

3.4 EXAMPLE. a. In **Theorem 3.1**, $\{E_{na} T_{mb} \hat{g}\}$ is a tight frame with frame bounds $A = B = 1$ in the case $2T\Omega = 1$, where $a = T$ and $b = 2\Omega$. Clearly, $\langle E_{na} T_{mb} \hat{g}, E_{qa} T_{pb} \hat{g} \rangle$ is 1 if $(m, n) = (p, q)$ and is 0 if $m \neq p$. If $m = p$ and $n \neq q$ then this inner product is

$$\frac{e^{2\pi i (2T\Omega) m (n-q)}}{2T\Omega\pi(n-q)} \sin(2T\Omega\pi(n-q)).$$

Thus, $\{E_{na} T_{mb} \hat{g}\}$ is an orthonormal sequence if and only if $2T\Omega = 1$. Consequently, by **Theorem 2.2b**, $\{E_{na} T_{mb} \hat{g}\}$ is an orthonormal basis if and only if $2T\Omega = 1$.

b. Suppose $2T\Omega < 1$. To construct $g \in \mathcal{S}(\mathbf{R})$ satisfying the conditions of **Theorem 3.3** we proceed as follows, cf., [H] for a different construction depending on the Pythagorean theorem.

We begin in the standard "distributional way" by defining

$$\psi_\epsilon(\gamma) = \frac{\phi(\epsilon - |\gamma|)}{\int \phi(\epsilon - |\gamma|)\, d\gamma},$$

where $\phi \in C^\infty(\hat{\mathbf{R}})$ vanishes on $(-\infty, 0]$ and equals $e^{-1/\gamma}$ on $[0, \infty)$. Thus, $\psi_\epsilon \in C_c^\infty(\hat{\mathbf{R}})$ is an even function satisfying the conditions, *supp* $\psi_\epsilon = [-\epsilon, \epsilon]$ and $\int \psi_\epsilon(\gamma)\, d\gamma = 1$. Next set

$$\psi_{U,V} = \frac{1}{|V|} \mathbf{1}_V * \mathbf{1}_{U-V}, \quad U,\ V \subseteq \hat{\mathbf{R}},$$

so that $\psi_{U,V}$ is 1 on U and vanishes off of $U + V - V$. The function g will be defined in terms of $\hat{g}$ as $\hat{g} = \psi_{U,V} * \psi_\epsilon$, where we shall now specify ϵ, U, and V given $2T\Omega < 1$. Let $U = [-u, u]$, where $u \in (\Omega, \frac{1}{2T})$ is arbitrary, and let $\epsilon = u - \Omega$. Choose $V = [-v, v]$ by setting $v = \frac{w-u}{2}$, where $w = \frac{1}{2T} + \epsilon$. These choices are necessitated by a simple geometrical argument, and the resulting function $\hat{g}$ satisfies the desired properties.

4. Weighted Fourier frames for irregular lattices.

In the case of irregular lattices, the following result is the analogue of **Theorem 2.7** for $\hat{\mathbf{R}}$.

4.1 THEOREM. *Given $\Omega > 0$ and let $g \in PW_\Omega$. Assume that $\{a_n\}$, $\{b_m\}$ are real sequences for which*

$$\{E_{a_n}\} \text{ is a frame for } L^2[-\Omega, \Omega], \tag{4.1}$$

and that there exist $A, B > 0$ such that

$$0 < A \le G(\gamma) \le B < \infty \ \text{ a.e. on } \hat{\mathbf{R}}, \tag{4.2}$$

where

$$G(\gamma) = \sum |\hat{g}(\gamma - b_m)|^2.$$

Then $\{E_{a_n} T_{b_m} \hat{g}\}$ *is a frame for* $L^2(\hat{\mathbf{R}})$*; and* $\{E_{a_n} T_{b_m} \hat{g}\}$ *is a tight frame for* $L^2(\hat{\mathbf{R}})$ *if and only if* $\{E_{a_n}\}$ *is a tight frame for* $L^2[-\Omega, \Omega]$ *and* G *is a constant a.e. on* $\hat{\mathbf{R}}$.

PROOF: $I = [-\Omega, \Omega]$ and set $I_m = I + b_m$. For fixed m, $\{T_{b_m} E_{a_n}\}$ is a frame for $L^2(I_m)$ with frame bounds A_I, B_I independent of m. Thus, for all $h \in L^2(\mathbf{R})$ for which $supp\, \hat{h} \subseteq I_m$, we have

$$A_I \|\hat{h}\|^2_{L^2(I_m)} \leq \sum_n |\langle \hat{h}, T_{b_m} E_{a_n} \rangle_{I_m}|^2 \leq B_I \|\hat{h}\|^2_{L^2(I_m)}, \tag{4.3}$$

Take any $f \in L^2(\mathbf{R})$. Because of (4.2), $\hat{g} \in L^\infty(\hat{\mathbf{R}})$; and, hence, $\hat{h}_{m,f} = \hat{f}\, T_{b_m} \overline{\hat{g}} \in L^2(I_m)$. Also, since g is Ω–bandlimited, $\hat{h}_{m,f}$ vanishes off of I_m. Substituting $\hat{h}_{m,f}$ into (4.3) and summing over m, we obtain

$$A_I \sum_m \|\hat{f}\, T_{b_m} \overline{\hat{g}}\|^2_{L^2(I_m)} \leq \sum_m \sum_n |\langle \hat{f}, (T_{b_m} \hat{g})\, T_{b_m} E_{a_n} \rangle|^2 \leq B_I \sum_m \|\hat{f}\, T_{b_m} \overline{\hat{g}}\|^2_{L^2(I_m)}. \tag{4.4}$$

We now compute

$$\langle \hat{f}, (T_{b_m} \hat{g})\, T_{b_m} E_{a_n} \rangle = \langle \hat{f}, T_{b_m} (\hat{g} E_{a_n}) \rangle$$

and, using the fact that g is Ω–bandlimited,

$$\sum_m \|\hat{f}\, T_{b_m} \overline{\hat{g}}\|^2_{L^2(I_m)} = \int |\hat{f}(\gamma)|^2 (\sum |\hat{g}(\gamma - b_m)|^2)\, d\gamma.$$

By these calculations, as well as (4.2) and (4.4), we obtain

$$AA_I \|\hat{f}\|_2^2 \leq \sum_m \sum_n |\langle \hat{f}, T_{b_m} E_{a_n} \hat{g} \rangle|^2 \leq BB_I \|\hat{f}\|_2^2. \tag{4.5}$$

Thus, $\{E_{a_n} T_{b_m} \hat{g}\}$ is a frame for $L^2(\hat{\mathbf{R}})$. The characterization of $\{E_{a_n} T_{b_m} \hat{g}\}$ as a tight frame follows immediately from (4.5). ∎

4.2 COROLLARY. *Given the hypotheses of* **Theorem 4.1** *and set* $I_m = [-\Omega, \Omega] + b_m$. *For each fixed* m, $\{T_{b_m} E_{a_n}\}$ *is a frame for* $L^2(I_m)$ *with frame operator* S_m, *cf., (4.3),* $\{E_{a_n} T_{b_m} \hat{g}\}$ *is a frame for* $L^2(\hat{\mathbf{R}})$ *with frame operator* S, *and*

$$\forall\, h \in L^2(\mathbf{R}), \quad S\hat{h} = \sum T_{b_m} \hat{g}\, S_m(\hat{h}\, T_{b_m} \overline{\hat{g}}).$$

PROOF: We compute

$$\begin{aligned} S\hat{h} &= \sum_m \sum_n \langle \hat{h}, E_{a_n} T_{b_m} \hat{g} \rangle E_{a_n} T_{b_m} \hat{g} \\ &= \sum_m T_{b_m} \hat{g} \Big(\sum_n \langle \hat{h}, E_{a_n} T_{b_m} \hat{g} \rangle_{\hat{\mathbf{R}}}\, E_{a_n} \Big) \mathbf{1}_{I_m} \\ &= \sum_m T_{b_m} \hat{g} \Big(\sum_n \langle \hat{h}\, T_{b_m} \overline{\hat{g}}, E_{a_n} \rangle_{I_m} E_{a_n} \mathbf{1}_{I_m} \Big) \\ &\equiv \sum_m T_{b_m} \hat{g}\, S_m(\hat{h}\, T_{b_m} \overline{\hat{g}}). \ \blacksquare \end{aligned}$$

If "$\hat{g}$" is any Borel measurable function for which $G(\gamma) \le B$ *a.e.* on $\hat{\mathbf{R}}$, then $\hat{g} \in L^\infty(\hat{\mathbf{R}})$. The converse is a part of the following result.

4.3 THEOREM. *Given* $\Omega > 0$. *Assume that* $\{a_n\}$, $\{b_m\}$ *are real sequences for which* $\{E_{a_n}\}$ *is a frame for* $L^2[-\Omega, \Omega]$, *and that there exist* d, $D > 0$ *such that*

$$\forall\, m, \quad 0 < d \le b_{m+1} - b_m \le D < 2\Omega, \tag{4.6}$$

where $\lim_{m \to \pm\infty} b_m = \pm\infty$. *Suppose* $g \in PW_\Omega$ *has the properties that* $\hat{g} \in L^\infty(\hat{\mathbf{R}})$ *and* $A = \inf\{|\hat{g}(\lambda)|^2 : \lambda \in I\} > 0$ *for some interval* $I \subseteq [-\Omega, \Omega]$ *having measure* $|I| = D$. *Then* $\{E_{a_n} T_{b_m} \hat{g}\}$ *is a frame for* $L^2(\hat{\mathbf{R}})$.

PROOF: It suffices to verify condition (4.2) of **Theorem 4.1.**

For each γ, $G(\gamma)$ is a finite sum; and, in fact, this sum has at most $[\frac{2\Omega}{d}] + 1$ terms. Thus,

$$\forall\, \gamma, \quad G(\gamma) \le ([\frac{2\Omega}{d}] + 1)\|\hat{g}\|_\infty \equiv B < \infty,$$

and the upper bound is obtained.

For each $\gamma \in \hat{\mathbf{R}}$ there is a b_m such that $\gamma - b_m \in I$. Thus,

$$G(\gamma) \ge |\hat{g}(\gamma - b_m)|^2 \ge A > 0,$$

and the lower bound is obtained. ∎

4.4 REMARK. a. Consider condition (4.1), used in both **Theorem 4.1** and **4.3**.

a.i. A sequence $\{a_n\} \subseteq \mathbf{R}$ has **uniform density** $\Delta > 0$ if there exist constants L and d such that

$$\forall\, n, \quad |a_n - \frac{n}{\Delta}| \leq L$$

and

$$\forall\, n \neq m, \quad |a_n - a_m| \geq d > 0.$$

Duffin and Schaeffer [DS] proved that *if* $\{a_n\}$ *has uniform density* $\Delta > 0$ *and* $0 < 2\Omega < \Delta$ *then* $\{E_{a_n}\}$ *is a frame for* $L^2[-\Omega, \Omega]$. For a given sequence $\{a_n\} \subseteq \mathbf{R}$ let $\Omega_{\mathbf{R}}$ be the least upper bound of all Ω for which $\{E_{a_n}\}$ is a frame for $L^2[-\Omega, \Omega]$; $\Omega_{\mathbf{R}}$ is the **frame radius** of $\{a_n\}$. Duffin and Schaeffer's theorem can be rephrased and refined as follows: if $\{a_n\}$ has uniform density $\Delta > 0$ then $\Omega_{\mathbf{R}} \geq \frac{\Delta}{2}$.

Important work on this topic is due to [La; J], cf., [H]. We mention the following fact which follows from [DS; J]. *Suppose* $\{E_{a_n}\}$ *is an exact frame for* $L^2[-\Omega, \Omega]$. *Then* $\{E_{a_n}\}$ *is not a frame for* $L^2[-\Omega_1, \Omega_1]$ *for any* $\Omega_1 > \Omega$, *and* $\{E_{a_n}\}$ *is an inexact frame for* $L^2[-\Omega_1, \Omega_1]$ *for every* $0 < \Omega_1 < \Omega$. In this latter case we can remove any finite number of arbitrarily selected elements of $\{a_n\}$ and still have a frame for $L^2[-\Omega_1, \Omega_1]$.

a.ii. If $a_n = na$ and $a = \frac{1}{2\Omega}$ then $\{E_{na}\}$ is an orthonormal basis of $L^2[-\Omega, \Omega]$. The sequence $\{na\}$ has uniform density $\Delta = \frac{1}{a}$.

b.i. *Given the hypotheses of* **Theorem 4.1** *in the case* $a_n = na$ *and* $a = \frac{1}{2\Omega}$. *Then*

$$\forall\, f \in L^2(\mathbf{R}), \quad S^{-1}\hat{f} = \frac{1}{2\Omega}\frac{\hat{f}}{G}. \tag{4.7}$$

To verify (4.7) note that $\{\frac{1}{(2\Omega)^{1/2}} E_{na}\}$ is an orthonormal basis of each $L^2(I_m)$ and that $\hat{f}\, T_{b_m}\overline{\hat{g}} \in L^2(I_m)$. Since $\{E_{na} T_{b_m} \hat{g}\}$ is a frame for $L^2(\hat{\mathbf{R}})$ we have

$$\begin{aligned} S\hat{f} &\equiv \sum_m \sum_n \langle \hat{f}, E_{na} T_{b_m} \hat{g} \rangle E_{na} T_{b_m} \hat{g} \\ &= \sum_m (T_{b_m}\hat{g}) \Big(\sum_n \langle \hat{f}\, T_{b_m}\overline{\hat{g}}, E_{na} \rangle E_{na} \Big) \qquad (4.8) \\ &= 2\Omega\, \hat{f}\, G. \end{aligned}$$

Using $S^{-1}\hat{f}$ instead of $\hat{f}$ in (4.8) we obtain (4.7).

b.ii. In **Theorem 3.3** we used the commutativity of the operators S^{-1} and $E_{na}T_{mb}$ in proving the sampling formula.

Now suppose we have the hypotheses of **Theorem 4.1** in the case $a_n = na$ and $a = \frac{1}{2\Omega}$. Then by part **b.i.** we have (4.7), so that

$$S^{-1}(E_{na}T_{b_m}\hat{g}) = \frac{1}{2\Omega}\frac{E_{na}T_{b_m}\hat{g}}{G}.$$

On the other hand,

$$E_{na}T_{b_m}S^{-1}\hat{g} = \frac{1}{2\Omega}\frac{E_{na}T_{b_m}\hat{g}}{T_{b_m}G},$$

so that *the operators* S^{-1} *and* $E_{na}T_{b_m}$ *are not commutative for irregular sequences* $\{b_m\}$.

4.5 EXAMPLE. Given the hypotheses of **Theorem 4.1**. Then

$$\forall\, f \in L^2(\mathbf{R}), \quad \hat{f} = \sum \langle \hat{f}, S^{-1}(E_{a_n}T_{b_m}\hat{g})\rangle E_{a_n}T_{b_m}\hat{g} \quad \textit{in } L^2(\hat{\mathbf{R}}),$$

and so

$$\forall\, f \in L^2(\mathbf{R}), \quad f(t) = \sum c_n(t)\, T_{-a_n}g(t) \quad \textit{in } L^2(\mathbf{R}), \tag{4.9}$$

where

$$c_n(t) = \sum_m \langle \hat{f}, e^{-2\pi i a_n b_m} S^{-1}(E_{a_n}T_{b_m}\hat{g})\rangle E_{b_m}(t).$$

With various further hypotheses, (4.9) will be a "sampling" formula, cf., **Theorem 6.2**. The point we make now is that *the frequencies for Fourier frames on* $\hat{\mathbf{R}}$ *provide the translation points on* $\mathbf{R}$ *for sampling formulas.*

5. Irregular sampling —sampled coefficients and exact frames.

The theory of non-harmonic Fourier series was developed by Paley and Wiener [PW, Chapters 6 and 7] and Levinson [L, Chapter 4]. Related work preceding [PW] is due to G.

D. Birkhoff (1917), J.L. Walsh (1921), and Wiener (1927). The Paley–Wiener and Levinson theory has been reformulated and analyzed in terms of irregular sampling by Beutler [Be1; Be2] for completeness and Yao and Thomas [YT] for expansions. The Yao and Thomas expansion was discovered independently by Higgins [Hi1] using reproducing kernels; there is also the interesting new work by Rawn [R]. In this section we shall state and prove this irregular sampling expansion by frame methods. The coefficients in the expansion are the values of the given signal at the given irregularly spaced sampling points, cf., **Section 6**.

Whereas we implemented S^{-1} as a multiplier in **Section 3**, in this section we shall invoke a formula, viz., (5.1), related to the fact that $\{S^{-1}g_n\}$ is the unique biorthonormal sequence associated to a given exact frame $\{g_n\}$, cf., **Theorem 2.2c**.

5.1 PROPOSITION. *Let H be a separable Hilbert space and let $\{g_n\} \subseteq H$ be an exact frame with inverse frame operator S^{-1}. Then*

$$\forall\, f \in H, \quad S^{-1}f = \sum \langle f, h_n \rangle\, h_n \quad \text{in } H, \tag{5.1}$$

where $\{h_n\}$ is the unique biorthonormal sequence associated with $\{g_n\}$. In particular, $\{S^{-1}g_n\} = \{h_n\}$, and so S^{-1} is the frame operator of the dual frame $\{S^{-1}g_n\}$.

PROOF: Since $\{g_n\}$ is exact, $\{g_n\}$ and $\{S^{-1}g_n\}$ are biorthonormal (**Theorem 2.2c**); and since $\{g_n\}$ is complete, we see that $\{S^{-1}g_n\}$ is the unique biorthonormal sequence associated with $\{g_n\}$. (5.1) follows immediately from **Theorem 2.2a**. ∎

5.2 THEOREM. *Given $\Omega > 0$ and $\{a_n\} \subseteq \mathbf{R}$, let $t_n = -a_n$, and assume $\{E_{a_n}\}$ is an exact frame for $L^2[-\Omega, \Omega]$. Define $s_n(t)$ in terms of its involution $\tilde{s}_n(t) = \overline{s_n(-t)}$, where*

$$\forall\, t \in \mathbf{R}, \quad \tilde{s}_n(t) = \int_{-\Omega}^{\Omega} \overline{h}_n(\gamma)\, e^{2\pi i t\gamma} d\gamma, \tag{5.2}$$

and where $\{h_n\}$ is the unique biorthonormal sequence associated with $\{E_{a_n}\}$. (In particular, $\tilde{s}_n \in PW_\Omega$.) Then

$$\forall\, f \in PW_\Omega, \quad f = \sum f(t_n)\, s_n \ \text{ in } L^2(\mathbf{R}). \tag{5.3}$$

where $s_n(t) = \overline{\tilde{s}_n(-t)} \in PW_\Omega$.

PROOF: Let $g = \frac{1}{(2\Omega)^{1/2}} d_{2\pi\Omega}$ and set $b_m = 2\Omega m$. Note that

$$G(\gamma) \equiv \sum |\hat{g}(\gamma - b_m)|^2 = \frac{1}{2\Omega} \quad a.e.$$

and $supp\, \hat{g} \subseteq [-\Omega, \Omega]$. Thus, since $\{E_{a_n}\}$ is a frame, we can apply **Theorem 4.1** to obtain that $\{E_{a_n} T_{b_m} \hat{g}\}$ is a frame for $L^2(\hat{\mathbf{R}})$ with frame operator S. In particular,

$$(5.4) \qquad \forall\, h \in L^2(\mathbf{R}), \quad \hat{h} = \sum \langle \hat{h}, E_{a_n} T_{b_m} \hat{g} \rangle\, S^{-1}(E_{a_n} T_{b_m} \hat{g}) \;\; in\; L^2(\hat{\mathbf{R}}).$$

Similarly to (3.4), we obtain

$$\langle \hat{f}, E_{a_n} T_{2\Omega m} \hat{g} \rangle = \begin{cases} \frac{1}{(2\Omega)^{1/2}} f(-a_n), & if\ m = 0 \\ 0, & if\ m \neq 0 \end{cases}$$

for $f \in PW_\Omega$.

Let S_m be the frame operator for the frame $\{T_{b_m} E_{a_n}\}$ for $L^2(I_m)$, where $I_m = [-\Omega, \Omega] + b_m$. By **Corollary 4.2**, we have

$$(5.6) \qquad \forall\, h \in L^2(\mathbf{R}), \quad S\hat{h} = \sum_m T_{b_m} \hat{g}\, S_m(\hat{h}\, T_{b_m} \overline{\hat{g}}) \;\; in\; L^2(\hat{\mathbf{R}}).$$

From (5.6) and the definition of g we compute

$$\begin{aligned} S\hat{f} &= \frac{1}{2\Omega} \sum_m \mathbf{1}_{(\Omega)}(\gamma - 2\Omega m)\, S_m(\hat{f}(\cdot)\, \mathbf{1}_{(\Omega)}(\cdot - 2\Omega m))(\gamma) \\ &= \frac{1}{2\Omega} S_0 \hat{f}(\gamma) \end{aligned}$$

for $f \in PW_\Omega$, where the second equality follows since $supp\, \hat{f} \subseteq [-\Omega, \Omega]$. Thus,

$$(5.7) \qquad \forall\, f \in PW_\Omega, \quad S\hat{f} = \frac{1}{2\Omega} S_0 \hat{f},$$

i.e., the action of S on $L^2[-\Omega, \Omega]$ can be realized by the action of $\frac{1}{2\Omega} S_0$ on that same subspace of $L^2(\hat{\mathbf{R}})$.

Using (5.7), we can write

$$\forall\, f \in PW_\Omega, \quad \hat{f} = S^{-1} S \hat{f}$$

$$= \frac{1}{2\Omega}\, S^{-1} S_0 \hat{f},$$

so that, if we replace $\hat{f}$ by $S_0^{-1}\hat{f}$, we obtain

$$S^{-1} = 2\Omega\, S_0^{-1} \quad \text{on } L^2[-\Omega, \Omega].$$

Therefore, since $g \in PW_\Omega$,

$$S^{-1}(E_{a_n} T_{b_0} \hat{g}) \equiv S^{-1}(E_{a_n} \hat{g})$$

$$= 2\Omega\, S_0^{-1}(E_{a_n} \hat{g})$$

$$= (2\Omega)^{1/2} S_0^{-1}(E_{a_n} \mathbf{1}_{(\Omega)})$$

$$= (2\Omega)^{1/2} \sum_m \langle E_{a_n}, h_m \rangle_{[-\Omega,\Omega]}\, h_m$$

$$= (2\Omega)^{1/2} h_n,$$

where the penultimate equality depends on the exactness hypothesis and **Proposition 5.1**. Substituting this information into (5.4) and (5.5) gives the reconstruction,

$$\forall\, f \in PW_\Omega, \quad \hat{f} = \sum_n \frac{1}{(2\Omega)^{1/2}}\, f(-a_n)(2\Omega)^{1/2} h_n \quad \text{in } L^2(\hat{\mathbf{R}}),$$

which, in turn, yields (5.3). ∎

5.3 REMARK. Levinson [L, Theorem 18] proved that if $\Omega > 0$ and $\{a_n\} \subseteq \mathbf{R}$ satisfy

$$\sup_n |n - 2\Omega\, a_n| < \frac{1}{4}, \tag{5.8}$$

then $\{E_{a_n}\}$ is complete in $L^2[-\Omega, \Omega]$ and has a unique biorthonormal sequence $\{h_n\}$. Kadec (1964) [Ka] provided the direct calculation proving that $\{E_{a_n}\}$ is a Riesz basis, i.e., exact frame, if (5.8) holds, cf., [Y, pp. 34-36] for a characterization to ensure that complete sets with associated biorthonormal sequences are Riesz bases.

The bound "1/4" in (5.8) is best possible [L, Theorem 19].

The explicit formulas in the following result are proved in [PW, pp. 89-90 and pp. 114-116] and [L, pp. 4 ff]. The calculations by Paley and Wiener were refined by Young (1979), e.g., [Y, pp. 148-150]. The remainder of the proof is referenced in **Remark 5.3**.

5.4 THEOREM. *Given $\Omega > 0$ and $\{a_n\} \subseteq \mathbf{R}$, and assume (5.8). Then $\{E_{a_n}\}$ is an exact frame for $L^2[-\Omega,\Omega]$ with unique biorthonormal sequence $\{h_n\}$; and $\tilde{s}_n$, defined by (5.2), is*

$$\tilde{s}_n(t) = \frac{\tilde{s}(t)}{\tilde{s}'(a_n)(t-a_n)} \tag{5.9}$$

where

$$\tilde{s}(t) = (t-a_0)\prod_{n=1}^{\infty}(1-\frac{t}{a_n})(1-\frac{t}{a_{-n}}). \tag{5.10}$$

5.5 EXAMPLE. Note that the sampling functions s_n, defined in **Theorem 5.2**, are given by

$$s_n(t) = \frac{s(t)}{s'(t_n)(t-t_n)},$$

where

$$s(t) = (t-t_0)\prod_{n=1}^{\infty}(1-\frac{t}{t_n})(1-\frac{t}{t_{-n}});$$

and they have the property that

$$\forall\ m,\ n \quad s_n(t_m) = \langle h_n, E_{a_m}\rangle_\Omega = \delta_{mn}. \tag{5.11}$$

This property of sampling functions is usually described by saying that $\{s_n\}$ is a sequence of Lagrangia interpolating functions.

6. Irregular sampling —irregular translates of a sampling function.

Our basic result in this section, **Theorem 6.2**, is dual to the sampling theorem, **Theorem 5.2**, in the following way. Exact frames were required in **Section 5** and the sampled values of the signal were explicit in the *dual* frame expansion. **Theorem 6.2** will use general frames, and the frame expansion will only require the irregular translates of a single sampling function. The dual frame expansion was global in **Section 5** and the frame expansion is local in this section.

The following fact is clear.

6.1 LEMMA. *Given $f, f_n \in L^2(\mathbf{R})$, and assume $f = \sum f_n$ in $L^2(\mathbf{R})$. If $g \in L^\infty(\mathbf{R})$ then $fg = \sum f_n g$ in $L^2(\mathbf{R})$.*

6.2 THEOREM. *Given $\Omega > 0$ and $\{a_n\} \subseteq \mathbf{R}$, let $t_n = -a_n$, and assume $\{E_{a_n}\}$ is a frame for $L^2[-\Omega_1, \Omega_1]$, for some $\Omega_1 > \Omega$, with frame operator S. Let $g \in \mathcal{S}(\mathbf{R})$ have the properties that $supp\, \hat{g} \subseteq [-\Omega_1, \Omega_1]$ and $\hat{g} = 1$ on $[-\Omega, \Omega]$. Then*

$$(6.1) \qquad \forall\, f \in PW_\Omega, \quad f = \sum \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{-t_n}\rangle_{[-\Omega_1,\Omega_1]}\, T_{t_n} s \;\in\; L^2(\mathbf{R}).$$

where $s \equiv g$. (We choose "s" since it represents the "sampling" function.)

PROOF: Since $\{E_{a_n}\}$ is a frame for $L^2[-\Omega_1, \Omega_1]$ and $supp\, \hat{f} \subseteq [-\Omega, \Omega]$, we have

$$\begin{aligned} \hat{f} &= \hat{f}\,\mathbf{1}_{(\Omega_1)} \\ &= \sum \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{a_n}\rangle_{[-\Omega_1,\Omega_1]}\, E_{a_n}\mathbf{1}_{(\Omega_1)} \quad in\ L^2(\hat{\mathbf{R}}). \end{aligned} \tag{6.2}$$

In this expression, we note that S^{-1}, being positive, is self-adjoint so that the frame expansion in **Theorem 2.2a** gives rise to (6.2). Also, the $L^2[-\Omega_1, \Omega_1]$ convergence from our frame hypothesis can be taken to be in $L^2(\hat{\mathbf{R}})$ by extending all functions to be zero outside $[-\Omega_1, \Omega_1]$.

We have $\hat{f} = \hat{f}\hat{g}$ on $\hat{\mathbf{R}}$ since $\hat{g} = 1$ on $[-\Omega, \Omega]$ and $\hat{f} = 0$ off of $[-\Omega, \Omega]$. Also,

$$\begin{aligned} &\hat{g} \sum \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{a_n}\rangle_{[-\Omega_1,\Omega_1]}\, E_{a_n}\mathbf{1}_{(\Omega_1)} \\ &= \sum \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{a_n}\rangle_{[-\Omega_1,\Omega_1]}\, E_{a_n}\mathbf{1}_{(\Omega_1)}\hat{g} \quad in\ L^2(\hat{\mathbf{R}}) \end{aligned}$$

by **Lemma 6.1**. Thus, since $supp\, \hat{g} \subseteq [-\Omega_1, \Omega_1]$, we obtain

$$\begin{aligned} \hat{f} &= \hat{f}\,\hat{g} \\ &= \sum_n \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{a_n}\rangle_{[-\Omega_1,\Omega_1]}\, E_{a_n}\hat{g} \quad in\ L^2(\hat{\mathbf{R}}). \end{aligned}$$

Taking the inverse transform gives (6.1). ∎

The following result allows us to be more explicit about the coefficients in (6.1) in the case of exact frames and the Levinson (and Kadec) condition (5.8).

6.3 THEOREM. *Given $\Omega > 0$ and $\{a_n\} \subseteq \mathbf{R}$, let $t_n = -a_n$, and assume*

$$\sup_n |n - 2\Omega_1 a_n| < \frac{1}{4}$$

for some $\Omega_1 > \Omega$. Then $\{E_{a_n}\}$ is an exact frame for $L^2[-\Omega_1, \Omega_1]$ with frame operator S and unique biorthonormal sequence $\{h_n\}$. Further, if we define $\tilde{s}_n$ and $\tilde{s}$ on $[-\Omega_1, \Omega_1]$ by (5.9) and (5.10), then $(\tilde{s}_n)\hat{} = \overline{h}_n$ (where $h_n \equiv 0$ off of $[-\Omega_1, \Omega_1]$) and the coefficients of (6.1) are

$$\forall\, n, \quad \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{-t_n}\rangle_{[-\Omega_1,\Omega_1]} = \langle f(t), s_n(t)\rangle,$$

where $f \in PW_\Omega$.

PROOF: The exact frame and biorthonormal sequence conclusions follow from **Theorem 5.4**, as well as (5.9), (5.10), and the relation $(\tilde{s}_n)\hat{} = \overline{h_n}$. Letting $H = L^2[-\Omega_1, \Omega_1]$ in **Proposition 5.1**, we have

$$\forall\, F \in L^2[-\Omega_1, \Omega_1], \quad S^{-1}(F) = \sum \langle F, h_m\rangle_{[-\Omega_1,\Omega_1]}\, h_m.$$

Consequently, by orthonormality,

$$\begin{aligned} \forall\, n, \quad S^{-1}(E_{a_n}) &= \sum \langle E_{a_n}, h_m\rangle_{[-\Omega_1,\Omega_1]}\, h_m \\ &= \langle E_{a_n}, h_n\rangle_{[-\Omega_1,\Omega_1]}\, h_n \\ &= h_n. \end{aligned}$$

Therefore, setting $h_n = 0$ off of $[-\Omega_1, \Omega_1]$ and noting that $\tilde{s}_n$ is real-valued, we compute

$$\begin{aligned} \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{-t_n}\rangle_{[-\Omega_1,\Omega_1]} &= \langle \hat{f}, S^{-1}(E_{a_n})\rangle_{[-\Omega_1,\Omega_1]} \\ &= \langle \hat{f}, h_n\rangle_{[-\Omega_1,\Omega_1]} \\ &= \langle \hat{f}, h_n\rangle_{\hat{\mathbf{R}}} \\ &= \langle f, \tilde{h_n}\rangle_{\mathbf{R}} \\ &= \langle f(t), \overline{\tilde{s}_n(-t)}\rangle \\ &= \langle f(t), s_n(t)\rangle, \end{aligned}$$

for each $f \in PW_\Omega$. ∎

6.4 ALGORITHM. It is possible to estimate the coefficients in (6.1) without dealing with exact frames. In so doing, we shall see to what extent these coefficients contain information from the sampled values $f(t_n)$.

Let $\{E_{a_n}\}$ be a frame for $L^2[-\Omega_1, \Omega_1]$ with frame operator S and frame bounds A and B. Since

$$\|I - \frac{2}{A+B}S\| \leq \frac{B-A}{A+B} < 1, \tag{6.3}$$

we have

$$S^{-1} = \frac{2}{A+B}\sum_{k=0}^{\infty}(I - \frac{2}{A+B}S)^k, \tag{6.4}$$

where $I : L^2[-\Omega_1, \Omega_1] \to L^2[-\Omega_1, \Omega_1]$ is the identity map, the norm in (6.3) is the operator norm, and the convergence in (6.4) is the operator norm topology on the space of continuous linear operators on $L^2[-\Omega_1, \Omega_1]$ (into itself).

Setting $t_n = -a_n$ and using (6.4), the coefficients in (6.1) become

$$\begin{aligned} c_n &\equiv \langle S^{-1}(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{-t_n}\rangle_{[-\Omega_1,\Omega_1]} \\ &= \frac{2}{A+B}\sum_{k=0}^{\infty}\langle (I - \frac{2}{A+B}S)^k(\hat{f}\,\mathbf{1}_{(\Omega_1)}),\, E_{-t_n}\rangle_{[-\Omega_1,\Omega_1]} \end{aligned} \tag{6.5}$$

for $f \in PW_\Omega$, $\Omega < \Omega_1$. If we truncate the expansion (6.5) after the $k = 0$ term, then

$$c_n = \frac{2}{A+B}f(t_n).$$

Notation.

The Fourier transform $\hat{f}$ of $f \in L^1(\mathbf{R})$ is defined as $\hat{f}(\gamma) = \int f(t)e^{-2\pi i t\gamma}\,dt$, where "$\int$" designates integration over the real line $\mathbf{R}$; $\hat{f}$ is defined on $\hat{\mathbf{R}}$ $(= \mathbf{R})$ and $\check{f}$ is the inverse Fourier transform of f. The Fourier transform is defined on $L^2(\mathbf{R})$, and, for fixed $\Omega > 0$,

$$PW_\Omega \equiv \{f \in L^2(\mathbf{R}) : supp\,\hat{f} \subseteq [-\Omega, \Omega]\},$$

where $supp\,\hat{f}$ is the support of $\hat{f}$. A function (or distribution) f, whose Fourier transform exists, is Ω-**bandlimited** if $supp\,\hat{f} \subseteq [-\Omega, \Omega]$.

Besides the $L^p(\mathbf{R})$-spaces and the Schwartz space $\mathcal{S}(\mathbf{R})$, we deal with the space $C^\infty(\mathbf{R})$ of infinitely differentiable functions and its subspace $C_c^\infty(\mathbf{R})$ whose elements have compact support.

"$\sum$" designates summation over the whole discrete group in question, e.g., over $\mathbf{Z} \times \mathbf{Z}$ where $\mathbf{Z}$ is the group of integers. The function $\mathbf{1}_S$ is the characteristic function of $S \subseteq \mathbf{R}$, $|S|$ is the Lebesgue measure of S, and $\mathbf{1}_{(\Omega)} \equiv \mathbf{1}_{[-\Omega,\Omega]}$. The function δ_{mn} is defined as 0 if $m \neq n$ and as 1 if $m = n$. The dilation f_λ of the function f is $f_\lambda(t) = \lambda f(\lambda t)$, and the translation $T_{t_0} f$ is $T_{t_0} f(t) = f(t - t_0)$. Finally, the exponential function E_a is $E_a(t) = e^{2\pi i a t}$.

Acknowledgement.

We appreciate the exchange and sharing of ideas with Professor Hans Feichtinger on the subject of irregular sampling. We had numerous fruitful discussions while he was Visiting Professor at the University of Maryland during the 1989–1990 academic year. We are also pleased to acknowledge Professor Gilbert Walter's interesting recent work on irregular sampling. His manuscript came to our attention after completion of this manuscript, and there is some overlap of his contributions with our **Section 5**.

REFERENCES

[B] J. Benedetto, "Real variable and integration," B. G. Teubner, Stuttgart, 1976.

[BM] A. Beurling and P. Malliavin, *On the closure of characters and the zeros of entire functions*, Acta Math. **118** (1967), 79–93.

[Be1] F. Beutler, *Sampling theorems and bases in Hilbert space*, Inform. Contr. **4** (1961), 97–117.

[Be2] _____, *Error-free recovery of signals from irregularly spaced samples*, SIAM Review **8** (1966), 328–335.

[BSS] P. Butzer, W. Splettstösser, and R. Stens, *The sampling theorem and linear prediction in signal analysis*, Jber. d. Dt. Math.-Verein. **90** (1988), 1–70.

[DS] R. Duffin and A. Schaeffer, *A class of nonharmonic Fourier series*, Trans. Am. Math. Soc. **72** (1952), 341–366.

[FG] H. Feichtinger and K. Gröchenig, *A real analysis approach to the irregular sampling theorem*, SIAM Annual Meeting Chicago, 1990.

[G] D. Gabor, *Theory of communication*, J. IEE London **93** (1946), 429–457.

[HW] C. Heil and D. Walnut, *Continuous and discrete wavelet transforms*, SIAM Review **31** (1989), 628–666.

[H] W. Heller, *Frames of exponentials and applications*, Ph.D. thesis, University of Maryland, College Park, MD, 1991.

[Hi1] J. Higgins, *A sampling theorem for irregularly spaced sample points*, IEEE Trans. Inf. Theory **IT-22** (1976), 621–622.

[Hi2] _____, *Five short stories about the cardinal series*, Bull. AMS **12** (1985), 45–89.

[Hö] L. Hörmander, "The analysis of linear partial differential operators," I and II, Springer-Verlag, N. Y., 1983.

[J] S. Jaffard, *A density criterion for frames of complex exponentials.*

[Ka] M. Kadec, *The exact value of the Paley–Wiener constant*, Sov. Math. Dokl. **5** (1964), 559–561.

[K] G. Köthe, *Das Trägheitsgesetz der quadratischen Formen im Hilbertschen Raum*, Math. Z. **41** (1936), 137–152.

[La] H. Landau, *Necessary density conditions for sampling and interpolation of certain entire functions*, Acta Math. **117** (1967), 37–52.

[L] N. Levinson, "Gap and density theorems," Am. Math. Soc. Colloq. Publ. vol 26, Am. Math. Soc., 1940.

[PW] R. Paley and N. Wiener, "Fourier transforms in the complex domain," Am. Math. Soc. Colloq. Publ. vol 19, Am. Math. Soc., 1934.

[R] M. Rawn, *A stable nonuniform sampling expansion involving derivatives*, IEEE Trans. Inf. Theory **IT-36** (1990).

[S] L. Schwartz, "Théorie des distributions," Hermann, Paris, 1966.

[SW] E. Stein and G. Weiss, "Fourier analysis on Eucildean spaces," Princeton University Press, 1971.

[V] G. Vitali, *Sulla condizione di chiusura di un sistema di funzioni ortogonali*, Atti R. Accad. Naz. Lincei, Rend. Cl. Sci. Fis. Mat. Nat. **30** (1921), 498–501.

[YT] K. Yao and J. Thomas, *On some stability and interpolatory properties of nonuniform sampling expansions*, IEEE Trans. Circuit Theory **CT-14** (1967), 404–408.

[Y] R. Young, "An introduction to nonharmonic Fourier series," Academic Press, N. Y., 1980.

©1992 IEEE. Reprinted, with permission, from *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, San Franisco, CA, March 23-26, 1992, Vol. IV, pp. 381-384.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commerical advantage, the IEEE copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Institute of Electrical and Electronics Engineers. To copy otherwise, or to republish, requires a fee and specific permission.

MULTIRESOLUTION REPRESENTATIONS USING THE AUTO-CORRELATION FUNCTIONS OF COMPACTLY SUPPORTED WAVELETS

Naoki Saito[1]* *Gregory Beylkin*[2]

[1]Schlumberger-Doll Research, Old Quarry Road, Ridgefield, CT 06877
[2]Program in Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309-0526

ABSTRACT

We propose a shift-invariant multiresolution representation of signals or images using dilations and translations of the auto-correlation functions of compactly supported wavelets. Although this set of functions does not form an orthonormal basis, a number of properties of the auto-correlation functions of the compactly supported wavelets makes them useful for signal and image analysis. Unlike wavelet-based orthonormal representations, our representation has (1) symmetric analyzing functions, (2) shift-invariance, (3) natural and simple iterative interpolation schemes, (4) a simple algorithm for finding the locations of the multiscale edges as zero-crossings.

We also develop a non-iterative method for reconstructing signals from their zero-crossings (and slopes at these zero-crossings) in our representation. This method reduces the problem to that of solving a system of linear equations.

1. INTRODUCTION

The information about the local behavior of a function is hidden in the decay (or growth) from scale to scale of the coefficients of the orthonormal wavelet expansions. Exploiting this property in applications, however, is not a straightforward exercise in part due to the fact that the coefficients of the orthonormal wavelet expansions are not shift invariant. In implementing multiresolution algorithms for image processing and signal analysis, redundant representations are being used in order to simplify the analysis of coefficients from scale to scale.

Another difficulty in utilizing the orthonormal wavelets for the analysis of signals in image processing is associated with the asymmetric shape of compactly supported wavelets [4]. On the one hand, using compactly supported wavelets implies that the associated exact quadrature mirror filters are of finite size which is advantageous in computer implementations. On the other hand, the symmetric basis functions are preferred in image processing since their use simplifies finding zero-crossings (or extrema) corresponding to the locations of edges in images at later stages of processing. There are two approaches for dealing with this problem. The first approach consists in constructing approximately symmetric orthonormal wavelets and gives rise to approximate quadrature mirror filters [5]. The second consists in using biorthogonal bases [3] so that the basis functions may be chosen to be exactly symmetric.

In this paper, we propose a "hybrid" multiresolution representation which utilizes dilations and translations of the auto-correlation functions of compactly supported wavelets. In this representation, the exact filters for the decomposition (similar to the quadrature mirror filters) are symmetric. The auto-correlation functions of the compactly supported wavelets may be viewed as pseudo-differential operators of even order and behave, essentially, as derivative operators of the same order. This allows us to relate the zero-crossings in this representation to the locations of edges at different scales in the signal. The recursive definition of compactly supported wavelets and, therefore, of their auto-correlation functions, allows us to construct fast recursive algorithms to generate the multiresolution representations. Though it is not an orthogonal representation, there is a simple relation with the wavelet-based orthogonal representations on each scale. We describe a simple reconstruction algorithm to recover functions from such expansions.

2. AN ORTHONORMAL SHELL

In this section, we introduce a shift-invariant representation using orthonormal wavelets. We refer to the set of functions

$$\left\{\widetilde{\psi}_{j,k}(x)\right\}_{1\le j\le n_0,\ 0\le k\le N-1} \quad \text{and} \quad \{\widetilde{\varphi}_{n_0,k}(x)\}_{0\le k\le N-1}$$

as a *shell* of the orthonormal wavelets (an *orthonormal shell* for short), where

$$\widetilde{\varphi}_{j,k}(x) = 2^{-j/2}\varphi(2^{-j}(x-k)),$$

$$\widetilde{\psi}_{j,k}(x) = 2^{-j/2}\psi(2^{-j}(x-k)),$$

$\varphi(x)$ and $\psi(x)$ are the scaling function and the basic wavelet. Note that for each j, $\{\widetilde{\varphi}_{j,k}\}$ and $\{\widetilde{\psi}_{j,k}\}$ are 2^j times more redundant than $\{\varphi_{j,k}\}$ and $\{\psi_{j,k}\}$. Let V_0 be the vector space representing the finest scale of interest. The orthonormal shell coefficients of a function $f \in V_0$, $f = \sum_{k=0}^{N-1} s_k^0 \varphi_{0,k}$, are

$$\{d_k^j\}_{1\le j\le n_0,\ 0\le k\le N-1} \quad \text{and} \quad \{s_k^{n_0}\}_{0\le k\le N-1},$$

where the coefficients s_k^j and d_k^j are defined as

$$s_k^j = \int f(x)\widetilde{\varphi}_{j,k}(x)dx,$$

*Also with Department of Mathematics, Yale University, New Haven, CT 06520

$$d_k^j = \int f(x)\widetilde{\psi}_{j,k}(x)dx.$$

Essentially, we do not subsample at each scale. We note that the computational diagram of this algorithm is essentially identical to the Hierarchical Discrete Correlation scheme (HDC) [2] of P. Burt, which was designed for efficient correlation of images at multiple scales. This representation is redundant but contains all orthonormal wavelet coefficients of all circular shifts of the original signal [1], and its computational cost is $O(N \log_2 N)$. However, due to the asymmetric shape of the compactly supported orthonormal wavelets, this representation might still be inconvenient for signal analysis purposes.

3. AN AUTO-CORRELATION SHELL

Instead of the compactly supported wavelets we use their auto-correlation functions, i.e.

$$\Phi(x) = \int_{-\infty}^{+\infty} \varphi(y)\varphi(y-x)dy,$$

$$\Psi(x) = \int_{-\infty}^{+\infty} \psi(y)\psi(y-x)dy,$$

to overcome a number of difficulties associated with the representations in the orthonormal shell.

3.1. Properties of the Auto-Correlation Functions

Let us summarize some useful properties of the auto-correlation functions of compactly supported wavelets. Orthogonality of the wavelet bases implies that

$$\Phi(k) = \delta_{0k} \quad \text{and} \quad \Psi(k) = \delta_{0k},$$

and

$$\hat{\Phi}(\xi) + \hat{\Psi}(\xi) = \hat{\Phi}(\xi/2),$$

or equivalently,

$$\Psi(x) = 2\Phi(2x) - \Phi(x).$$

(Compare this, e.g., with the approximation of the Mexican Hat function by the difference of two Gaussians functions). It is easy to derive the following *two-scale difference equations* for the functions $\Phi(x)$ and $\Psi(x)$,

$$\Phi(x) = \Phi(2x) + \frac{1}{2}\sum_{l=-L/2+1}^{L/2} a_{|2l-1|}\Phi(2x+2l-1),$$

$$\Psi(x) = \Phi(2x) - \frac{1}{2}\sum_{l=-L/2+1}^{L/2} a_{|2l-1|}\Phi(2x+2l-1),$$

where $\{a_k\}$ are the auto-correlation coefficients of the quadrature mirror filter $H = \{h_k\}_{0\le k\le L-1}$,

$$a_k = 2\sum_{l=0}^{L-1-k} h_l h_{l+k} \qquad \text{for} \quad k = 1,\ldots,L-1,$$

$$a_{2k} = 0 \qquad \text{for} \quad k = 1,\ldots,L/2-1.$$

The coefficients $\{a_{2k-1}\}_{1\le k\le L/2}$ were used in [1] for computing representations of derivatives and convolution operators in the bases of compactly supported wavelets. They also have compact supports and vanishing moments:

$$\text{supp } \Phi(x) = \text{supp } \Psi(x) = [-L+1, L-1].$$

$$\int_{-\infty}^{+\infty} x^m \Psi(x)dx = 0, \qquad \text{for} \quad 0 \le m \le L,$$

$$\int_{-\infty}^{+\infty} x^m \Phi(x)dx = 0, \qquad \text{for} \quad 1 \le m \le L,$$

$$\int_{-\infty}^{+\infty} \Phi(x)dx = 1.$$

In addition to these properties, we have

- a complete symmetry of the functions $\Phi(x)$ and $\Psi(x)$.
- $\hat{\Psi}(\xi) \sim O(\xi^L)$, which means that the operator of convolution with $\Psi(x)$ behaves essentially as a differential operator $(d/dx)^L$.
- an iterative interpolation scheme induced by the function $\Phi(x)$. By choosing appropriate wavelets, the corresponding auto-correlation function $\Phi(x)$ produce a whole range of interpolation schemes starting from the linear interpolation and up to the band-limited interpolation (generated by the sinc function) [8].

3.2. An Auto-Correlation Shell

Now we define a multiresolution representation using the auto-correlation functions of wavelets. We refer to the set of functions

$$\left\{\widetilde{\Psi}_{j,k}(x)\right\}_{1\le j\le n_0,\ 0\le k\le N-1} \quad \text{and} \quad \left\{\widetilde{\Phi}_{n_0,k}(x)\right\}_{0\le k\le N-1}$$

as a *shell* of the auto-correlation functions of orthonormal wavelets (an *auto-correlation shell* for short).

$$\widetilde{\Phi}_{j,k}(x) = 2^{-j/2}\Phi(2^{-j}(x-k)),$$

$$\widetilde{\Psi}_{j,k}(x) = 2^{-j/2}\Psi(2^{-j}(x-k)).$$

The relation to the orthonormal shell is derived as follows. First, we define functions based on orthonormal shell coefficients.

$$f_s^j(x) = \sum_{k=0}^{N-1} s_k^j \varphi(x-k),$$

$$f_d^j(x) = \sum_{k=0}^{N-1} d_k^j \varphi(x-k)$$

Then, convolving $f_s^j(x)$ and $f_d^j(x)$ with $2^{-j}\varphi(2^{-j}x)$ and $2^{-j}\psi(2^{-j}x)$ respectively, we obtain

$$\mathcal{A}_s^j f(x) = \int f_s^j(y)\ 2^{-j}\varphi(2^{-j}(y-x))dy,$$

$$\mathcal{A}_d^j f(x) = \int f_d^j(y)\ 2^{-j}\psi(2^{-j}(y-x))dy.$$

Finally, we set

$$\mathcal{A}_s^j f(x) = \sum_{k=0}^{N-1} S_k^j \Phi(x-k),$$

$$\mathcal{A}_d^j f(x) = \sum_{k=0}^{N-1} D_k^j \Phi(x-k).$$

The coefficients S_k^j and D_k^j are defined as "samples" of $\mathcal{A}_s^j f(x)$ and $\mathcal{A}_d^j f(x)$, i.e.,

$$S_k^j = \mathcal{A}_s^j f(k) \quad \text{and} \quad D_k^j = \mathcal{A}_d^j f(k).$$

The *auto-correlation shell coefficients* of a function $f \in \mathbf{V}_0$, $f = \sum_{k=0}^{N-1} s_k^0 \varphi_{0,k}$, are

$$\{D_k^j\}_{1 \le j \le n_0,\ 0 \le k \le N-1} \quad \text{and} \quad \{S_k^{n_0}\}_{0 \le k \le N-1}.$$

As easily seen, $\{\Phi_{j,k}\}$ and $\{\Psi_{j,k}\}$ are not orthonormal bases of $\mathbf{V}_j$ and $\mathbf{W}_j$ anymore. However, the representation of functions in the auto-correlation shell, has the following features:

- it is shift-invariant and contains the coefficients of all circular shifts of the original signal,
- it is convertible to the orthonormal shell representation,
- it is completely symmetric,
- zero-crossings of the auto-correlation shell coefficients correspond to the multiscale edges,
- the computational cost is $O(N \log_2 N)$.

We also point out an important relation between the original coefficients $\{s_k^0\}$ and the auto-correlation shell coefficients,

Proposition 1

$$\sum_{k=0}^{N-1} S_k^j \widetilde{\Phi}_{0,k} = \sum_{k=0}^{N-1} s_k^0 \widetilde{\Phi}_{j,k}$$

$$\sum_{k=0}^{N-1} D_k^j \widetilde{\Phi}_{0,k} = \sum_{k=0}^{N-1} s_k^0 \widetilde{\Psi}_{j,k}.$$

See [8] for the proof.

3.3. Fast Decomposition and Reconstruction Algorithms

Rewriting the two-scale difference equations for the auto-correlation functions, we have

$$\frac{1}{\sqrt{2}} \Phi(x/2) = \sum_{k=-L+1}^{L-1} p_k \Phi(x-k),$$

$$\frac{1}{\sqrt{2}} \Psi(x/2) = \sum_{k=-L+1}^{L-1} q_k \Psi(x-k),$$

where

$$p_k = \begin{cases} 2^{-1/2} & \text{for } k = 0, \\ 2^{-3/2} a_{|k|} & \text{for } k = \pm 1, \pm 3, \ldots, \pm(L-1), \\ 0 & \text{for } k = \pm 2, \pm 4, \ldots, \pm(L-2), \end{cases}$$

and

$$q_k = \begin{cases} 2^{-1/2} & \text{for } k = 0, \\ -p_k & \text{otherwise.} \end{cases}$$

We view these coefficients as filters $P = \{p_k\}_{-L+1 \le k \le L-1}$ and $Q = \{q_k\}_{-L+1 \le k \le L-1}$ which are symmetric and have only $L/2+1$ distinct non-zero coefficients. Although this pair of filters is not a quadrature mirror filter pair, their role and use in the numerical algorithms is similar. For example, for the Haar functions, we have

$$\{p_k\} = 2^{-1/2} \left\{ \frac{1}{2}, 1, \frac{1}{2} \right\},$$

For the Daubechies's wavelet with $L = 2M = 4$, we have

$$\{p_k\} = 2^{-1/2} \left\{ -\frac{1}{16}, 0, \frac{9}{16}, 1, \frac{9}{16}, 0, -\frac{1}{16} \right\},$$

and for the "sinc" functions, we have

$$\{p_k\} = 2^{-1/2} \{\mathrm{sinc}(k/2)\}_{k=-\infty}^{+\infty}.$$

Using these filters P and Q, we compute

$$S_k^j = \sum_{l=-L+1}^{L-1} p_l S_{k+2^{j-1}l}^{j-1},$$

$$D_k^j = \sum_{l=-L+1}^{L-1} q_l S_{k+2^{j-1}l}^{j-1}.$$

As for the reconstruction, we immediately obtain a simple formula,

$$S_k^{j-1} = \frac{1}{\sqrt{2}} \left(S_k^j + D_k^j \right),$$

for $j = 1, \ldots, n_0$, $k = 0, \ldots, N-1$. Given the auto-correlation shell coefficients $\{D_k^j\}_{1 \le j \le n_0,\ 0 \le k \le N-1}$ and $\{S_k^{n_0}\}_{0 \le k \le N-1}$,

$$S_k^0 = 2^{-n_0/2} S_k^{n_0} + \sum_{j=1}^{n_0} 2^{-j/2} D_k^j,$$

for $k = 0, \ldots, N-1$.

4. RECONSTRUCTION FROM ZERO-CROSSINGS

Since the operator of convolution with $\Psi(x)$ behaves, essentially, as a derivative operator of the even order, zero-crossings in our representation are related to the multiscale edges of the original signal. We also have an efficient iterative algorithm to "zoom in" at these zero-crossings (Dubuc's symmetric iterative interpolation [6],[7],[8]). In this section, we briefly describe our reconstruction algorithm from zero-crossings (and slopes at these zero-crossings).

4.1. Computation of Zero-Crossings and Slopes

Using the symmetric iterative interpolation scheme mentioned above, we compute the zero-crossing locations of the set of functions $\{\sum_{k=0}^{N-1} D_k^j \Phi(x-k)\}_{1\le j\le n_0}$ within the prescribed numerical accuracy, e.g., $\epsilon = 10^{-14}$. To compute the locations of zero-crossings, we recursively subdivide the unit interval bracketing the zero-crossing until the length of the subdivided interval bracketing that zero-crossing becomes less than the accuracy ϵ. The iterative interpolation scheme allows us to zoom in as much as we want around the zero-crossing. This process requires at most $O(-L\log_2\epsilon)$ operations per zero-crossing. Once the zero-crossing is found, the computation of the slope requires values at only $2(L-2)$ points around the zero-crossing [8].

4.2. The Problem of Reconstruction

We address the following problem: *Given the coarsest subsampled coefficients* $\{S^{n_0}_{2^{n_0}k}\}_{0\le k\le 2^{n-n_0}-1}$, *and the zero-crossings and the slopes at these zero-crossings* $\{x_m^j, v_m^j\}_{1\le j\le n_0,\ 0\le m\le N_z^j-1}$, *where* N_z^j *is the number of zero-crossings of the function* $\sum_{k=0}^{N-1} D_k^j\Phi(x-k)$, *reconstruct the original vector* $\{s_k^0\}_{0\le k\le N-1}$.

Proposition 1 provides a simple mechanism for defining a linear system which relates the unknown signal $\{s_k^0\}$ and the values of the function $\Phi(x)$ and its derivative at the integer translates of zero-crossings.

It follows from Proposition 1, that the zero-crossing coordinate x_k^j satisfies

$$\sum_{k=0}^{N-1} s_k^0 \widetilde{\Psi}_{j,k}(x_m^j) = 0,$$

$$\sum_{k=0}^{N-1} s_k^0\ 2^{-j}\widetilde{\Psi}'_{j,k}(x_m^j) = v_m^j.$$

We also have

$$\sum_{k=0}^{N-1} s_k^0 \widetilde{\Phi}_{n_0,k}(2^{n_0}l) = S^{n_0}_{2^{n_0}l},$$

for $l = 0,1,\ldots,N_s-1$, where $N_s = 2^{n-n_0}$. Using these equations, we set up a system of linear algebraic equations for the unknown vector $\{s_k^0\}$,

$$\mathbf{A}\ \mathbf{s} = \mathbf{v},$$

where $\mathbf{s}\in\mathbf{R}^N$ is a shorthand notation of the original signal $\{s_k^0\}$, and $\mathbf{v}\in\mathbf{R}^{2N_s+N_s}$ is a data vector including the available coefficients. Matrix $\mathbf{A}$ is a $(2N_s+N_s)\times N$ matrix and has the following structure:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^{n_0} \\ \mathbf{S}^{n_0} \end{pmatrix},$$

where $\mathbf{A}^j$ is a $2N_z^j\times N$ submatrix whose entries are

$$(\mathbf{A}^j)_{2k,l} = \widetilde{\Psi}_{j,l}(x_k^j),$$

$$(\mathbf{A}^j)_{2k+1,l} = 2^{-j}\widetilde{\Psi}'_{j,l}(x_k^j),$$

for $k = 0,\ldots,N_z^j-1$ and $l = 0,\ldots,N-1$ and $\mathbf{S}^{n_0}$ is a $N_s\times N$ submatrix where

$$(\mathbf{S}^{n_0})_{k,l} = \widetilde{\Phi}_{n_0,l}(2^{n_0}k),$$

for $k = 0,\ldots,N_s$ and $l = 0,\ldots,N-1$.

Finally, the constraints such as the maximum distance between zero-crossings should be imposed for some "sparse" original signals (e.g., an impulse, a boxcar). This constraints may be expressed as

$$\mathbf{B}\ \mathbf{s} = \mathbf{0},$$

where $\mathbf{B}\in\mathbf{R}^{(2N_s+N_s)\times N}$.

The problem may now be stated as follows:

$$\textit{Minimize}\quad \|\mathbf{A}\ \mathbf{s}-\mathbf{v}\| \quad \textit{subject to}\quad \mathbf{B}\ \mathbf{s} = 0.$$

We obtain the least square solution

$$\hat{\mathbf{s}} = (\mathbf{A}^T\mathbf{A}+\lambda\mathbf{B}^T\mathbf{B})^{-1}\mathbf{A}^T\mathbf{v}.$$

We note that our formulation is completely linear except for the process of the zero-crossing detection. See [8] for the details and examples.

REFERENCES

[1] G. Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM J. Numer. Anal.*, 1991. to appear.

[2] P. J. Burt. Fast filter transforms for image processing. *Comput. Graphics and Image Processing*, 16:20–51, 1981.

[3] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *preprint*, 1990.

[4] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, 41:909–996, 1988.

[5] I. Daubechies. Orthonormal bases of compactly supported wavelets. II. variations on a theme. *preprint*, 1990.

[6] G. Deslauriers and S. Dubuc. Symmetric iterative interpolation processes. *Constructive Approximation*, 5:49–68, 1989.

[7] S. Dubuc. Interpolation through an iterative scheme. *J. Math. Anal. and Appl.*, 114:185–204, 1986.

[8] N. Saito and G. Beylkin. Multiresolution representations using the auto-correlation functions of compactly supported wavelets. Technical report, Schlumberger-Doll Research, 1991.

THIS PAGE IS BLANK

DUE TO A

PAGE NUMBERING ERROR

Non-Separable Bidimensional Wavelet Bases

Albert Cohen and Ingrid Daubechies

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

We build orthonormal and biorthogonal wavelet bases of $L^2(\mathbb{R}^2)$ with dilation matrices of determinant 2. As for the one dimensional case, our construction uses a scaling function which solves a two-scale difference equation associated to a FIR filter. Our wavelets are generated from a single compactly supported mother function. However, the regularity of these functions cannot be derived by the same approach as in the one dimensional case. We review existing techniques to evaluate the regularity of wavelets, and we introduce new methods which allow to estimate the smoothness of non-separable wavelets and scaling functions in the most general situations. We illustrate these with several examples.

I Introduction

In the most general sense, wavelet bases are discrete families of functions obtained by dilations and translations of a finite number of well chosen mother functions. The most well known are certainly dyadic orthonormal bases of $L^2(\mathbb{R})$, of the type

$$\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k) \quad j \in \mathbb{Z}, \quad k \in \mathbb{Z} . \tag{1.1}$$

These constructions have found many interesting applications, both in mathematics because they form Riesz bases for many functional spaces and in signal processing because wavelet expansions are more appropriate than Fourier series to represent the abrupt changes in non-stationary signals.

Several examples have been given by Meyer [Me1], Lemarié [Le] and Daubechies [Dau1], generalizing the classic Haar basis in which the mother wavelet $\psi = \chi_{[0,1/2]} - \chi_{[1/2,1]}$ suffers from a lack of regularity since it is not even continuous. All are based on the concept of multiscale analysis, i.e. a ladder of closed subspace $\{V_j\}_{j\in\mathbb{Z}}$ which approximates $L^2(\mathbb{R})$,

$$\{0\} \rightarrow \ldots V_1 \subset V_0 \subset V_{-1} \ldots \rightarrow L^2(\mathbb{R}) , \tag{1.2}$$

(note that in some papers and in the Meyer's book, the converse convention is used, i.e. $V_j \subset V_{j+1}$) and satisfies the following properties,

$$f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j-1} \Longleftrightarrow f(2^j x) \in V_0 , \tag{1.3}$$

(1.4) there exists a function $\varphi(x)$ in V_0 such that the set $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0 .

Since $V_0 \subset V_1$, the scaling function $\varphi(x)$ has to be the solution of a two scale difference equation,

$$\varphi(x) = 2\sum_{n\in\mathbb{Z}} c_n \, \varphi(2x - n) . \tag{1.5}$$

The associated wavelet is then derived from the scaling function by the formula

$$\psi(x) = 2\sum_{n\in\mathbb{Z}} (-1)^n \, \bar{c}_{1-n} \, \varphi(2x - n) . \tag{1.6}$$

In the standard interpretation of a multiresolution analysis, the projections of a function f on the spaces V_j are viewed as successive approximations to f, with finer and finer resolution as j decreases. The wavelets can then be used to express the additional details needed to go from one resolution to the next finer level, since the $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ constitute an orthonormal basis for W_0, the orthogonal complement of V_0 in V_{-1}. The whole set $\{\psi_k^j(x)\}_{j,k\in\mathbb{Z}}$ forms then an orthonormal basis of $L^2(\mathbb{R})$.

We are here interested in similar constructions adapted to functions or signals of more than one variable.

The most commonly used method to build a multiresolution analysis and wavelet bases in $L^2(\mathbb{R}^n)$ is the tensor product of a multiresolution analyses of $L^2(\mathbb{R})$. In $L^2(\mathbb{R}^2)$ it leads to a ladder of spaces $\mathcal{V}_j = V_j \otimes V_j \subset \mathcal{V}_{j-1}$ generated by the families,

$$\Phi_{k\ell}^j(x,y) = 2^{-j}\varphi(2^{-j}x - k)\,\varphi(2^{-j}y - \ell), \quad k,\ell \in \mathbb{Z} . \tag{1.7}$$

Three wavelets are then necessary to construct the orthogonal complement of V_0 in V_{-1}, namely,

$$\Psi_a(x,y) = \varphi(x)\psi(y) \tag{1.8}$$

$$\Psi_b(x,y) = \psi(x)\varphi(y) \tag{1.9}$$

$$\Psi_c(x,y) = \psi(x)\psi(y)\,. \tag{1.10}$$

Actually, the theory of multiresolution analysis, as it was introduced by S. Mallat and Y. Meyer (see [Mal] and [Mel]) was first motivated by the possibility of building these separable wavelets for the analysis of digital picture.

It is clear, however, that this choice is restrictive and that it gives a particular importance to the x and y directions, since Ψ_a and Ψ_b match respectively the horizontal and vertical details.

A more general way of extending multiresolution analysis to n dimensions consist in replacing the axioma (1.3) and (1.4) by

$$f(x) \in V_j \Longleftrightarrow f(Dx) \in V_{j-1} \tag{1.11}$$

(1.12) There exists a function ϕ in V_0 such that the set $\{\phi(x-k)\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis for V_0

where D is a $n \times n$ dilation matrix.

All the singular values $\lambda_1,\ldots,\lambda_n$ of D must satisfy

$$|\lambda_m| > 1\,, \tag{1.13}$$

to ensure that the approximation gets finer in every direction as j goes to $-\infty$. Furthermore, we require D to have integer entries. This condition means that the action of D on the translation grid $\mathbb{Z}^n$ leads to a sublattice $\Gamma \subset \mathbb{Z}^n$.

The number of basic wavelets required to characterize the orthogonal complement of V_0 in V_{-1} is in that case trivially given by the following heuristic argument. This complement should be generated by the action of $\mathbb{Z}^n$ on the basic wavelets, in the same way that V_0 is generated by the action of $\mathbb{Z}^n$ on ϕ, whereas V_{-1} is generated by the action of $D^{-1}\mathbb{Z}^n$. Consequently, each of the generating functions can be associated with an elementary coset of $D^{-1}\mathbb{Z}^n/\mathbb{Z}^n \sim \mathbb{Z}^n/D\mathbb{Z}^n$ except one which corresponds to the scaling function (see figure 1). Therefore, $d = |\det D| - 1$ different wavelets are needed. Note that it is not strictly necessary that the entries of D be integer to build wavelet bases using D as the elementary dilation. However, the condition seems to be necessary for the existence of a multiresolution analysis based on a single, real valued, compactly supported scaling function.

In this work we shall indeed focus on real valued, compactly supported scaling functions and wavelets. They have the advantage that the sequence $\{c_n\}_{n\in\mathbb{Z}}$ introduced in the two scale difference equation (1.5) is real and finite. These coefficients play an important part in the numerical applications because they are used directly in the Fast Wavelet Transform algorithm as decomposition and reconstruction filters. They constitute in that case an FIR (finite impulse response) filter which can be implemented very easily. Furthermore, this finite set of coefficients contains all the information about the multiresolution analysis since

the functions φ and ψ can be constructed as solutions of (1.5) and (1.6). Our starting point to build wavelet bases will thus be a finite set of coefficients and the associate two-scale difference equation, rather than the approximation spaces V_j themselves.

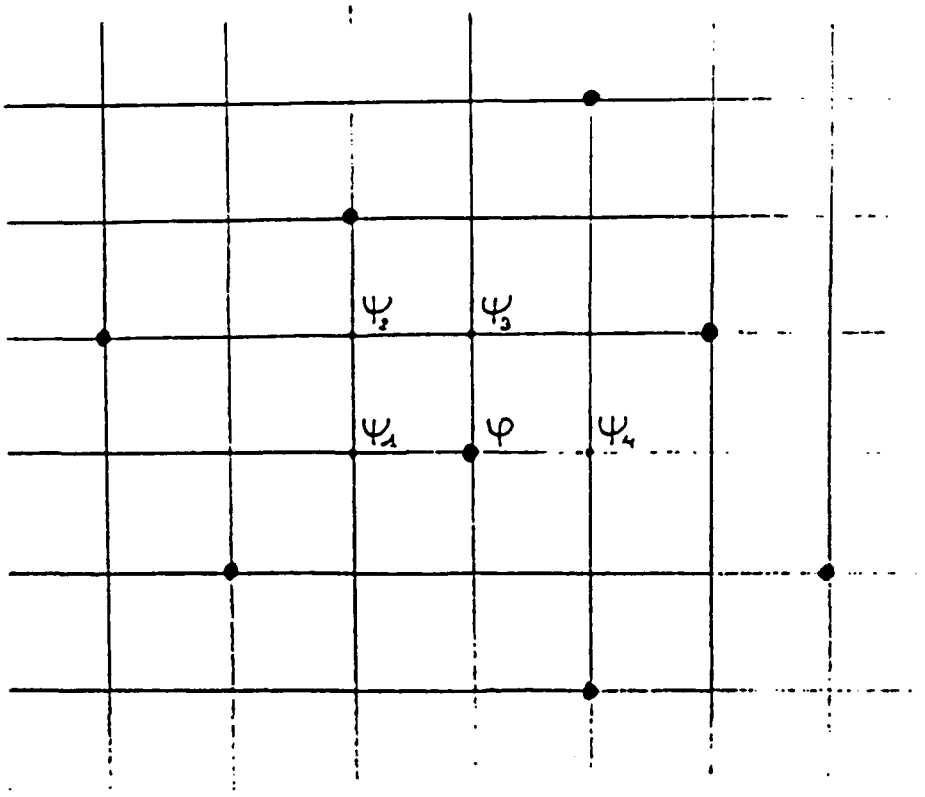


Figure 1

$\mathbb{Z}^2$ and $D\mathbb{Z}^2$ in the case where $D = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$

The scaling function and the four basic wavelets are indexed by an element of $\mathbb{Z}^2/D\mathbb{Z}^2$

The main difficulty in this approach is the design of the FIR filter $\{c_n\}_{n=0\ldots N}$ in such a way that φ and ψ are smooth and have orthonormal translates.

In the one dimensional case, it is shown in [Dau1] that orthonormal wavelets can be constructed by choosing a filter which corresponds to a particular case of exact reconstruction subband coding schemes, and which can be made arbitrarily regular by increasing the number of taps in a proper way. Several contributions have followed, giving supplementary information on the type of filter which has to be used (see [Me2], [DL], [Co1], [Dau2], [Co2], [Dau3]).

In the present bidimensional case, the design of filters associated to "nice wavelet bases" turns out to be more difficult because some of the one-dimensional techniques do not generalize trivially (or do not generalize at all!) to higher dimensions and new methods have to be introduced. This article concentrates on the situation where D is a 2×2 matrix with $|\det D| = 2$.

We deliberately restrain ourself to this set of matrices for two reasons:

- These dilations have already been considered by electrical engineers and seem to have interesting applications in signal analysis and image processing. For example, since only one basic wavelet is required, one may hope for a more isotropic analysis than with the separable construction. Subband coding schemes with decimation on the quincunx sublattice have been studied in the works of J. C. Feauveau [Fea] and M. Vetterli and J. Kovacevic [KV]. Our work is complementary to their signal processing approach since we investigate here the mathematical properties, such as the Hölder regularity of the wavelet bases associated to these schemes. This regularity is important when one asks that the reconstruction of the signal from the coarse scales has a smooth aspect (see section II.2).

- These dilations are simple and our study will be reduced to the case of two basic matrices. However, the difficulties which appear in the evaluation of the regularity of the corresponding wavelets are common to all the non-separable constructions, and the techniques that we develop to solve this problem can be used for other types of dilations. We believe that the set of integer matrices with $|\det D| = 2$ constitutes an interesting "laboratory case" in the general framework of multidimensional wavelets.

In the next section of this paper, we shall give an overview of different techniques which can be used in the construction of one dimensional compactly supported wavelets. Some new tools will be introduced specifically to be generalized and used in the multidimensional situation.

The third section examines the possible subband coding schemes with decimation on the quincunx sublattice and their general relations with non-separable wavelet bases.

In the fourth section, orthonormal bases of wavelets are constructed from such coding schemes. We show that for the same filters, different bases with widely differing regularity can be obtained, depending on the choice of the dilation matrix. Finally, we use a biorthogonal approach, in section five, to construct more symmetrical wavelet bases corresponding to linear phase filters and allowing a more isotropic analysis. We show that arbitrarily high regularity can be attained and we give some asymptotical results.

II The Construction of Compactly Supported Wavelets in One Dimension: A Complete Toolbox

The purpose of this section is to review, in the one dimensional case, many different techniques that can be used to build regular wavelets from subband coding schemes, theoretically and numerically. Some of these techniques, like the Littlewood-Paley estimation of smoothness, are not frequently used in the one dimensional case, but they turn out to be very useful for the non-separable bidimensional wavelets. For more details, the reader can also consult [Dau1], [Me1], [Ma1], [Ve1], [Dau2], [Me2], [Co2] ...

II.1 Wavelet bases and subband coding schemes

II.1.a The orthonormal case

Let $\{V_j\}_{j\in\mathbb{Z}}$ be a multiresolution analysis of $L^2(\mathbb{R})$. We can use the discrete Fourier transform of the finite sequence $\{c_n\}_{n=N_1}^{N_2}$, i.e. the transfer function

$$m_0(\omega) = \sum_{n\in\mathbb{Z}} c_n e^{-in\omega} = \sum_{n=N_1}^{N_2} c_n e^{-in\omega} , \tag{2.1}$$

to rewrite the two scale difference equation (1.5) that characterizes $\varphi(x)$. We suppose that the c_n are real. Taking the Fourier transform of (1.5) and (1.6) we obtain

$$\hat{\varphi}(2\omega) = m_0(\omega)\hat{\varphi}(\omega) \tag{2.2}$$

$$\hat{\psi}(2\omega) = e^{-i\omega}\,\overline{m_0(\omega+\pi)}\,\hat{\varphi}(\omega) = m_1(\omega)\hat{\varphi}(\omega) . \tag{2.3}$$

Two fundamental properties of $m_0(\omega)$ can be derived from the multiresolution analysis properties:

- Since $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_0, the Fourier transform $\hat{\varphi}(\omega)$ satisfies a Poisson identity
  $$\sum_{n\in\mathbb{Z}} |\hat{\varphi}(\omega+2n\pi)|^2 = 1 . \tag{2.4}$$
  Combined with (2.2) this leads to
  $$|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1 \tag{2.5}$$
  which may also be written as
  $$2\sum_{n\in\mathbb{Z}} c_n c_{n+2k} = \delta_{k,0} \quad (= 1 \text{ if } k=0,\ 0 \text{ otherwise}) . \tag{2.6}$$

- The denseness of $\{V_j\}_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ is equivalent to $\hat{\varphi}(0) = \int\varphi(x)dx = 1$ (see [Me1], [Ma1] or [Co1]).

Consequently, we have
$$m_0(0) = 1 \quad \text{and} \quad m_0(\pi) = 0 , \tag{2.7}$$
which may also be written as

$$\sum_{n=N_1}^{N_2} c_n = 1 \quad \text{and} \quad \sum_{n=N_1}^{N_2} (-1)^n c_n = 0 . \tag{2.8}$$

The subband coding scheme associated to our multiresolution analysis appears clearly in the Fast Wavelet Transform Algorithm of S. Mallat [Ma2]. Let us recall how it works. The initial data are considered as the approximation of a continuous function at the scale $j = 0$,

$$S_k^0 = \langle f|\varphi(x-k)\rangle, \quad k\in\mathbb{Z} . \tag{2.9}$$

This allows the computation of the approximations and the details at coarser scale i.e.

$$S_k^j = 2^{-j/2}\,(f|\varphi_k^j) \quad \text{and} \quad D_k^j = 2^{-j/2}\,(f|\psi_k^j), \quad j > 0\,. \tag{2.10}$$

(The coefficients are normalized in such way that if $f \equiv 1$ locally, then $S_k^j = 1$ in that area). The sequence $\{S_k^j\}_{k\in\mathbb{Z}}$ (resp. $\{D_k^j\}_{k\in\mathbb{Z}}$) is then derived from $\{S_k^{j-1}\}_{k\in\mathbb{Z}}$ by a convolution with the filter $m_0(\omega)$ (resp. $m_1(\omega)$) followed by a decimation of one sample out of two to keep the same total amount of information, i.e.

$$S_k^j = \sum_n c_{n-2k}\, S_n^{j-1}, \quad D_k^j = \sum_n (-1)^{n-1}\, c_{2k+1-n}\, S_n^{j-1}\,.$$

The algorithm then iterates on $\{S_k^j\}_{k\in\mathbb{Z}}$. Conversely, the sequence $\{S_k^{j-1}\}_{k\in\mathbb{Z}}$ can be recovered by applying the same filters $m_0(\omega)$ and $m_1(\omega)$ on $\{S_k^j\}_{k\in\mathbb{Z}}$ and $\{D_k^j\}_{k\in\mathbb{Z}}$ after inserting a zero between every pair of consecutive samples, and summing the two components (multiplied by two for normalization purposes), i.e.

$$S_n^{j-1} = 2\sum_k c_{n-2k}\, S_k^j + (-1)^{n-1}\, c_{2k+1-n}\, D_k^j\,.$$

All these operations, decomposition - decimation - interpolation - reconstruction, constitute a complete subband coding scheme as shown on figure 2. The property of exact reconstruction can now be derived in two ways. It is a natural consequence of the multiresolution approach, since $V_j = V_{j+1} \overset{\perp}{\oplus} W_{j+1}$ but it can also be viewed as a consequence of formula (2.5) for the filter m_0. This type of filter pair (m_0, m_1) is known as a pair of "conjugate quadrature filters" (CQF); they were first discovered by Smith and Barnwell in 1983 [SB1]. The design of FIR pairs, with real coefficients and perfect reconstruction, has been generalized in [Daul]. It also appears in [ASH], [SB2], [Vel].

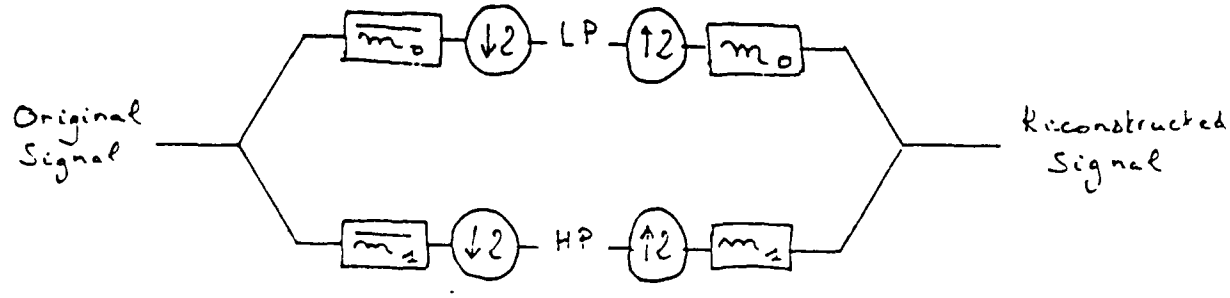


Figure 2
Subband coding scheme corresponding to the FWT algorithm.
The sign $2\downarrow$ stands for "decimation of one sample out of two" and $2\uparrow$ for the insertion of zeros at the intermediate values.

Since $m_0(\omega)$ is regular (it is a trigonometric polynomial) and since $m_0(0) = 1$, we can iterate (2.2) to obtain

$$\hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega) \,. \tag{2.11}$$

Given a conjugate quadrature filter $m_0(\omega)$ (i.e. a trigonometric polynomial satisfying (2.5) and (2.7)), it is thus possible to define the scaling function, either as a solution of the two scale difference equation (1.5), or explicitly with the above infinite product. However, this does not always lead to a multiresolution analysis: the function $\varphi(x) = \frac{1}{3}\chi_{[0,3]}$ generated by the CQF $m_0(\omega) = \frac{1+e^{3i\omega}}{2}$, for example, does not satisfy the orthonormality of the translates. Orthonormality of the $\varphi(x-k)$ turns out to be equivalent to the L^2 convergence of the truncated products $\hat{\varphi}_n(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega)\chi_{[-2^n\pi, 2^n\pi]}(\omega)$ to $\hat{\varphi}(\omega)$ (because $\{\varphi_n(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal set as soon as (2.5) is satisfied).

More precisely, the following result characterizes the subclass of CQF filter leading to a multiresolution analysis and orthonormal basis of wavelets.

Theorem 2.1 *Let $m_0(\omega)$ be a Conjugate Quadrature Filter. Then, the infinite product (2.11) leads to a multiresolution analysis if and only if there exist a compact set $K \subset \mathbb{R}$ such that,*

i) K contains a neighbourhood of the origin,
ii) $|K| = 2\pi$ and for all ω in $[-\pi, \pi]$, there exist $n \in \mathbb{Z}$ such that $\omega + 2n\pi \in K$,
iii) for all $n > 0$, $m_0(2^{-n}\omega)$ does not vanish on K.

The set K is said to be "congruent to $[-\pi, \pi]$ modulo 2π" (figure 3). The proof of this result can be found in [Col]. It exploits the continuity of m_0, the compactness of K and $m_0(0) = 1$ to show that (iii) is equivalent to $\hat{\varphi}(\omega) \geq c > 0$ on K. This is then sufficient to derive the L^2 convergence of the φ_n by Lebesgue's theorem. We shall use a multidimensional generalization of Theorem 2.1 in the fourth section.

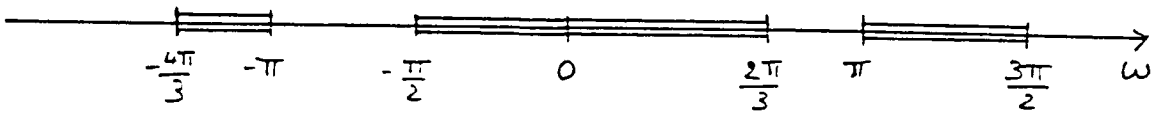


Figure 3
Example of compact set congruent to $[-\pi, \pi]$ modulo 2π.

II.1.b The biorthogonal case

The conjugate quadrature filters are a very particular case of subband coding scheme with perfect reconstruction, because identical filters (up to a complex conjugation) are used for both the decomposition and the reconstruction stages. If we don't impose this restriction, then the scheme uses four different filters: $\overline{m_0(\omega)}$ and $\overline{m_1(\omega)}$ for the decomposition, $m_0(\omega)$

and $m_1(\omega)$ for the reconstruction. Perfect reconstruction for any discrete signal is then ensured if,

$$\text{(2.12)} \qquad \begin{cases} \overline{m_0(\omega)}\,\tilde{m}_0(\omega) + \overline{m_1(\omega)}\,\tilde{m}_1(\omega) = 1 \\ \tilde{m}_0(\omega+\pi)\,\overline{m_0(\omega)} + \tilde{m}_1(\omega+\pi)\,\overline{m_1(\omega)} = 0 \end{cases}$$

$\tilde{m}_0(\omega)$ and $\tilde{m}_1(\omega)$ may thus be regarded as the solutions of a linear system. However, to avoid the infinite impulse response solutions, we shall force the determinant of this system to be $\alpha e^{ik\omega}$, $\alpha \neq 0$, $k \in \mathbb{Z}$. For sake of convenience we take $\alpha = -1$ and $k = 1$ (a change of these values would only mean a shift and a scalar multiplication on the impulse response of our filters). This leads to

$$\text{(2.13)} \qquad \overline{m_0(\omega)}\,\tilde{m}_0(\omega) + \overline{m_0(\omega+\pi)}\,\tilde{m}_0(\omega+\pi) = 1\,,$$

and

$$\text{(2.14)} \qquad m_1(\omega) = \epsilon^{-i\omega}\,\overline{\tilde{m}_0(\omega+\pi)},\quad \tilde{m}_1(\omega) = \epsilon^{-i\omega}\,\overline{m_0(\omega+\pi)}\,.$$

The formulas (2.13) and (2.14) are thus the most general setting for finite impulse response subband coders with exact reconstruction (in the two channels case). The functions $m_0(\omega)$ and $\tilde{m}_0(\omega)$ are called "dual filters". It is clear that the special case $m_0(\omega) = \tilde{m}_0(\omega)$ corresponds to the conjugate quadrature filters of II.1.a. However, dual filters are easier to design than CQF's. For example, if m_0 is fixed, $\tilde{m}_0$ can be found as the solution of a Bezout problem which is equivalent to a linear system. The coefficients of these filters can be very simple numerically (in particular they can have finite binary expansion which is very useful for practical implementation), furthermore they can be chosen symmetrical ("linear phase filter"), a property which is impossible to satisfy in the CQF case.

We can mimic, in this more general framework, the construction of orthonormal wavelets from CQF. Assuming that $m_0(0) = \tilde{m}_0(0) = 1$ and $m_0(\pi) = \tilde{m}_0(\pi) = 0$, we define

$$\text{(2.15)} \qquad \hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega)$$

$$\text{(2.16)} \qquad \hat{\psi}(2\omega) = m_1(\omega)\hat{\varphi}(\omega)$$

$$\text{(2.17)} \qquad \hat{\tilde{\varphi}}(\omega) = \prod_{k=1}^{+\infty} \tilde{m}_0(2^{-k}\omega)$$

$$\text{(2.18)} \qquad \hat{\tilde{\psi}}(2\omega) = \tilde{m}_1(\omega)\hat{\tilde{\varphi}}(\omega)\,.$$

In [CDF], the following theorem was proved,

Theorem 2.2

- *If* $\hat{\varphi}_n(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega)\chi_{[-2^n\pi,\,2^n\pi]}(\omega)$ *and* $\hat{\tilde{\varphi}}_n(\omega) = \prod_{k=1}^{n} \tilde{m}_0(2^{-k}\omega)\,\chi_{[-2^n\pi,\,2^n\pi]}(\omega)$ *converge in* $L^2(\mathbb{R})$ *respectively to* $\hat{\varphi}(\omega)$ *and* $\hat{\tilde{\varphi}}(\omega)$, *then the following duality relations are satisfied:*

$$\text{(2.19)} \qquad \langle \varphi(x-k) | \tilde{\varphi}(x-k') \rangle = \delta_{k,k'}$$

$$\text{(2.20)} \qquad \langle \psi_k^j | \tilde{\psi}_{k'}^{j'} \rangle = \delta_{j,j'}\,\delta_{k,k'}$$

and for all f in $L^2(\mathbb{R})$ one has the unique decomposition

$$f = \lim_{J \to +\infty} \sum_{j=-J}^{J} \sum_{k \in \mathbb{Z}} \langle f | \psi_k^j \rangle \, \tilde{\psi}_k^j \tag{2.21}$$

(in the L^2 sense).

- *If φ and $\tilde{\varphi}$ satisfy $|\hat{\varphi}(\omega)| + |\hat{\tilde{\varphi}}(\omega)| \leq C(1+|\omega|)^{-1/2-\epsilon}$ for some $\epsilon > 0$, then the families $\{\psi_k^j\}_{j,k \in \mathbb{Z}}$ and $\{\tilde{\psi}_k^j\}_{j,k \in \mathbb{Z}}$ are frames of $L^2(\mathbb{R})$.*
- *When these two properties hold, then $\{\psi_k^j, \tilde{\psi}_k^j\}_{j,k \in \mathbb{Z}}$ are biorthogonal (or dual) Riesz bases of $L^2(\mathbb{R})$.*

Many examples of these systems can be found in [CDF] and a sharper analysis of the frame conditions is developed in [CD]. We now recall a practical way of constructing φ and ψ numerically from a given subband coding scheme.

II.2 The cascade algorithm

In the last section we saw that the scaling function $\varphi(x)$ could be approximated, at least in $L^2(\mathbb{R})$, by a sequence of band limited functions $\{\varphi_n\}_{n>0}$ defined by

$$\hat{\varphi}_n(\omega) = \prod_{j=1}^{n} m_0(2^{-j}\omega) \, \chi_{[-2^n\pi, 2^n\pi]}(\omega) \, . \tag{2.22}$$

These functions are characterized by their sampled values at the points $2^{-n}k$ ($k \in \mathbb{Z}$), i.e.,

$$s_k^n = \varphi_n(2^{-n}k) \, . \tag{2.23}$$

This sequence can also be considered as the impulse response of the transfer function

$$S_n(\omega) = 2^n \prod_{j=1}^{n-1} m_0(2^j\omega) \tag{2.24}$$

$S_n(\omega)$ can be obtained recursively by the formula

$$S_{n+1}(\omega) = 2m_0(\omega) \, S_n(2\omega) \, . \tag{2.25}$$

In the time domain, (2.25) becomes an interpolation scheme; the sequence s_k^n is dilated by insertion of zeros ($S_n(\omega) \rightarrow S_n(2\omega)$) before being filtered (multiplication by $2m_0(\omega)$). We have thus,

$$s_p^{n+1} = 2 \sum_{k \in \mathbb{Z}} c_{p-2k} \, s_k^n \, . \tag{2.26}$$

This iterative process, which computes the $\{s_k^n\}_{k \in \mathbb{Z}}$ sequences from an initial Dirac sequence $\delta_{0,k}$ is called the "cascade algorithm". We illustrate it on figure 4 (our sequences are represented by piecewise constant functions).

Note that it identifies exactly with the reconstruction stage in the FWT algorithm described in II.1.a. The scaling function is thus approached by the reconstructed signal from a single approximation coefficient at a coarse scale. Similarly, the wavelet will be obtained by starting the reconstruction from a detail coefficient at a coarse scale (and thus applying $m_1(\omega)$ at the first step of the cascade).

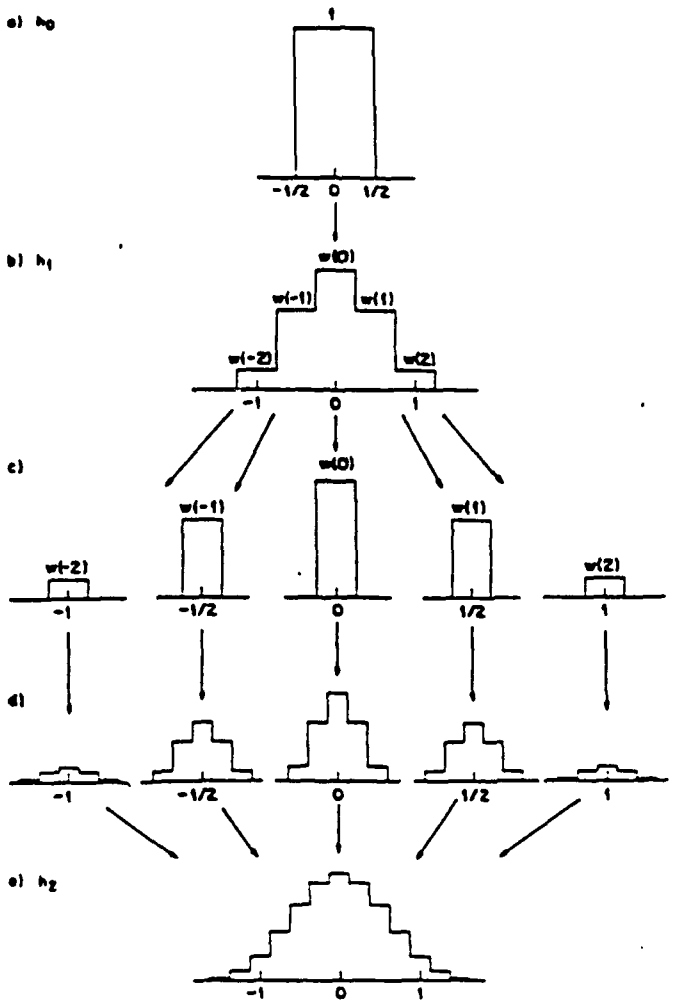


Figure 4
The cascade algorithm (from [Dau1])

This explains why subband coding schemes associated with regular wavelets are particularly interesting: the smoothness of the wavelet determines the appearance of the coarse scale components of the reconstructed signal. A smooth appearance is important for many applications such as compression where a big part of the finer scale information is thrown away.

In the biorthogonal case, the analysis and the synthesis wavelets (ψ and $\tilde{\psi}$) need not have the same regularity. As just discussed, smoothness is important for the reconstructing function; the analyzing function needs only to be sufficiently regular to ensure that the wavelet bases are unconditional, so that the FWT algorithm is stable. Note that an important property on the analyzing wavelet is cancellation, i.e. vanishing moments, ensuring small high scale coefficients for smooth regions in the function or signal to be analyzed.

Let us finally mention that this type of "refinement method" is well known in approximation theory as "stationary subdivision" (e.g. [CDM], [DyL]). Most of these papers are motivated by interpolation problems, where smooth curves or surfaces need to be constructed, connecting (or close to) given sparse data points. Consequently, they are mainly concerned with what we call the reconstruction stage and they do not study the existence of an associated subband coding scheme. This also means that they do not care about an easy way of encoding or representing the extra "detail information" ($\leftarrow W_j$) that can be added in going from one refinement level to the next one ($V_j \leftarrow V_{j-1}$). On the other hand, the subband coding literature seldom mentions the importance of the smoothness appearing in the cascade of the reconstruction from the low scales. Orthonormal and biorthogonal wavelet bases lead to an elegant combination of these two approaches.

We now present several different methods to estimate the regularity of the wavelets associated to a given subband coding scheme. We shall concentrate on the regularity of the scaling function which determines the regularity of the wavelet itself because $\psi(x)$ is a finite linear combination of translates of $\varphi(2x)$. Whatever the method used, if a global regularity of order r is achieved, then the cascade algorithm also converges uniformly up to this order (see [Dau1], [DL], [Co2]).

II.3 Regularity: the spectral approach

II.3.a A Fourier estimation of the Hölder exponent

Let us denote by C^α the Hölder space defined as follows. For $\alpha = n + \beta$, $\beta \in [0, 1[$, $f \in C^\alpha$ if and only if it is n times continuously differentiable and for all $x \neq y$, $\frac{|f^n(x) - f^n(y)|}{|x-y|^\beta} \leq C(f)$. Define also

$$\mathcal{F}_p^\alpha = \{f | (1 + |\omega|)^\alpha \hat{f}(\omega) \in L^p\} \quad (\alpha \geq 0,\ p \geq 1)\ . \tag{2.27}$$

It is well known (and easy to check) that $\mathcal{F}_\infty^{\alpha+1+\epsilon} \subset \mathcal{F}_1^\alpha \subset C^\alpha$, for $\epsilon > 0$. For compactly supported functions f, we also have

$$f \in C^\alpha \Longrightarrow f \in \mathcal{F}_\infty^\alpha \tag{2.28}$$

so that the decay of the Fourier transform can be used to evaluate the global regularity. To estimate this decay in the case of the scaling function, it is possible to use the factorization of $m_0(\omega)$: due to its cancellation at $\omega = \pi$, we have indeed

$$m_0(\omega) = \left(\frac{1 + e^{i\omega}}{2}\right)^N p(\omega)\ . \tag{2.29}$$

The infinite product (2.11) is thus divided in two parts. The first part, which comes from the factor $\left(\frac{1+e^{i\omega}}{2}\right)^N$ gives decay, since

$$\left|\prod_{k=1}^{+\infty}\left(\frac{1+e^{i2^{-k}\omega}}{2}\right)\right| = \left|\prod_{k=2}^{+\infty}\cos(2^{-k}\omega)\right| = \left|\frac{2}{\omega}\sin(\omega/2)\right| . \tag{2.30}$$

The second part, which involves the factor $p(\omega)$, can be controlled by a polynomial expression. Indeed, since $p(0) = 1$ and p is a regular function, the infinite product generated by the second factor satisfies

$$\left|\prod_{k=1}^{+\infty} p(2^{-k}\omega)\right| \le C \prod_{1\le k<\log(1+|\omega|)/\log 2} |p(2^{-k}\omega)| . \tag{2.31}$$

Defining, for $j > 0$,

$$B_j = \sup_{\omega\in\mathbb{R}}\left|\prod_{k=0}^{j-1} p(2^k\omega)\right| \tag{2.32}$$

and

$$b_j = \frac{\log B_j}{j\log 2} , \tag{2.33}$$

we obtain

$$\left|\prod_{k=1}^{+\infty} p(2^{-k}\omega)\right| \le C(B_j)^{\log(1+|\omega|)/\log 2} \le C(1+|\omega|)^{b_j} \tag{2.34}$$

and

$$|\hat{\varphi}(\omega)| \le C(1+|\omega|)^{b_j-N} . \tag{2.35}$$

Consequently, φ is in $\mathcal{F}_1^\alpha$ and C^α if $\alpha < N - b_j - 1$ for some $j > 0$. We see here that N must be large to allow high regularity since b_j is always positive. In fact, one can prove that if the wavelet is r times continuously differentiable then it has at least $r+1$ vanishing moments (see [Mel], [Dau1]), i.e. $\left(\frac{d}{d\omega}\right)^n(\hat{\psi})(0) = \left(\frac{d}{d\omega}\right)^n(m_0)(\pi) = 0$, for $n = 0,\ldots,r+1$ and thus $N \ge r+1$. These cancellations are also known as the Fix-Strang conditions [SF]; they are equivalent to the property that the polynomials of order $N-1$ can be expressed as linear combinations of the $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$. However, these conditions are necessary but not sufficient to ensure the regularity of the scaling function since the effect of N may be killed by a large value of b_j. Fortunately, this can be avoided by a careful choice of the filter $m_0(\omega)$ (and, in the biorthogonal case, additionally $\tilde{m}_0(\omega)$).

In the CQF-orthonormal case, a particular family of FIR filters indexed by N has been constructed in [Dau1]. This construction uses the polynomial

$$P_N(y) = \sum_{j=0}^{N-1}\binom{N-1+j}{j} y^j \tag{2.36}$$

(with the shorthand notation $y = \sin^2(\omega/2)$), which is the lowest degree solution of the Bezout problem

$$P_N(y)(1-y)^N + y^N P_N(1-y) = 1 . \tag{2.37}$$

The corresponding filters are defined by

$$m_0^N(\omega) = \left(\frac{1+e^{i\omega}}{2}\right)^N p_N(\omega) \tag{2.38}$$

with

$$|p_N(\omega)|^2 = P_N(y) = P_N\left(\frac{1-\cos\omega}{2}\right) . \tag{2.39}$$

The Fejer-Riesz lemma guarantees that there exists a FIR filter $p_N(\omega)$ which satisfies (2.39). It is clear that the CQF condition (2.5) is equivalent to (2.36) and the conditions in Theorem 2.1 are trivially satisfied with $K = [-\pi, \pi]$. For large values of N, the regularity $\alpha(N)$ of the associated scaling function is approximately $0.2\,N$ and the exact asymptotic ratio between $\alpha(N)$ and N can be determined. Intuitively speaking, this means that the contribution of $p_N(\omega)$ removes "eighty percent of the regularity" brought by the factor $\left(\frac{1+e^{i\omega}}{2}\right)^N$. For this estimation, we need to optimize the inequality (2.35), i.e. find the best possible exponent for the decay of $\hat{\varphi}(\omega)$.

II.3.b Optimal and asymptotical Fourier estimation: The role of fixed points

We start by defining "the critical exponent of $m_0(\omega)$":

$$b = \inf_{j>0} b_j = \inf_{j>0} \max_{\omega\in\mathbb{R}} \left(\frac{1}{j\log 2} \log\left|\prod_{k=0}^{j-1} p(2^k\omega)\right|\right) . \tag{2.40}$$

Then, it was proved in [Co2] that under the hypothesis $|p(\pi)| > |p(0)| = 1$ (satisfied in the present case (2.39)), $\hat{\varphi}(\omega)$ cannot have a better decay at infinity than $|\omega|^{b-N}$. If the infimum b is attained for some finite j, $b = b_j$, then this estimate is optimal.

How can we estimate the critical exponent? A first method consists in evaluating b_j for large values of j. Indeed, b is also the limit of the sequence b_j because the boundedness of p implies $b_J \le b_j + O(j/J)$. This may however require heavy computations.

In several cases, it is possible to use a more powerful method based on the transformation $\tau : \omega \mapsto 2\omega$ modulo 2π and the fixed points of its powers τ^n, $n > 0$. Indeed, let ω_0 be a fixed point of τ^n for $n > 0$ and define its orbit $\omega_j = \tau^j\omega_0$, for $j = 0, \ldots, n-1$. Since $p(\omega)$ has period 2π, we have

$$p(2^{nk}\omega_j) = p(\omega_j), \quad \text{for all } k > 0 \tag{2.41}$$

and consequently

$$b_{nk} \ge \frac{1}{n\log 2} \log\left|\prod_{j=0}^{n-1} p(\omega_j)\right| . \tag{2.42}$$

Letting k go to $+\infty$, this leads to

$$b \ge \frac{1}{n\log 2} \log\left|\prod_{j=0}^{n-1} p(\omega_j)\right| . \tag{2.43}$$

Fixed points of τ lead therefore to lower bounds for b and upper bounds for the regularity index. In fact they can do much better and provide optimal estimates for certain types of filters. Let us consider the smallest orbit of τ different from $\{0\}$, namely the pair $\left\{-\frac{2\tau}{3}, \frac{2\tau}{3}\right\}$. Note that, because our filters have real coefficients, $|m_0(\omega)|$ and $|p(\omega)|$ are even functions so that $|p(\frac{2\tau}{3})| = |p(-\frac{2\tau}{3})|$. The following result associates the value $|p(\frac{2\tau}{3})|$ and the critical exponent b.

Theorem 2.3 *Suppose that $p(\omega)$ satisfies*

$$|p(\omega)| \leq \left|p\left(\frac{2\pi}{3}\right)\right| \quad \text{if } |\omega| \leq \frac{2\pi}{3} \tag{2.44}$$

$$|p(\omega)p(2\omega)| \leq \left|p\left(\frac{2\pi}{3}\right)\right|^2 \quad \text{if } \frac{2\pi}{3} \leq |\omega| \leq \pi\ . \tag{2.44'}$$

Then

$$b = \frac{1}{\log 2} \log\left|p\left(\frac{2\pi}{3}\right)\right|\ . \tag{2.45}$$

Proof:

We already know from (2.43) that $b \geq \frac{1}{\log 2} \log\left|p\left(\frac{2\pi}{3}\right)\right|$. We now use the bounds on p to find an upper bound for b_j, $j > 0$. We can regroup the factors in (2.32) by packets of one or two elements in order to apply either (2.44) or (2.44') on each block. Since only the last factor can miss one of these two inequalities, we obtain

$$\left|\prod_{k=0}^{j-1} p(2^k\omega)\right| \leq \left|p\left(\frac{2\pi}{3}\right)\right|^{j-1} \sup|p|\ , \tag{2.46}$$

and thus,

$$b_j \leq \frac{1}{\log 2}\left[\frac{j-1}{j} \log\left|p\left(\frac{2\pi}{3}\right)\right| + \frac{\sup[\log|p|]}{j}\right]\ , \tag{2.47}$$

which leads to

$$b \leq \frac{1}{\log 2} \log\left|p\left(\frac{2\pi}{3}\right)\right| \tag{2.48}$$

and to (2.45). ∎

The equality (2.45) means that the worst decay of $\hat{\varphi}(\omega)$ occurs for the sequence $\omega_k = \frac{2\pi\tau}{3}$, $n > 0$. This is interesting, because (2.44) and (2.44') turn out to be satisfied in many cases and in particular for the whole family of CQF defined by (2.38), (2.39). This is easy to check directly for small values of N, since the inequalities can be rewritten as

$$P_N(y) \leq P_N\left(\frac{3}{4}\right) \quad \text{if } y \leq \frac{3}{4} \tag{2.49}$$

$$P_N(y)\ P_N(4y(1-y)) \leq \left(P_N\left(\frac{3}{4}\right)\right)^2 \quad \text{if } \frac{3}{4} \leq y \leq 1\ . \tag{2.49'}$$

The discussion for general N is more difficult and we refer to [CC] for a complete proof of (2.49), (2.49'). However, a similar result can be obtained in a simple way. To characterize the asymptotical behavior of the critical exponent when N goes to $+\infty$, one doesn't need the full force of (2.44), (2.44'), however. It can also be derived from a weaker, asymptotically valid inequality, as proved by H. Volkner in [V]:

Theorem 2.4 *Let $b(N)$ be the critical exponent associated to $m_0^N(\omega)$ and $\alpha(N)$ the Hölder exponent of the corresponding scaling function. Then*

$$\lim_{N\to+\infty} \frac{b(N)}{N} = \frac{\log 3}{2\log 2} \tag{2.50}$$

and

$$\lim_{n\to+\infty} \frac{\alpha(N)}{N} = \lim_{N\to+\infty} \frac{N - b(N)}{N} = 1 - \frac{\log 3}{2\log 2} \simeq 0.2075 . \tag{2.50'}$$

Proof:

This result can be viewed as a consequence of Theorem 2.3, but it can also be proved directly by using some properties of $P_N(y)$. Let us write (2.36) in the following form:

$$P_N(y) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} \left(\frac{1}{2}\right)^j (2y)^j . \tag{2.51}$$

From (2.36) we see that $P_N\left(\frac{1}{2}\right) = 2^{N-1}$; since P_N is an increasing function between 0 and 1, we have

$$P_N \le [\max(4y, 2)]^{N-1} = |g'(y)|^{N-1} . \tag{2.52}$$

It is now trivial to check that (2.49) and (2.49') are satisfied if we replace $P_N(y)$ by $g(y)$. The same argument used in the proof of Theorem 2.3 leads then to

$$b(N) \le \frac{N-1}{2\log 2} \log\left|g\left(\frac{3}{4}\right)\right| = \frac{N-1}{2\log 2} \log 3 \tag{2.53}$$

but from (2.43) we get

$$\begin{aligned} b(N) \ge \frac{1}{2\log 2} \log\left|P_N\left(\frac{3}{4}\right)\right| &\ge \frac{1}{2\log 2} \log\left|\binom{2N-2}{N-1}\left(\frac{3}{4}\right)^{N-1}\right| \\ &\ge \frac{N-2}{2\log 2} \log 3 . \end{aligned} \tag{2.54}$$

This proves the limit (2.50), and consequently (2.50') since the decay index of the Fourier transform is equivalent to the Hölder exponent when both tend to $+\infty$. ∎

The use of fixed points for optimal estimations of the spectral decay is thus very efficient when one is looking for arbitrarily high regularity since a sharp asymptotical result is obtained. For small filters, this method does not give a good result because the error on the exact regularity may have the same order as the value of the Hölder exponent itself. For such filters, other methods, which take advantage of the small number of taps in the filter, can be used to derive more precise estimations. We now describe these methods; they are typically based on matrix computations.

II.4 Regularity: Matrix based sharper estimates

II.4.a The Littlewood-Paley approach

We first recall some aspects of the Littlewood-Paley theory. Let $\gamma(x)$ be a real-valued, symmetrical function of the Schwartz class $\mathcal{S}(\mathbb{R})$, which satisfies

$$\text{(2.55)} \qquad \begin{cases} \hat{\gamma}(\omega) = 0 & \text{if } |\omega| \leq \frac{1}{2} \text{ or } |\omega| \geq \frac{5}{2} \\ \hat{\gamma}(\omega) > 0 & \text{if } \frac{1}{2} < |\omega| < \frac{5}{2} \end{cases}$$

so that the frequency axis is covered by the dyadic dilations of γ. Indeed, we have

$$\text{(2.56)} \qquad 0 < C_1 \leq \sum_{j=-\infty}^{+\infty} \hat{\gamma}(2^j\omega) \leq C_2 \text{ if } \omega \neq 0 .$$

Define for any f in $\mathcal{S}'(\mathbb{R})$ the dyadic blocks $\Delta_j(f)$ by

$$\text{(2.57)} \qquad \Delta_j(f) = 2^j\gamma(2^j\cdot) * f \Longleftrightarrow \hat{\Delta}_j(f) = \hat{\gamma}(2^{-j}\cdot)\hat{f}$$

The Littlewood-Paley theory tells us that several functional spaces can be characterized by examining only the L^p norm of these blocks. This is the case in particular for the Sobolev spaces $W^{p,s}$ and the Hölder spaces C^α, $\alpha > 0$. To do this, it is necessary to change slightly the definition of C^α when α is an integer; we shall say that a bounded function f is in C^n if and only if f^{n-1} belongs to the Zygmund class Λ, i.e. there exists a constant C such that, for all x and y, we have

$$\text{(2.58)} \qquad \left| f^{n-1}(x+y) + f^{n-1}(x-y) - 2f^{n-1}(x) \right| \leq C|y| .$$

With this convention, the Hölder space C^α is characterized by the following conditions,

$$\text{(2.59)} \qquad \|\Delta_j(f)\|_{L^\infty} \leq C2^{-\alpha j} \quad \text{when } j \geq 0$$

$$\text{(2.59')} \qquad f \text{ is a bounded continuous function.}$$

Note that the choice (2.55) for γ is arbitrary and that more general functions could be chosen to divide the Fourier domain into dyadic blocks. To derive these types of estimates on the scaling function φ, we introduce a tool which will be very useful in the bidimensional case.

Definition 2.1 *Let $L^2[0, 2\pi]$ be the space of 2π periodic, square integrable functions on $[0, 2\pi]$, and $C[0, 2\pi]$ the space of 2π periodic continuous functions. Then, for any $m(\omega)$ in $C[0, 2\pi]$, we define the transition operator T_m associated to $m(\omega)$ by*

$$\text{(2.60)} \qquad \begin{cases} T_m : \ L^2[0, 2\pi] \longrightarrow L^2[0, 2\pi] \\ f \longmapsto T_m f(\omega) = m\left(\frac{\omega}{2}\right) f\left(\frac{\omega}{2}\right) + m\left(\frac{\omega}{2} + \pi\right) f\left(\frac{\omega}{2} + \pi\right) . \end{cases}$$

Note that when $m(\omega)$ is a trigonometric polynomial, the study of T_m can be made in a finite dimensional space. More precisely, if we define

$$E(N_1, N_2) = \left\{ \sum_{n=N_1}^{N_2} h_n e^{in\omega} \middle| (h_{N_1}, \ldots, h_{N_2}) \in \mathbb{C}^{N_2-N_1+1} \right\} \tag{2.61}$$

then we have clearly

$$(f, m) \in [E(N_1, N_2)]^2 \Longrightarrow T_m f \in E(N_1, N_2) \, . \tag{2.62}$$

This is due to the contraction $\omega \mapsto \frac{\omega}{2}$ which appears in the definition (2.60) of T_m. If c_n is the n^{th} Fourier coefficient of $m(\omega)$, then the matrix of P_m in the complex exponentials basis is given by

$$T_{\ell,n} = (2c_{2\ell-n}) \, . \tag{2.63}$$

The size of this matrix P in $E(N_1, N_2)$ is $L \times L$ with $L = N_2 - N_1 + 1$. This operator has been studied by J. P. Conze and A. Raugi and several ideas presented below are due to their work [CR], [Con]. We shall use it to derive Littlewood-Paley type of estimations for the Hölder continuity of the scaling function. For this, we need the following result:

Lemma 2.5 *For all* $n > 0$,

$$\int_{-\pi}^{\pi} (T_m)^n f(\omega) d\omega = \int_{-2^n\pi}^{2^n\pi} f(2^{-n}\omega) \prod_{k=1}^{n} m(2^{-k}\omega) d\omega \, . \tag{2.64}$$

Proof:

We prove it by induction. It is clear for $n = 1$ since

$$\begin{aligned} \int_{-\pi}^{\pi} T_m f(\omega) d\omega &= \int_{-\pi}^{\pi} \left[m\left(\frac{\omega}{2}\right) f\left(\frac{\omega}{2}\right) + m\left(\frac{\omega}{2}+\pi\right) f\left(\frac{\omega}{2}+\pi\right) \right] d\omega \\ &= 2\int_{-\pi/2}^{\pi/2} [m(\omega)f(\omega) + m(\omega+\pi)f(\omega+\pi)] \, d\omega \\ &= 2\int_{-\pi}^{\pi} m(\omega)f(\omega)d\omega = \int_{-2\pi}^{2\pi} m\left(\frac{\omega}{2}\right) f\left(\frac{\omega}{2}\right) d\omega \, . \end{aligned}$$

Assuming (2.64) for n, we obtain at the next step,

$$\begin{aligned} \int_{-\pi}^{\pi} (T_m)^{n+1} f(\omega) \, d\omega &= \int_{-\pi}^{\pi} (T_m)^n T_m f(\omega) \, d\omega \\ &= \int_{-2^n\pi}^{2^n\pi} \left[\prod_{k=1}^{n} m(2^{-k}\omega) \right] \left[m(2^{-n-1}\omega) f(2^{-n-1}\omega) + \right. \\ &\qquad \left. m(2^{-n-1}\omega + \pi) f(2^{-n-1}\omega + \pi) \right] d\omega \\ &= 2^{n+1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left[\prod_{k=1}^{n} m(2^k\omega) \right] [m(\omega)f(\omega) + m(\omega+\pi)f(\omega+\pi)] \, d\omega \\ &= \int_{-2^{n+1}\pi}^{2^{n+1}\pi} \left[\prod_{k=1}^{n+1} m(2^{-k}\omega) \right] f(2^{-n-1}\omega) \, d\omega \, . \end{aligned}$$

This concludes the proof. ▪

We now suppose that $m(\omega)$ is a positive trigonometric polynomial in $E_M = E(-M, M)$ and that $m(0) = 1$ and $m(\pi) = 0$. Then m can be factorized as

$$m(\omega) = \cos^{2N}\left(\frac{\omega}{2}\right) p(\omega) \tag{2.65}$$

where $p(\omega)$ is a trigonometric polynomial that does not vanish for $\omega = \pi$. Note that necessarily $N \leq M$. From this cancellation property, we can derive,

Lemma 2.6 *$\left\{1, \frac{1}{2}, \ldots, 2^{-2N+1}\right\}$ are eigenvalues of T_m. The row vectors $p_j = (n^j)_{n=-M,\ldots,M}$, for $0 \leq j \leq 2N-1$ generate a subspace which is left invariant by T_m and contains one eigenvector for each of these $2N$ eigenvalues.*

Consequently, the orthogonal subspace defined by

$$F_N = \left\{ \sum_{n=-M}^{M} h_n e^{-in\omega} \,\middle|\, \sum_{n=-M}^{M} n^j h_n = 0, \quad j = 0, \ldots, 2N-1 \right\} \tag{2.66}$$

is right invariant by T_m.

Proof:

The factorization in (2.65) is equivalent to the cancellation rules

$$\sum_{n=-M}^{M} (-1)^n n^j c_n = 0 \text{ for } j = 0, \ldots, 2N-1 . \tag{2.67}$$

In particular, for $j = 0$, we have

$$\sum_n c_{2n} = \sum_n c_{2n+1} = \frac{1}{2} \quad (\text{because } m(0) = 1) . \tag{2.68}$$

This means that the sums of each column in the matrix of T (2.63) are equal to 1 and that $p_0 = (1, \ldots, 1)$ is a left eigenvector for the eigenvalue 1. For $0 < j \leq 2N-1$ we define $q_j = p_j P = \left(q_j^{-M}, \ldots, q_j^{M}\right)$; we have,

$$q_j^\ell = \sum_n n^j c_{2n-\ell} . \tag{2.69}$$

Thus, if ℓ is even

$$q_j^\ell = \sum_n \left(n + \frac{\ell}{2}\right)^j c_{2n} \tag{2.70}$$

and if ℓ is odd

$$q_j^\ell = \sum_n \left(n + \frac{1}{2} + \frac{\ell}{2}\right)^j c_{2n+1} . \tag{2.70'}$$

Using the binomial formula and the cancellation rules (2.67), we see that q_j is a linear combination of p_k for $k = 0, \ldots j$. The coefficient of p_j is given by the last term of the binomial and is thus equal to 2^{-j}. Consequently $\{p_j\}_{j=0,\ldots,2N-1}$ is a triangular basis for the left action of T_m and the eigenvalues are $\{2^{-j}\}_{j=0,\ldots,2N-1}$. ∎

We now come back to the scaling function φ, given by the infinite product

$$\hat{\varphi}(\omega) = \prod_{k=0}^{+\infty} m(2^{-k}\omega) . \tag{2.71}$$

Theorem 2.7 *Let F_N be the invariant subspace of T_m defined by (2.66). If λ is the largest eigenvalue of T_m restricted to F_N and if $|\lambda| < 1$, then, defining $\alpha = -\frac{1}{\log 2}\log(\lambda)$ (> 0), we have,*

- *φ is in $C^{\alpha-\epsilon}$ for all $\epsilon > 0$*
- *φ is in C^{α} if the restriction of T_m to the invariant subspace F_λ of eigenvalue λ is purely diagonal (i.e. $= \lambda I$).*

These two estimates are optimal if $\hat{\varphi}(\omega)$ does not vanish on $[-\pi, \pi]$.

Proof:

Consider the trigonometric polynomial

$$C_N(\omega) = (1 - \cos\omega)^N . \tag{2.72}$$

It clearly belongs to F_N.

Consequently, for all $n > 0$,

$$\begin{aligned} \int_{-\pi}^{\pi} (T_m)^n\, C_N(\omega) d\omega &\leq (2\pi)^{1/2} \left(\int_{-\pi}^{\pi} |(T_m)^n\, C_N(\omega)|^2\, d\omega \right)^{1/2} \\ &\leq C(\lambda + \epsilon)^n \text{ or } C\lambda^n \text{ if } T_m|_{F_\lambda} = \lambda I . \end{aligned} \tag{2.73}$$

We now use Lemma 2.5 combined with the inequality

$$C_N(\omega) \geq 1 \text{ when } \frac{\pi}{2} \leq |\omega| \leq \pi . \tag{2.74}$$

This leads us to

$$\begin{aligned} \int_{2^{n-1} \leq |\omega| \leq 2^n \pi} \hat{\varphi}(\omega)\, d\omega &\leq C \int_{2^{n-1}\pi \leq |\omega| \leq 2^n \pi} \prod_{k=1}^{n} m(2^{-k}\omega)\, d\omega \\ &\leq C \int_{-2^n\pi}^{2^n\pi} C_N(2^{-n}\omega) \prod_{k=1}^{n} m(2^{-k}\omega)\, d\omega \\ &= C \int_{-\pi}^{\pi} (T_m)^n\, C_N(\omega) d\omega . \end{aligned}$$

Consequently the Littlewood-Paley blocks satisfy the inequality

$$\|\hat{\Delta}_j(\varphi)\|_{L^1} \leq C2^{-(\alpha-\epsilon)j},\ \epsilon > 0,\ \alpha = -\frac{1}{\log 2}\log(\lambda) \tag{2.75}$$

$$\|\hat{\Delta}_j(\varphi)\|_{L^1} \leq C2^{-\alpha j}, \text{ if } T_m|_{F_\lambda} \text{ is purely diagonal.} \tag{2.75'}$$

Since $\|\Delta_j(\varphi)\|_{L^\infty} \leq \|\hat{\Delta}_j(\varphi)\|_{L^1}$ we obtain the announced regularity.

To prove that these estimates are optimal, we need to reverse all the inequalities which have been used. First, note that since $m(\omega)$ and $\hat{\varphi}(\omega)$ are positive, we have $\|\Delta_j(\varphi)\|_{L^\infty} = \|\hat{\Delta}_j(\varphi)\|_{L^1}$.

Let f_λ be an eigenfunction in F_λ. We have

$$\int_{-2^n\pi}^{2^n\pi} f_\lambda(2^{-n}\omega)\prod_{k=1}^{n} m(2^{-k}\omega)d\omega = \int_{-\pi}^{\pi}(T_m)^n f_\lambda(\omega)d\omega = \lambda^n\int_{-\pi}^{\pi} f_\lambda(\omega)d\omega \geq C\lambda^n. \tag{2.76}$$

Note that we have supposed that $\int_{-\pi}^{\pi} f_\lambda(\omega)\,d\omega \neq 0$. If $\int_{-\pi}^{\pi} f_\lambda(\omega)\,d\omega = 0$, then the argument has to be modified slightly; see below (after (2.78)). Since we have supposed that $\hat{\varphi}(\omega)$ does not vanish on $[-\pi, \pi]$, we have

$$\hat{\varphi}(\omega) \geq C\prod_{k=1}^{n} m(2^{-k}\omega) \text{ for all } n > 0 \text{ and } |\omega| \leq 2^n\pi\ . \tag{2.77}$$

Note that this hypothesis corresponds to the condition of Theorem 2.1 with $K = [-\pi, \pi]$. In a more general setting, we could replace the integrals on $[-2^n\pi, 2^n\pi]$ by integrals on 2^nK and the same results would hold. Combining (2.76) and (2.77) gives

$$\int_{-2^n\pi}^{2^n\pi} |\hat{\varphi}(\omega)|\,|f_\lambda(2^{-n}\omega)|d\omega \geq C\lambda^n\ . \tag{2.78}$$

(If $\int_{-\pi}^{\pi} f_\lambda(\omega)\,d\omega = 0$, then a slightly more sophisticated argument will do the trick. Lemma 2.5 still holds if the measure $d\omega$ is replaced by any other measure of the type $g(\omega)\,d\omega$ where g is a 2π-periodic, strictly positive, continuous function. We can always choose g such that

$$\int_{-\pi}^{\pi} f_\lambda(\omega)\,g(\omega)\,d\omega \neq 0\ ;$$

(2.76) then holds if $d\omega$ is replaced everywhere by $g(\omega)\,d\omega$. Since g is strictly positive, this modified version of (2.76) combined with (2.77), still implies (2.78).)

Since f_λ has a zero of order $2N$ at the origin, the function $\gamma(x)$, defined by $\hat{\gamma}(\omega) = |f_\lambda(\omega)|\,\chi_{[-\pi,\pi]}(\omega)$ is convenient for the Littlewood-Paley analysis of Hölder regularity less than $2N$. This is the case for φ since $2N+1$ vanishing moments would be necessary for a higher Hölder exponent than $2N$ (see [SF], [DL] or [DyL]). Consequently (2.78) tells us that φ cannot be more regular than C^α. To prove the optimality of $C^{\alpha-\epsilon}$ when $T_m|_{F_\lambda}$ is

not purely diagonal, it suffices to replace f_λ by a function g_λ such that $T_m g_\lambda = \lambda g_\lambda + \mu f_\lambda$ with $\mu \neq 0$. This leads to

$$\int_{-2^n\pi}^{2^n\pi} |\hat{\varphi}(\omega)|\,|g_\lambda(2^{-n}\omega)|\,d\omega \geq Cn\lambda^n \tag{2.78'}$$

which proves the optimality of $C^{\alpha-\epsilon}$.

The theorem is thus completely proved. ∎

Remarks:

- The estimates (2.75) and (2.75′) can be found by an equivalent technique, using the transition operator T_p corresponding to the factor $p(\omega)$ in (2.65). We simply consider the largest eigenvalue λ_p and iterate T_p on $f \equiv 1$. This leads to

$$\begin{aligned}\int_{2^{j-1}\pi\leq\omega\leq 2^j\pi} \hat{\varphi}(\omega)d\omega &\leq C\int_{2^{j-1}\pi\leq|\omega|\leq 2^j\pi} |\omega|^{-2N}\left[\prod_{k=1}^{j} p(2^{-k}\omega)\right] d\omega \\ &\leq C2^{-2Nj}\int_{-\pi}^{\pi}(T_p)^j 1\; d\omega \\ &\leq C(\lambda_p+\epsilon)^j\, 2^{-2Nj} \quad \left(\text{or } C\lambda_p^j\, 2^{-2Nj} \text{ if } T_p/F_{\lambda_p} = \lambda_p I\right)\end{aligned}$$

  and thus $\varphi \in C^{\alpha-\epsilon}$ with $\alpha = 2N - \frac{\log(\lambda_p)}{\log 2}$. This estimate is in fact the same as (2.75). Indeed, if μ is an eigenvalue of T_m in F_N, then its associated eigenfunction can be written as

$$f_\mu = \left(\sin^2\left(\frac{\omega}{2}\right)\right)^N g_\mu(\omega)\,. \tag{2.79}$$

  Replacing $m(\omega)$ by its factorized form in

$$\mu f_\mu(\omega) = f_\mu\left(\frac{\omega}{2}\right) m\left(\frac{\omega}{2}\right) + f_\mu\left(\frac{\omega}{2}+\pi\right) m\left(\frac{\omega}{2}+\pi\right) \tag{2.80}$$

  we obtain, after dividing by $[\sin^2(\frac{\omega}{2})\,\cos^2(\frac{\omega}{2})]^N$,

$$\mu 2^{2N} g_\mu(\omega) = g_\mu\left(\frac{\omega}{2}\right) p\left(\frac{\omega}{2}\right) + g_\mu\left(\frac{\omega}{2}+\pi\right) p\left(\frac{\omega}{2}+\pi\right)\,. \tag{2.81}$$

  We see here that the eigenvalues of T_p are exactly given by $\mu_p = 2^{2N}\mu$. This proves the equivalence between the two techniques.

- In general $m(\omega)$ is not a positive function. One can then define $M(\omega) = |m(\omega)|^2$ and use the operator T_M associated to $M(\omega)$. The result is an estimate of the L^2 norms of $\Delta_j(\varphi)$. Using the Cauchy-Schwarz inequality, we derive the following corollary,

Corollary 2.8 *Suppose that $M(\omega) = |m(\omega)|^2$ has a zero of order $2N$ at $\omega = \pi$. Define λ, the largest eigenvalue of T_M on F_N and $\alpha = -\frac{1}{2\log 2}\log(\lambda)$. Then, $\varphi \in H^{\alpha-\epsilon} \subset C^{\alpha-\frac{1}{2}-\epsilon}$ where H^s is the Sobolev space of index s. The value α is attained if $T_M|_{F_\lambda} = \lambda I$.*

Note that the Hölder exponent has no chance of being optimal because we have used the Cauchy-Schwarz inequality and $\hat{\varphi}(\omega)$ is not a positive function. The Sobolev exponent however is optimal. The regularity of compactly supported wavelets was estimated with this method in [Dau1].

The transition operator plays also a crucial role in the biorthogonal wavelet theory: we show in Appendix A how it can be used to prove that the families $\{\psi_k^j\}_{j,k\in\mathbb{Z}}$ and $\{\tilde{\psi}_k^j\}_{j,k\in\mathbb{Z}}$ are unconditional bases, with weaker assumption than the boundedness of $(1+|\omega|)^{1/2+\epsilon}(|\hat{\varphi}(\omega)|+|\hat{\tilde{\varphi}}(\omega)|)$ imposed in Theorem 2.2.

The optimal estimate for the global and local Hölder regularity of any wavelet can be estimated by another method developed by I. Daubechies and J. Lagarias in [DL]. We now recall its main points.

II.4.b The time domain approach

Let $m(\omega) = \sum_{n=0}^{N} c_n e^{in\omega}$ be a trigonometric polynomial such that $m(0) = 1$ and $m(\pi) = 0$. We do not require that $m(\omega)$ be positive. Let $\varphi(x)$ be the scaling function defined by the infinite product (2.71). It is at least a compactly supported distribution in $[0, N]$.

In the time domain approach, we represent $\varphi(x)$ by its "vector" form $w(x) : [0,1] \to \mathbb{R}^N$

$$[w(x)]_n = \varphi(x+n-1) \quad n = 1, \ldots, N \,. \tag{2.82}$$

From the two scale difference equation (1.5) we get

$$w(x) = \begin{cases} T_0 \; w(2x) & \text{if } \; x \le \frac{1}{2} \\ T_1 \; w(2x-1) & \text{if } \; x \ge \frac{1}{2} \end{cases} \tag{2.83}$$

where T_0 and T_1 are $N \times N$ matrices defined by

$$(T_0)_{i,j} = c_{2i-j-1} \qquad 1 \le i,j \le N \tag{2.84}$$

$$(T_1)_{i,j} = c_{2i-j} \qquad 1 \le i,j \le N \,. \tag{2.84'}$$

Using the notations

$$d_n(x) = n^{\text{th}} \text{ binary digit of } x \in [0,1]$$

$$\tau(x) = \begin{cases} 2x & \text{if } \; x \le \frac{1}{2} \\ 2x-1 & \text{if } \; x \ge \frac{1}{2} \end{cases} \qquad \text{(binary shift)} \,,$$

we can rewrite (2.83) as a "fixed point" equation

$$w(x) = T_{d_1(x)} \, w(\tau(x)) \,. \tag{2.85}$$

This leads to an evaluation of $w(x)$ and its derivative by an iterative process. The regularity of the result depends of course on the spectral properties of T_0 and T_1. Note that when

$m(\omega)$ has a zero of order L (as for the transition operator studied in the previous section), then the space F_L orthogonal to the vector $p_j = (n^j)_{n=1,\ldots,N}$ for $j = 0, \ldots, L-1$ is invariant by T_0 and T_1. This method gives sharp estimates on the local regularity in x by considering the products $T_{d_1(x)} \ldots T_{d_n(x)}$ for all $n \geq 0$. The main result on global regularity proved in [DL; Theorem 3.1] is the following

Theorem 2.9 *Suppose that there exist $\rho < 1$ such that, for all binary sequence $(d_j)_{j \in \mathbb{Z}}$ and all $m > 0$, we have*

(2.86) $$\|T_{d_1} T_{d_2} \ldots T_{d_m} |_{F_L}\| \leq C \rho^m \ .$$

Define $\alpha = -\frac{\log \rho}{\log 2}$. Then,

- *if α is not an integer, φ belongs to C^α*
- *if α is an integer, $\varphi^{\alpha-1}$ is almost Lipschitz:*

$$\text{for all } (x,t),\ |\varphi^{\alpha-1}(x+t) - \varphi^{\alpha-1}(x)| \leq C|t|\ |\log|t||\ .$$

Remark:

- The "generalized spectral norm"

(2.87) $$\rho(T_0, T_1) = \limsup_{m \to \infty} \left[\max_{\substack{d_j = 0 \text{ or } 1 \\ j=1,\ldots,m}} \|T_{d_1} T_{d_2} \ldots T_{d_m} |_{F_L}\|^{1/m} \right]$$

  gives a sharp estimate of the global regularity. Note that it is in general superior to the spectral radius of T_0 and T_1. When N is not too large it is possible to compute the exact value of $\rho(T_1, T_2)$. For example, in the case of orthonormal wavelets, the optimal Hölder exponent was found in [DL] for $N = 4$, 6 and 8. The same evaluation becomes more difficult for larger filters.

- The generalization of this approach in higher dimensions is not trivial. In particular, it involves nonstandard binary expansions depending on the dilation matrix which is used. We describe these techniques in Appendix B.

As a conclusion of this review of regularity estimators, we could say that these three approach are complementary: the time domain method gives sharp results but it is only practicable for small filters, the Littlewood-Paley estimates can be derived for longer filters but they will be optimal only if $m(\omega)$ is a positive function and finally, the Fourier approach is less precise but appropriate to asymptotical results on very large filters. Let us also mention that another method recently developed by O. Rioul [Ri] and based on $\ell^1(\mathbb{Z})$ norms estimates of the iterated filters leads to interesting results; in particular, it is still manageable for larger filters than the time domain method of [DL].

We are now ready to deal with the bidimensional wavelets. We start by examining the different subband coding schemes that can be used to build these non-separable multiscale bases.

III Two Channel Bidimensional Subband Coding Schemes

As mentioned previously, we shall concentrate on the dilation matrices of determinant equal to 2 or -2. In such conditions, the subband coding scheme that we consider split the signal in two channels (instead of four in the separable case) and only one wavelet is then necessary to characterize the detail coefficients at each scale. We first present a short summary of the equations satisfied by these filter. They are immediate generalizations of the results presented in II.1.

III.1 General conditions for exact reconstruction

As in the one dimensional case, the scheme that we are considering here is based on four fundamental operations:

- The action of two analyzing filters, one low pass $\tilde{M}_0(\omega) = \tilde{M}_0(\omega_1, \omega_2)$ and one high pass $\tilde{M}_1(\omega) = \tilde{M}_1(\omega_1, \omega_2)$
- Decimation on each channel by keeping only the samples on the sublattice $\Gamma = D\mathbb{Z}^2$
- Insertion of zero values at the intermediate points of $\mathbb{Z}^2/\Gamma$
- Interpolation by two synthesis filters, one lowpass $M_0(\omega) = M_0(\omega_1, \omega_2)$ and one high pass $M_1(\omega) = M_1(\omega_1, \omega_2)$, followed by reconstruction of the original signal by summation.

We see here that the conditions for perfect reconstruction will not depend on the dilation matrix D but only on the sublattice $\Gamma = D\mathbb{Z}^2$ that is generated (different matrices may lead to the same Γ). More precisely, there exist only two types of grid corresponding to a decimation of a factor 2 in $\mathbb{Z}^2$:

- The quincunx sublattice, shown on figure 5, is generated by the integer combinations of $(1,1)$ and $(1,-1)$.
- The column sublattice, shown on figure 6, is generated by the integer combinations of $(0,1)$ and $(2,0)$. It is of course equivalent to the row sublattice, by exchange of the coordinates.

The same arguments that were used in II.1.b show that perfect reconstruction is achieved by FIR filters, if and only if they satisfy (up to a shift) the following equations, which are similar to (2.13) and (2.14).

- In the quincunx case,

$$\overline{\tilde{M}_0(\omega)}\,\tilde{M}_0(\omega) + \overline{\tilde{M}_0(\omega+(\pi,\pi))}\,\tilde{M}_0(\omega+(\pi,\pi)) = 1 \tag{3.1}$$

and

$$M_1(\omega) = \epsilon^{-1(\omega_1+\omega_2)}\,\overline{\tilde{M}_0(\omega+(\pi,\pi))},\quad \tilde{M}_1(\omega) = \epsilon^{-1(\omega_1+\omega_2)}\,\overline{M_0(\omega+(\pi,\pi))}. \tag{3.2}$$

- In the column case,

(3.3) $$\overline{M_0(\omega)}\,\hat{M}_0(\omega) + \overline{M_0(\omega+(\pi,0))}\,\hat{M}_0(\omega+(\pi,0)) = 1$$

and

(3.4) $$M_1(\omega) = \epsilon^{-i\omega_1}\,\overline{\hat{M}_0(\omega+(\pi,0))},\quad \hat{M}_1(\omega) = \epsilon^{-i\omega_1}\,\overline{M_0(\omega+(\pi,0))}\,.$$

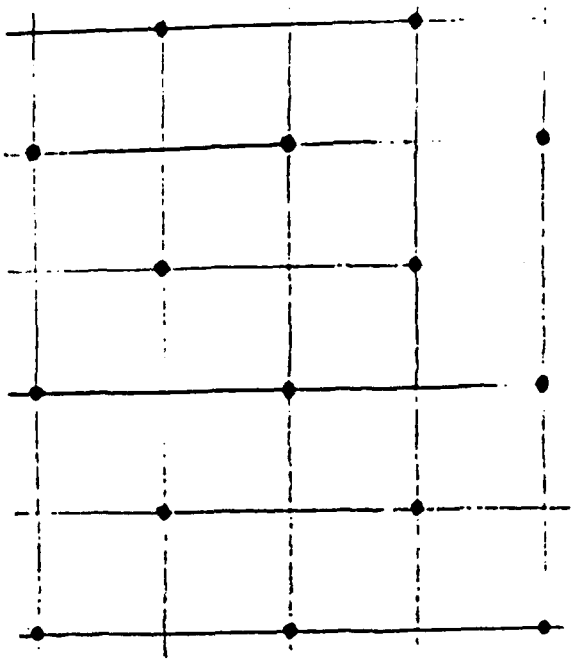


Figure 5
Quincunx decimation

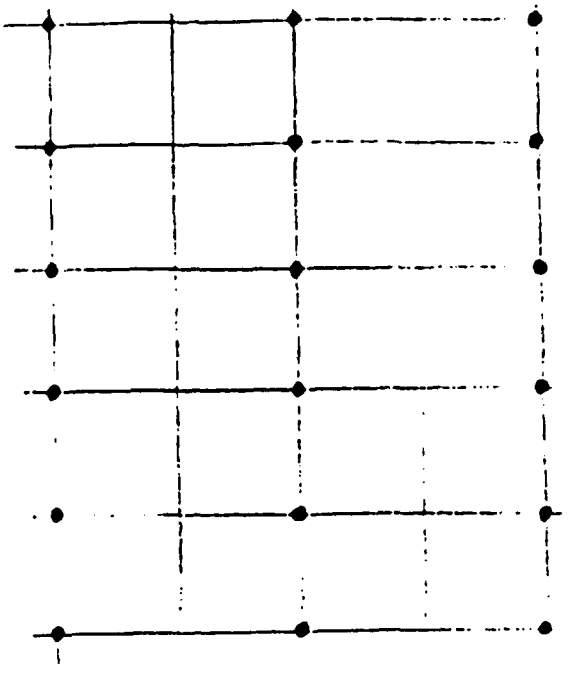


Figure 6
Column decimation

If the analysis and synthesis filters are equal, we find two generalization of the CQF condition (2.5). The formulas (3.1) and (3.2) become

(3.5) $$|M_0(\omega)|^2 + |M_0(\omega+(\pi,\pi))|^2 = 1,\quad M_1(\omega) = \epsilon^{-i(\omega_1+\omega_2)}\,\overline{M_0(\omega+(\pi,\pi))}\,;$$

whereas (3.3) and (3.4) become

(3.6) $$|M_0(\omega)|^2 + |M_0(\omega+(\pi,0))|^2 = 1,\quad M_1(\omega) = \epsilon^{-i\omega_1}\,\overline{M_0(\omega+(\pi,0))}\,.$$

As in the one dimensional situation, we want to build from these schemes the associated scaling function which can be viewed as the limit of the cascade-reconstruction algorithm.

III.2 Non-separable scaling function and wavelets

If c_{mn} are the Fourier coefficients of $M_0(\omega)$, i.e.

(3.7) $$M_0(\omega) = M_0(\omega_1,\omega_2) = \sum_{m,n} c_{mn}\,\epsilon^{-i(m\omega_1+n\omega_2)}\,.$$

Then, the associated scaling function $\phi(x) = \phi(x_1, x_2)$ satisfies a two scale difference equation

$$\phi(x) = 2\sum_{m,n} c_{mn}\, \phi(Dx - (m,n)) \tag{3.8}$$

and its Fourier transform can be expressed as an infinite product

$$\hat{\phi}(\omega) = \prod_{k=1}^{+\infty} M_0(D^{-k}\omega) \tag{3.9}$$

which is convergent if and only if $M_0(0) = 1$.

This scaling function has compact support if and only if $M_0(\omega)$ is an FIR filter. We see from (3.9) that ϕ will be highly dependent on the choice of D. For the same sublattice and the same filter, the results can be completely different for different D. The column sublattice for example is generated by both matrices $D_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $D_2 = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$, but the first one cannot lead to an L^2 scaling function. Indeed, we would have

$$\hat{\phi}_1(0, 2n\pi) = \prod_{k=1}^{+\infty} M_0(D_1^{-k}(0, 2n\pi)) = 1\ ,$$

for all $n \geq 0$. But since ϕ_1 is compactly supported and belongs to $L^2(\mathbb{R})$, it is also in $L^1(\mathbb{R})$ and its Fourier transform should tend to zero at infinity. We can also remark that only the eigenvalues of D_2 have their modulus strictly superior to 1.

The choice of the dilation matrix is thus very important. In fact, although the equations (3.1)–(3.2) are different from (3.3)–(3.4), the choice of the sublattice is less important: Indeed, for any dilation matrix D_1 such that $D_1\mathbb{Z}^2$ is the column sublattice, we can define

$$D_2 = P\, D_1 P^{-1} \quad \text{with} \quad P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\ . \tag{3.10}$$

Clearly, the image of $\mathbb{Z}^2$ by D_2 is now the quincunx sublattice. Then, for any filter $M_0^1(\omega)$ satisfying the column-CQF condition (3.6), the corresponding scaling function ϕ_1 can be written in the following way,

$$\begin{aligned} \hat{\phi}_1(\omega) &= \prod_{k=1}^{+\infty} M_0^1(D_k^{-k}\omega) \\ &= \prod_{k=1}^{+\infty} M_0^1(P^{-1}D_2^{-k}P\omega) \\ &= \hat{\phi}_2(P\omega) \end{aligned}$$

where $\hat{\phi}_2$ is also a scaling function defined by

$$\begin{cases} \hat{\phi}_2(\omega) = \prod_{k=1}^{+\infty} M_0^2(D_2^{-k}\omega) \\ M_0^2(\omega) = M_0^1(P^{-1}\omega)\ . \end{cases} \tag{3.11}$$

Since $P^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, we have

$$|M_0^2(\omega)|^2 + |M_0^2(\omega + (\pi, \pi))|^2 = |M_0^1(\omega_1, \omega_1 + \omega_2)|^2 + |M_0^1(\omega_1 + \pi, \omega_1 + \omega_2 + 2\pi)|^2 = 1 .$$

And thus M_0^2 satisfies the quincunx-CQF condition (3.5). A similar result holds of course if we start from two dual filters M_0^1 and $\tilde{M}_0^1$ which satisfy (3.3). This shows that the scaling functions associated to D_1 and D_2 are linked by the simple relation $\phi_2(x) = \phi_1(Px)$. Consequently we can restrain our study to the quincunx case. More generally, if D_1 and D_2 satisfy

(3.12) $$D_2 = PD_1P^{-1}$$

where P is a matrix having integer entries and determinant equal to 1, then we also have the same type of equivalence between the scaling functions. For this reason, we shall only consider the two simplest dilation matrices of determinant 2, which cannot be related as in (3.12) since they do not have the same eigenvalues:

(3.13) $$R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \left(\text{Rotation of } \frac{\pi}{4} \text{ and dilation of } \sqrt{2}\right)$$

and

(3.13′) $$S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \left(\text{Symmetry around } \frac{\pi}{8} \text{ and dilation of } \sqrt{2}\right) .$$

In both of these cases the image of $\mathbb{Z}^2$ is the quincunx sublattice. The wavelet ψ is then defined by

(3.14) $$\hat{\psi}(D\omega) = M_1(\omega)\hat{\phi}(\omega) \qquad D = R \text{ or } S ,$$

where $M_1(\omega)$ is defined by (3.5) in the orthogonal case, and by (3.2) in the biorthogonal case where we also have a dual wavelet defined by

(3.15) $$\hat{\tilde{\psi}}(D\omega) = \tilde{M}_1(\omega)\hat{\tilde{\phi}}(\omega) \qquad D = R \text{ or } S .$$

The goal is now to design filters leading to regular scaling functions and wavelets. We end this section by presenting two important families of filters. The regularity of the associated ϕ, ψ, $\tilde{\phi}$ and $\tilde{\psi}$ will be estimated in section IV and V by different techniques which all are natural generalizations of the one dimensional tools that we introduced previously.

III.3 Filter design

III.3.a The orthonormal case

Recall (see [Dau1]) that in $1D$, the CQF filter can be designed in the following way, in order to obtain wavelets with an arbitrarily high regularity:

1) For a given number N of vanishing moments, define m_0 by

(3.16) $$|m_0(\omega)|^2 = \left[\cos^2\left(\frac{\omega}{2}\right)\right]^N P_N\left[\sin^2\left(\frac{\omega}{2}\right)\right]$$

where $P_N(y)$ is a polynomial, solution of the Bezout problem

$$y^N P_N(1-y) + (1-y)^N P_N(y) = 1 . \tag{3.17}$$

The minimal degree choice is given by $P_N(y) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} y^j$.

2) Find the function $m_0(\omega)$ by using the Riesz lemma which guarantees that there exist a trigonometric polynomial solving (3.16).

Unfortunately, this last result does not generalize to higher dimensions. We thus have to find other means to build trigonometric polynomials which satisfy (3.5). One possible method is the "polyphase component" construction used by Vaidyanathan [Va] and M. Vetterli [Ve], [VK]. It is based on the remark that $M_0(\omega)$ satisfies (3.5) if and only if the polyphase matrix

$$H_0(\omega) = \frac{1}{\sqrt{2}} \begin{pmatrix} M_0(\omega) + M_0(\omega + (\pi, \pi)) & M_1(\omega) + M_1(\omega + (\pi, \pi)) \\ M_0(\omega) - M_0(\omega + (\pi, \pi)) & M_1(\omega) - M_1(\omega + (\pi, \pi)) \end{pmatrix} \tag{3.18}$$

is unitary for all ω. Since the product of two polyphase matrices is also a polyphase matrix for a third pair of filter, infinite families can be constructed by multiplying elementary building blocks of the type (3.18) as soon as we know some simple filters which satisfy (3.5). The disadvantage of this method is that it does not furnish the vanishing moments in a natural way. Recall (see [Me1]) that the N times differentiability of the function ψ implies

$$|\hat{\psi}(\omega)| \leq C \left(|\omega_1|^{N+1} + |\omega_2|^{N+1} \right) \quad (|\omega| \to 0) \tag{3.19}$$

and thus $M_0(\omega)$ has necessarily a zero or order $N+1$ at the frequency $\omega = (\pi, \pi)$. This can also be viewed as the Strang-Fix condition (see [SF]) for the regularity of the scaling function ϕ.

The simplest way to build such filters with N arbitrarily high is to remark that if $m_0(\omega)$ is a $1D$ solution of the CQF equation (2.5), then the $2D$ filter defined by

$$M_0(\omega) = M_0(\omega_1, \omega_2) = m_0(\omega_1) \tag{3.20}$$

satisfies the equation (3.5). It is apparently a good candidate for building regular wavelets since it has the same order of cancellation in (π, π) as $m_0(\omega)$ in π. This allows us to build an infinite family of filters with an arbitrarily high number of vanishing moments by posing

$$M_0^N(\omega) = m_0^N(\omega_1) \tag{3.21}$$

where $\{m_0^N(\omega)\}_{N>0}$ is the family of filters designed in [Dau1], defined by (2.35), (2.37) and (2.38). Note that the filter (3.21) has a unidimensional structure but since the dilation D contains either a rotation or a symmetry, the final analysis (using iterates of the filter) is performed in all the directions of the plane. In section IV, we shall take a closer look at the associated wavelets and their regularity. If $D = R$, then one can also derive another family of "almost" one-dimensional filters M_0 from unidimensional m_0 (they get again fanned out

to other directions by applying R^{-1}). Explicitly,

$$M_0(\omega_1,\omega_2) = \frac{1}{2}\left[m_0\left(\frac{\omega_1-\omega_2}{2}\right) + m_0\left(\frac{\omega_1-\omega_2}{2}+\pi\right)\right] + \frac{1}{2}\left[m_0\left(\frac{\omega_1-\omega_2}{2}\right) - m_0\left(\frac{\omega_1-\omega_2}{2}+\pi\right)\right] e^{i(\omega_1+\omega_2)/2} .$$

This construction corresponds to a filter with taps on two diagonals, $h_{n_1,n_2} = 0$ if $n_1 \neq n_2$ and $n_1 \neq -n_2+1$. It is easy to check that this M_0 satisfies (3.5) if m_0 satisfies $|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1$. If $m_0(0) = 1$, $m_0(\pi) = 0$, then $M_0(\pi,\pi) = 0$ follows, so that M_1, as defined in (3.5), satisfies $M_1(0,0) = 0$, as it should. One easily checks, however, that $\partial_{\omega_1} M_0(\pi,\pi)$ and $\partial_{\omega_2} M_0(\pi,\pi)$ cannot both be zero for these examples, so that the corresponding bases cannot possibly by C^1. Only the small examples are therefore of any interest; it seems possible (numerical experiment) to construct a continuous ϕ corresponding to a 4-tap filter in this way.

III.3.b The biorthogonal case

The filter design is clearly easier in the biorthogonal situation. One can start from a given filter $M_0(\omega)$ and find the dual $\tilde{M}_0(\omega)$ by solving linear equations.

In particular we can look for filters which have more isotropy than those of the family (3.21). Here, again, the one dimensional theory can help us to build families of filters in a simple way. Several examples of real and symmetrical dual filters have been designed by the authors and J. C. Feauveau in [CDF].

In these one dimensional construction the symmetry allows us to use the variable $y = \sin^2(\frac{\omega}{2})$ and to write the transfer functions as

$$m_0(\omega) = p(y) \text{ and } \tilde{m}_0(\omega) = \tilde{p}(y) \tag{3.22}$$

where p and $\tilde{p}$ are two polynomial satisfying

$$p(y)\tilde{p}(y) + p(1-y)\tilde{p}(1-y) = 1 . \tag{3.23}$$

In two dimensions, consider the variables $y_1 = \sin^2(\frac{\omega_1}{2})$ and $y_2 = \sin^2(\frac{\omega_2}{2})$. If the filters are symmetrical with respect to the vertical and the horizontal axes, the duality condition in (3.3) can be rewritten as

$$P(y_1, y_2)\tilde{P}(y_1, y_2) + P(1-y_1, 1-y_2)\tilde{P}(1-y_1, 1-y_2) = 1 , \tag{3.24}$$

where $$P(y_1, y_2) = M_0(\omega_1, \omega_2),\ \tilde{P}(y_1, y_2) = \tilde{M}_0(\omega_1, \omega_2) .$$

We see that a possible choice for P and $\tilde{P}$ is given by

$$P(y_1, y_2) = p(\alpha y_1 + (1-\alpha)y_2) \tag{3.25}$$

$$\tilde{P}(y_1, y_2) = \tilde{p}(\alpha y_1 + (1-\alpha)y_2) \tag{3.25'}$$

where α is in $[0,1]$. For an optimal isotropy it is natural to choose $\alpha = \frac{1}{2}$; in this case the diagonals are also symmetry axes. This choice is known in signal processing as the McClellan transform of the 1D filters p and $\tilde{p}$. Using the variable $z = \frac{1}{2}(y_1 + y_2)$ we can thus write

$$M_0(\omega) = p(z) \text{ and } \tilde{M}_0(\omega) = \tilde{p}(z) \tag{3.26}$$

where p and $\tilde{p}$ are polynomials satisfying (3.24). These polynomials must also satisfy

$$p(0) = \tilde{p}(0) = 1 \text{ and } p(1) = \tilde{p}(1) = 0 \tag{3.27}$$

which are necessary for the construction of wavelet bases. Note that we have

$$z = \frac{1}{2}\left(\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right) = \frac{1}{8}\left(4 - e^{i\omega_1} - e^{i\omega_2} - e^{-i\omega_1} - e^{i\omega_2}\right) \tag{3.28}$$

and thus z can be regarded as the transfer function of the filter which computes the discrete Laplacian with the formula

$$(\Delta_d x)_{m,n} = \frac{1}{8}(4x_{m,n} - x_{m-1,n} - x_{m+1,n} - x_{m,n-1} - x_{m,n+1}) . \tag{3.29}$$

Since a Laplacian scheme has frequently been proposed in image processing to detect the edges with a maximum isotropy (see [AB], [M]), it seems tempting to use z or one of its power as a high pass analyzing filter (and thus $1 - z$ as the corresponding low pass synthesis filter). This can be achieved in a very simple way, by a method already used to build biorthogonal bases in $L^2(\mathbb{R})$. Recall that $P_N(z) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} z^j$ is the lowest degree solution of the Bezout problem

$$z^N P_N(1-z) + (1-z)^N P_N(z) = 1 . \tag{3.30}$$

If we fix the reconstruction low pass as $M_0^N(\omega) = (1-z)^N$ (so that the analyzing high pass is, up to a shift, the N^{th} power of the Laplacian), then a possible choice for the dual filter is given by

$$\tilde{M}_0^{N,L}(\omega) = (1-z)^L P_{N+L}(z) \tag{3.31}$$

where L is a positive integer indicating the cancellation order of $\tilde{M}_0$ at $\omega = (\pi, \pi)$. L has to be chosen large enough so that both functions $\varphi(x)$ and $\tilde{\varphi}(x)$ satisfy the necessary conditions to generate a pair $\{\psi_k^j, \tilde{\psi}_k^j\}_{j\in\mathbb{Z}, k\in\mathbb{Z}^2}$ of unconditional Riesz bases (see Theorem 2.1 and Appendix A). We shall examine the properties of these functions and give an estimate of the minimal value of L in Section V.

We have now at hand two families of filters, orthonormal and biorthogonal with an arbitrarily high number of vanishing moments. We still have to know if these filters allow us to build wavelet bases with an arbitrarily high regularity like in the one dimensional case ([Dau1], [Co2]). As we shall see in the two next sections, the results of our investigations are very surprising and show that the multidimensional situation contains a lot of new difficulties from this point of view.

IV Orthonormal Bases of Non-separable Wavelets

Let us consider the family of CQF filters defined by

$$M_0^N(\omega_1, \omega_2) = m_0^N(\omega_1) \tag{4.1}$$

with

$$|m_0^N(\omega)|^2 = \left[\cos^2\left(\frac{\omega}{2}\right)\right]^N \sum_{j=0}^{N-1} \binom{N-1+j}{j} \left[\sin^2\left(\frac{\omega}{2}\right)\right]^j \tag{4.2}$$

and the associated scaling functions for the dilations S and R:

$$\hat{\phi}_{N,S}(\omega) = \prod_{k=1}^{\infty} M_0^N(S^{-k}\omega) \tag{4.3}$$

$$\hat{\phi}_{N,R}(\omega) = \prod_{k=1}^{\infty} M_0^N(R^{-k}\omega)\,. \tag{4.4}$$

IV.1 Orthonormality of the translates

A first requirement is that the $\mathbb{Z}^2$-translates of $\phi_{N,S}$ and $\phi_{N,R}$ are orthonormal. This is a necessary and sufficient condition to generate multiresolution analyses and orthonormal bases of wavelets.

Theorem 4.1 *For all $N > 0$, the functions $\phi_{N,S}$ and $\phi_{N,R}$ have orthonormal translates and generate wavelet bases of the type $2^{-j/2}\psi(D^{-j}x - k)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}^2$, $D = S$ or R.*

Proof:

By a trivial generalization of Theorem 2.1, this orthonormality is ensured if and only if $|\hat{\phi}(\omega)| \geq C > 0$ on a compact set K congruent to $[-\pi, \pi]^2$ modulo $2\pi\mathbb{Z}^2$ which contains a neighbourhood of the origin.

It is clear that $M_0^N(\omega)$ vanishes only on the vertical lines $\omega_1 = (2k+1)\pi$, $k \in \mathbb{Z}$. Consequently we see that the simple choice $K = [-\pi, \pi]^2$ is not convenient since for both dilations, we have

$$D^{-1}(\pi, \pi) = (\pi, 0) \tag{4.5}$$

and thus

$$\hat{\phi}(\pi, \pi) = 0\,. \tag{4.6}$$

Recall that in the one dimensional case, the trivial choice $K = [-\pi, \pi]$ was convenient for the family $m_0^N(\omega)$. Here we have to use a compact set K slightly different from $[-\pi, \pi]^n$ so that $D^{-j}K \cap \{\omega_1 = (2k+1)\pi\}$ is empty for all $j > 0$ and for all k in $\mathbb{Z}$. This can be done very easily by removing small neighbourhoods of (π, π) and $(-\pi, -\pi)$ and translating them by $(-2\pi, 0)$ and $(2\pi, 0)$ as shown in figure 7.

One checks easily that all the sets $D^{-j}K$ for $j > 0$ are contained in the strip $|\omega_1| \leq \pi - \epsilon$, $\epsilon > 0$ where $M_0^N(\omega)$ does not vanish.

We now have to check the regularity of the scaling functions which have been obtained. We shall see that the results are completely different depending on whether one chooses S or R as the dilation matrix.

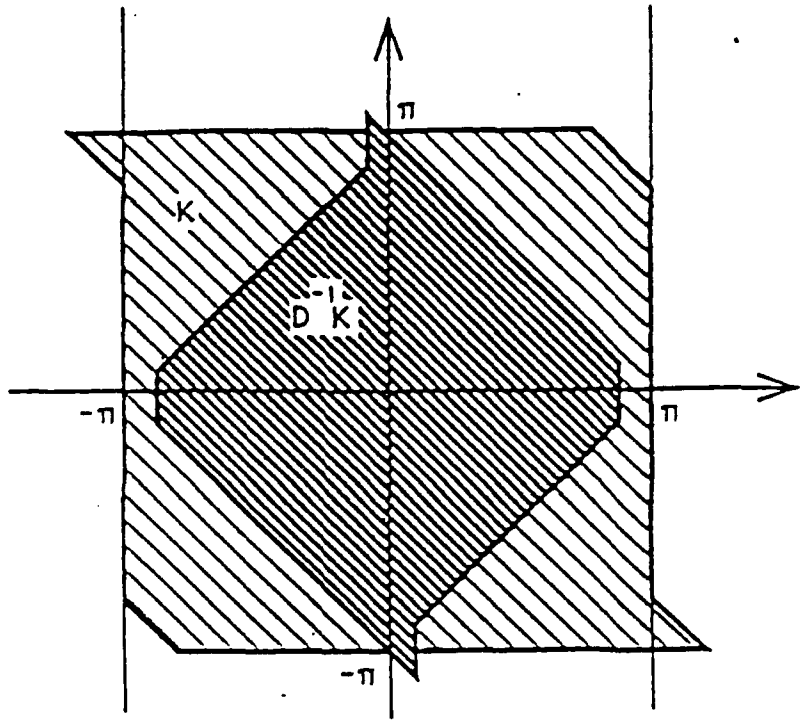


Figure 7
The convenient compact set K congruent to $[-\pi, \pi]^2$:
Neighbourhoods of (π, π) and $(-\pi, -\pi)$ have been shifted so that $\hat{\varphi}$ does not vanish on K

IV.2 The symmetry dilation case

In this case the dilation matrix is $S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and its inverse is $S^{-1} = \frac{1}{2}S$. Since $M_0^N(\omega) = m_0^N(\omega_1)$, we have to consider the sequence $\{[S^{-j}\omega]_1\}_{j>0}$ for a given $\omega = (\omega_1, \omega_2)$. Clearly, it has the following form:

$$\frac{1}{2}(\omega_1 + \omega_2),\ \frac{1}{2}\omega_1,\ \frac{1}{4}(\omega_1 + \omega_2),\ \frac{1}{4}\omega_1, \ldots, 2^{-j}(\omega_1 + \omega_2),\ 2^{-j}\omega_1, \ldots$$

Since $S^{-2} = \frac{1}{2}I$, the odd and the even parts are simple dyadic sequences and this leads to:

$$\hat{\phi}_{N,S}(\omega) = \hat{\varphi}_N(\omega_1 + \omega_2)\ \hat{\varphi}_N(\omega_1) \tag{4.7}$$

or

$$\phi_{N,S}(x) = \varphi_N(x_2)\ \varphi_N(x_1 - x_2) \tag{4.8}$$

where φ_N is the one dimensional scaling function. The associated wavelet is defined by

$$\hat{\psi}_{N,S}(\omega) = M_1(\omega)\hat{\phi}_{N,S}(\omega) = \hat{\psi}_N(\omega_1 + \omega_2)\ \hat{\varphi}_N(\omega_1) \tag{4.9}$$

or

$$\psi_{N,S}(x) = \psi_N(x_2)\,\varphi_N(x_1 - x_2)\,. \tag{4.10}$$

We see here that the scaling function and wavelet are in this case separable in the sense that they can be expressed directly in terms of the one dimensional functions φ_N and ψ_N. This separability can be explained by the fact that S is similar to the matrix $\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, which is simply a dilation by a factor 2 (in one) direction, followed by an exchange of the axes. The regularity can of course be made arbitrarily high since it is directly given by the Hölder exponent of φ_N.

Remark:

Theorem 4.1 is not necessary here to prove the orthonormality of the translates since it is a trivial consequence of the separability formulas (4.7) and (4.8).

We now consider the case of the matrix R which is by far less trivial.

IV.3 The rotation dilation case

We now have $R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and $R^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. The sequence $\{[R^{-j}\omega]_1\}_{j>0}$ is then,

$$\frac{1}{2}(\omega_1+\omega_2),\ \frac{1}{2}\omega_2,\ \frac{1}{4}(\omega_2-\omega_1),-\frac{1}{4}\omega_1,-\frac{1}{8}(\omega_1+\omega_2),$$

$$-\frac{1}{8}\omega_2,\frac{1}{16}(\omega_1-\omega_2),\frac{1}{16}(\omega_1),\frac{1}{32}(\omega_1+\omega_2),\frac{1}{32}\omega_2,\ldots$$

Here the first power of R^{-1} proportional to the identity is $R^{-4} = -\frac{1}{4}I$. Consequently, it is not possible to use the one dimensional scaling functions and wavelets to express the ϕ_N and ψ_N in a separable way. We first consider the case $N = 1$ which corresponds to the Haar filter. The result of the cascade algorithm with this filter shows how different the situation is when R is used instead of S.

IV.3.a The twin dragon

For $M_0^1(\omega) = \frac{1+e^{-i\omega_1}}{2}$, the function $\phi_{1,R}$ satisfies

$$\phi_{1,R}(x) = \phi_{1,R}(Rx) + \phi_{1,R}(Rx - (1,0)) \tag{4.11}$$

and

$$\hat{\phi}_{1,R} = \prod_{k=1}^{\infty} M_0^1(R^{-k}\omega)\,. \tag{4.12}$$

By iteration of the cascade algorithm, one finds that ϕ is the characteristic function of a well known fractal set called the "twin dragon" (see [K]) shown in figure 8. This set can be

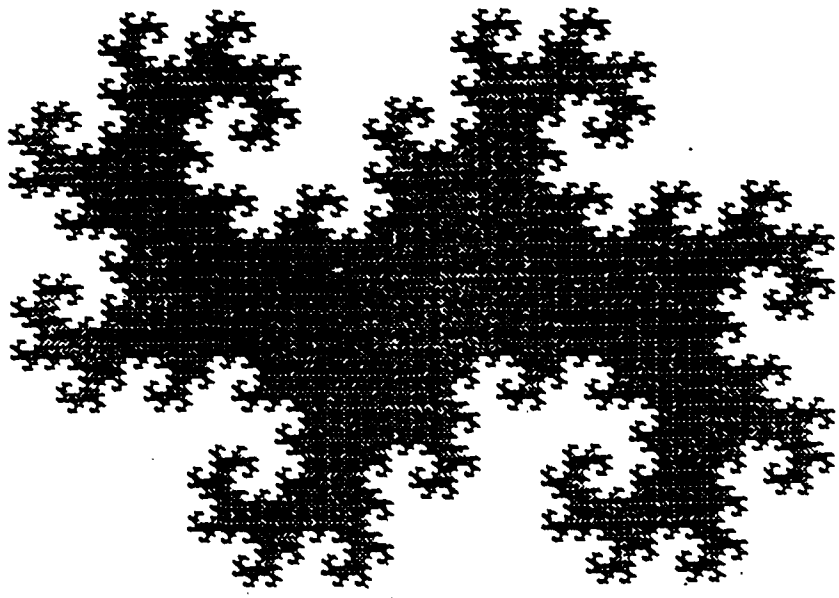


Figure 8
The "twin dragon" set Δ

defined directly in the complex plane as

$$\Delta = \left\{ \sum_{n=1}^{\infty} \epsilon_n \left(\frac{1-i}{2} \right)^n \Big/ \{\epsilon_n\}_{n \in \mathbb{N}} \in \{0,1\}^{\mathbb{N}} \right\} \tag{4.13}$$

and it is clear that $\phi_{1,R} = \chi_\Delta$ solves (3.41) since we have

$$\Delta = \left(\frac{1-i}{2} \right) \Delta \cup \left(\frac{1-i}{2} \right) (\Delta + 1) \sim R^{-1}\Delta \cup R^{-1}(\Delta + (0,1)) . \tag{4.14}$$

The self similarity of Δ is thus expressed by the two scale difference equation (4.11), but furthermore, since the family $\{\phi_{1,R}(x-k)\}_{k \in \mathbb{Z}^2}$ is orthonormal (by Theorem 4.1) and since $|\Delta| = \hat{\phi}_{1,R}(0) = 1$, these integer translates constitute a fractal tiling of the whole plane $\mathbb{R}^2$ (similarly to the squares obtained in tensor product situation with the same filter). This beautiful property has been remarked independently by W. Madych and K. Gröchenig [MG] and W. Lawton and H. Resnikoff [LR]. More generally, such tilings can be derived by considering a two scale difference equation of the type

$$\phi(x) = \sum_{i=1}^{d} \phi(Dx + e_i) \tag{4.15}$$

where D is a dilation matrix and $\{e_i\}_{i=1,\dots,d}$ are d representatives of $\mathbb{Z}^n/D\mathbb{Z}^n$ ($d = |\det D|$). This scaling function and the corresponding wavelet do not seem however of great interest for image processing: not only are they discontinuous but the set of discontinuity is a very chaotic fractal curve. Nevertheless the twin dragon is important in estimating the regularity (local and global) of the wavelets with dilation matrix R. Indeed, if we want to generalize the method of [DL] (see Section II.4.6), it is necessary to consider the expansion of any point in $\mathbb{C}$ in terms of the power of $\left(\frac{1\pm i}{2}\right)$ ($\sim R^{-1}$), which also means that the point is considered as the limit of a "dragonic sequence" $\{\Delta_j\}_{j\in\mathbb{Z}}$ with $\Delta_j \subset \Delta_{j-1}$ and $|\Delta_j| = 2^{-j}$. These "dragonic expansion" techniques are described in Appendix B.

Let us now examine the functions obtained with higher order filters which have more vanishing moments.

IV.3.b Higher order filters

We are interested in the family of scaling function $\phi_{N,R}$, $N > 1$.

Recall that in the one dimensional case, the asymptotic result ensuring arbitrarily high regularity (Theorem 2.4, Section II.3.6) is based on the value of $\left|m_0\left(\pm\frac{2\pi}{3}\right)\right|$ since $\left\{-\frac{2\pi}{3}, \frac{2\pi}{3}\right\}$ is a cyclic orbit of $\omega \mapsto 2\omega$ modulo 2π. In the present case similar considerations for a fixed orbit of $\omega \mapsto R\omega$ modulo $2\pi\mathbb{Z}^2$, lead to an opposite result: arbitrarily high regularity cannot be obtained by increasing the number of vanishing moments. More precisely, we have

Theorem 4.2 *For all $N > 0$, the function $\phi_{N,R}$ is not in $C^1(\mathbb{R}^2)$.*

Proof:

This is of course true for $N = 1$ since we obtain the twin dragon. For $N > 1$, we shall prove a stronger result: the decay at infinity of $\hat{\phi}_{N,R}(\omega)$ cannot be majorated by $C|\omega|^{-1}$ (which is a necessary condition for $\phi_{N,R}$ to be in C^1 because it is a compactly supported function). For this we consider the orbit of $\omega \mapsto R\omega$ modulo $2\pi\mathbb{Z}^2$ given by the four points $\left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)$, $\left(\frac{2\pi}{5}, -\frac{4\pi}{5}\right)$, $\left(-\frac{2\pi}{5}, -\frac{4\pi}{5}\right)$ and $\left(-\frac{2\pi}{5}, \frac{4\pi}{5}\right)$. Let us denote $v_0 = \left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)$ and $v_j = R^j v_0$. One checks easily that

$$|\hat{\phi}_{N,R}(v_0)| = C_N \neq 0 \text{ for all } N > 0 . \tag{4.16}$$

We then have, for all $N > 0$,

$$|\hat{\phi}_{N,R}(v_j)| = C_N \left|m_0^N\left(\frac{2\pi}{5}\right)\right|^j . \tag{4.17}$$

From the definition of m_0^N we have

$$\left|m_0^N\left(\frac{2\pi}{5}\right)\right|^2 = \left[\cos^2\left(\frac{\pi}{5}\right)\right]^N P_N\left(\sin^2\left(\frac{\pi}{5}\right)\right) \tag{4.18}$$

and we know from (2.51) that

$$P_N(y) \leq (4y)^{N-1}, \text{ if } \frac{1}{2} \leq y \leq 1 . \tag{4.19}$$

Because $\cos^2\left(\frac{\pi}{5}\right) > \cos^2\left(\frac{\pi}{4}\right) = \frac{1}{2}$, we can write

$$\begin{aligned}\left|m_0^N\left(\frac{2\pi}{5}\right)\right|^2 &= 1 - \left[\sin^2\left(\frac{\pi}{5}\right)\right]^N P_N\left[\cos^2\left(\frac{\pi}{5}\right)\right] \\ &\geq 1 - \left[\sin^2\left(\frac{\pi}{5}\right)\right]^N \left[4\cos^2\left(\frac{\pi}{5}\right)\right]^{N-1} \\ &= 1 - \sin^2\left(\frac{\pi}{5}\right)\left[\sin^2\left(\frac{2\pi}{5}\right)\right]^{N-1}\end{aligned}$$

and thus, since $|v_j| \geq 2^{j/2}$,

$$\begin{aligned}|\hat{\phi}_{N,R}(v_j)| &\geq C_N\left[1 - \sin^2\left(\frac{\pi}{5}\right)\left[\sin^2\left(\frac{2\pi}{5}\right)\right]^{N-1}\right]^{j/2} \\ &\geq C_N|v_j|^{-\alpha_N}\end{aligned}$$

with $\alpha_N = \frac{1}{\log 2}\left|\log\left(1 - \sin^2\left(\frac{\pi}{5}\right)\left(\sin^2\left(\frac{2\pi}{5}\right)\right)^{N-1}\right)\right|$. Clearly α_N is decreasing with N. Since $\alpha_1 \simeq 0.6115 < 1$, this ends the proof. ∎

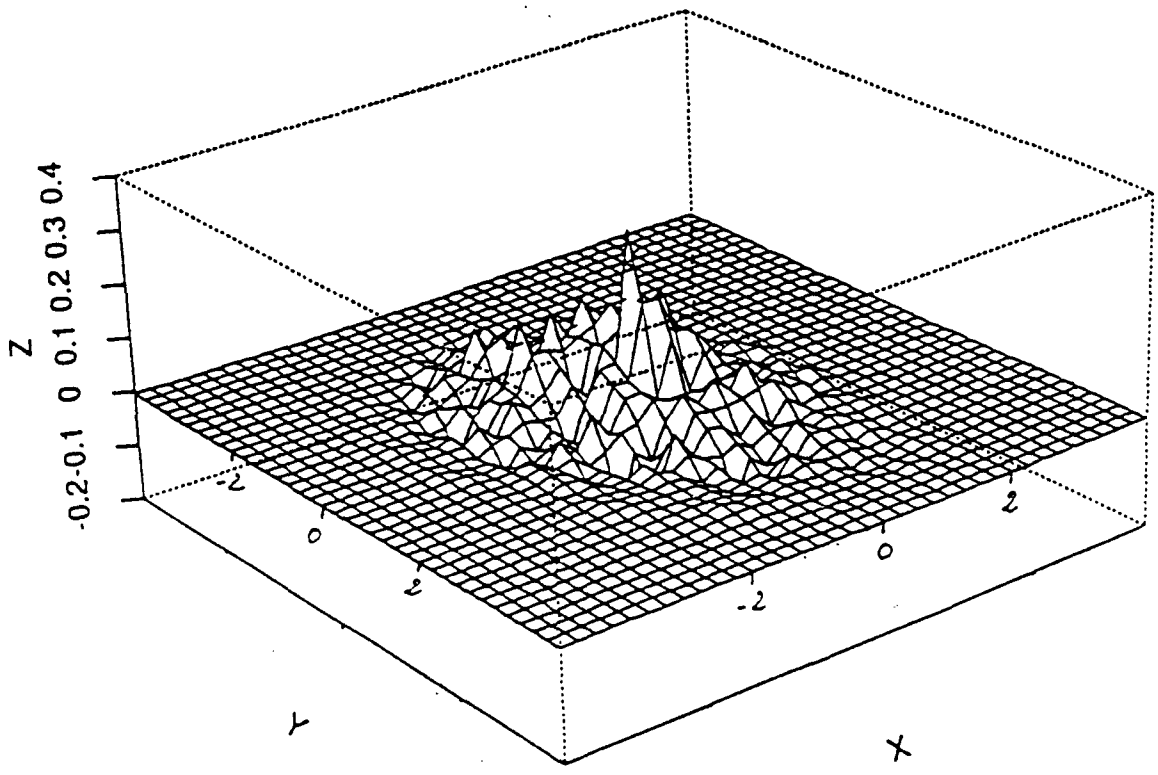


Figure 9
Approximation of the scaling function ϕ_{2R}

In fact, these wavelets do not even seem continuous although we have no mathematical proof of it. A simple look at the result of the cascade algorithm for the 4 taps filter (which

corresponds to a .55 Hölder continuous one dimensional wavelet) shows how chaotic the functions $\varphi_{R,N}$ can be (figure 9). The design of FIR filters leading to regular wavelet bases with R as the dilation matrix seems to be a difficult problem. Using a polyphase component approach M. Vetterli and J. Kovacevic ([KV], p. 32) have constructed a filter for which the result of the cascade looks continuous but no infinite family with arbitrarily high regularity has been designed so far.

The main difficulty which makes this design unpracticable is the absence of the Riesz lemma in more than one dimension and thus the impossibility to start by designing the square modulus of $M_0(\omega)$ in an appropriate way. Apart from this problem, the CQF filters (in particular the family (3.21) that we have introduced) cannot be symmetrical. We must keep in mind that one of the interests of the quincunx grid decimation is to have a more isotropic analysis; this is only achieved if the filter coefficients are themselves symmetrical around the horizontal, vertical and diagonal directions.

These two reasons encourage us to construct biorthogonal bases of wavelets from dual filters for which the Riesz lemma is not necessary and linear phase can be achieved.

V Biorthogonal Bases of Nonseparable Wavelets

Let us recall the family of dual filters introduced in III.3.b. It is based on the variable $z = \frac{1}{2}\left(\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right)$. We have chosen,

$$M_0^N(\omega) = (1-z)^N \tag{5.1}$$

and

$$\tilde{M}_0^{N,L}(\omega) = (1-z)^L \, P_{N+L}(z) \tag{5.2}$$

where L is still to be fixed.

A first remark is that the action of the dilation matrices R and S on the variable z are equivalent. This is due to the fact that z is invariant if we exchange ω_1 and ω_2 or if we change the sign of one of these variable. We shall thus consider a dilation matrix D which can be equal to R or S. To express its action on z we still need the two variables

$$y_1 = \sin^2\left(\frac{\omega_1}{2}\right) \quad \text{and} \quad y_2 = \sin^2\left(\frac{\omega_2}{2}\right) . \tag{5.3}$$

We then have,

$$\begin{aligned} z = \frac{1}{2}(y_1 + y_2) \quad &\xleftarrow{D} \quad z = \frac{1}{2}(y_1 + y_2 - 2y_1y_2) \\ &\xleftarrow{D} \quad z = \frac{1}{2}(4y_1(1-y_1) + 4y_2(1-y_2)) \;=\; \frac{1}{2}(y_1' + y_2') \\ &\xleftarrow{D} \quad z = \frac{1}{2}(y_1' + y_2' - 2y_1'y_2') \ldots \end{aligned}$$

We shall at first study the scaling function φ_1 associated to the filter $M_0^1(\omega) = 1 - z$, because it is the elementary building block for the family φ_N $(= (*)^N \varphi_1)$.

V.1 The quincunx Laplacian scheme

The coefficients of $M_0^1(\omega)$ are centered around the origin and have the following form:

$$\frac{1}{8} \times \begin{pmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{pmatrix} . \tag{5.4}$$

Note that this is the simplest symmetrical filter (with respect to the horizontal, vertical and diagonal directions) which satisfies the cancellation condition $M_0^1(\pi, \pi) = 0$. To estimate the decay of $\hat{\phi}_1(\omega)$ we could hope for a bidimensional formula equivalent to

$$\prod_{k=1}^{+\infty} \cos(2^{-k}\omega) = \frac{\sin \omega}{\omega} , \tag{5.5}$$

used in the one dimensional case. Note that (5.5) is based on the iteration of $\sin \omega = 2 \sin\left(\frac{\omega}{2}\right) \cos\left(\frac{\omega}{2}\right)$. Unfortunately, similar relations do not exist in the bidimensional case for the dilation matrix D. In particular the infinite product

$$\hat{\phi}_1(\omega) = \prod_{k=1}^{+\infty} M_0(D^{-j}\omega) \tag{5.6}$$

has no simple expression and one checks easily that, unlike (5.5), it does not have uniform decay at infinity. Indeed, let us consider the sets $\left\{\left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)\right\}$ and $\left\{\left(\frac{2\pi}{3}, \frac{2\pi}{3}\right), \left(\frac{2\pi}{3}, 0\right)\right\}$. These are two cyclic orbits of $\omega \mapsto D\omega$ *modulo* $2\pi\mathbb{Z}^2$ and modulo the exchange of coordinates and sign changes which do not affect the variable z. Consequently, if we define $v_j = D^j\left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)$ and $\mu_j = D^j\left(\frac{2\pi}{3}, \frac{2\pi}{3}\right)$, we have, when j goes to $+\infty$,

$$\hat{\phi}_1(v_j) \sim C \left[\frac{\cos^2\left(\frac{\pi}{5}\right) + \cos^2\left(\frac{2\pi}{5}\right)}{2}\right]^j \sim C|v_j|^{-\alpha_v} \tag{5.7}$$

and

$$\hat{\phi}_1(\mu_j) \sim C \left[\left(\frac{\cos^2\left(\frac{\pi}{3}\right) + 1}{2}\right) \cos^2\left(\frac{\pi}{3}\right)\right]^{j/2} \sim C|\mu_j|^{-\alpha_\mu} \tag{5.8}$$

with

$$\alpha_v = -\frac{2}{\log 2} \log\left[\frac{\cos^2\left(\frac{\pi}{5}\right) + \cos^2\left(\frac{2\pi}{5}\right)}{2}\right] \simeq 2.83 \tag{5.9}$$

and

$$\alpha_\mu = -\frac{1}{\log 2} \log\left[\left(\frac{\cos^2\left(\frac{\pi}{3}\right) + 1}{2}\right) \cos^2\left(\frac{\pi}{3}\right)\right] \simeq 2.68 \neq \alpha_v . \tag{5.10}$$

Still we would like to find a global exponent for the decay of $\hat{\phi}_1(\omega)$ at infinity. For this we shall introduce an "artificial" function which will play the same role as $\cos\omega$ in (5.5). We define.

$$C(\omega) = \frac{\sin^2\left(\frac{\omega_1+\omega_2}{2}\right) + \sin^2\left(\frac{\omega_1-\omega_2}{2}\right)}{2\left[\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right]}, \quad C(0) = 1 . \tag{5.11}$$

Contrarily to $M_0^1(\omega)$, $C(\omega)$ is not a trigonometric polynomial, but it is a bounded regular function which vanishes at the point (π, π) with the same order of cancellation as $M_0^1(\omega)$. Moreover, it satisfies by construction,

$$\prod_{k=1}^{+\infty} C(D^{-j}\omega) = \frac{2\left[\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right]}{\omega_1^2 + \omega_2^2} \leq C(1+|\omega|)^{-2} . \tag{5.12}$$

The decay of this infinite product is now uniform and, for this reason, $C(\omega)$ will play an important role in the construction of our dual bases. For the moment, by comparing $C(\omega)$ and $M_0^1(\omega)$, we obtain the following result,

Proposition 5.1 *The decay of $\hat{\phi}_1(\omega)$ at infinity is controlled by*

$$|\hat{\phi}_1(\omega)| \leq C(1+|\omega|)^{-2} . \tag{5.13}$$

Furthermore, this exponent is globally optimal, i.e. there exists a sequence $\{\omega_j\}_{j>0}$ such that $\lim_{j\to+\infty}|\omega_j| = +\infty$ and $|\hat{\phi}_1(\omega_j)| \sim C|\omega_j|^{-2}$.

Proof:

Using the variables $y_1 = \sin^2\frac{\omega_1}{2}$ and $y_2 = \sin^2\frac{\omega_2}{2}$ we can rewrite $C(\omega)$.

$$C(\omega) = \frac{y_1 + y_2 - 2y_1y_2}{y_1 + y_2} = \frac{(1-y_1)y_2 + (1-y_2)y_1}{y_1 + y_2} . \tag{5.14}$$

We thus have

$$\begin{aligned} C(\omega) - M_0^1(\omega) &= \frac{(1-y_1)y_2 + (1-y_2)y_1}{y_1 + y_2} - \frac{(1-y_1) + (1-y_2)}{2} \\ &= \frac{(1-y_1)(y_2-y_1) + (1-y_2)(y_1-y_2)}{2(y_1+y_2)} \\ &= \frac{(y_1-y_2)^2}{2(y_1+y_2)} \geq 0 . \end{aligned}$$

Thus $M_0^1(\omega) \leq C(\omega)$ and by (5.12) $|\hat{\phi}_1(\omega)| \leq C(1+|\omega|)^{-2}$. To prove that this exponent is optimal we consider a small vector $\rho \neq 0$ in $\mathbb{R}^2$ and define

$$\omega_j = D^j(\pi, \pi) + \rho , \tag{5.15}$$

so that

$$\hat{\phi}_1(\omega_j) = \prod_{k=1}^{+\infty} M_0^1\left(D^{j-k}(\pi,\pi) + D^{-k}\rho\right) . \tag{5.16}$$

Let us divide this product in three parts:

$$\begin{aligned} \hat{\phi}_1(\omega_j) &= \left[\hat{\phi}_1\left((\pi,\pi) + D^{-j}\rho\right)\right]\left[\prod_{k=1}^{j-1} M_0^1\left(D^{j-k}(\pi,\pi) + D^{-k}\rho\right)\right] \\ &\times \left[M_0^1\left((\pi,\pi) + D^{-j}\rho\right)\right] \\ &= A(j)\,B(j)\,C(j) . \end{aligned} \tag{5.17}$$

One checks easily that $\hat{\phi}_1(\pi, \pi) \neq 0$ and thus, for j large enough or choosing ρ small enough, we have $0 < C_1 \leq A(j) \leq 1$. It is also clear that for $1 \leq k \leq j-1$, $M_0^1\left(D^{j-k}(\pi, \pi)\right) = 1$ and that for $\ell \geq 1$, $M_0^1(D^\ell(\pi, \pi)+\sigma) \geq 1 - C\|\sigma\|$ for σ small enough, with $C > 0$. Consequently, if ρ has been chosen small enough, $1 \geq B(j) \geq \prod_{\ell=1}^{\infty} [1 - C2^{-\ell}\|\rho\|] \geq C_2 > 0$. Finally since (π, π) is a second order zero of $M_0(\omega)$, the third factor satisfies

$$(5.18) \qquad 2^{-j}C_3\|\rho\|^2 = C_3\|D^{-j}\rho\|^2 \leq C(j) \leq C_4\|D^{-j}\rho\|^2 = 2^{-j}C_4\|\rho\|^2 .$$

This shows that $\hat{\phi}_1(\omega_j)$ behaves like $2^{-j} \sim |\omega_j|^{-2}$ when j goes to $+\infty$ and the proposition is proved. ▪

Note that from the decay of $\hat{\phi}_1(\omega)$ we cannot even conclude that it belongs to $L^1(\mathbb{R}^2)$ or that $\phi_1(\cdot)$ is a continuous function. Yet both are true; we are going to prove this by the Littlewood-Paley method exposed in II.4.a. The filter $M_0^1(\omega)$ and the scaling function $\hat{\phi}_1(\omega)$ are particularly well adapted for this approach since they are positive so that the regularity estimation is optimal (because $\|\Delta_j(\phi_1)\|_{L^\infty} \sim \|\widehat{\Delta_j(\phi_1)}\|_{L^1}$; see §II.4.a).

Proposition 5.2 *The optimal global Hölder exponent for $\phi_1(x)$ is* $\alpha = \frac{2}{\log 2}\log\left(\frac{1+\sqrt{5}}{4}\right) \simeq .61$.

Proof:

We consider the transition operator defined by

$$(5.19) \qquad TF(D\omega) = M_0^1(\omega)\, F(\omega) + M_0^1(\omega + (\pi, \pi))\, F(\omega + (\pi, \pi)) .$$

As in the one dimensional case T can be studied in a finite dimensional space but this subspace cannot be defined as simply as $E(N_1, N_2)$ in (2.61). One way of finding an invariant subspace is to apply T to the constant 1 and then iterate it on the characters $e^{i(k_1\omega_1 + k_2\omega_2)}$ which are obtained until a stable set is attained. With M_0^1 corresponding to (5.4), this subspace is trivial, since $T1 = 1$. Lemma 2.5 then guarantees the integrability of $\hat{\phi}_1$, hence the continuity of ϕ_1. To estimate the Hölder exponent of ϕ_1 we need a larger subspace, which we obtain by iterating T on 1 and on $\cos\omega_1 + \cos\omega_2$. The size of the matrix representing the action of T on this subspace can be seriously reduced by exploiting the symmetries, i.e. the invariance under $\omega_1 \mapsto -\omega_1$, $\omega_2 \mapsto -\omega_2$ and $\omega_1 \mapsto \omega_2$.

Using the subspace E generated by the basis

$$(5.20) \qquad \epsilon_1 = 1, \quad \epsilon_2 = \cos\omega_1 + \cos\omega_2, \quad \epsilon_3 = \cos(\omega_1 + \omega_2) + \cos(\omega_2 - \omega_1)$$

we obtain the following matrix

$$(5.21) \qquad T = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$$

which has the eigenvalues $\left\{1, \frac{1+\sqrt{5}}{4}, \frac{1-\sqrt{5}}{4}\right\}$. The two last eigenvalues correspond to the subspace $E_0 \subset E$ defined by

$$E_0 = \{F(\omega) \in E,\ F(0) = 0\} \ . \tag{5.22}$$

Similarly to the one dimensional case, we iterate T on the positive function $\epsilon_1 - \frac{1}{2}\epsilon_2$ which is clearly in E_0 and this leads us to

$$\left\|\Delta_{j/2}(\phi_1)\right\|_{L^\infty} \sim \left\|\hat{\Delta}_{j/2}(\phi_1)\right\|_{L^1} \sim C\left(\frac{1+\sqrt{5}}{4}\right)^j \ , \tag{5.23}$$

where $\Delta_{j/2}(\phi_1)$ is the Littlewood-Paley block corresponding to the region $D^j([-\pi,\pi]^2)/D^{j-1}([-\pi,\pi]^2)$, situated at a distance $2^{j/2}$ of the origin. Consequently, if we define

$$\alpha = -\frac{2}{\log 2}\log\left(\frac{1+\sqrt{5}}{4}\right) \simeq 0.61 \tag{5.24}$$

it follows from (5.23) that

$$(1+|\omega|)^\alpha\ \hat{\phi}_1(\omega) \in L^1(\mathbb{R}^2) \ \text{ and } \ \phi_1(x) \in C^\alpha(\mathbb{R}^2) \ . \tag{5.25}$$

Consequently ϕ_1 is Hölder continuous with regularity 0.61. ▪

This property appears in the graph of ϕ_1 on figure 10 (obtained by the cascade algorithm) which presents a smooth aspect with several pointwise cusps. Note that this regularity is not sufficient to derive a better decay of $\hat{\phi}_1(\omega)$ than $|\omega|^{-0.61}$; Propositions 5.1 and 5.2 are thus complementary.

Remarks:

- Note that, since we have

$$M_0^1(\omega) + M_0^1(\omega+(\pi,\pi)) = 1 \ , \tag{5.26}$$

we can derive the L^1 convergence of the truncated products $\hat{\phi}_{1n} = \prod_{j=1} M_0(D^{-j}\omega)\ \chi_{D^n([-\pi,\pi]^2)}(\omega)$ with the same method as in the orthonormal case for the L^2 convergence (Theorem 2.1). This leads us to a Poisson summation formula

$$\sum_{k\in\mathbb{Z}^2} \hat{\phi}_1(\omega+2k\pi) = 1 \tag{5.27}$$

which is equivalent to

$$\phi_1(n_1,n_2) = 1 \ \text{ if } \ n_1 = n_2 = 0, \ \ 0 \ \text{if } (n_1,n_2) \in \mathbb{Z}^2/\{0\} \ . \tag{5.28}$$

This interpolating property of ϕ_1 has been noticed in approximation theory by Deslaurier and Dubuc [DD]. It explains the four cusps surrounding the center at the points $(0,1)$, $(1,0)$, $(0,-1)$ and $(-1,0)$ which are visible on figure 10. However, a sharper analysis shows that the isolated points where $\phi_1(x) = 0$ are an infinite family.

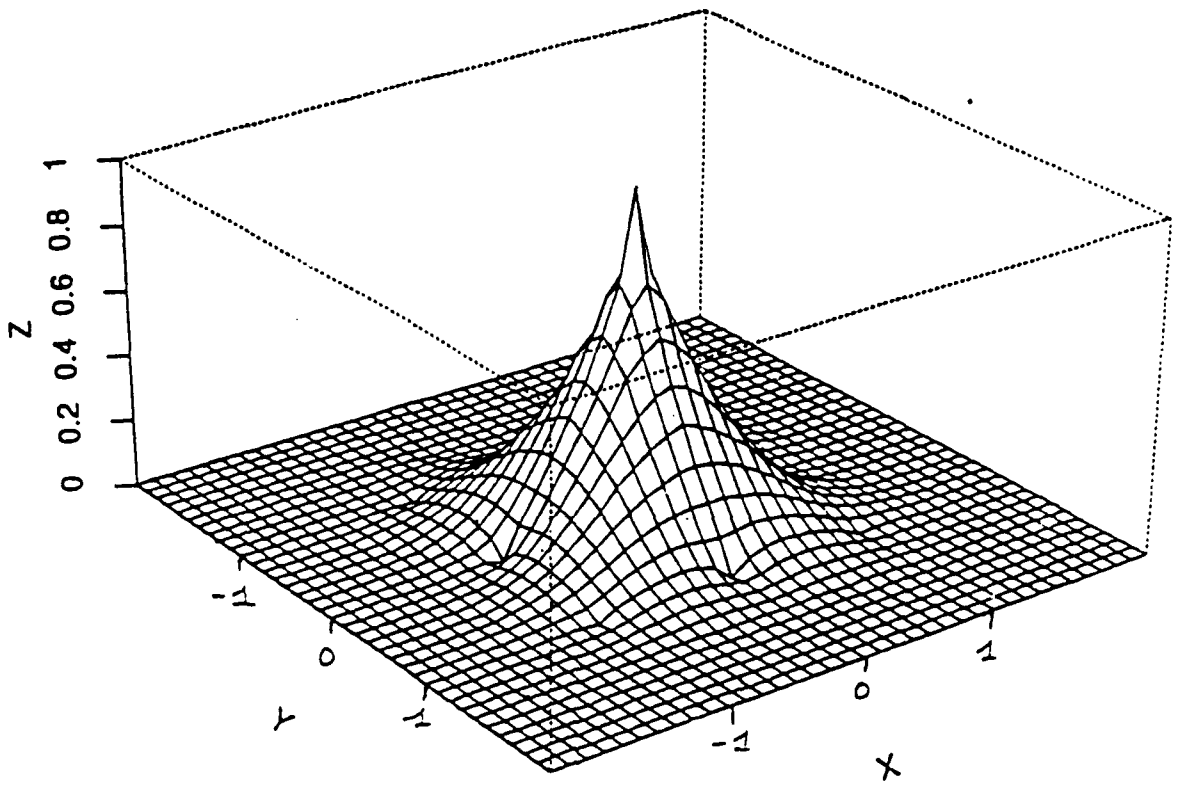


Figure 10
The scaling function $\phi_1(x)$

- As mentioned in section III.3.b, the variable $z = \frac{1}{2}(y_1 + y_2)$ can be replaced by, more generally, $z_\lambda = \lambda y_1 + (1-\lambda)y_2$ with $\lambda \in [0,1]$; $M_0^1(\omega) = 1 - z_\lambda$ is still positive. Let us now distinguish the dilation matrices R and S. Then, a similar analysis in the case of $D = R$ leads to a 5×5 matrix in the basis

$$(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5) = (1, \cos\omega_1, \cos\omega_2, \cos(\omega_1 + \omega_2), \cos(\omega_1 - \omega_2))$$

$$T_\lambda = \frac{1}{2}\begin{pmatrix} 2 & \lambda & 1-\lambda & 0 & 0 \\ 0 & 1-\lambda & \lambda & 0 & 2 \\ 0 & 1-\lambda & \lambda & 2 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 1-\lambda & 0 & 0 \end{pmatrix} \tag{5.29}$$

and numerical computations show that the "isotropic value" $\lambda = \frac{1}{2}$ gives the highest index of regularity. The lowest index of regularity is attained for $\lambda = 0$ or 1. Note that $\lambda = 1$ corresponds to the convolution product $g(x) = \chi_\Delta * \chi_\Delta$ where Δ is the twin dragon introduced in IV.3.a. The Hölder exponent is then $\alpha \simeq 0.47$.

- To estimate the decay of $\hat{g}(\omega)$ $(= (\hat{\chi}_\Delta(\omega))^2)$, one can again use the function $C(\omega)$ of Proposition 5.1, in a slightly different way. Remark that, if we define

$G(\omega) = 1 - z_1 = 1 - y_1$. then

$$C(\omega) - G(\omega) = \frac{(1-y_1)y_2 + (1-y_2)y_1}{y_1 + y_2} - (1 - y_1)$$

$$= \frac{y_1(y_1 - y_2)}{y_1 + y_2} \geq 0 \quad \text{if } y_1 \geq y_2$$

and

$$2C(\omega) - G(\omega) = \frac{2[(1-y_1)y_2 + (1-y_2)y_1]}{y_1 + y_2} - (1 - y_1)$$

$$= \frac{(1-y_1)(y_2 - y_1) + 2y_1(1-y_2)}{y_1 + y_2} \geq 0 \quad \text{if } y_2 \geq y_1 .$$

On the other hand

(5.30) $$|\hat{g}(\omega)| = \prod_{k=1}^{\infty} G(R^{-k}\omega) ;$$

to majorate $|\hat{g}(\omega)|$ for $2^{\frac{j}{2}} \leq |\omega| \leq 2^{\frac{j+1}{2}}$ we only need to majorate the j first factors in (5.30). Since R rotates of $\frac{\pi}{4}$, half of the factors can be majorated by $C(\omega)$ and the others by $2C(\omega)$. This leads to

(5.31) $$|\hat{g}(\omega)| \leq C 2^{\frac{\log(1+|\omega|)}{\log 2}} \prod_{1 \leq k \leq \frac{2\log(1+|\omega|)}{\log 2}} C(R^{-k}\omega)$$

and thus

(5.32) $$\hat{g}(\omega) \leq C(1 + |\omega|)^{-1} .$$

It is easy to check (in a similar way as for $\hat{\phi}_1(\omega)$) that this estimate is optimal. An immediate consequence is that the Fourier transform of the twin dragon characteristic *function* χ_Δ *satisfies*

(5.33) $$\hat{\chi}_\Delta(\omega) \leq C(1 + |\omega|)^{-1/2}$$

which was not obvious since we did not have a formula similar to (5.5) for $\hat{\chi}_\Delta$.

We now return to the construction of our biorthogonal bases and attack the problem of obtaining isotropic wavelet bases with arbitrarily high regularity.

V.2 Biorthogonal wavelet bases with arbitrarily high regularity

We now consider the whole family of filter $\left\{M_0^N(\omega),\ \tilde{M}_0^{N,L}(\omega)\right\}_{N,L>0}$ defined by (5.1) and (5.2).

A first remark is that the regularity of the functions ϕ_N increases linearly with N. More precisely, since

(5.34) $$\phi_N(x) = (*)^N \phi_1(x) ,$$

we can use the characterization of the optimal decay exponent for $\hat{\phi}_1(\omega)$ established in Proposition (5.1) to estimate the regularity index $\alpha(N)$ of $\phi_N(x)$. This leads to

$$2N - 2 \leq \alpha(N) \leq 2N \tag{5.35}$$

and thus

$$\lim_{N \to +\infty} \frac{\alpha(N)}{N} = 2 . \tag{5.36}$$

The estimate (5.36) is of course more interesting for large values of N than for small values where the error is comparable with the regularity.

For $N = 1$, we have seen that $\alpha \simeq 0.61$.

For $N = 2$, the Littlewood-Paley approach is still reasonable; using the symmetries reduces the size of the matrix to 9×9. Analyzing the eigenvalues, one find that ϕ_2 is in C^α with $\alpha \simeq 2.93$. The function $\phi_2 = \phi_1 * \phi_1$ looks very smooth indeed on Figure 13.

For $N = 3$, the matrix becomes too large to tackle by hand. In all cases the regularity of the wavelet $\psi_{N,L}$ will of course be the same as that of ϕ_N. The problem is now to find the appropriate dual function for the analysis. More precisely we want to design the filter

$$\tilde{M}_0^{N,L}(\omega) = (1 - z)^L P_{N+L}(z) \tag{5.37}$$

by choosing the number L in such way that the hypothesis of Theorem 2.2 (in its bidimensional generalization) are satisfied, i.e. that we have at least

$$\left|\hat{\tilde{\phi}}_{N,L}(\omega)\right| \leq C(1 + |\omega|)^{-1-\epsilon}, \quad \epsilon > 0 . \tag{5.38}$$

To show that such a choice is possible for any value of N (i.e. for an arbitrarily regular synthesis function), we need an asymptotical result of the same nature as Theorem 2.4. We want to be sure that the regularizing action of the factor $(1 - z)^L$ can compensate the inverse effect of P_{N+L} if L is large enough.

Using a similar approach, we consider the simplest fixed point of $\omega \longmapsto D\omega$ modulo $2\pi\mathbb{Z}^2$, and *modulo sign changes* and the exchange of ω_1 and ω_2. This fixed point is $\omega_0 = \left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)$ which corresponds to $z_0 = z(\omega_0) = \frac{5}{8}$.

We now decompose $\tilde{M}_0^{N,L}$ into three factors, by introducing the function $C(\omega)$ defined by (5.11):

$$\tilde{M}_0^{N,L}(\omega) = [C(\omega)]^L [Q(\omega)]^L P_{N+L}(z) = [C(\omega)]^L B_{N,L}(\omega) \tag{5.39}$$

with

$$Q(\omega) = \frac{M_0^1(\omega)}{C(\omega)} = \frac{(y_1 + y_2)(2 - y_1 - y_2)}{2(y_1 + y_2 - 2y_1 y_2)} .$$

We already know from section II.3.b that

$$P_N(z) \leq (4z)^{N-1} \text{ if } z \geq \frac{1}{2} . \tag{5.40}$$

From the Bezout relation (3.30), we also have

$$P_N(z) \leq \left(\frac{1}{1 - z}\right)^N . \tag{5.41}$$

Consequently, we can roughly majorate $P_N(z)$ by

$$P_N(z) \leq \left[\min\left(\frac{1}{1-z},\ \max(4z,2)\right)\right]^N \quad \text{if } z \in [0,1]\,. \tag{5.42}$$

Defining $H(\omega) = \min\left(\frac{1}{1-z},\ \max(4z,2)\right)$ and $G(\omega) = H(\omega)Q(\omega)$, (5.39) leads us to

$$\tilde{M}_0^{N,L}(\omega) \leq [C(\omega)]^L\,[G(\omega)]^L\,[H(\omega)]^N\,. \tag{5.43}$$

We are now facing a similar situation as in Theorem 2.4 where we had shown that the function $g(y) = \max(2,4y) = h(\omega)$ satisfied

$$\begin{cases} h(\omega) = g(y) \leq g\left(\frac{3}{4}\right) & \text{if } y \leq \frac{3}{4} \\ \\ h(\omega)h(2\omega) = g(y)g(4y(1-y)) \leq \left[g\left(\frac{3}{4}\right)\right]^2 & \text{if } \frac{3}{4} \leq y \leq 1 \end{cases} \tag{5.44}$$

In the present case, although we do not dispose of any simple mathematical proof, numerical evidence shows that we have

$$\begin{cases} G(\omega)G(D\omega) \leq [G(\omega_0)]^2 & \text{or if not,} \\ \\ G(\omega)G(D\omega)G(D^2\omega) \leq [G(\omega_0)]^3 \end{cases} \tag{5.45}$$

and similarly

$$\begin{cases} H(\omega)H(D\omega) \leq [H(\omega_0)]^2 & \text{or if not,} \\ \\ H(\omega)H(D\omega)H(D^2\omega) \leq [H(\omega_0)]^3 \end{cases} \tag{5.46}$$

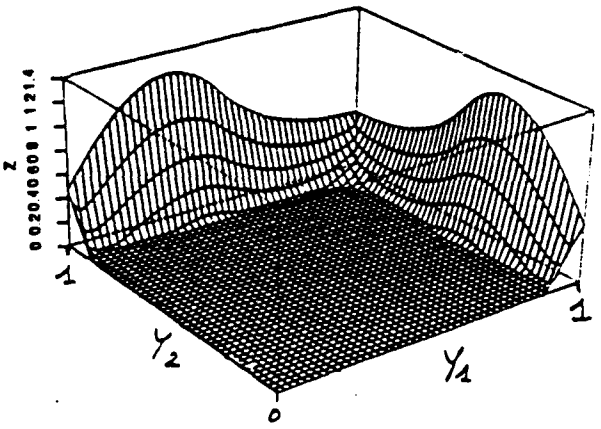
These two statements are illustrated respectively in Figures 11 and 12. On a) and b) of each of these figures we have plotted the functions $\max\left(F(\omega)F(D\omega), [F(\omega_0)]^2\right) - [F(\omega_0)]^2$ and $\max(F(\omega)F(D\omega), F(D^2\omega);\ [F(\omega_0)]^3) - [F(\omega_0)]^3$ for $F = G$ and H (the coordinates are $(y_1, y_2) \in [0,1]^2$). On c) the support of a) and b) are shown to be disjoint regions in $[0,1]^2$.

We now estimate $\hat{\tilde{\phi}}_{N,L}(\omega)$. From(5.39) and (5.43) we get

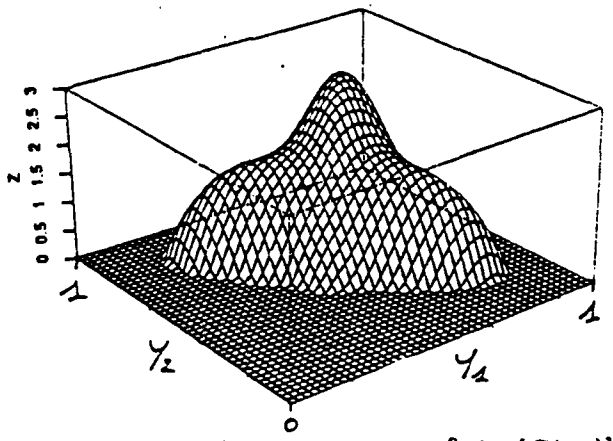
$$\begin{aligned} \hat{\tilde{\phi}}_{N,L}(\omega) &= \prod_{k=1}^{+\infty} \left[C(D^{-k}\omega)\right]^L B_{N,L}(D^{-k}\omega) \\ &\leq C(1+|\omega|)^{-2L} \prod_{1 \leq k \leq \frac{2\log(1+|\omega|)}{\log 2}} B_{N,L}(D^{-k}\omega) \\ &\leq C(1+|\omega|)^{-2L} \left[\prod_{1 \leq k \leq \frac{2\log(1+|\omega|)}{\log 2}} G(D^{-k}\omega)\right]^L \left[\prod_{1 \leq k \leq \frac{2\log(1+|\omega|)}{\log 2}} H(D^{-k}\omega)\right]^N . \end{aligned}$$

Using (5.45) and (5.46) to divide these products in groups of two of three factors which satisfy one of the inequalities, this leads to

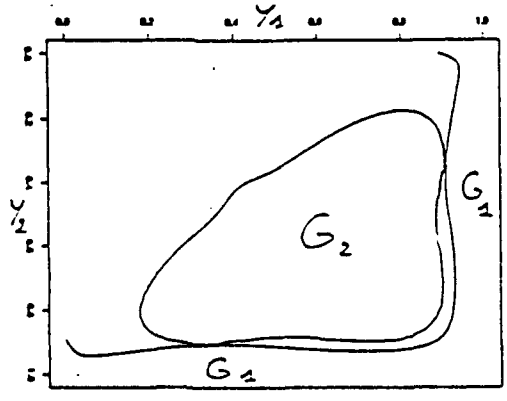
$$\hat{\tilde{\phi}}_{N,L}(\omega) \leq C(1+|\omega|)^{-2L + \frac{2L\log(G(\omega_0))}{\log 2} + \frac{2N\log(H(\omega_0))}{\log 2}} \tag{5.47}$$



a) Graph of $G_1(y_1, y_2) = \max\,(G(\omega)G(D\omega);\ [G(\omega_0)]^2)\ -\ [G(\omega_0)]^2$

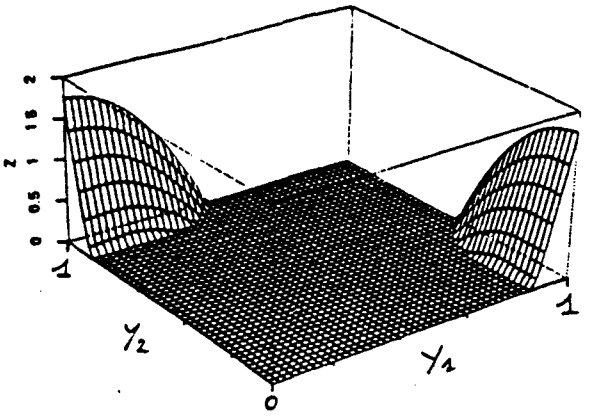


b) Graph of $G_2(y_1, y_2) = \max\,(G(\omega)G(D\omega)G(D^2\omega),\ [G(\omega_0)]^3)\ -\ [G(\omega_0)]^3$

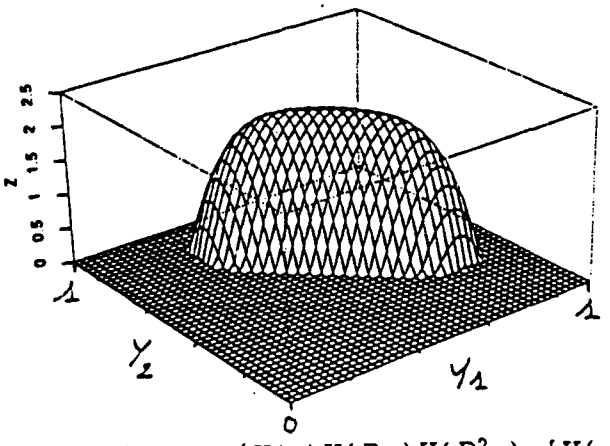


c) Compared supports of $G_1(y_1, y_2)$ and $G_2(y_1, y_2)$

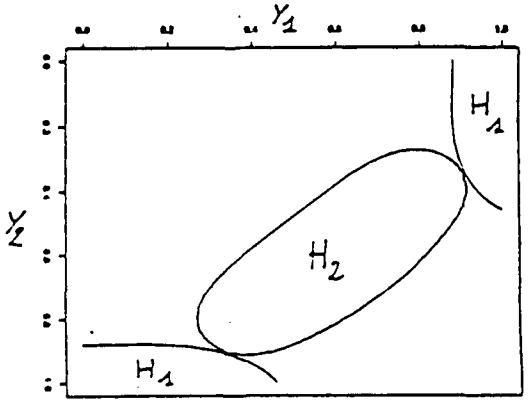
Figure 11
Graphic proof of (5.45)



a) Graph of $H_1(y_1, y_2) = \max\,(H(\omega)H(D\omega),\ [H(\omega_0)]^2)\ -\ [H(\omega_0)]^2$



b) Graph of $H_2(y_1, y_2) = \max\,(H(\omega)H(D\omega)H(D^2\omega),\ [H(\omega_0)]^3)\ -\ [H(\omega_0)]^3$



c) Compared supports of $H_1(y_1, y_2)$ and $H_2(y_1, y_2)$

Figure 12
Graphic proof of (5.46)

or

$$\hat{\tilde{\phi}}_{N,L}(\omega) \leq C(1+|\omega|)^{2L(\alpha-1)+2N\beta} \tag{5.47'}$$

with

$$\alpha = \frac{\log(G(\omega_0))}{\log 2} \simeq 0.907 \quad \text{and} \quad \beta = \frac{\log(H(\omega_0))}{\log 2} \simeq 1.322 .$$

Fortunately $\alpha < 1$. This means $\tilde{\phi}_{L,N}(x)$ can be made arbitrarily regular by choosing L large enough. In particular, (5.38) will be satisfied if we have

$$2L(\alpha-1) + 2N\beta < -1 . \tag{5.48}$$

The smallest L such that M_0^N and $\tilde{M}_0^{N,L}$ generate unconditional multiscale bases is therefore given asymptotically by

$$L(N) \simeq \frac{\beta N}{1-\alpha} = \frac{\log 5 - \log 2}{\log 16 - \log 15} N \simeq 14.2\, N . \tag{5.49}$$

This asymptotical estimate is moreover optimal. Indeed define $\omega_j = D^j\omega_0 = D^j\left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)$. Because of the fixed point property of ω_0, we clearly have

$$\hat{\tilde{\phi}}^{N,L}(\omega_j) \sim C\left[\tilde{M}_0^{N,L}(\omega_0)\right]^j \sim C|\omega_j|^{-\gamma} \tag{5.50}$$

with

$$\gamma = \frac{-2\,\log\left(\tilde{M}_0^{N,L}(\omega_0)\right)}{\log 2} . \tag{5.51}$$

From the definition (5.2) of $\tilde{M}_0^{N,L}$, we get

$$\begin{aligned} \tilde{M}_0^{N,L}(\omega_0) &= \left(\frac{3}{8}\right)^L P_{N+L}\left(\frac{5}{8}\right) \\ &\geq \left(\frac{3}{8}\right)^L \binom{2(N+L-1)}{N+L-1} \left(\frac{5}{8}\right)^{N+L-1} \\ &\geq C\left(\frac{3}{8}\right)^L \left(\frac{5}{2}\right)^{N+L} = C\left(\frac{15}{16}\right)^L \left(\frac{5}{2}\right)^N \end{aligned}$$

and thus

$$\begin{aligned} \gamma &\leq C + 2L\,\frac{\log\left(\frac{16}{15}\right)}{\log 2} - 2N\,\frac{\log\left(\frac{5}{2}\right)}{\log 2} \\ &= C + 2L(1-\alpha) - 2N\beta . \end{aligned}$$

It follows that the estimate (5.49), if true, is certainly optimal. While we expect (5.45), (5.46), hence (5.49), to be true, we have unfortunately no rigorous proof. However, we *can* prove inequalities which are slightly less strong than (5.45), (5.46), leading to a non-optimal

but rigorous estimate for $L(N)$. More precisely, we can prove that $\Omega = [-\pi, \pi]^2$ can be split up as $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$, with

$$(5.52) \qquad \begin{cases} G(\omega) \le \xi & \omega \in \Omega_1 \\ G(\omega)\, G(D\omega) \le \xi^2 & \omega \in \Omega_2 \\ G(\omega)\, G(D\omega)\, G(D^2\omega) \le \xi^3 & \omega \in \Omega_3 \end{cases}$$

with $\xi/2 \simeq .9588 < 1$, resulting in (5.47') with

$$\alpha = \frac{\log \xi}{\log 2} \simeq .93982 .$$

If we use the crude estimate $H(\omega) \le 4$ for all $\omega \in [-\pi, \pi]^2$, corresponding to $\beta = 2$, then this leads to

$$L(N) \simeq \frac{\beta}{1-\alpha} N \simeq 32.959\, N ;$$

this factor is about twice as large as in (5.49). The detailed proof of this estimate is in Appendix C.

All these results can be summarized in the following theorem:

Theorem 5.3 *The family of dual filters* $\left\{ M_0^N(\omega),\ \tilde{M}_0^{N,L}(\omega) \right\}_{N,L>0}$ *generates biorthogonal bases of compactly supported wavelets with arbitrarily high regularity. For large values of* N, *the Hölder exponent of* $\phi_N(x)$ *is equivalent to* $2N$ *and the minimal choice for* L *is asymptotically proportional to* N,

$$(5.53) \qquad L(N) \simeq \mathcal{K} N ,$$

with $14.215 \le \mathcal{K} \le 32.959$.

Here the upper bound on $\mathcal{K}$ is not tight, and we expect $\mathcal{K} = 14.215$ to hold, as indicated above.

Remark:

By taking L larger than $L(N)$, $\tilde{\phi}^{N,L}$ can also be made arbitrarily regular. However, in many applications such as coding, approximation, data storage and compression, we do not really care about the regularity of the analyzing functions $\tilde{\psi}$ and $\tilde{\phi}$; only the synthesis function ψ and ϕ have to be smooth since this property is important for the cascade-reconstruction algorithm. This justifies the choice of the minimal value $L(N)$ such that the families $\left\{ 2^{j/2}\, \psi_{N,L}(D^j x - k) \right\}_{j \in \mathbb{Z},\ k \in \mathbb{Z}^2}$ and $\left\{ 2^{j/2}\, \tilde{\psi}_{N,L}(D^j x - k) \right\}_{j \in \mathbb{Z},\ k \in \mathbb{Z}^2}$ are unconditional dual bases of $L^2(\mathbb{R})$. Recall that the existence of frame bounds is essential for the stability of the subband coding scheme.

We end this section by taking a closer look at the size of these dual filters.

V.3 Size and optimal implementation of the dual filters

The asymptotical ratio $\frac{L(N)}{N} \simeq 14.2$ is big in the sense that the filter $\tilde{M}_0^{N,L(N)}$ may have a very large number of taps. More precisely, a polynomial $P(z)$ of degree p corresponds to a filter with $p^2 + (p+1)^2$ nonzero coefficients. For example, if $N = 3$,

$$\tilde{M}_0^{N,L(N)}(\omega) = (1 - z)^{L(N)}\, P_{N+L(N)}(z)$$

is according to (5.49) a polynomial of degree $p = N + 2L(N) \simeq 87$ in z. Consequently it is the transfer function of a filter with approximately 1350 taps!

It seems thus that the dual filter is, even for small values of N, much too large for a realistic implementation. This is not quite true for several reasons.

First, one can factorize the polynomial $P_{N+L(N)}(z)$ and express $\tilde{M}_0^{N,L}$ as a product of p monomials in z. By applying successively these monomial filters instead of using directly their product, the number of multiplications per sample in the filtering process is reduced from order p^2 to p. Note that this complexity reduction associated with the factorization is due to the multidimensional situation and does not occur in the $1D$ case.

Second, the filter corresponding to the variable z, i.e. the laplacian discrete scheme, has coefficients $c_{0,0} = \frac{1}{2}$ and $c_{1,0} = c_{-1,0} = c_{0,1} = c_{0,-1} = -\frac{1}{8}$. It can thus be implemented by using binary shifts instead of multiplications. This is very important since a binary shift is usually performed 10 times faster than an addition and 100 times faster than a multiplication in most processors. This shows that only the additions count here. If t is the time for one multiplication, each monomial filter will generate one sample in approximately $\frac{3t}{5}$ and the same operation will take $\frac{3pt}{5}$ for the whole filter. For $N = 3$ and $p = 87$, this corresponds to the complexity of a 52 tap filter which is much more reasonable than the first estimation.

Finally, for small values of N, it is clear that the asymptotical estimate (5.49) of $L(N)$ is far from sharp, just as, in $1D$, the asymptotical estimate on regularity of section II.3.b was ill-suited to small filters.

A better estimate for $L(N)$ can be found by checking that the optimal decay exponent for $\hat{\tilde{\phi}}(\omega)$ is exactly determined by the value of $\tilde{M}_0^{N,L}$ at $\omega_0 = \left(\frac{2\pi}{5}, \frac{4\pi}{5}\right)$. More precisely recall that we have

$$\tilde{M}_0^{N,L}(\omega) \;=\; [C(\omega)]^L\, B_{N,L}(\omega)\,.$$

For the small values of N and L considered below, one can check by the same graphical arguments that the inequalities (5.45) or (5.46) are also satisfied by $B_{N,L}(\omega)$, i.e.

$$(5.54) \qquad \begin{cases} B_{N,L}(\omega)\, B_{N,L}(D\omega) \le [B_{N,L}(\omega_0)]^2 \qquad \text{or if not,} \\ \\ B_{N,L}(\omega)\, B_{N,L}(D\omega)\, B_{N,L}(D^2\omega) \le [B_{N,L}(\omega_0)]^3\,. \end{cases}$$

In order for (5.38) to be satisfied, we therefore only need

$$(5.55) \qquad \tilde{M}_0^{N,L}(\omega_0) \le \frac{\sqrt{2}}{2}$$

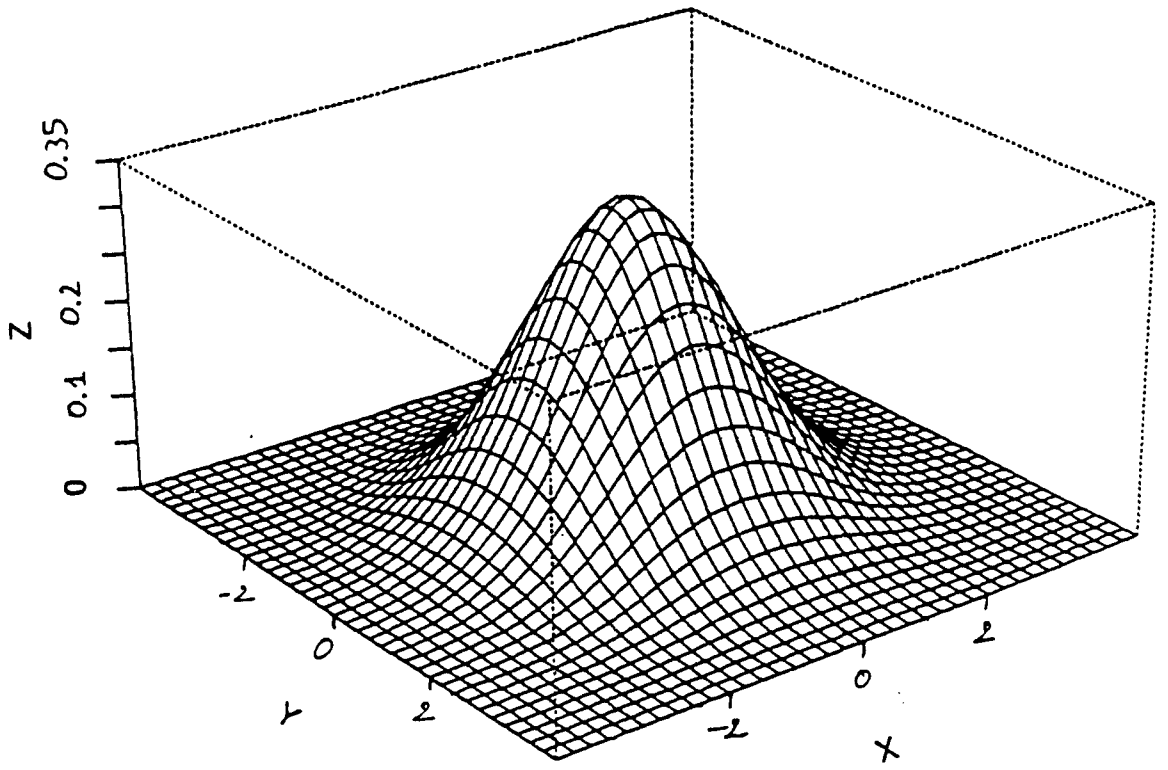
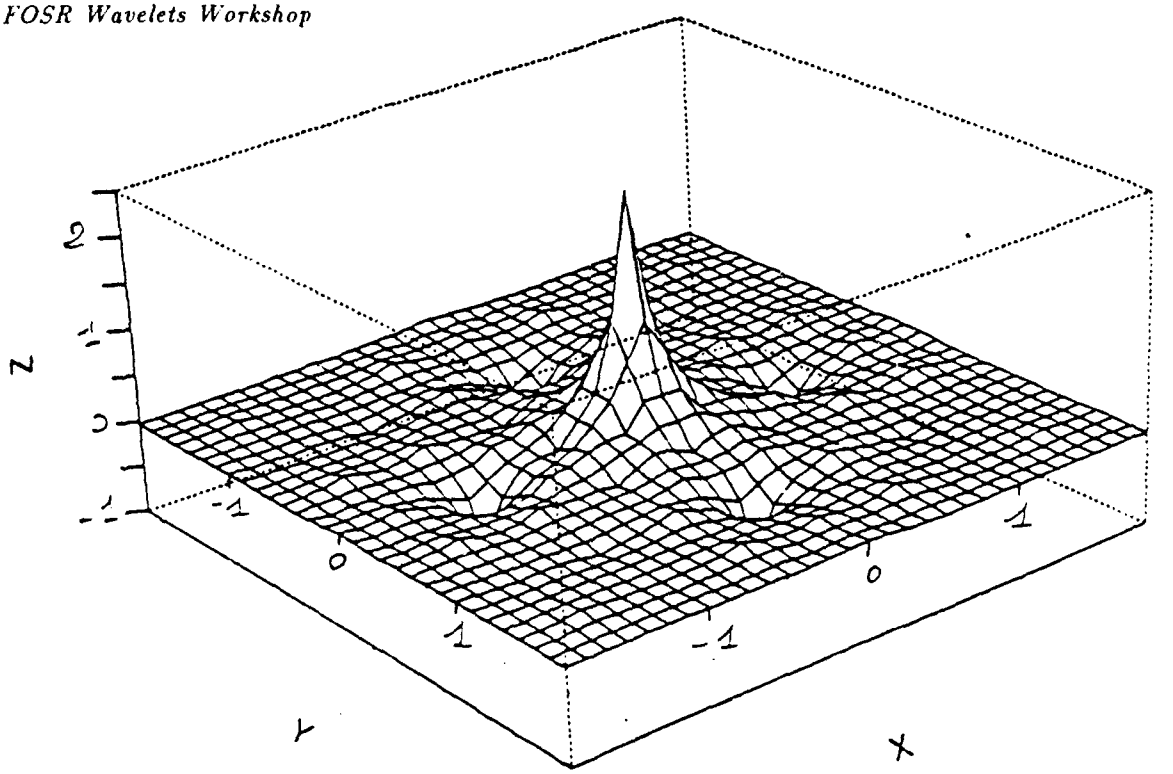
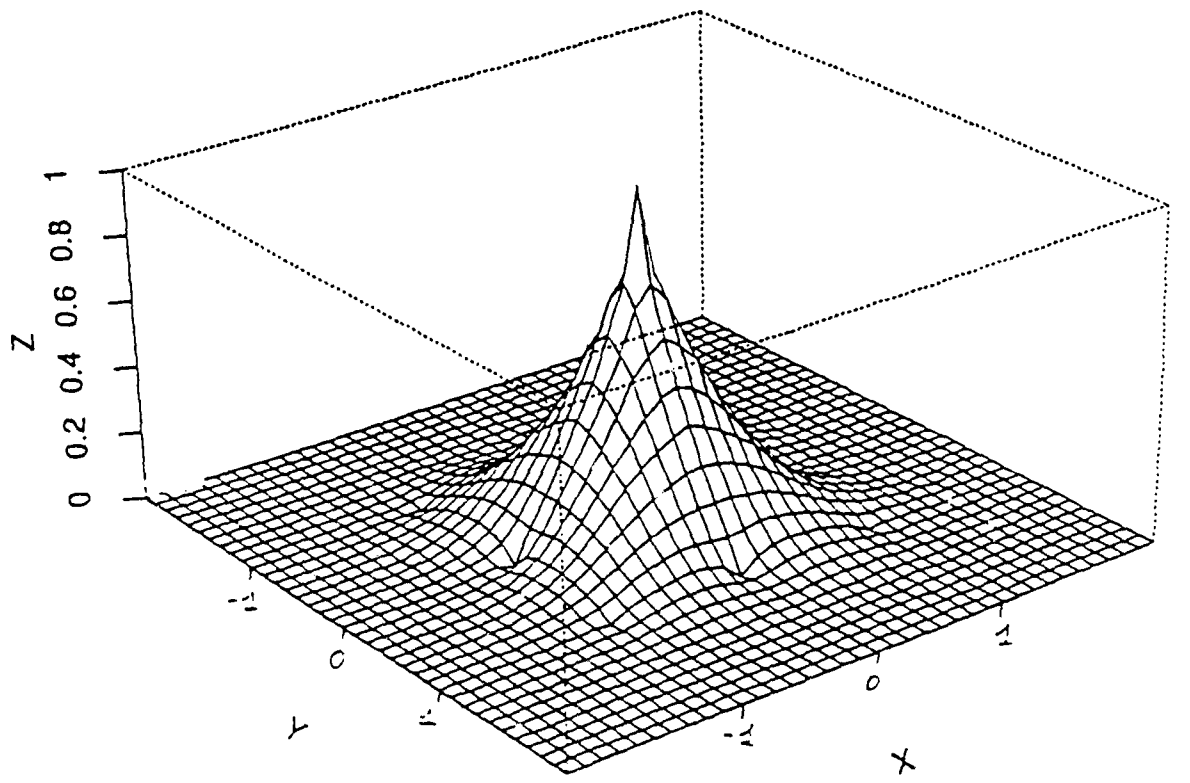


Figure 13
The scaling function ϕ_2 $(= \phi_1 * \phi_1)$

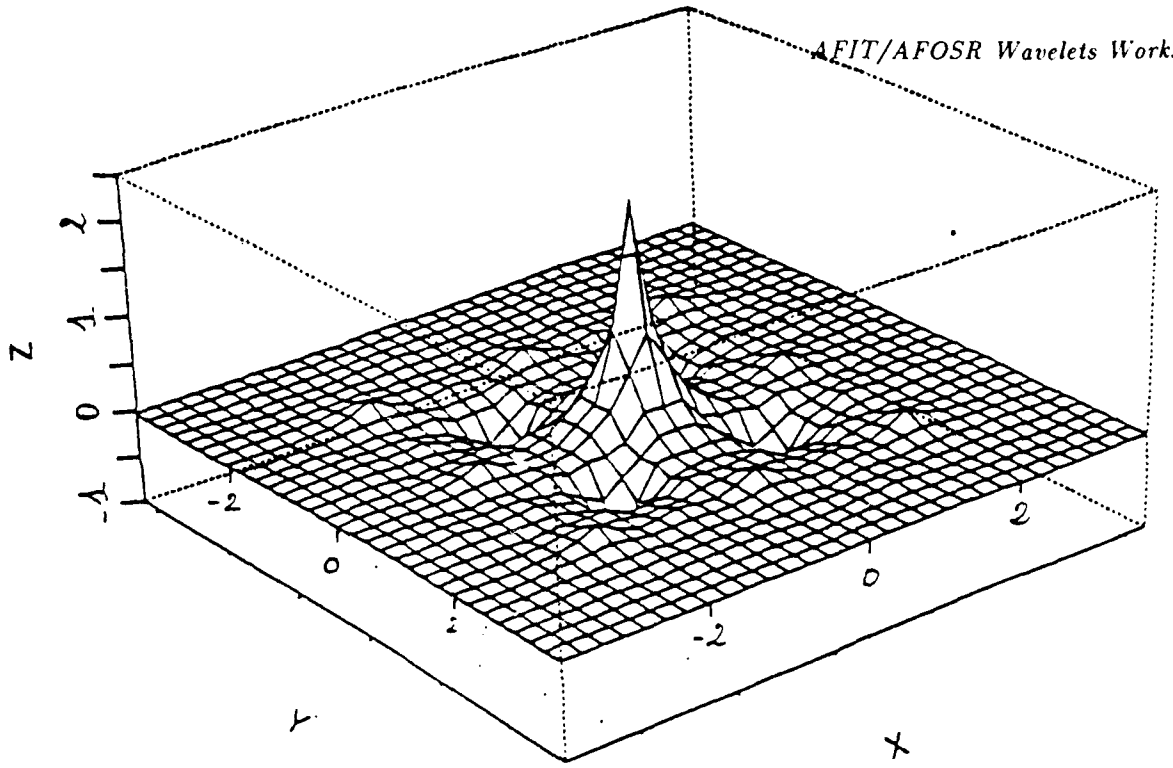


a) $\tilde{\phi}_{12}$

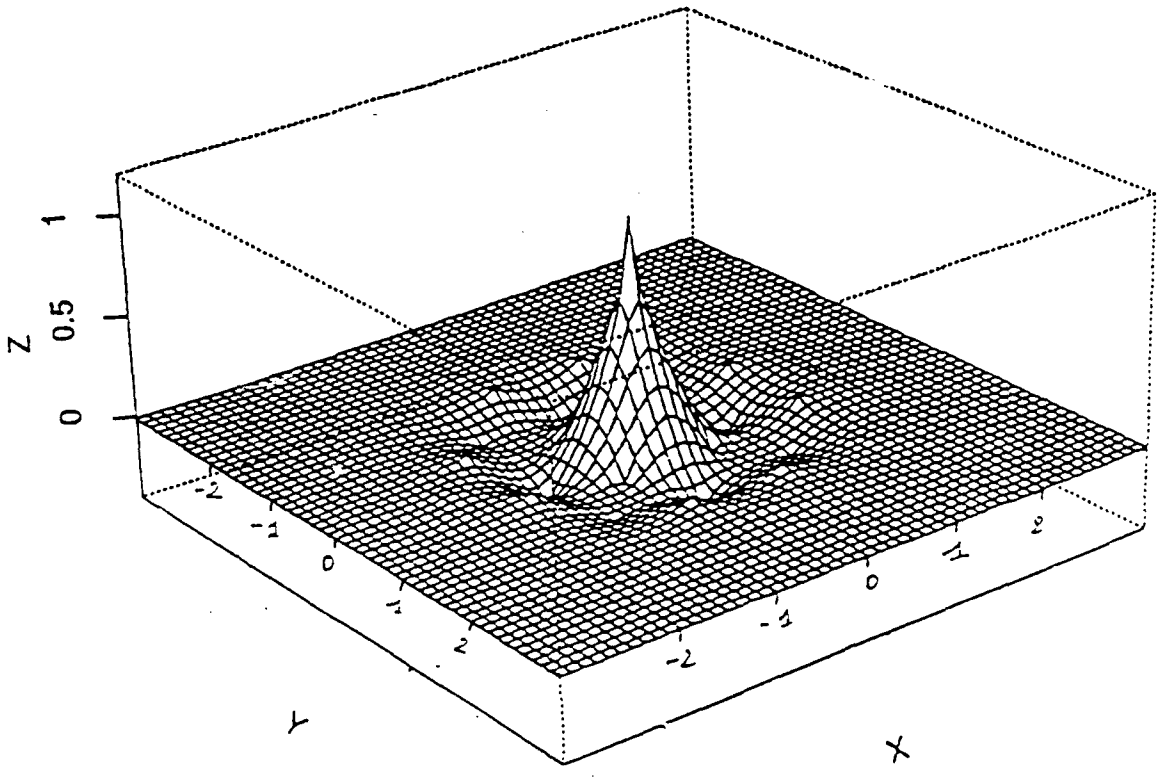


b) ϕ_1

Figure 14
Analysis and synthesis scaling function for $N = 1$ and $L = 2$



a) $\dot{\psi}_{12}$



b) ψ_{12}

Figure 15
Analysis and synthesis wavelets for $N = 1$ and $L = 2$

and this will be sufficient for these small values of N and L for which (5.52) holds. Using the definition (5.2) of $\tilde{M}_0^{N,L}$ we obtain:

- for $N = 1$, $L(1) = 3$
- for $N = 2$, $L(2) = 12$
- for $N = 3$, $L(3) = 22$

Clearly these estimates are much better than (5.49). Finally, $L(N)$ can be even more reduced, for small values of N, if an even sharper criterion that the frequency decay (5.38) is used to ensure the existence of frame bounds. We show indeed in Appendix A that the spectral analysis of the transition operators T_0 and $\tilde{T}_0$ corresponding to the functions $|M_0|^2$ and $|\tilde{M}_0|^2$ can be used to derive both the frame property and the L^2 convergence needed to have a pair of dual Riesz bases. In [CD] we prove that this criterion is sharp so that the value of $L(N)$ is here optimal. Unfortunately the matrices of T_0 and $\tilde{T}_0$ can be very big, even for small N and L.

For $N = 1$, we now obtain $L(N) = 2$ so that the two filters M_0^1 and $\tilde{M}_0^{12}$ are of small size. We show on figure 14 and 15 the scaling functions and wavelets obtained from such a choice.

Acknowledgments

The authors are grateful to K. Gröchenig, W. Madych and W. Lawton for introducing them to fractal tilings and several related problems. They are also endebted to J. Kovacevic and M. Vetterli for fruitful discussions and exchange of ideas.

References

[AB] E. Adelson and P. Burt, "The Laplacian Pyramid as a compact image code", IEEE Trans. Comm. 31 (1983) 482-540.

[ASH] E. Adelson, E. Simoncelli and R. Hingorani, "Orthogonal pyramid transform for image coding", SPIE 845 (1987) 50-58.

[CC] A. Cohen and J. P. Conze, "Régularité des bases d'ondelettes et mesures ergodiques", to appear in Revista Matemática Iberoamericana (1991).

[CD] A. Cohen and I. Daubechies, "A stability criterion for biorthogonal wavelet bases and their related subband coding schemes", preprint AT&T Bell Laboratories (1991).

[CDF] A. Cohen, I. Daubechies and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets", to appear in Comm. Pure & Appl. Math. (1991).

[CDM] A. Caveretta, W. Dahmen and C. Micchelli, "Stationary Subdivision", to appear in memoirs of AMS (1989).

[CR] J. P. Conze and A. Raugi, "Fonction Harmonique pour un operateur de transition et application", preprint Dept. de Math., Université de Rennes (France) (1990).

[Co1] A. Cohen, "Ondelettes, analyses multiresolutions et filtres miroirs en quadrature", Ann. Inst. H. Poincaré, Analyse non lineaire 7 (1990), 439-459.

[Co2] A. Cohen, "Construction de bases d'ondelettes α-Hölderiennes", revista matemática Iberoamericana 6 (1990), 91-108.

[Con] J. P. Conze, "Sur le calcul de la norme de Sobolev des fonctions d'échelles", preprint, Dept. de Math., Université de Rennes (France) (1990).

[Dau1] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure & Appl. Math. 41 (1989) 909-996.

[Dau2] I. Daubechies, "Ten Lectures on Wavelets", notes of the CBMS Conference (Lowell), ed. Jones & Bartlett (1991).

[Dau3] I. Daubechies, "Orthonormal bases of compactly supported wavelets. Part II & III: variation on a theme", preprint AT&T Bell Laboratories (1990).

[DD] G. Deslauriers and S. Dubuc, "Interpolation dyadique" in "Fractals, dimensions non entieres et applications", ed. Masson, Paris (1987) 44-45.

[DL] I. Daubechies and J. Lagarias, "Two scale difference equations. Part I & II", to appear in SIAM Journ. Math. Anal. (1989).

[DyL] N. Dyn and D. Levin. "Interpolating subdivision schemes for the generation of curves and surfaces", preprint Math. Dept. Tel Aviv Univ. (Israel, 1989).

[Fea] J. C. Feauveau, "Analyse multirésolution par ondelettes non orthogonale et benes de filtres numériques", PhD. Thesis, Université de Paris Sud. France (1990).

[K] D. Knuth, "The art of computer programming, vol. II", ed. Addison Wesley (1968).

[KV] J. Kovacevic and M. Vetterli, "Non separable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathbb{R}^n$", preprint Columbia Univ. (1991).

[Le] P. G. Lemarié, "Ondelettes à localisation exponentielle", J. Math. Pures et Appl. 67 (1988) 227-236.

[LR] W. Lawton and M. Resnikoff, "Multidimensional wavelet bases", preprint AWARE (1990).

[M] D. Marr, "Vision", ed. Freeman & Co. (1982).

[Ma1] S. Mallat, "Multiresolution approximation and wavelets orthonormal bases of $L^2(\mathbb{R})$", Trans. of AMS 315 (1989) 69-87.

[Ma2] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", IEEE PAMI 2 no. 7 (1989).

[Me1] Y. Meyer, "Ondelettes et Opérateurs", ed. Hermann, Paris (1990).

[Me2] Y. Meyer, "Ondelettes, fonctions splines et analyses graduées", CEREMADE Report 8703 (1987).

[MG] W. Madych and K. Gröchenig, "Multiresolution analysis, Haar bases and self similar tilings of $\mathbb{R}^n$", preprint, Univ. of Connecticut (1990).

[Mo] J. P. Mongeau, "Proprietés de l'interpolation iterative", PhD. Thesis, Université de Montréal (1990).

[Ri] O. Rioul, "Dyadic Up-Scaling Schemes: Simple Criteria for Regularity", submitted to SIAM J. Math. Anal. (1991).

[SB1] M. J. Smith and T. P. Barnwell, "Exact reconstruction techniques for tree structured subband coders", IEEE ASSP 34 (1986) 434-441.

[SB2] M. J. Smith and T. P. Barnwell, "A new filter bank theory for time frequency representation", IEEE ASSP 35 (1987) 314-326.

[SF] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method" in "Constructive aspect of functional analysis", ed. Geymonant (1973) 793-840.

[V] M. Volkner, "On the regularity of wavelets", preprint, Dept. of Math., Univ. of Wisconsin, Milwaukee (1990).

[Ve] M. Vetterli, "Filter bank allowing perfect reconstruction", Signal Processing 10 (1986) 219-244.

Appendix A: A sharp criterion for frame bounds

We want to give here a better result than Theorem 2.2 to characterize the dual filter pair $(m_0, \tilde{m}_0)$ which lead to biorthogonal Riesz bases of wavelets. The method that we show here uses the transition operators associated to the positive functions $|m_0|^2$ and $|\tilde{m}_0|^2$ (see Section II.4.a).

First, recall that the φ, $\tilde{\varphi}$, ψ and $\tilde{\psi}$ are defined by

$$\text{(A.1)} \qquad \begin{cases} \hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega)\,, \quad \hat{\psi}(2\omega) = m_1(\omega)\,\hat{\varphi}(\omega) \\ \hat{\tilde{\varphi}}(\omega) = \prod_{k=1}^{+\infty} \tilde{m}_0(2^{-k}\omega)\,, \quad \hat{\tilde{\psi}}(2\omega) = \tilde{m}_1(\omega)\,\hat{\tilde{\varphi}}(\omega) \end{cases}$$

As mentioned in Theorem 2.2 the duality relations (2.19), (2.20) and the decomposition formula (2.21) are ensured as soon as the partial products $\hat{\varphi}_n(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega)\,\chi_{[-2^n\pi,\,2^n\pi]}(\omega)$ and $\hat{\tilde{\varphi}}_n(\omega) = \prod_{k=1}^{n} \tilde{m}_0(2^{-k}\omega)\,\chi_{[2^n\pi,\,2^n\pi]}(\omega)$ converge in $L^2(\mathbb{R})$ respectively to $\hat{\varphi}(\omega)$ and $\hat{\tilde{\varphi}}(\omega)$.

The main difficulty is then to obtain the frame bounds A, B, $\tilde{A}$ and $\tilde{B}$ all strictly positive such that for all f in $L^2(\mathbb{R})$,

$$\text{(A.2)} \qquad \begin{cases} A\,\|f\|^2 \le \sum_{j,k\in\mathbb{Z}} |\langle f|\psi_k^j\rangle|^2 \le B\,\|f\|^2 \\ \tilde{A}\,\|f\|^2 \le \sum_{j,k\in\mathbb{Z}} |\langle f|\tilde{\psi}_k^j\rangle|^2 \le \tilde{B}\|f\|^2\,. \end{cases}$$

It is sufficient to obtain the two upper bounds of (A.2) because the lower bounds are then obtained by using (2.21) and the Schwarz inequality which give

$$\text{(A.3)} \qquad \|f\|^2 \le \left(\sum_{j,k} |\langle f|\psi_k^j\rangle|^2\right)^{1/2} \left(\sum_{j,k} |\langle f|\tilde{\psi}_k^j\rangle|^2\right)^{1/2}\,.$$

In [CDF] we used the following assumption

$$\text{(A.4)} \qquad |\hat{\psi}(\omega)|^2 + |\hat{\tilde{\psi}}(\omega)|^2 \le C(1+|\omega|)^{-\frac{1}{2}-\epsilon}$$

which can also be formulated with φ and $\tilde{\varphi}$ instead of ψ and $\tilde{\psi}$. Here, we shall prove the L^2 convergence of $\{\varphi_n, \tilde{\varphi}_n\}_{n>0}$ and the frame inequalities (A.2) using weaker assumptions. More precisely let T_0 and $\tilde{T}_0$ be the two transition operators associated to the functions $|m_0|^2$ and $|\tilde{m}_0|^2$, as defined in Section II.4.a. They both operate in two spaces of trigonometric polynomials E_N and $E_{\tilde{N}}$. We have proved in Lemma 2.2 that the subspaces $F_N = \{f \in E_N | f(0) = 0\}$ and $F_{\tilde{N}} = \{f \in E_{\tilde{N}} | f(0) = 0\}$ are invariant under the action of T_0 and $\tilde{T}_0$. The following result gives a sharp characterization of the dual filter pairs associated to biorthogonal wavelet bases.

Theorem A.3 *Let λ (resp. $\tilde{\lambda}$) be the largest eigenvalue of T_0 (resp. $\tilde{T}_0$) in the subspace F_N (resp. $F_{\tilde{N}}$). Then if $|\lambda|$ and $|\tilde{\lambda}|$ are both strictly inferior to 1, the functions ψ and $\tilde{\psi}$ defined by (A.1) generates biorthogonal Riesz bases of wavelets $\{\psi_k^j, \tilde{\psi}_k^j\}_{j,k\in\mathbb{Z}}$.*

Proof:

We shall prove here that this condition on the eigenvalues of T_0 and $\tilde{T}_0$ is sufficient to obtain biorthogonal wavelet bases. In fact, it is also a necessary condition. This result is detailed in [CD].

We first show that φ and $\tilde{\varphi}$ are in $L^2(\mathbb{R})$. As in Theorem 2.7, we apply T_0^n to the function $C_1(\omega) = 1 - \cos\omega$ which is in F_N and by using Lemma 2.5, we obtain

$$\begin{aligned}\int_{2^{n-1}\pi \le |\omega| \le 2^n\pi} |\hat{\varphi}(\omega)|^2 d\omega &\le C \int_{2^{n-1}\pi \le |\omega| \le 2^n\pi} |\hat{\varphi}_n(\omega)|^2 d\omega \\ &\le C \int_{-2^n\pi}^{2^n\pi} C_1(2^{-n}\omega) |\hat{\varphi}_n(\omega)|^2 d\omega \\ &\le C \left(\frac{\gamma+1}{2}\right)^n\end{aligned}$$

because $\frac{\gamma+1}{2} > \gamma$. Since we also have $\frac{\gamma+1}{2} < 1$, it follows that the dyadic blocks in the Littlewood-Paley decomposition of φ satisfy the inequality

$$\|\Delta_j(\varphi)\|_{L^2} \le C 2^{-\epsilon j} \quad \text{for some } \epsilon > 0 . \tag{A.5}$$

This proves that φ and ψ are even better than L^2: They belong to a Besov space $B_2^{\epsilon,\infty}$ ($\subset L^2(\mathbb{R})$) for some $\epsilon > 0$. We shall use this property to prove the frame inequalities. Similarly $\tilde{\varphi}$ and $\tilde{\psi}$ belong to $B_2^{\tilde{\epsilon},\infty}$ for some $\tilde{\epsilon} > 0$. To prove the L^2 convergence of the sequence φ_n to φ, we remark that since $m_0(0) = 1$, for α in $]0, \pi]$ small enough we have

$$|\omega| \le \alpha \Rightarrow |\hat{\varphi}(\omega)| \ge C > 0 . \tag{A.6}$$

We now introduce the sequence φ_n^o defined by

$$\hat{\varphi}_n^o(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega) \chi_{[-2^n\alpha, 2^n\alpha]}(\omega) . \tag{A.7}$$

It is clear that $\hat{\varphi}_n^o(\omega)$ converges pointwise to $\hat{\varphi}(\omega)$, but (A.6) also implies $|\hat{\varphi}_n^o(\omega)| \le \frac{|\hat{\varphi}(\omega)|}{C}$ for all $n > 0$. By the Lebesgue dominated convergence theorem we get

$$\lim_{n \to \infty} \|\varphi_n^o - \varphi\|_{L^2} = 0 . \tag{A.8}$$

We now use the hypothesis on the eigenvalues to evaluate the L^2 norm of the difference $\varphi_n - \varphi_n^o$

$$\begin{aligned}\int |\hat{\varphi}_n(\omega) - \hat{\varphi}_n^o(\omega)|^2 d\omega &= \int_{|\omega| > \alpha} |\hat{\varphi}_n(\omega)|^2 d\omega \\ &\le \frac{1}{C_1(\alpha)} \int_{-2^n\pi}^{2^n\pi} C_1(2^{-n}\omega) |\hat{\varphi}_n(\omega)|^2 d\omega \\ &\le C 2^{-\epsilon n} \xrightarrow[n \to +\infty]{} 0 .\end{aligned}$$

Consequently φ_n converges to φ in $L^2(\mathbb{R})$ and the same holds for $\tilde{\varphi}_n$ and $\tilde{\varphi}$.

It remains to establish the upper frame inequalities in (A.1). We shall obtain them by using the following lemma.

Lemma A.2 *Let ψ be a function in $L^2(\mathbb{R})$ such that for some $\sigma > 0$,*

$$\sum_{k\in\mathbb{Z}} |\hat{\psi}(\omega + 2k\pi)|^{2-\sigma} \leq C_1 \tag{A.9}$$

$$\sum_{j\in\mathbb{Z}} |\hat{\psi}(2^{-j}\omega)|^{\sigma} \leq C_2 \tag{A.10}$$

uniformly in ω. Define, for j, k in $\mathbb{Z}$, $\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k)$. Then, for all f in $L^2(\mathbb{R})$,

$$\sum_{j,k\in\mathbb{Z}} |\langle f|\psi_k^j\rangle|^2 \leq C_1 C_2 \|f\|^2 . \tag{A.11}$$

Let us first assume that this result is true to conclude the proof of the theorem. We thus have to check that there exist a $\sigma > 0$ such that (A.9) and (A.10) are satisfied for ψ and $\tilde{\psi}$.

To check (A.10), we define $I_j = [-2^{j+1}\pi, -2^j\pi] \cup [2^j\pi, 2^{j+1}\pi]$. For $j \leq 1$, we can use the cancellation of $\hat{\psi}(\omega)$ at the origin to obtain

$$\max_{\omega\in I_j} |\hat{\psi}(\omega)| \leq C2^j \quad \text{for } j \leq 1 . \tag{A.12}$$

For $j \geq 2$, we know that $\hat{\psi}(\pm 2^j\pi) = 0$ since $\hat{\varphi}(2k\pi) = 0$ for $k \in \mathbb{Z} \setminus \{0\}$. We thus have

$$\begin{aligned}
\max_{\omega\in I_j} |\hat{\psi}(\omega)|^2 &\leq \int_{I_j} \frac{d}{d\omega}\left(|\hat{\psi}|^2\right) d\omega \\
&\leq 2\int_{I_j} \left|\hat{\psi}(\omega)\frac{d\hat{\psi}}{d\omega}(\omega)\right| d\omega \\
&\leq 2\left[\int_{I_j} |\hat{\psi}(\omega)|^2 \, d\omega\right]^{1/2}\left[\int_{\mathbb{R}} \left|\frac{d\hat{\psi}}{d\omega}(\omega)\right|^2 d\omega\right]^{1/2} .
\end{aligned}$$

The first factor can be majorated by 2^{-cj} because we have proved that ψ belongs to $B_2^{c,\infty}$. The second factor is finite since it is proportional to $\int |x\psi(x)|^2 \, dx$ and ψ is a compactly supported L^2 function and consequently

$$\max_{\omega\in I_j} |\hat{\psi}(\omega)| \leq C2^{-\frac{c}{2}j} \quad \text{for } j \geq 2 . \tag{A.13}$$

Combining (A.12) and (A.13) we see that (A.10) holds for all $\sigma > 0$, since we have

$$\begin{aligned}
\max_{\omega\in\mathbb{R}} \sum_{j\in\mathbb{Z}} |\hat{\psi}(2^j\omega)|^{\sigma} &\leq \sum_{j\in\mathbb{Z}} \left[\max_{\omega\in I_j} |\hat{\psi}(\omega)|\right]^{\sigma} \\
&\leq C\left[\sum_{j\leq 1} 2^{\sigma j} + \sum_{j\geq 2} 2^{-\frac{c}{2}\sigma j}\right] \\
&\leq C_2(\sigma) .
\end{aligned}$$

We now check that (A.9) is satisfied for some $\sigma > 0$. Because the wavelet satisfies $\dot{\psi}(4k\pi) = 0$ for all $k \in \mathbb{Z}$, we can derive

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} |\dot{\psi}(\omega + 2k\pi)|^{2-\sigma} &\leq \sum_{k\in\mathbb{Z}} \int_{2k\pi}^{2k\pi+2\pi} \left| \frac{d}{d\omega}(|\dot{\psi}|^{2-\sigma}) \right| d\omega \\
&\leq \int_{\mathbb{R}} \left| \frac{d}{d\omega} \left[(|\dot{\psi}|^2)^{1-\frac{\sigma}{2}} \right] \right| d\omega \\
&\leq |2-\sigma| \int_{\mathbb{R}} |\dot{\psi}(\omega)|^{1-\sigma} \left| \frac{d\dot{\psi}}{d\omega} \right| d\omega \\
&\leq |2-\sigma| \left[\int_{\mathbb{R}} |\dot{\psi}(\omega)|^{2-2\sigma} \right]^{1/2} \left[\int_{\mathbb{R}} \left| \frac{d\dot{\psi}}{d\omega} \right|^2 d\omega \right]^{1/2} .
\end{aligned}$$

We already saw that the second factor was finite (in the proof of (A.10)). The first factor is also finite for σ small enough. Indeed, using $\psi \in B_2^{\epsilon,\infty}$ and the Hölder inequality, we obtain

$$\begin{aligned}
\int_{I_j} |\dot{\psi}(\omega)|^{2-2\sigma} d\omega &\leq \left[\int_{I_j} |\dot{\psi}(\omega)|^2 \, d\omega \right]^{1-\sigma} (2^{j+1}\pi)^{\sigma} \\
&\leq C\, 2^{j(\sigma - 2\epsilon(1-\sigma))} .
\end{aligned}$$

We thus have to choose $\sigma > 0$ such that $\sigma - 2\epsilon(1-\sigma) < 0$, i.e. $\sigma < \frac{2\epsilon}{1+2\epsilon}$. Since the same results also hold for the dual wavelet $\tilde{\psi}$, the theorem is proved modulo the Lemma A.2 that we now tackle. Using the Plancherel and the Poisson formulas, we derive for any f in $L^2(\mathbb{R})$

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} |\langle f | \psi_k^j \rangle|^2 &= \frac{1}{4\pi^2} \sum_{k\in\mathbb{Z}} 2^j \left| \int_{\mathbb{R}} \hat{f}(\omega) \, \overline{\hat{\psi}(2^j\omega)} \, e^{-i2^j k\omega} \, d\omega \right|^2 \\
&= \frac{1}{4\pi^2} \sum_{k\in\mathbb{Z}} 2^{-j} \left| \int_{\mathbb{R}} \hat{f}(2^{-j}\omega) \, \overline{\hat{\psi}(\omega)} \, e^{-ik\omega} \, d\omega \right|^2 \\
&= \frac{1}{2\pi} \int_0^{2\pi} 2^{-j} \left| \sum_{\ell\in\mathbb{Z}} \hat{f}(2^{-j}(\omega + 2\ell\pi)) \, \overline{\hat{\psi}(\omega + 2\ell\pi)} \right|^2 d\omega \\
&\leq \frac{2^{-j}}{2\pi} \int_0^{2\pi} \left(\sum_{\ell\in\mathbb{Z}} |\hat{f}(2^{-j}(\omega + 2\ell\pi)| \, |\dot{\psi}(\omega + 2\ell\pi)|^{\frac{\sigma}{2}} \, |\dot{\psi}(\omega + 2\ell\pi)|^{1-\frac{\sigma}{2}} \right)^2 d\omega \\
&\leq \frac{2^{-j}}{2\pi} \int_0^{2\pi} \left(\sum_{\ell\in\mathbb{Z}} |\hat{f}(2^{-j}(\omega + 2\ell\pi))|^2 \, |\dot{\psi}(\omega + 2\ell\pi)|^{\sigma} \right) \left(\sum_{\ell\in\mathbb{Z}} |\dot{\psi}(\omega + 2\ell\pi)|^{2-\sigma} \right) d\omega \\
&\leq C_1 \frac{2^{-j}}{2\pi} \int_{\mathbb{R}} |\hat{f}(2^{-j}\omega)|^2 \, |\dot{\psi}(\omega)|^{\sigma} \, d\omega \\
&= \frac{1}{2\pi} C_1 \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \, |\dot{\psi}(2^j\omega)|^2 \, d\omega .
\end{aligned}$$

Summing on all the scales $j \in \mathbb{Z}$ and using (A.10), we get

$$\sum_{j,k\in\mathbb{Z}} |\langle f|\psi_k^j\rangle|^2 \le \frac{C_1C_2}{2\pi} \int_{\mathbb{R}} |\hat{f}(\omega)|^2 d\omega = C_1C_2 \|f\|^2 \tag{A.11}$$

and this concludes the proof. ∎

Appendix B. Dragonic expansions.

In this Appendix we want to show how the one-dimensional techniques in [DL] can be extended to multidimensional situations. As an example we discuss the two-dimensional case, with the dilation matrix $R = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$.

A first multidimensional extension of [DL] can be found in [Mo]. Even though he looks at general matrices, Mongeau effectively reduces his analysis to pure dilations by considering the smallest n such that $\overline{D} = D^n$ is a multiple of the identity, and rewriting (by iteration) the two-scale equation so that it involves only $\overline{D}$. This procedure can drastically increase the number of different terms in the equation. We choose here to work directly with $D = R$ itself.

When the two-scale equation is one-dimensional, and the dilation factor is 2, the regularity at x of the function ϕ solving the equation is regulated by the binary expansion of x (for dilation factor p, the same role is played by the p-ary expansion). Moreover, $\mathbb{R}$ and in particular $\mathrm{supp}(\phi)$ is tiled with integer translates of the interval $[0,1]$, which can be viewed as the set of numbers equal to the decimal part only of their dyadic expansion; if N such tiles are needed to cover support ϕ, then the two-scale functional equation can be rewritten as an equation for an N-dimensional vector-valued function involving two matrices T_0 and T_1. The spectral properties of T_0, T_1 then determine the regularity of ϕ, both local and global [DL].

In the two-dimensional case with dilation matrix R, the role of elementary tile is now played by the twin dragon set Δ. It is defined by

$$\text{(B.1)} \quad \Delta = \left\{ x \in \mathbb{R}^2;\ x = \sum_{j=1}^{\infty} R^{-j} p_j \text{ where } p_j \in L = \mathbb{Z}^2/R\mathbb{Z}^2 = \{(0,0),\ (1,0)\} \right\} .$$

Under the standard identification of $\mathbb{R}^2$ with $\mathbb{C}$, with $(x,y) \simeq x + iy$, Δ can also be written as

$$\text{(B.2)} \qquad \Delta = \left\{ z \in \mathbb{C};\ z = \sum_{j=1}^{\infty} d_j \left(\frac{1+i}{2}\right)^j \text{ where } d_j = 0 \text{ or } 1 \right\} .$$

This set Δ is compact, has fractal boundary, is selfsimilar, and its $\mathbb{Z}^2$-translates tile the plane. The indicator function of Δ is the solution to the two-scale equation

$$\phi(x) = \phi(Rx) + \phi(Rx - (1,0))$$

(see [GM]). Δ is called the twin dragon set [K]. We shall give the name *dragonic expansions* to expansions of x or z as in (B.1), (B.3). Note that (as in the binary case) some points may have two different dragonic expansions. e.g. $.01000\ldots = \left(\frac{1+i}{2}\right)^2 = \frac{i}{2} = .1011111\ldots$ (This example also illustrates that addition follows rules very different from the binary case, since $.0100\ldots + .0100\ldots = .1111\ldots$.)

Suppose we are interested in various regularity properties of L^1-solutions ϕ of

$$\text{(B.3)} \qquad \phi(x) = \sum_{k \in \Lambda} c_k\, \phi(Rx - k) ,$$

where Λ is a finite subset of $\mathbb{Z}^2$. Such solutions are uniquely defined up to normalization and have necessarily compact support. One can determine the minimal set $\Gamma \subset \mathbb{Z}^2$ so that $R^{-1}(\Gamma + \Lambda - L) \subset \Gamma$; then support $\phi \subset \cup_{\ell \in \Gamma}(\Delta + \ell)$. The equation (B.3) for ϕ can be rewritten by defining the $|\Gamma|$-dimensional vector $v(x)$ by

$$v_j(x) = \phi(x + j) \qquad j \in \Gamma,\ x \in \Delta\ ; \tag{B.4}$$

we have

$$v_j(x) = \sum_k c_{Rj+d_1(x)-k}\, v_k(\tau x)$$

where $d_1(x)$ is the first digit in the dragonic expansion of x ,

and τx is the point obtained by dropping $d_1(x)$ from the (same) dragonic expansion of x, $\tau x = \sum_{j=1}^{\infty} d_{j+1}(x) \left(\frac{1+i}{2}\right)^j$.

Equation (B.4) can be recast as

$$v(x) = T_{d_1(x)}\, v(\tau x)\ , \tag{B.5}$$

where $(T_0)_{jk} = c_{Rj-k}$, $(T_1)_{jk} = c_{Rj-k+(1,0)}$.

We have completed a setup analogous to that of [DL]. The question is now whether the proof techniques of [DL] still work in this case. The answer is basically yes. For instance, we still have

Theorem B.3 *Assume that the c_k in (B.3) satisfy*

$$\sum_n c_{Rn} = \sum_n c_{Rn+(1,0)} = 1\ .$$

Then $\epsilon_1 = (1, 1, \ldots, 1)$ is a common lefteigenvector of T_0, T_1 with eigenvalue 1 for both matrices. Define E_1 to be the one-dimensional subspace orthogonal to ϵ_1. If there exist $\lambda < 1$, $C > 0$ so that

$$\|T_{d_1} \ldots T_{d_m}|_{E_1}\| \le C\lambda^m \tag{B.6}$$

for all possible $d_j = 0$ or 1, all $m \in \mathbb{N}$, then the L^1-solution ϕ to (B.3) is Hölder continuous with exponent $\alpha = |\log \lambda| / \log \sqrt{2}$.

This is the analog of Theorem 2.3 in [DL]. Two different strategies of proof are given in [DL]. The first one involves piecewise linear spline approximants; this technique would be hard to generalize here because of the fractal boundary of our domain building blocks $\Delta + k$. A second strategy, which does not use splines at all, but leads to longer proofs, is explained in the Appendix in [DL]: this strategy generalizes to the present case. The main point we have to check to make sure the proof carries over is whether elements that are close necessarily have dragonic expansions with the same starting digits. In the one-dimensional,

binary case, if two dyadic rationals x, y are closer than 2^{-m}, $|x-y| < 2^{-m}$, then x and y have binary expansions with coinciding first m digits. (If e.g. $x \le y < x + 2^{-m}$, then the expansion "from above" of x — ending in all zeros — has the same first m digits as the expansion "from below" of y — ending in all ones.) This is crucial in the proof, and allows to extract Hölder continuity from the condition (B.6). We therefore have to check whether a similar property holds in the "dragonic" case.

By analogy we shall call *dragonic rationals* all the points in Δ for which a terminating dragonic expansion can be written. Typically dragonic rationals also have other, non-terminating dragonic expansions. For each dragonic rational x the terminating expansion is unique; we denote its digits by $d_j^0(x)$, $j \in \mathbb{N}$.

Let us also introduce the notations R_0, R_1,

$$R_0 y = Ry, \quad R_1 y = Ry + (1,0),$$

or
$$R_d y = Ry + d(1,0), \quad \text{with } d = 0 \text{ or } 1 .$$

Take now a fixed dragonic rational x, and assume that $d_j^0(x) = 0$ for $j > J$. All the $y \in \Delta$ that have the same first J digits $d_j^0(x)$, $j \le J$, constitute a little dragon $\Delta_J(x)$ themselves,

$$\Delta_J(x) = R^{-1}_{d_J(x)} \dots R^{-1}_{d_1(x)} \Delta ;$$

x itself is the image of $(0,0)$ under the same map $R^{-1}_{d_J(x)} \dots R^{-1}_{d_1(x)}$. The set Δ is tiled by 2^J little dragons of the same size as Δ_J, all translates of Δ_J. For every such little dragon, we call the point corresponding to $(0,0)$ the "bottom", and the point corresponding to $(0,1)$ (the only other point in $\mathbb{Z}^2 \cap \Delta$) the "top". If x is a dragonic rational with at most N nonzero digits, then x is the bottom of $\Delta_J(x)$ for all $J > N$. (But note that the "orientation" of $\Delta_J(x)$, as indicated by the line connecting bottom and top, changes with J!). It follows that x is on the border of these $\Delta_J(x)$. If x is not at the edge of Δ itself, then there must exist another little dragon $\Delta_J(y)$ so that x is the top of $\Delta_J(y)$ (since Δ is the union of all the 2^J possible dragons Δ_J). Since the top $(0,1)$ of Δ is given by the expansion $.111111\ldots$, we can therefore find another dragonic expansion for x, ending in all one's, and with the same J first digits as y,

$$d_j^1(x) = d_j^0(y) \text{ for } j \le J, \quad d_j^1(x) = 1 \text{ for } j > J .$$

We have seen how to obtain the two expansions for a dragonic rational x. We now want to show that if another dragonic rational y is "close" to x, then at least one of its expansions starts with the same digits as one of the expansions for x. Define

$$\rho = \max \{r;\ B((0,0); r) \subset \Delta \cup (\Delta - (0,1))\} ,$$

where $B(y;\lambda)$ is the open Euclidean ball centered at y with radius λ. Suppose x is a dragonic rational with $d_j^0(x) = 0$ for $j > J$. Take $m > J$, and consider the set

$$B_m = \{y \in \Delta;\ |y - x| \le \rho 2^{-m/2}\} .$$

There are two possibilities: either x is on the border $\partial\Delta$ of Δ, or it isn't. If $x \in \partial\Delta$, then

$$R^{-1}_{d^0_m(x)} \dots R^{-1}_{d^0_1(x)}(\Delta - (0,1))$$

has no common interior points with Δ, so that $B_m \subset \Delta_m(x)$, and the terminating expansions of all $y \in B_m$ has the same first m digits $d^0_j(x)$, $j = 1, \dots, m$. If $x \notin \partial\Delta$, then $R^{-1}_{d^0_m(x)} \dots R^{-1}_{d^0_1(x)}(\Delta - (0,1)) \subset \Delta$: this set is then a little dragon $\Delta_m(x)$ of which x is the top. In this case $B_m \subset \Delta_m(x) \cup \Delta_m(z)$, so that every point $y \in B_m$ has a dragonic expansion with either the same first m digits as $d^0(x)$ (if $y \in \Delta_m(x)$) or as $d^1(x)$ (if $y \in \Delta_m(z)$). This is the main ingredient needed to make the proof of Theorem 2.3, as sketched in the Appendix in [DL], work in the present case.

One other point that needs checking is whether the existence of two different dragonic expansions for x doesn't lead to inconsistencies for the definition of $v(x)$. If $d^0_j(x) = 0$ for $j > J$, then $d^0(x)$, $d^1(x)$ are linked by

$$x = \sum_{j=1}^{N} d^0_j(x)\, R^{-j}(1,0) = \sum_{j=1}^{N} d^1_j(x)\, R^{-j}(1,0) + R^{-N}(0,1)$$

for $N \geq J$ arbitrary. One can then compute $v(x)$ in two ways, using the two expansions. The following computation shows that they lead to the same result: for $k \in \Gamma$,

$$\begin{aligned}
\left[v\left(\sum_{j=1}^{N} d^0_j(x) R^{-j}(1,0) \right) \right]_k &= \left[T_{d^0_1(x)} \dots T_{d^0_n(x)} v(0,0) \right]_k \\
&= \sum_{j_1 \dots j_N} c_{Rk + d^0_1(x)(1,0) - j_1}\, c_{Rj_1 + d^0_2(x)(1,0) - j_2} \cdots c_{Rj_N + d^0_N(x)(1,0) - j_N} [v(0,0)]_{j_N} \\
&= \sum_{m_1 \dots m_N} c_{Rk + d^1_1(x)(1,0) - m_1} \cdots c_{Rm_{N-1} + d^1_n(x)(1,0) - m_N} \\
&\qquad [v(0,0)]_{m_N + \sum_{k=1}^{N} d^0_k(x) R^{N-k}(1,0) - \sum_{k=1}^{N} d^1_k(x) R^{N-k}(1,0)} \\
&= \sum_{m_1 \dots m_N} c_{Rk + d^1_1(x)(1,0) - m_1} \cdots c_{Rm_{N-1} + d^1_n(x)(1,0) - m_N} \, [v(0,0)]_{m_N + (0,1)} \\
&= \left[T_{d^1_1(x)} \dots T_{d^1_N(x)}\, v(0,1) \right]_k .
\end{aligned}$$

The reader can now check that the proof in [DL] indeed carries over to prove Theorem B.1. Similarly, one can prove differentiability of ϕ under stronger conditions on T_0, T_1, similar to Theorem 3.1 in [DL]. Finally, the same techniques can also be used for local regularity estimates, but these are a bit more tricky, and require further study of the properties of dragonic expansions. In practice, the matrices $T_0|_{E_1}$, $T_1|_{E_1}$ are often too large to permit a rigorous estimate of λ in (B.6). However, λ is bounded below by the quantities $\rho(T_{d_1} \dots T_{d_m}|_{E_1})^{1/m}$, and this leads to upper bounds for the Hölder exponent α.

Examples.

1. $g(x) = \frac{1}{2}\, g(Rx + (1,0)) + g(Rx) + \frac{1}{2}\, g(Rx - (1,0))$
   The solution to this equation is the convolution $\chi_\Delta * \chi_\Delta$, where χ_Δ is the indicator

function of the dragon set Δ (see also the second remark following Proposition 5.2). In this case Γ has 10 elements. The largest spectral radius of $T_d|_{E_*}$ is obtained for $d = 0$, $\rho(T_0|_{E_1}) = .847810\ldots$, corresponding to a lower bound $\lambda \geq \rho(T_0|_{E_1})$ in ( ), or a Hölder exponent $\alpha \leq .47637\ldots$. Via other methods (using the transition operator T of (5.19)) one also derives that this value is a lower bound. This global Hölder exponent is attained in dragonic rationals, in particular in $(0,0)$.

Note that when M_0 is positive, as in this case, the transition operator T is already known to give optimal results. One easily checks that the matrix representing T is in fact a submatrix of T_0, so that it is not surprising that they have a common eigenvalue!

2. $\phi(x) = h_0\phi(Rx) + h_1\phi(Rx-(1,0)) + h_2\phi(Rx-(-1,1)) + h_3\phi(Rx-(0,1))$, with $h_0 = .506970418225$, $h_1 = -.207072424345$, $h_2 = .493029581775$, $h_3 = 1.20707242435$. This is an example from the family described at the very end of §III.3.a. It leads to an orthonormal wavelet basis. In this case $|\Gamma| = 14$; the parameters have been chosen so that $\rho(T_0|_{E_1} \simeq \rho(T_1|_{E_1}) \simeq .714$. Plots of approximations to ϕ seem to suggest that ϕ might be continuous, but we have no proof. If it is, then its Hölder exponent is bounded above by $\log[\rho(T_0T_1|_{E_1})^{1/2}]/\log\sqrt{2} \simeq \log(.90649)/\log\sqrt{2} \simeq .28327$.

Appendix C. Proof of the inequalities (5.52) for G(ω).

The function G is defined as

$$G(\omega) = \left[\cos^2\frac{\omega_1}{2} + \cos^2\frac{\omega_2}{2}\right]\left[\sin^2\frac{\omega_1}{2} + \sin^2\frac{\omega_2}{2}\right]\left[\sin^2\frac{\omega_1+\omega_2}{2} + \sin^2\frac{\omega_1-\omega_2}{2}\right]^{-1} H(\omega).$$

with
$$H(\omega) = h\left(\frac{1}{2}\left(\sin^2\frac{\omega_1}{2} + \sin^2\frac{\omega_2}{2}\right)\right),$$

and
$$h(t) = \begin{cases} \dfrac{1}{1-t} & 0 \le t \le 1/2 \\ \\ 4t & 1/2 \le t \le 1 . \end{cases}$$

We want to prove inequalities for $G(\omega)$, $G(\omega)G(D\omega)$ and $G(\omega)G(D\omega)G(D^2\omega)$, where $D(\omega_1,\omega_2)$ is either $(\omega_1+\omega_2,\ \omega_1-\omega_2)$ or $(\omega_1-\omega_2,\ \omega_1+\omega_2)$. (Since G is invariant for the interchange of ω_1,ω_2, it does not matter which definition of D is taken, $D = R$ of $D = S$.) To prove these inequalities it is convenient to use different variables,

$$s = s(\omega) = \frac{1}{2}\left(\sin^2\frac{\omega_1}{2} + \sin^2\frac{\omega_2}{2}\right),\ p = p(\omega) = \sin^2\frac{\omega_1}{2}\sin^2\frac{\omega_2}{2}.$$

We then have
$$G(\omega) = \frac{s(1-s)}{s-p}h(s) = 2\eta(s,p).$$

Moreover,
$$s(D\omega) = 2(s-p),\ \ p(D\omega) = 4(s^2-p).$$

As ω ranges over $[-\pi,\pi]$, (s,p) fill out the domain Δ defined by

$$\Delta = \{(s,p);\ 0 \le s \le 1,\ \max(0,2s-1) \le p \le s^2\}.$$

In terms of these new variables, we therefore want to study $\eta(s,p)$, $\eta(s,p)\,\eta(\tilde{D}(s,p))$ and $\eta(s,p)\,\eta(\tilde{D}(s,p))\,\eta(\tilde{D}^2(s,p))$, for all $(s,p) \in \Delta$, where $\tilde{D}$ is defined by

$$\tilde{D}(s,p) = (\tilde{s},\tilde{p}) = (2(s-p),\ 4(s^2-p)).$$

Note that $\tilde{D}$ maps Δ twice onto itself (both $\Delta \cap \{s \le 1/2\}$ and $\Delta \cap \{s \ge 1/2\}$ get mapped to all of Δ). Moreover $\tilde{D}$ has one fixed point, $(s_0,p_0) = \left(\frac{5}{8},\ \frac{5}{16}\right)$, corresponding to $\eta(s_0,p_0) = \frac{15}{16}$.

We shall prove that $\Delta = \Delta_1 \cup \Delta_2 \cup \Delta_3$, where

(C.1)
$$\eta(s,p) \le \zeta \quad \text{on } \Delta_1$$

(C.2)
$$\eta(s,p)\,\eta(\tilde{D}(s,p)) \le \zeta^2 \quad \text{on } \Delta_2$$

(C.3) $$\eta(s,p)\,\eta(\tilde{D}(s,p))\,\eta(\tilde{D}^2(s,p)) \le \zeta^3 \quad \text{on } \Delta_3 .$$

The value of ζ will be fixed by our estimates below; our goal is to obtain $\zeta < 1$.

Choose $a = \sqrt{.9}$, and define the region Δ_1 by

$$\Delta_1 = \left\{(s,p) \in \Delta:\ p \le \left(1 - \frac{1}{2a}\right)s \ \text{ if } \ s \le 1/2\,,\right.$$

$$\left. p \le s - \frac{2}{a}\,s^2(1-s) \ \text{ if } \ s \ge 1/2\right\}\,.$$

Since

$$\eta(s,p) = \frac{s}{2(s-p)} \ \text{ if } \ s \le 1/2, \quad \frac{2s^2(1-s)}{s-p} \ \text{ if } \ s \ge 1/2\,,$$

we automatically have

(C.4) $$\eta(s,p) \le a \ \text{ on } \ \Delta_1 .$$

By the definition of η and $\tilde{D}$, we have to distinguish four different regions when studying $\eta_2(s,p) = \eta(s,p)\eta(\tilde{D}(s,p))$:

$$\eta_2(s,p) = \begin{cases} \dfrac{s}{2(\tilde{s}-\tilde{p})} = \dfrac{s}{4(s-2s^2+p)} & \text{if } s \le 1/2,\ p \le s - 1/4 \\[2ex] \dfrac{2s\tilde{s}(1-\tilde{s})}{\tilde{s}-\tilde{p}} & \text{if } s \le 1/2,\ p \ge s - 1/4 \\[2ex] \dfrac{s^2(1-s)}{\tilde{s}-\tilde{p}} = \dfrac{s^2(1-s)}{s-2s^2+p} & \text{if } s \ge 1/2,\ p \ge s - 1/4 \\[2ex] \dfrac{4s^2(1-s)(1-\tilde{s})}{\tilde{s}-\tilde{p}} = \dfrac{8s^2(1-s)(s-p)(1-2(s-p))}{s+p-2s^2} & \text{if } s \ge 1/2,\ p \le s - 1/4\,. \end{cases}$$

We define Δ_2 by

$$\Delta_2 = \left\{(s,p) \in \Delta;\ s \le 1/2,\ \ p \ge \left(1 - \frac{1}{2a}\right)s\right\}$$

$$\cup\ \{(s,p) \in \Delta:\ \ s \ge 1/2,\ \ p \ge 1.8s - .81\}$$

$$= \Delta_{2,1} \cup \Delta_{2,2}\,.$$

Since $\tilde{s}(1-\tilde{s}) \le 1/4$ for all $\tilde{s} \in [0,1]$, we have $\eta_2(s,p) \le \frac{s}{2(\tilde{s}-\tilde{p})} = \frac{s}{4(s-2s^2+p)}$ on all of $\Delta_{2,1}$.

Since moreover $p \ge \left(1 - \frac{1}{2a}\right)s$, we have

$$\eta_2(s,p) \le \left[4\left(2 - \frac{1}{2a} - 2s\right)\right]^{-1} \le \left[4\left(1 - \frac{1}{2a}\right)\right]^{-1} < a^2$$

on $\Delta_{2,1}$.

On $\Delta_{2,2} \cap \{(s,p) \in \Delta;\ \ p \geq s - 1/4\}$, one easily checks that $\eta_2(s,p) = \frac{s^2(1-s)}{s-2s^2+p}$ satisfies $\partial_p \eta_2 \neq 0$ everywhere. It follows that η_2 achieves its maximum on the boundary of this domain, given by the three pieces $p = s^2$, with $1/2 \leq s \leq .9$, $p = s - 1/4$ with $1/2 \leq s \leq .7$, and $p = 2\alpha s - \alpha^2$ with $.7 \leq s \leq .9$. One easily checks that the maximum value of η_2 on this boundary is .9.

Similarly one checks that η_2 achieves its maximum on $\Delta_{2,2} \cap \{(s,p) \in \Delta;\ \ p \leq s - 1/4\}$ on the boundary of this set; again this leads to $\eta_2 \leq .9$.

It follows that

$$\eta_2(s,p) \leq .9 = \alpha^2 \text{ on all of } \Delta_2 . \tag{C.5}$$

It remains to determine an upper bound on $\eta_3(s,p) = \eta(s,p)\eta(\tilde{D}(s,p))\eta(\tilde{D}^2(s,p))$ on $\Delta \setminus (\Delta_1 \cup \Delta_2) = \{(s,p);\ \ 2s-1 \leq p \leq s^2,\ p \geq s - \frac{2}{\alpha}s^2(1-s),\ \ p \leq 1.8s - .81\}$ Since $s - \frac{2}{\alpha}s^2(1-s)$ is strictly increasing, we have $\Delta \setminus (\Delta_1 \cup \Delta_2) \subset \Delta_3 = \{(s,p),\ 2s-1 \leq p \leq s^2,\ p_1 = 1.8s_1 - .81 \leq p \leq 1.8s - .81\}$, where s_1 is the solution to $s - \frac{2}{\alpha}s^2(1-s) = 1.8s - .81$. In Δ_3 one has to distinguish 4 subdomains, corresponding to different expressions for η_3, namely $\Delta_{3,1} = \Delta_3 \cap \{p \geq p_1,\ p \geq 2s-1,\ p \leq 2(s-1/4)^2\}$, $\Delta_{3,2} = \Delta_3 \cap \{p \geq 2(s-1/4)^2,\ p \leq s - 1/4\}$, $\Delta_{3,3} = \Delta_3 \cap \{p \geq s - 1/4,\ p \geq 2(s-1/4)^2\}$ and $\Delta_{3,4} = \Delta_3 \cap \{p \geq s - 1/4,\ p \leq s(s-1/4)^2\}$. On $\Delta_{3,1}$, $\Delta_{3,3}$ and $\Delta_{3,4}$ one checks explicitly that $\partial_p \eta_3 \neq 0$. On $\Delta_{3,2}$, the exact expression for η_3 is too complicated, but one can replace it by an upper bound,

$$\begin{aligned} \eta_3(s,p) &= \frac{2s^2(1-s)}{s-p}\ \frac{2\tilde{s}^2(1-\tilde{s})}{\tilde{s}-\tilde{p}}\ \frac{2\tilde{\tilde{s}}^2(1-\tilde{\tilde{s}})}{\tilde{\tilde{s}}-\tilde{\tilde{p}}} \\ &\leq \frac{2s^2(1-s)}{s-p}\ \frac{\tilde{s}}{\tilde{s}-\tilde{p}}\ \frac{2\tilde{\tilde{s}}^2(1-\tilde{\tilde{s}})}{\tilde{\tilde{s}}-\tilde{\tilde{p}}} = \frac{16s^2(1-s)\tilde{\tilde{s}}(1-\tilde{\tilde{s}})}{\tilde{\tilde{s}}-\tilde{\tilde{p}}} = \overline{\eta}_3(s,p) . \end{aligned}$$

This upper bound again satisfies $\partial_p \overline{\eta}_3 \neq 0$ on $\Delta_{3,2}$. It follows that η_3 on Δ_3 is bounded by the maximum of η_3 on the boundaries of $\Delta_{3,1}$, $\Delta_{3,3}$, $\Delta_{3,4}$ and of $\overline{\eta}_3$ on the boundary of $\Delta_{3,2}$. Explicitly, for all $(s,p) \in \Delta_3$,

$$\eta_3(s,p) \leq \overline{\eta}_3(s_1,p_1) = .88145650226\ldots . \tag{C.6}$$

This numerical upper bound is larger than $(.9)^{3/2}$; it follows therefore from (C.4) and (C.5) that we have proved (C.1)–(C.2) for $\zeta = [\overline{\eta}_3(s_1,p_1)]^{1/3} = .958812370442\ldots$.

Scale and Inverse Frequency Representations

LEON COHEN*

CAIP Center, Rutgers University, Piscataway, New Jersey 08855-1390

It is shown that there are two basic ideas regarding scale. One is inverse frequency and the other is scaling of frequency functions. An explicit expression for the scaling operator is given and its general properties are derived. A general approach for obtaining joint scale representations is presented. Joint representations of scale and time are obtained and a method to generate an infinite number of such distributions is given. The results of Bertrand and Bertrand, Marinovic and Altes are obtained as special cases of joint distributions involving scale and inverse frequency. The expression for instantaneous scale is derived. The uncertainty principle involving scale and time, and the minimum time-scale uncertainty signal is obtained.

CONTENTS

Introduction and General Approach ..... Sec. 1
Scale Operators 2
Method of Evaluation 3
Joint Distributions With Scale 4
General Class 5
Instantaneous Scale 6
Eigenfunctions and Eigenvalues 7
Uncertainty Principle for Scale 8
Inverse Frequency 9
Distributions of Time and Inverse Frequency 10
Wavelet Transform 11
Conclusion 12

1. INTRODUCTION AND GENERAL APPROACH

Our aim is to study scale and describe a general method for obtaining joint representations of scale with time or frequency. The fundamental idea is to use the characteristic function method which we now describe. The characteristic function of the joint distribution, $P(a,b)$, is

$$M(\xi,\zeta) = \iint e^{j\xi a + j\zeta b}\, P(a,b)\, da\, db\,. \qquad (1.1)$$

The distribution is obtained by

$$P(a,b) = \frac{1}{4\pi^2} \iint M(\xi,\zeta)\, e^{-j\xi a - j\zeta b}\, d\xi\, d\zeta\,. \quad (1.2)$$

Therefore, if there were a way to obtain the characteristic function we could find the distribution function. The characteristic function is an average, namely the average of $e^{j\xi a + j\zeta b}$,

$$M(\xi,\zeta) = \langle e^{j\xi a + j\zeta b} \rangle \qquad (1.3)$$

and generally can not be obtained without knowing the distribution. However, the essence of the method we present is that indeed one can calculate the characteristic function directly from the signal. This is done by associating operators with physical quantities, a method first proposed by Gabor and Ville for time and frequency and generalized by others.

One calculates averages by "sandwiching" the operator between the signal and its complex conjugate. In particular, if a quantity is represented by the operator $C(\mathcal{T},\mathcal{W})$ where $\mathcal{T}$ and $\mathcal{W}$ are the time and frequency operators then its average value is obtained by way of

$$\langle C \rangle = \int s^*(t)\, C(\mathcal{T},\mathcal{W})\, s(t)\, dt \qquad (1.4)$$

where $s(t)$ is the signal. (Operators will be denoted in calligraphic letters. We generally assume that they are Hermitian and that their eigenfunctions are normalized to a delta function.)

To apply this to our case suppose we have two quantities a and b and suppose these two quantities are represented by the operators $\mathcal{A}$ and $\mathcal{B}$. We define the characteristic function operator by

* Permanent address: Hunter College and Graduate Center of CUNY, New York, NY, 10021.

$$\mathcal{M}(\xi,\zeta) = e^{j\xi\mathcal{A}+j\zeta\mathcal{B}} \tag{1.5}$$

and the characteristic function is then the average of

$$M(\xi,\zeta) = \langle \mathcal{M}(\xi,\zeta) \rangle \, . \tag{1.6}$$

Using Eq. (1.4) we take

$$M(\xi,\zeta) = \int s^*(t)\, e^{j\xi\mathcal{A}+j\zeta\mathcal{B}}\, s(t)\, dt \, . \tag{1.7}$$

In Eq. (1.7) $\mathcal{A}$ and $\mathcal{B}$ have to be expressed in the time representation. However any other representation can be used. In particular, in the frequency representation,

$$M(\xi,\zeta) = \int S^*(\omega)\, e^{j\xi\mathcal{A}+j\zeta\mathcal{B}}\, S(\omega)\, d\omega \tag{1.8}$$

where $S(\omega)$ is the Fourier transform of $s(t)$

$$S(\omega) \;=\; \frac{1}{\sqrt{2\pi}} \int s(t)\, e^{-j\omega t}\, dt \, . \tag{1.9}$$

In that case we have to express the operators in terms of frequency variables.

Once $M(\xi,\zeta)$ is obtained the distribution is calculated by Eq. (1.2). That is the basis of our method. We now face two issues. First we must find the scale operator and secondly we must have ways of evaluating quantities like Eq. (1.7). We address these two issues in the next two Sections.

We also point out that because $\mathcal{A}$ and $\mathcal{B}$ are operators there is an ordering problem. For example, one can take the characteristic function operator to be

$$\mathcal{M}(\xi,\zeta) = e^{j\xi\mathcal{A}/2}\, e^{j\zeta\mathcal{B}}\, e^{j\xi\mathcal{A}/2} \tag{1.10}$$

Since in general operators do not commute this will give a different answer than Eq. (1.7). For each different ordering a different characteristic function is obtained and hence a different distribution. This is the reason why we have an infinite number of distributions [6].

2. SCALE OPERATORS

The fundamental operators are the time and frequency operators, $\mathcal{T}$ and $\mathcal{W}$, which in the time and frequency representation are given by [7]

$$\mathcal{T} \to t \quad ; \quad \mathcal{W} \to -j\frac{d}{dt} \qquad \text{(time domain)} \tag{2.1}$$

$$\mathcal{T} \to \; j\frac{d}{d\omega} \quad ; \quad \mathcal{W} \to \omega \qquad \text{(frequency domain)} \, . \tag{2.2}$$

They satisfy the commutation relation

$$[\mathcal{T},\mathcal{W}] = \mathcal{T}\mathcal{W} - \mathcal{W}\mathcal{T} = j \, . \tag{2.3}$$

The frequency operator translates a time function, $f(t)$ according to

$$e^{j\tau\mathcal{W}} f(t) = f(t+\tau) \tag{2.4}$$

where τ is any number.

A. Time scale operator

We propose the following operator for time scale

$$\mathcal{S} = \tfrac{1}{2}\,(\mathcal{T}\mathcal{W} + \mathcal{W}\mathcal{T}) \, . \tag{2.5}$$

The reason for calling $\mathcal{S}$ a scale operator is because of the following action on time functions, $f(t)$,

$$e^{j\sigma\mathcal{S}} f(t) = e^{\sigma/2} f(e^{\sigma} t) \, , \tag{2.6}$$

where σ is any number parameter. That is, $e^{j\sigma\mathcal{S}}$ scales time functions in analogy with $e^{j\tau\mathcal{W}}$ which translates functions in time. The scaling factor is e^{σ}. Note that the $\mathcal{S}$ is Hermitian and that $e^{j\sigma\mathcal{S}}$ is unitary as is the case with the frequency operator.

We list some properties of the time scaling operator which we will subsequently use [8]. First we note that it can be written in the following alternate ways

$$\mathcal{S} = \mathcal{T}\mathcal{W} - \tfrac{1}{2}j = \mathcal{W}\mathcal{T} + \tfrac{1}{2}j \tag{2.7}$$

The commutation relation of the scale operator with the time and frequency operators are

$$[\mathcal{T},\mathcal{S}] = j\mathcal{T} \tag{2.8}$$

$$[\mathcal{W},\mathcal{S}] = -j\mathcal{W}\,. \tag{2.9}$$

In addition the following commutation relations will turn out to be important

$$[\mathcal{T},[\mathcal{T},\mathcal{S}]] = 0\,, \tag{2.10}$$

$$[\mathcal{W},[\mathcal{W},\mathcal{S}]] = 0\,, \tag{2.11}$$

$$[\mathcal{S},[\mathcal{T},\mathcal{S}] = \mathcal{T} \tag{2.12}$$

$$[\mathcal{S},[\mathcal{W},\mathcal{S}] = \mathcal{W}\,. \tag{2.13}$$

The time scale operator has the following effect on functions of frequency, $F(\omega)$,

$$e^{j\sigma\mathcal{S}}F(\omega) = e^{-\sigma/2}F(e^{-\sigma}\omega) \tag{2.14}$$

B. Frequency Scale Operator

Eq. (2.14) suggests that the frequency scale operator is appropriately defined by

$$\mathcal{S}_F = -\mathcal{S} \tag{2.15}$$

$$= -\tfrac{1}{2}(\mathcal{TW} + \mathcal{WT}) \tag{2.16}$$

$$= \tfrac{1}{2}\,j - \mathcal{TW} = -\mathcal{WT} - \tfrac{1}{2}j \tag{2.17}$$

Its operation on time and frequency functions are

$$e^{j\sigma\mathcal{S}_F}F(\omega) = e^{\sigma/2}F(e^{\sigma}\omega) \tag{2.18}$$

$$e^{j\sigma\mathcal{S}_F}f(t) = e^{-\sigma/2}f(e^{-\sigma}t)\,. \tag{2.19}$$

The time and frequency scale operators are simply related to each other and hence once relations are obtained for one of the operators it will be easy to transcribe to the other.

The relevant commutation relations for the frequency operator are

$$[\mathcal{T},\mathcal{S}_F] = -j\mathcal{T}\,, \tag{2.20}$$

$$[\mathcal{W},\mathcal{S}_F] = j\mathcal{W}\,, \tag{2.21}$$

$$[\mathcal{T},[\mathcal{T},\mathcal{S}_F]] = 0\,, \tag{2.22}$$

$$[\mathcal{W},[\mathcal{W},\mathcal{S}_F]] = 0\,, \tag{2.23}$$

$$[\mathcal{S}_F,[\mathcal{T},\mathcal{S}_F] = \mathcal{T}\,, \tag{2.24}$$

$$[\mathcal{S}_F,[\mathcal{W},\mathcal{S}_F] = \mathcal{W}\,. \tag{2.25}$$

3. METHOD OF EVALUATION

There are two general methods that have been devised to evaluate expressions like Eq. (1.7). The first is by direct evaluation as per the following prescription [16,8]

$$e^{j\xi\mathcal{A}+j\zeta\mathcal{B}}s(t) = \int G(t,t')\;s(t')\;dt' \tag{3.1}$$

where

$$G(t,t') = \int e^{j\lambda}\;u^*(\lambda,t')\;u(\lambda,t)\;d\lambda\,. \tag{3.2}$$

and where λ and u are the eigenvalue and eigenfunctions of

$$\{\;\xi\mathcal{A}+\zeta\mathcal{B}\;\}\;u(\lambda,t) = \lambda\;u(\lambda,t) \tag{3.3}$$

The second approach is simplification of the operator $e^{j\xi\mathcal{A}+j\zeta\mathcal{B}}$. This problem arises in many fields and has not been solved generally. However simplification is possible for special cases. The most common special case is when $\mathcal{A}$ and $\mathcal{B}$ commute with their commutator, $[\mathcal{A},\mathcal{B}]$. That is the case for time and frequency. However that will not be the case for the scale operator. However, as can be seen from Eq. (2.12-2.13) the scale operator and time operator satisfy a commutation relation of the form

$$[\mathcal{A},\mathcal{B}] = \beta + \alpha\mathcal{A} \tag{3.4}$$

For this case [8]

$$e^{j\xi A+j\zeta B} = e^{j\mu\xi B/\alpha}\, e^{j\xi\mu A}\, e^{j\zeta B}\, e^{j\xi A}\,. \qquad (3.5)$$

where

$$\mu = \frac{1}{j\zeta\alpha}\left\{1-(1+j\zeta\alpha)\, e^{-j\zeta\alpha}\right\} \qquad (3.6)$$

4. JOINT DISTRIBUTIONS WITH SCALE

The concept of a time-scale representation was first considered by Bertrand and Bertrand [1-3] in their pioneering work. They derived a particular distribution using group theory and generalized to the wide band ambiguity function. Subsequently Marinovic [13] and Altes [1] considered the issue from different perspectives and also derived a distribution. It is different from the one obtained by Bertrand and Bertrand. We shall show that these two distributions are fundamentally different. One deals with scaling as defined by the scaling operator and the other deals with inverse frequency. Inverse frequency will be discussed in Section (9).

We now derive a distribution of time and scale using Eq. (1.7) for the characteristic function operator. We shall use t and a to denote time and scale respectively and use θ and σ for the corresponding variables in the characteristic function. The characteristic function is

$$M(\theta,\sigma) = \int s^*(t)\, e^{j\theta T+j\sigma S_T} s(t)\, dt\,. \qquad (4.1)$$

Using Eq. (3.5) this becomes

$$M(\theta,\sigma) = \int s^*(t)\, e^{j\theta\mu T}\, e^{j\sigma S_T}\, e^{j\theta T} s(t)\, dt \qquad (4.2)$$

with

$$\mu = \frac{1}{\sigma}\left\{1-(1+\sigma)\, e^{-\sigma}\right\}. \qquad (4.3)$$

Straightforward algebra reduces this to

$$M(\theta,\sigma) = \int s^*(e^{\sigma/2}t)\, e^{2j\theta t\sinh(\sigma/2)/\sigma} s(e^{-\sigma/2}t)\, dt\,. \qquad (4.4)$$

To obtain the distribution we use

$$P(t,a) = \frac{1}{4\pi^2}\iint M(\theta,\sigma)\, e^{-j\theta t-j\sigma a}\, d\theta\, d\sigma\,. \qquad (4.5)$$

Substituting Eq. (4.4) for the characteristic function, we obtain, after some simplification, that

$$P(t,a) = \frac{1}{2\pi}\int \frac{\sigma}{2\sinh(\sigma/2)}\, e^{-j\sigma a}$$

$$s^*(e^{\sigma/2}\frac{\sigma t}{2\sinh(\sigma/2)})\, s(e^{-\sigma/2}\frac{\sigma t}{2\sinh(\sigma/2)})\, d\sigma\,. \qquad (4.6)$$

This is a joint representation for time and scale.

A. Marginals

The marginals of a joint distribution are the densities of the individual variables. To obtain the time marginal we integrate out scale

$$\int P(t,a)\, da = |s(t)|^2 \qquad (4.7)$$

This result is expected since $|s(t)|^2$ is the time density. To obtain the density of scale we integrate out time. The result is

$$\int P(t,a)\;\; dt = \left|\frac{1}{\sqrt{2\pi}}\int s(e^x)\, e^{x/2}\, e^{jax}\, dx\right|^2. \qquad (4.8)$$

This density for scale will be discussed further in Section (7).

B. Other Distributions

Suppose we use the following characteristic function operator,

$$\mathcal{M}(\theta,\sigma) = e^{j\sigma S_F/2}\, e^{j\theta T}\, e^{j\sigma S_F/2}\,. \qquad (4.9)$$

to obtain a distribution. The characteristic function is then

$$M(\theta,\sigma) = \int s^*(t)\, e^{j\sigma S_F/2}\, e^{j\theta T}\, e^{j\sigma S_F/2}\, s(t)\, dt \qquad (4.10)$$

and direct evaluation leads to

$$M(\theta,\sigma) = \int s^*(e^{\sigma/2}t)\, e^{j\theta t} s(e^{-\sigma/2}t)\, dt\,. \qquad (4.11)$$

The distribution is

$$P(t,a) = \frac{1}{4\pi^2}\iint M(\theta,\sigma)\, e^{-j\theta t - j\sigma a}\, d\theta\, d\sigma \qquad (4.12)$$

which gives

$$P(t,a) = \frac{1}{2\pi}\int s^*(e^{\sigma/2}t)\, e^{-j\sigma a}\, s(e^{-\sigma/2}t)\, d\sigma\,. \qquad (4.13)$$

This distribution was derived by Marinovic [13] and Altes [1]. It also satisfies the marginals, Eqs. (4.7-4.8).

How is it that we may obtain two different distributions? In fact, we will obtain an infinite number of distributions. The reason is that different orderings of the operators produce different distributions. However all these distributions are connected as we discuss in the next Section.

5. GENERAL CLASS WITH SCALE

As in the time frequency case we can obtain a general class of time scale distributions. One defines a generalized characteristic function by

$$M_G(\theta,\sigma) = \phi(\theta,\sigma)\, M(\theta,\sigma) \qquad (5.1)$$

where $M(\theta,\sigma)$ is any particular characteristic function and $\phi(\theta,\sigma)$ is the kernel. Choosing different kernels produces different distributions. The general class is [6,7,5]

$$P(t,a) = \frac{1}{4\pi^2}\iint M_G(\theta,\sigma)\, e^{-j\theta t - j\sigma a}\, d\theta\, d\sigma \qquad (5.2)$$

$$= \frac{1}{4\pi^2}\iint \phi(\theta,\sigma)\, M(\theta,\sigma)\, e^{-j\theta t - j\sigma a}\, d\theta\, d\sigma\,. \qquad (5.3)$$

Which particular characteristic function is chosen as the base is not important, but rather motivated by convenience. Suppose we chose the one given by Eq. (4.4).

$$P(t,a) = \frac{1}{4\pi^2}\iiint s^*(\, e^{\sigma/2}u)\phi(\theta,\sigma)$$

$$e^{-j\theta t - j\sigma a + 2j\theta u \sinh(\sigma/2)/\sigma} s(e^{-\sigma/2}u)\, d\theta\, dt\, d\sigma \qquad (5.4)$$

If we choose the one given by Eq. (4.11) then

$$P(t,a) = \frac{1}{4\pi^2}\iiint s^*(e^{\sigma/2}u)\, e^{-j\theta t - j\sigma a + j\theta u}$$

$$\phi(\theta,\sigma) s(e^{-\sigma/2}u)\, d\theta\, dt\, d\sigma \qquad (5.5)$$

Eq. (5.4) or (5.5) may be considered a general class and any other distribution is obtained by appropriate choice of the kernel.

6. INSTANTANEOUS SCALE

Since Eq. (1.4) allows us to calculate the average of any function we can use it to calculate the average scale,

$$\langle S_F \rangle = \int s^*(t)\, S_F\, s(t)\, dt \qquad (6.1)$$

If we express the signal in terms of the amplitude and phase

$$s(t) = A(t)\, e^{j\varphi(t)}\,, \qquad (6.2)$$

then†

$$\mathcal{S}_F\, s(t) = \left[\left(tA'/A + \tfrac{1}{2}\right) j - t\varphi'(t)\right] s(t) \qquad (6.3)$$

Substituting in Eq. (6.1) one straightforwardly obtains

$$\langle \mathcal{S}_F \rangle = -\int t\, \varphi'(t)\, A^2(t)\; dt \qquad (6.4)$$

For the scale bandwidth, B, we have

$$B^2 = \int s^*(t)\, (\mathcal{S}_F - \langle \mathcal{S}_F \rangle)^2\, s(t)\, dt \qquad (6.5)$$

$$= \int |\, (\mathcal{S}_F - \langle \mathcal{S}_F \rangle)\, s(t)|^2\, dt \qquad (6.6)$$

$$= \langle \mathcal{S}_F^2 \rangle - \langle \mathcal{S}_F \rangle^2 \qquad (6.7)$$

Direct calculation leads to

$$B^2 = \int \left(t\,\frac{A'}{A} + \frac{1}{2}\right)^2 A^2(t)\, dt$$

$$+ \int (\, t\, \varphi'(t) + \langle \mathcal{S}_F \rangle)^2\, A^2(t)\; dt \qquad (6.8)$$

This equation can be interpreted in the following way. Consider any two variables, x and y, which have a joint density. The standard deviation of y can be expressed in terms of the conditional standard deviation, $\sigma_{y|x}$, and the conditional average $\langle\, y\, \rangle_x$. The relation is [9,10]

$$\sigma_y^2 = \int \sigma_{y|x}^2\, P(x)\, dx + \int (\, \langle\, y\, \rangle_x - \langle\, y\, \rangle\,)^2\; P(x)\, dx \qquad (6.9)$$

where $P(x)$ is the density of x. The conditional value is what is commonly called an instantaneous or local value. Comparing Eq. (6.8) with Eq. (6.9) suggests that we take for instantaneous scale

$$a_t = -t\; \varphi'(t) \qquad (6.10)$$

The average scale is given by averaging the instantaneous scale over all time as per Eq. (6.4).

Also, the conditional spread of scale for a given time may be taken to be

$$\sigma_{a|t}^2 = \left(t\frac{A'}{A} + \frac{1}{2}\right)^2 \qquad (6.11)$$

Similar to the above considerations one can show that scale for a given frequency is given by

$$a_\omega = \omega\; \psi'(\omega)\,. \qquad (6.12)$$

where ψ is the phase of the spectrum.

7. EIGENFUNCTIONS AND EIGENVALUES OF SCALE

The significance of solving the eigenvalue problem for an operator is that it gives a means of obtaining the density for that quantity. Also the eigenvalues give the range of the possible values. For scale, the eigenvalue problem is

$$\mathcal{S}_F\; \eta(a,t) = a\; \eta(a,t) \qquad (7.1)$$

Since the scale operator is Hermitian the eigenvalues will be real and eigenfunctions complete. Therefore any function $s(t)$ can be expanded in terms of the eigenfunctions

$$s(t) = \int F(a)\; \eta(a,t) da \qquad (7.2)$$

with the inverse transformation

$$F(a) = \int s(t)\; \eta^*(a,t)\; dt \qquad (7.3)$$

† Primes denote differentiation with respect to the argument.

The density of scale is then

$$P(a) = |F(a)|^2 \tag{7.4}$$

In the time representation the scale eigenvalue problem becomes

$$j\left(\tfrac{1}{2} + t\frac{d}{dt}\right)\eta(a,t) = a\,\eta(a,t). \tag{7.5}$$

and the solution is

$$\eta(a,t) = \frac{1}{\sqrt{2\pi}}\frac{e^{-ja\log t}}{\sqrt{t}}. \tag{7.6}$$

These satisfy the completeness relations

$$\int_0^\infty \eta^*(a',t)\,\eta(a,t)\,dt = \delta(a-a') \tag{7.7}$$

$$\int \eta^*(a,t')\,\eta(a,t)\,da = \delta(t-t'). \tag{7.8}$$

From Eq. (7.3) we have that for any signal

$$F(a) = \int s(t)\,\eta^*(a,t)\,dt \tag{7.9}$$

$$= \frac{1}{\sqrt{2\pi}}\int s(t)\,\frac{e^{ja\log t}}{\sqrt{t}}\,dt. \tag{7.10}$$

The density of scale is then

$$P(a) = |F(a)|^2 \tag{7.11}$$

$$= \left|\frac{1}{\sqrt{2\pi}}\int s(t)\,\frac{e^{ja\log t}}{\sqrt{t}}\,dt\right|^2 \tag{7.12}$$

$$= \left|\frac{1}{\sqrt{2\pi}}\int s(e^x)\,e^{x/2}\,e^{jax}\,dx\right|^2 \tag{7.13}$$

which is the same as Eq. (4.8) which was obtained as one of the marginals of the joint distribution of time and scale.

8. UNCERTAINTY PRINCIPLE FOR SCALE

For arbitrary operators the uncertainty principle is

$$\Delta\mathcal{A}\,\Delta\mathcal{B} \geq \tfrac{1}{2}|\langle[\mathcal{A},\mathcal{B}]\rangle|, \tag{8.1}$$

where $(\Delta\mathcal{A})^2$ is the standard deviation of $\mathcal{A}$ given by

$$(\Delta\mathcal{A})^2 = \int s^*(t)\,(\mathcal{A} - \langle\mathcal{A}\rangle)^2\,s(t)\;dt \tag{8.2}$$

$$= \int |(\mathcal{A} - \langle\mathcal{A}\rangle)\,s(t)|^2\,dt \tag{8.3}$$

$$= \langle\mathcal{A}^2\rangle - \langle\mathcal{A}\rangle^2. \tag{8.4}$$

Similarly for $\mathcal{B}$.

For the case of time and scale we have

$$\Delta\mathcal{T}\,\Delta\mathcal{S}_F \geq \tfrac{1}{2}|\langle[\mathcal{T},\mathcal{S}_F]\rangle| \tag{8.5}$$

Using Eq. (2.8) this gives

$$\Delta\mathcal{T}\,\Delta\mathcal{S}_F \geq \tfrac{1}{2}|\langle t\rangle| \tag{8.6}$$

where $\langle t\rangle$ is the mean time.

One can obtain the signal which minimizes the uncertainty product. In general one solves the following equation

$$(\mathcal{B} - \langle\mathcal{B}\rangle)\,s(t) = \lambda(\mathcal{A} - \langle\mathcal{A}\rangle)\,s(t) \tag{8.7}$$

where

$$\lambda = \frac{\langle[\mathcal{A},\mathcal{B}]\rangle}{2(\Delta\mathcal{A})^2}. \tag{8.8}$$

Specializing to time and scale

$$\left(\frac{j}{2} + jt\frac{d}{dt} - \langle\mathcal{S}_F\rangle\right)s(t) = \lambda(t - \langle t\rangle)\,s(t) \tag{8.9}$$

where

$$\lambda = \frac{\langle [\mathcal{T}, \mathcal{S}_F] \rangle}{2(\Delta t)^2} = -j \frac{\langle t \rangle}{2(\Delta t)^2} \tag{8.10}$$

and where $\langle t \rangle$ and Δt are the mean time and duration respectively. The quantity $\langle \mathcal{S}_F \rangle$ is the mean scale given by Eq. (6.4). The solution of Eq. (8.9) is

$$s(t) = c\, t^{\alpha_2}\, e^{-\alpha_1 t - j \langle \mathcal{S}_F \rangle \ln(t/\langle t \rangle)} \tag{8.11}$$

where c is a normalizing constant and

$$\alpha_1 = \frac{\langle t \rangle}{2(\Delta t)^2} \tag{8.12}$$

$$\alpha_2 = \frac{1}{2} \left(\frac{\langle t \rangle^2}{(\Delta t)^2} - 1 \right) . \tag{8.13}$$

We now address what minimizes frequency and frequency scale uncertainty. The equation we have to solve is

$$\left(\frac{j}{2} + jt \frac{d}{dt} - \langle \mathcal{S}_F \rangle \right) s(t) = \lambda \left(-j \frac{d}{dt} - \langle \omega \rangle \right) s(t) \tag{8.14}$$

with

$$\lambda = \frac{\langle [\mathcal{W}, \mathcal{S}_F] \rangle}{2(\Delta \omega)^2} = j \frac{\langle \omega \rangle}{2(\Delta \omega)^2} \tag{8.15}$$

The solution is

$$s(t) = (t + \lambda)^{-j \langle \mathcal{S}_F \rangle - 1/2 + \lambda \langle \omega \rangle} . \tag{8.16}$$

9. INVERSE FREQUENCY

We define the inverse frequency operator by

$$\mathcal{R} = \frac{\omega_0}{\mathcal{W}} \tag{9.1}$$

where ω_0 is some reference frequency. We shall use r to signify inverse frequency values to keep a distinction between inverse frequency and scale, where we have used a. We now give some properties of the operator $\mathcal{R}$.

The commutator of the inverse frequency operator with time is

$$[\mathcal{T}, \mathcal{R}] = -\frac{j}{\omega_0} \mathcal{R}^2 \tag{9.2}$$

and the uncertainty principle for time and inverse frequency is therefore

$$\Delta \mathcal{T}\, \Delta \mathcal{R} \geq \tfrac{1}{2} |\langle [\mathcal{T}, \omega_0 \mathcal{W}^{-1}] \rangle| = \frac{\omega_0}{2} \left\langle \frac{1}{\omega^2} \right\rangle \tag{9.3}$$

Also the minimum uncertainty signal is easily derived. We do it in the frequency representation. Using Eq. (8.7) we have that the minimum time-inverse frequency signal, $F(\omega)$, must satisfy

$$\left(j \frac{d}{d\omega} - \langle t \rangle \right) F(\omega) = \lambda \left(\frac{\omega_0}{\omega} - \langle \mathcal{R} \rangle \right) F(\omega) \tag{9.4}$$

where

$$\lambda = \frac{\langle [\mathcal{R}, \mathcal{T}] \rangle}{2(\Delta \mathcal{R})^2} = j \frac{\langle \mathcal{R}^2 \rangle}{2(\Delta \mathcal{R})^2} \tag{9.5}$$

The solution is

$$F(\omega) = c \omega^{\alpha_1} e^{-\alpha_2 \omega - j \langle t \rangle \omega} \tag{9.6}$$

with

$$\alpha_1 = \frac{\mathcal{R}^2}{(2\Delta \mathcal{R})^2} \quad ; \quad \alpha_2 = \frac{\langle R \rangle}{\omega_0}\, \alpha_1 . \tag{9.7}$$

The eigenfunctions of $\mathcal{R}$ in the frequency representation are obtained from

$$\frac{\omega_0}{\omega}\, u(r, \omega) = r\, u(r, \omega) \tag{9.8}$$

which gives

$$u(\tau,\omega) = \frac{\sqrt{\omega_0}}{\tau}\,\delta(\omega-\omega_0/\tau)\,. \qquad (9.9)$$

In the time representation they are

$$u(\tau,t) = \sqrt{\frac{\omega_0}{2\pi}}\,\frac{1}{\tau}\,e^{j\omega_0 t/\tau}\,. \qquad (9.10)$$

We note that in the time representation the inverse frequency operator is an integral operator,

$$\mathcal{R} = \omega_0 \int^t . \qquad (9.11)$$

10. DISTRIBUTION OF TIME AND INVERSE FREQUENCY

We now obtain joint distributions of time and inverse frequency by two different methods. The first method is a direct application of the general approach presented in the Introduction. However, since inverse frequency is a function of frequency, rather than an operator, there is a simpler method for this special case. This is presented as method two.

A. Method 1

We use the method presented in the Introduction with the two operators being $\mathcal{T}$ and $\mathcal{R}$. The characteristic function is

$$M(\tau,\sigma) = \int S^*(\omega)\,e^{j\theta\mathcal{T}+j\sigma\mathcal{R}}S(\omega)\,d\omega\,. \qquad (10.1)$$

One can show that

$$e^{j\theta\mathcal{T}+j\sigma\mathcal{R}}S(\omega) = \int S(\omega-\theta)\exp[j\frac{\omega_0\sigma}{\theta}\ln\frac{\omega}{\omega-\theta}]\,d\omega \qquad (10.2)$$

Therefore

$$M(\tau,\sigma) = \int S^*(\omega)S(\omega-\theta)\exp[j\frac{\omega_0\sigma}{\theta}\ln\frac{\omega}{\omega-\theta}]\,d\omega \qquad (10.3)$$

$$= \int S^*(\omega+\tfrac{1}{2}\theta)S(\omega-\tfrac{1}{2}\theta)\exp[j\frac{\omega_0\sigma}{\theta}\ln\frac{\omega+\frac{1}{2}\theta}{\omega-\frac{1}{2}\theta}]\,d\omega\,. \qquad (10.4)$$

The distribution is

$$P(t,\tau) = \frac{1}{4\pi^2}\iiint S^*(\omega+\tfrac{1}{2}\theta)S(\omega-\tfrac{1}{2}\theta)$$

$$\exp[j\frac{\omega_0\sigma}{\theta}\ln\frac{\omega+\frac{1}{2}\theta}{\omega-\frac{1}{2}\theta}]\,e^{-j\theta t-j\sigma a}d\theta\,d\sigma\,d\omega \qquad (10.5)$$

which after considerable simplification gives

$$P(t,\tau) = \frac{1}{8\pi\tau^3}\int\left(\frac{\omega_0 u}{\sinh u/2}\right)^2 e^{-j\omega_0 ut/\tau}$$

$$S^*(\frac{\omega_0 u}{2\tau}\frac{e^{u/2}}{\sinh u/2})\,S(\frac{\omega_0 u}{2\tau}\frac{e^{-u/2}}{\sinh u/2})\,du\,. \qquad (10.6)$$

This distribution was obtained by Bertrand and Bertrand by different methods [2,3,4].

B. Method 2

Since inverse frequency is functionally related to frequency we can use the standard methods for transformation of variables

$$P(t,\tau)\,dt\,d\tau = P(t,\omega)\,dt\,d\omega\,. \qquad (10.7)$$

where $P(t,\omega)$ is a time frequency distribution. In particular we take

$$\tau = \omega_0/\omega\,, \qquad (10.8)$$

$$d\tau = -\omega_0/\omega^2\,d\omega \qquad (10.9)$$

$$= -\frac{\tau^2}{\omega_0}\,d\omega \qquad (10.10)$$

and therefore

$$P(t,\tau) = \frac{\omega_0}{\tau^2}\,P(t,\omega_0/\tau)\,. \qquad (10.11)$$

What distribution should we use for the time frequency distribution? We can use any one, since they are all related by way of the kernel. If we use the Wigner distribution

$$P(t,\omega) = \frac{1}{2\pi}\int s^*(t-\tfrac{1}{2}\tau)\ \epsilon^{-j\tau\omega}\ s(t+\tfrac{1}{2}\tau)d\tau \tag{10.12}$$

we obtain

$$P(t,\tau) = \frac{\omega_0}{\tau^2}\ W(t,\omega_0/\tau) \tag{10.13}$$

$$= \frac{1}{2\pi}\frac{\omega_0}{\tau^2}\int s^*(t-\tfrac{1}{2}\tau)\ \epsilon^{-j\theta t/\tau}\ s(t+\tfrac{1}{2}\tau)d\tau \tag{10.14}$$

$$= \frac{1}{2\pi}\frac{\omega_0}{\tau^2}\int S^*(\omega_0/\tau-\theta/2)\ e^{-jt\theta}\ S(\omega_0/\tau-\theta/2)d\theta\ . \tag{10.15}$$

Any other time frequency distribution can be used. In fact the general class of distributions can be used to obtain the general class of time and inverse frequency distributions. The general class of time frequency distributions is [5,7]

$$P(t,\omega) = \frac{1}{4\pi^2}\iiint e^{-j\theta t-j\tau\omega+j\theta u}\phi(\theta,\tau)$$

$$s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,du\,d\tau\,d\theta\ . \tag{10.16}$$

Using Eq. (10.11) we have a general class for time and inverse frequency

$$P(t,\tau) = \frac{1}{4\pi^2}\frac{\omega_0}{\tau^2}\iiint \epsilon^{-j\theta t-j\tau\omega_0/\tau+j\theta u}\phi(\theta,\tau)$$

$$s^*(u-\tfrac{1}{2}\tau)\,s(u+\tfrac{1}{2}\tau)\,du\,d\tau\,d\theta\ . \tag{10.17}$$

In terms of the spectrum this is

$$P(t,\omega) = \frac{1}{4\pi^2}\frac{\omega_0}{\tau^2}\iiint\ \epsilon^{-j\theta t-j\tau\omega_0/\tau+j\tau u}\phi(\theta,\tau)$$

$$S^*(u+\tfrac{1}{2}\theta)S(u-\tfrac{1}{2}\theta)\ d\theta d\tau du\ . \tag{10.18}$$

A. Marginals

The marginals are easily obtained

$$\int P(t,\tau)\ d\tau = |s(t)|^2 \tag{10.19}$$

$$\int P(t,\tau)\ dt = \frac{\omega_0}{\tau^2}\,|S(\omega_0/\tau)|^2 \tag{10.20}$$

As before the time marginal is as expected, $|s(t)|^2$. The inverse frequency marginal can be understood from the following considerations. The distribution of frequency is

$$P(\omega) = |S(\omega)|^2\,, \tag{10.21}$$

hence the distribution of τ is

$$P(\tau) = P(\omega)\left|\frac{d\omega}{d\tau}\right|_{\omega=\omega_0/\tau} \tag{10.22}$$

$$= \frac{\omega_0}{\tau^2}\,|S(\omega_0/\tau)|^2\ . \tag{10.23}$$

which agrees with Eq. (10.20).

11. THE WAVELET TRANSFORM

The modulus squared of wavelet transform can be viewed as a joint representation of time and "scale". The wavelet transform is

$$WT(t,\tau) = \frac{1}{\sqrt{|\tau|}}\int s(u)\psi^*(\frac{u-t}{\tau})\ du \tag{11.1}$$

where we have used τ instead of the traditional a. The reason we have done so is that it will turn out that in the wavelet transform it is inverse frequency that comes in. The density of time and τ is

$$P_{WT}(t,\omega) = |WT(t,\tau)|^2\ . \tag{11.2}$$

We now address the relation of this distribution to the type of distributions discussed in the previous

sections. The first to consider a possible relationship with time frequency distributions were Jeong and Williams [11,12] and Rioul [14] using different approaches. Jeong and Williams [11,12] showed in a simple and direct way that the distribution given by Eq. (11.2) can be obtained from the general class of time frequency distributions by taking an appropriate kernel and evaluating the distribution at inverse frequency. Rioul [14] showed that it can be expressed as a convolution of two Wigner distributions. Posch [17] then generalized this to show P_{WT} can be expressed as a convolution of any two time frequency distributions. Rioul and Flandrin [15] used these relationships to define a general class of time and scale. We emphasize that "scale" as used in these works simply means inverse frequency. The distribution given by Eq. (11.2) is a member of the general class given by Eq. (10.17). To explicitly obtain the kernel one expands Eq. (11.2) and puts it in the form given by Eq. (10.17). The kernel is then

$$\phi(\theta,\tau) = 2\pi \frac{\tau^2}{\omega_0} e^{j\tau\omega_0/\tau} \phi_S(\tau\theta, \tau/\tau) \qquad (11.3)$$

where ϕ_S is the kernel that produces a spectrogram with window ψ,

$$\phi_S(\theta,\tau) = \int e^{-j\theta u} \psi^*(u+\tfrac{1}{2}\tau)\psi(u-\tfrac{1}{2}\tau)\,du \quad (11.4)$$

This is essentially the result obtained by Jeong and Williams. We note that it is not only a function of θ and τ but also depends on τ. Therefore we see that the distribution obtained from the wavelet transform is really a distribution of time and inverse frequency.

12. CONCLUSION

We have shown that scale as defined by the operator S and by the inverse frequency operator $\mathcal{R}$ both lead to a form of scaling and that these two different ideas have both been used in the literature for the concept of scale. It is important to keep them separate because although they do have many similarities they are quite different. They have very different densities, the density of scale being given by Eq. (4.8) and the density of inverse frequency by Eq. (10.20).

Also, a way to see that they are very different is to consider their commutator. The commutator of S_F and $\mathcal{R}$ is

$$[S_F, \mathcal{R}] = -j\mathcal{R} \qquad (12.1)$$

That is, scale and inverse frequency do not commute. That is an indication that they are quite different in their physical values. Which of the two mathematical constructs is better suited to describe our intuitive notion of "scale" remains to be seen. Perhaps both. Perhaps neither.

ACKNOWLEDGEMENT

I would like to thank Drs. Fineberg, Flanagan, and Mammone for their kind hospitality during my sabbatical at the CAIP center. Support from a grant awarded to CAIP by Rome Laboratory of the U. S. Air Force (F306 02-91-C 0120) is gratefully acknowledged.

REFERENCES

[1] R. A. Altes, "Wide-Band, Proportional-Bandwidth Wigner-Ville Analysis, *IEEE Trans. ASSP* , Vol. 38, pp. 1005-1012, 1990.

[2] J. Bertrand and P. Bertrand "Representation temps-frequence des signoux" C. R. Acad. Sc. vol. 299, serie 1 pp. 635-638 (1984).

[3] P. Bertrand and J. Bertrand "Time-Frequency Representations of Broad Band Signals" in *The Physics of Phase Space*, edited by Y.S. Kim and W. Zachary, Springer Verlag, pp. 250-252, 1987.

[4] J. Bertrand and P. Bertrand "Affine Time - Frequency Distributions", in: *Time Frequency Signal Analysis - Methods and Applications*, B. Boashash, editor, Longman-Cheshire, to appear.

[5] B. Boashash, "Time Frequency Signal Analysis", in: *Advances in Spectral Analysis*, S. Haykin, Editor, Prentice Hall, 1990.

[6] L. Cohen, "Generalized Phase Space Distribution Functions", *Jour. Math. Physics*, vol. 7, pp. 781-786, 1966.

[7] L. Cohen, "Time-Frequency Distributions - A Review", *Proc. of the IEEE*, 77, pp. 941-981, 1989.

[8] L. Cohen, "A General Approach For Obtaining Joint Representations in Signal Analysis And An Application to Scale", *Advanced Signal-Processing Algorithms, Architectures and Implementation II*, Franklin T. Luk, Editor, Proc. SPIE 1566, p. 109-133 (1991).

[9] L. Cohen and C. Lee , "Instantaneous Frequency, Its Standard Deviation and Multicomponent Signals", in: *Advanced Algorithms and Architectures for Signal Processing III*, Franklin T. Luk, Editor, Proc. SPIE 975, pp. 186-208, 1988.

[10] L. Cohen and C. Lee , "Instantaneous Bandwidth" in: *Time Frequency Signal Analysis - Methods and Applications*, B. Boashash, editor, Longman Cheshire, to appear.

[11] J. Jeong and W. J. Williams, "Variable Windowed Spectrograms: Connecting Cohen's Class and the Wavelet Transform", *ASSP Workshop on Spectrum Estimation and Modeling*, p. 270-273, 1990.

[12] J. Jeong "Time Frequency Signal Analysis and Synthesis Algorithms", Thesis, University of Michigan 1990.

[13] N. M. Marinovic, "The Wigner distribution and the Ambiguity Function: Generalizations, Enhancement, Compression and Some Applications", Ph. D. Thesis, The City university of New York, 1986. See also: G. Eichmann and N. M. Marinovic, "Scale Invariant Wigner Distribution and Ambiguity Functions" SPIE vol 519, pp. 18-24, 1984.

[14] O. Rioul, "Wigner-Ville Representations of Signal Adapted to Shifts and Dilations", *AT&T Bell Laboratories Tech. Memo.*,, 1988.

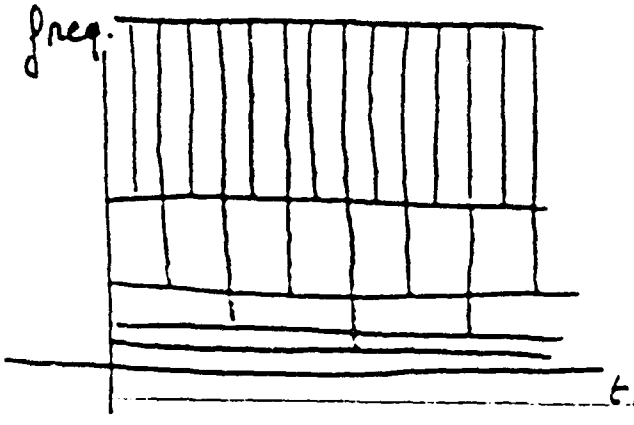
[15] O. Rioul and P. Flandrin "Time-Scale Energy Distributions: A General Class Extending Wavelet Transforms" to appear in *IEEE Trans. Acoust., Speech, Signal Processing*,

[16] M. Scully and L. Cohen, "Quasi-Probability Distributions for Arbitrary Operators" in *The Physics of Phase Space*, edited by Y.S. Kim and W.W. Zachary, Springer Verlag, 1987.

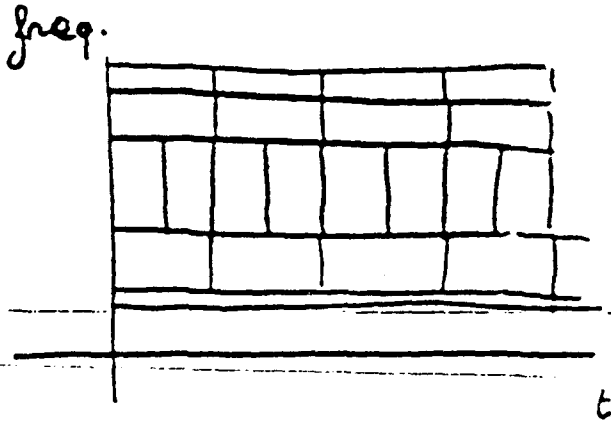
[17] T. E. Posch, "Wavelet Transform and Time-Frequency Distributions",*Advanced Algorithms and Architectures for Signal Processing IV*, Franklin T. Luk, Editor, Proc. SPIE 1152, pp. 477-482, 1989.

INGRID DAUBECHIES

Tiling time-frequency — in an adaptive way.

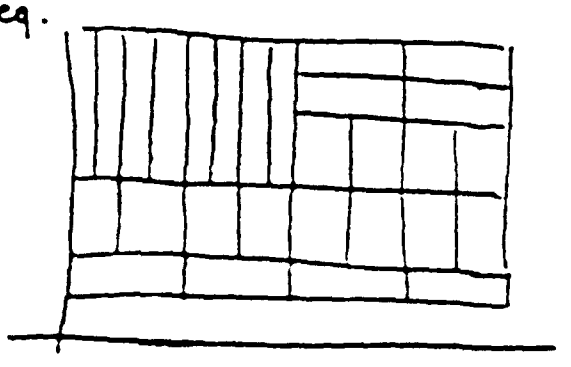


wavelets
or
subband filtering
(cascade, 2↓,
only low freq
gets split)

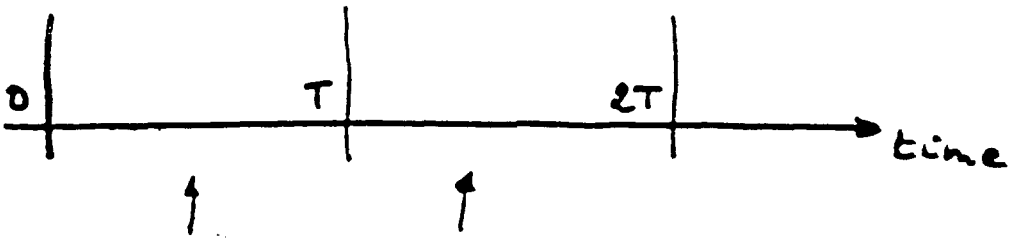


wavelet packets
or
subband filtering
with different tree
structure.

adaptive in time?



Can do this using wavelets on the interval.



on each interval : basis adapted to interval.

good description of smoothness

➡ no "penalty" (in large fine scale coeffs) for cutting.

However...

if you quantize → errors

in reconstruction: errors lead to difference with original

↳ combination of basic building

↓

at edges: abrupt cut-off.

⇒ bad.

→ Scheme with only "interior" scaling functions at left and all "straddling" scaling functions at right also has right number of scaling functions

Wavelets?

at left: all interior wavelets
+ N-1 adapted wavelets

at right: interior
+ N-1 adapted wvlts.

→ same "right" number.

Fewer wavelets at right than scaling functions?

Meyer: only straddling ψ with more than half their support within interval count.

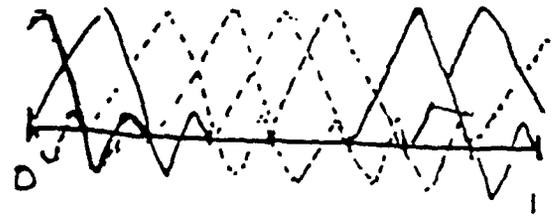
Need scheme with overlaps. (in time)

Ideally: larger overlaps for low freq.
than for high freq.

Construction of wavelets on interval by
A. Cohen, P. Vial & I.D. : convenient
for wavelet packets because

low freq. funct. = # high freq. funct.
$= 2^j$ at scale j.

Other way of achieving this.

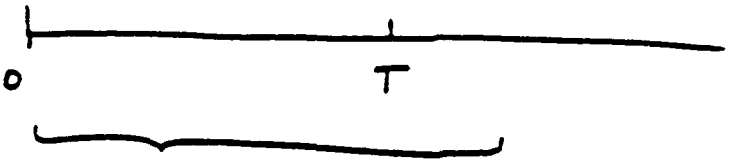


5 interior functions
2 straddle at left
2 straddle at right.

→ 8 function

↑
right number!

Scheme:



slightly larger than $[0, T]$

↳ find "best" basis

(best tree in subband filtering branching)

This sets the choice of tree to use on $[0, T]$.

Now: use this tree, with "adapted" filters at left; standard filters at right.

Except for one change:

at right, when using m_1 (or G) filter, "reallocate" the last $N-1$ taps to results from m_0.

Result:

decomposition into ∞ wavelet packets, many of which straddle right edge.

↳ recompose: original function reproduced between 0 and T

↳ + natural "tails" into $[T, 2T]$.

Now: on $[T, 2T]$ compute

f – tail from $[0, T] \longrightarrow f_1$.

$f_1(t)$ starts at $t = T$ with value 0, and many derivatives zero.

⇒ well adapted to treatment (at this end) with wavelet scheme with only interior functions.

↳ continue:
- find best tree on $[T, 2T]$
- decompose + reallocate ate
- compute tail (simple recursion.
- $f_2 = f|_{[27.377}$ – tail of f_1

End result:

- scheme with small overlap for high freq., larger overlap for low freq.
- quantization errors do not introduce discontinuities
- different tiling possible in every interval.

Decomposition into non orthonormal basis:

Still Riesz basis if finite number of splittings; condition number $\sim 2^J$ (J = number of splittings, or depth of tree)

1

On the Time-Scale Analysis of Self-Similar Processes

Patrick Flandrin

Ecole Normale Supérieure de Lyon
Laboratoire de Physique (URA 1325 CNRS)
46 allée d'Italie 69364 Lyon Cedex 07 France
e-mail: flandrin@ens-lyon.fr

Abstract

By construction, time-scale methods are well-suited for analyzing processes (such as $1/f$ processes) which are characterized simultaneously by nonstationarities in time and self-similarity properties in scale. In this respect, wavelets offer a natural possibility of time-scale analysis which is particularly powerful in the case of fractional Brownian motion (fBm): for instance, it is shown that mild conditions can ensure orthonormal wavelets to provide almost Karhunen-Loève expansions of fBm from which the scaling exponent can be estimated. Interesting scaling properties hold for continuous wavelet transforms too, especially when generalizing fBm to the case of locally self-similar processes. However, it is argued that, in the continuous case, a more general class of time-scale distributions (from which the wavelet transform is only a special case) can be used with possibly increased performance, e.g. for estimating local scaling exponents.

1 Introduction

In a large number of physical phenomena (e.g. in turbulence), self-similar processes with a $1/f$-type spectral behavior over wide ranges of frequencies are observed. Although of great importance, the study of processes of this kind is faced with a number of difficulties (such as the slow decay of the correlation structure) which call for specific models and adapted analysis tools.

One such modeling has been proposed in [21] and is referred to as *fractional Brownian motion* (fBm). Among other properties, it possesses that of being statistically *self-similar*, which means that any portion of a given fBm can be viewed (from a statistical point of view) as a scaled version of a larger part of the same process. This is of course in the spirit of *wavelets* [5], which can all be deduced from one elementary waveform by means of shifts and dilations.

2

On the other hand, recent works [13] [26] have emphasized the fact that wavelets are only a special case within a more general class of time-scale distributions, some of them with possibly better properties.

It is therefore the purpose of this paper to summarize a number of results concerning the usefulness of wavelets for analyzing (possibly modified) fBm [11] [12], as well as to present the more general framework into wavelets fit, suggesting hence companion ways of time-scale analysis for self-similar and $1/f$-type processes.

2 Self-similar and $1/f$ processes

2.1 Fractional Brownian motion

Fractional Brownian motion (fBm) is a natural extension of ordinary Brownian motion [21]. It is a Gaussian zero-mean nonstationary stochastic process $B_H(t)$, indexed by a single scalar parameter $0 < H < 1$, the usual Brownian motion being recovered from the specification $H = 1/2$.

It is a nonstationary process since [30]

$$\mathcal{E}(B_H(t)B_H(s)) = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$$

$$\Downarrow \tag{2.1}$$

$$\mathrm{var}(B_H(t)) = \sigma^2|t|^{2H},$$

where $\mathcal{E}$ stands for the expectation operator. As a nonstationary process, fBm only admits an *average* spectrum [30][8]

$$S_{B_H}(\omega) = \frac{\sigma^2}{|\omega|^{2H+1}} \tag{2.2}$$

which makes it well-suited for modeling $1/f$-type processes.

Increments of fBm are *stationary* and *self-similar* in the sense that the probability properties of the process $B_H(t+s) - B_H(t)$ only depend on the lag variable s with, moreover

$$(B_H(t+as) - B_H(t)) \stackrel{\mathrm{d}}{=} |a|^H (B_H(t+s) - B_H(t)), \tag{2.3}$$

where $\stackrel{\mathrm{d}}{=}$ means equality in distribution.

One of the key problems related to the analysis of fBm is the estimation of the self-similarity parameter H, which can be interpreted as a *scaling exponent* governing the fluctuations of the increment process.

2.2 Locally self-similar processes

It is clear that, in many situations, fBm can appear as a too restrictive model and more or less *ad hoc* modifications are required. One such modification is to drop the assumption of *global* self-similarity (characterized by only *one* value of H governing identical scaling properties at *all* scales) and to replace it by a milder requirement of *local* self-similarity. In such a case, only *small scale* behavior is concerned and the scaling exponent H is allowed to be *time-dependent*. This corresponds to processes $x(t)$ such that the local fluctuations of their increments satisfy

$$\mathcal{E}\left([x(t+\tau)-x(t)]^2\right) \sim |\tau|^{2H(t)} \quad , \quad \tau \to 0. \tag{2.4}$$

The $1/f$ spectral behavior of such processes is mainly dependent on the average value of $H(t)$ over time but the "spectrum" of all the possible values of $H(t)$ provides an additional information on a possible *multifractal* mechanism underlying the observed process, which is of considerable importance [14] [22].

3 Wavelets and self-similar processes

Two important features are to be taken into account when analyzing fBm or locally self-similar processes: *nonstationarity* and *self-similarity*. This suggests to look for some analysis which would be *time-dependent* and *scale-dependent*, respectively. As a result, wavelet analysis [5] which, by nature, is a time-scale method, appears as a natural possibility. First attempts for analyzing or synthesizing fBm *via* wavelets can be found in [8] [19] or [29]; most of the results given below provide a brief account of a more complete study reported in [12].

3.1 Orthonormal wavelets

Let us first consider some discrete orthonormal wavelet decomposition of a given fBm $B_H(t)$. By definition, wavelet coefficients are given by [19][23]

$$d_j[n] = 2^{-j/2} \int_{-\infty}^{+\infty} B_H(t)\psi(2^{-j}t-n)dt, \quad j \in \mathbb{Z}, n \in \mathbb{Z}, \tag{3.1}$$

where $\psi(t)$ is the basic "mother" wavelet, required to satisfy the admissibility condition [17]

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0. \tag{3.2}$$

The simplest family of orthonormal wavelets is the Haar system

4

$$\psi(t) = \begin{cases} +1 & 0 \le t < 1/2 \\ -1 & 1/2 \le t < 1. \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

For any given resolution 2^J, the wavelet mean-square representation of fBm is

$$B_H(t) = 2^{-J/2} \sum_{n=-\infty}^{+\infty} a_J[n]\phi(2^{-J}t-n) + \sum_{j=-\infty}^{J} 2^{-j/2} \sum_{n=-\infty}^{+\infty} d_j[n]\psi(2^{-j}t-n), \tag{3.4}$$

where equality is to be understood in a mean-square sense and where the *approximation* coefficients

$$a_j[n] = 2^{-j/2} \int_{-\infty}^{+\infty} B_H(t)\phi(2^{-j}t - n)dt \tag{3.5}$$

are computed with the help of the "scaling function" (or "father wavelet") $\phi(t)$ associated with $\psi(t)$ [19] [23]. In this picture, the wavelet coefficients are interpreted as *details*, i.e. as a measure of difference in information between two successive approximations.

For each scale 2^j, the wavelet coefficients $d_j[n]$ form a discrete sequence of random coefficients but, although the family $\{2^{-j/2}\psi(2^{-j}t - n), j \in \mathbb{Z}, n \in \mathbb{Z}\}$ is an orthonormal system, there is *a priori* no reason for them to be uncorrelated. The explicit correlation structure of the wavelet coefficients can be derived [12], one of its consequence being that, when normalized according to $\tilde{d}_j[n] = (2^j)^{-(H+1/2)}d_j[n]$, wavelet coefficients of fBm give rise to

- time sequences which are self-similar and stationary in the sense that, for any j, $\mathcal{E}(\tilde{d}_j[n]\tilde{d}_j[m])$ is a unique function of $n - m$, namely

$$\mathcal{E}(\tilde{d}_j[n]\tilde{d}_j[m]) = \frac{\sigma^2}{2}\left(-\int_{-\infty}^{+\infty} \gamma_\psi(\tau - (n-m))|\tau|^{2H}d\tau\right), \tag{3.6}$$

  with $\gamma_\psi(\tau) = \int_{-\infty}^{+\infty} \psi(t)\psi(t-\tau)d\tau$;

- scale sequences which are stationary in the sense that, for any n and $m = 2^{j-k}n$ associated with synchronous time instants, $\mathcal{E}(\tilde{d}_j[n]\tilde{d}_k[m])$ is a unique function of $j - k$, namely

$$\mathcal{E}(\tilde{d}_j[n]\tilde{d}_k[2^{j-k}n]) = \frac{\sigma^2}{2}\left(-\int_{-\infty}^{+\infty} A_\psi(2^{j-k},\tau)|\tau|^{2H}d\tau\right)(2^{j-k})^{-(H+1/2)}, \tag{3.7}$$

  with $A_\psi(\alpha,\tau) = \sqrt{\alpha}\int_{-\infty}^{+\infty} \psi(t)\psi(\alpha t - \tau)dt$.

(Both results have first been established in the case of continuous wavelet transforms, the former in [8] and the latter in [28].)

One key feature revealed by this structure is the stationarity of detail sequences at any resolution, nonstationarity of fBm being in fact encoded in the approximation sequences for which we get the time-dependent variance

$$\mathrm{var}(a_j[n]) = \frac{\sigma^2}{2}\left(-\int_{-\infty}^{+\infty}(\gamma_\phi(\tau) - 2\phi(\tau - n))|\tau|^{2H}d\tau\right)(2^j)^{2H+1}. \quad (3.8)$$

Variance of wavelet coefficients follows the power-law

$$\mathrm{var}(d_j[n]) = \frac{\sigma^2}{2}V_\psi(H)(2^j)^{2H+1}, \quad (3.9)$$

where $V_\psi(H)$ is defined by

$$V_\psi(H) = -\int_{-\infty}^{+\infty}\gamma_\psi(\tau)|\tau|^{2H}d\tau. \quad (3.10)$$

Therefore, the fBm index H can be easily obtained from the slope of this variance plotted as a function of scale in a log-log plot [19] [15]

$$\log_2(\mathrm{var}(d_j[n])) = (2H+1)j + \text{constant}. \quad (3.11)$$

An example is given in Figure 1.

In general, wavelet coefficients of fBm are correlated in both time and scale and more can be said about this correlation [27] [12]. The ideal situation would correspond to special cases for which perfect decorrelation could be achieved. In such a case, the wavelet decomposition would provide us with a Kahrunen-Loève-type expansion [29] and it would then play the role of a whitening filter especially adapted to self-similar processes (e.g. for the estimation of H according to eq. (15) *via* empirical variance estimates).

Although this goal cannot be strictly achieved, it turns out that the simplest orthonormal wavelet system, i.e. the Haar system, approaches such a doubly orthogonal decomposition when $H = 1/2$, i.e. for ordinary Brownian motion. More precisely [12], if we let $d_j[n], j \in \mathbb{Z}, n \in \mathbb{Z}$ be the Haar coefficients associated with ordinary Brownian motion (i.e. fBm with $H = 1/2$), the correlation of $d_j[n]$ with Haar coefficients at finer scales $k \leq j$ is such that

$$\mathcal{E}(d_j[n]d_k[m]) = 0 \quad (3.12)$$

outside of the (cone-shaped) time-scale domain defined by indexes (m, k) such that $2^{j-k}n \leq m \leq 2^{j-k}(n+1) - 1$. Moreover, we get for each scale

$$\mathcal{E}(d_j[n]d_j[m]) = \mathrm{var}(d_j[n])\delta_{nm} \quad (3.13)$$

and the correlation in scale varies as $2^{5(k-j)/2}$, for synchronous time instants.

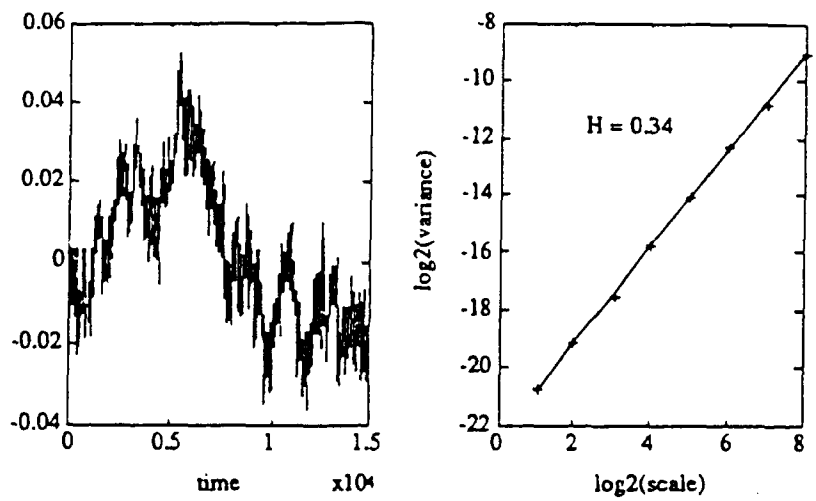


Figure 1. *Wavelet analysis of fractional Brownian motion*
The analyzed signal is a simulated fBm with H=1/3 (left) and empirical variance estimates of the corresponding wavelet coefficients (Daubechies 6) are plotted as a function of scale in a log-log plot (right). According to eq. (15), this is supposed to be a straight line whose slope leads to the estimate $H = 0.34$.

The Haar system does not provide, *stricto sensu*, a doubly orthogonal decomposition of ordinary Brownian motion. However, the correlation between distinct Haar coefficients decays as a power-law of scale and is zero for a given scale. Some extensions of this behavior can be considered by retaining the Haar system as the wavelet basis while replacing the particular value $H = 1/2$ by different values in the range $0 < H < 1$ [12]. It reveals in fact three different regimes associated to the three possible situations $0 < H < 1/2$, $H = 1/2$ and $1/2 < H < 1$. The first two cases correspond respectively to an approximate decorrelation (fast decay) and a perfect decorrelation (only one non-zero coefficient) whereas the last one is associated to a long-term correlation (slow decay)

$$\mathcal{E}(d_j[n]d_j[m]) \sim O\left(|n-m|^{2(H-1)}\right) \tag{3.14}$$

when $|n - m| \gg 1$.

From a frequency point of view, the decay properties of the coefficients correlation is governed by

$$\mathcal{E}(d_j[n]d_j[m]) =$$

$$\frac{\sigma^2}{2}\left(2\sin(\pi H)\Gamma(2H+1)\int_{-\infty}^{+\infty} e^{i\omega(n-m)}\frac{|\Psi(\omega)|^2}{|\omega|^{2H+1}}\frac{d\omega}{2\pi}\right)(2^j)^{2H+1}, \quad (3.15)$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. This heavily depends on the behavior of $\Psi(\omega)$ at the zero frequency, and hence on the number of vanishing moments of $\psi(t)$.

More precisely, if $\psi(t)$ has at most R vanishing moments, i.e. if

$$\int_{-\infty}^{+\infty} t^r\psi(t)dt = 0, \quad r \leq R, \quad (3.16)$$

then it is not possible to prevent divergence of $|\Psi(\omega)|^2|\omega|^{-(2H+1)}$ at the zero frequency if the fBm index H is such that $H > R - 1/2$. As a consequence, the correlation of the corresponding wavelet coefficients has a slow decay and the asymptotic behavior [12]

$$\mathcal{E}(d_j[n]d_j[m]) \sim O\left(|n-m|^{2(H-R)}\right) \quad (3.17)$$

when $|n-m| \gg 1$: this is exactly the Haar case ($R = 1$) for which the divergence situation corresponds to the fBm range $1/2 < H < 1$.

This drawback is easily overcome as soon as a wavelet with $R \geq 2$ is chosen, the quantity $R - 1/2$ being then ensured to exceed the maximum value of H within its range, i.e. 1. As expected, the pathological situation of a slow decay of the wavelet coefficients correlation, which results jointly from the low regularity of the Haar system and from a particular range of H, is no more encountered for more regular choices (e.g. any Daubechies' wavelet [6] such that $R \geq 2$), large R's leading to almost uncorrelated coefficients [27]. (A simulation illustrating this fact is given in [11].)

3.2 Continuous wavelets

Given a fBm of index H and its continuous wavelet transform T_{B_H} with wavelet $h(t)$ [17]

$$T_{B_H}(t,a) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} B_H(u)h^*\left(\frac{u-t}{a}\right)du, \quad (3.18)$$

it can be shown that, as in the orthonormal case, time stationarity is observed at any scale [8]

8

$$\mathcal{E}\left(T_{B_H}(t,a)T_{B_H}(s,a)\right) = \frac{\sigma^2}{2}\left(-\int_{-\infty}^{+\infty}\gamma_h\left(\tau-\frac{t-s}{a}\right)|\tau|^{2H}d\tau\right)a^{2H+1}, \tag{3.19}$$

as well as scale stationarity at any time [28]. Moreover, it turns out that the second-order stationarity of the wavelet transform (together with the self similarity of its correlation structure from scale to scale) is in fact a characteristic property of fBm, as pointed out in [25].

One of the most attractive properties of the continuous wavelet transform is that, in the case of a locally self-similar process $x(t)$, the small scale behavior of the increment process is directly mirrored by the small scale behavior of the wavelet transform of $x(t)$ [1], namely

$$\begin{aligned} &\mathcal{E}\left([x(t+\tau)-x(t)]^2\right) \sim |\tau|^{2H(t)} \quad , \quad \tau \to 0 \\ &\qquad\qquad \Rightarrow \\ &\mathcal{E}\left(|T_x(t,a)|^2\right) \sim a^{2H(t)+1} \quad , \quad a \to 0. \end{aligned} \tag{3.20}$$

Therefore, if we want to measure the local scaling exponents $H(t)$, what is required is a procedure ending up with

$$\hat{H}(t) = \frac{1}{2}\lim_{a\to 0^+}\frac{\log \mathcal{E}\left(|T_x(t,a)|^2\right)}{\log a} - \frac{1}{2}, \tag{3.21}$$

in which the unavailable ensemble average $\mathcal{E}\left(|T_x(t,a)|^2\right)$ is replaced by some efficient estimate.

It has been observed that the crude estimate $|T_x(t,a)|^2$ leads to a high variability for $\hat{H}(t)$, suggesting refined analyses such as maxima tracking in the wavelet representation prior to local least-squares fits "amplitude *vs* scale" in log-log coordinates [2]. Another approach is to consider time-scale analysis from a more general point of view, wavelets being only one possible solution.

4 Time-scale energy distributions

The purpose of a time-scale representation is to describe the information contained in a signal in terms of time and scale, simultaneously [10]. One useful information is *energy*, which in some cases can be obtained by integrating the representation (then referred to as an *energy distribution*) over the whole time-scale plane. This is especially the case for the *scalogram* (i.e. the squared modulus of the wavelet transform [13]) since we have for any finite energy signal $x(t)$

$$\int\int_{-\infty}^{+\infty}|T_x(t,a)|^2\,\frac{dtda}{a^2} = \int_{-\infty}^{+\infty}|x(t)|^2dt. \tag{4.1}$$

However, it turns out that the same property holds for other distributions too, as discussed below.

4.1 A general class

By construction, the role played by the scalogram in time-scale analysis is very similar to the one played by the *spectrogram* (i.e. the squared modulus of the short-time Fourier transform) in time-frequency analysis. We know that, within this latter framework, a much larger class of time-frequency energy distributions exist: it is referred to as the *Cohen's class* [4] and reads

$$C_x(t,\nu;\Pi) = \int\int_{-\infty}^{+\infty} W_x(u,n)\Pi(u-t,n-\nu)dudn, \tag{4.2}$$

where $W_x(t,\nu)$ is the so-called *Wigner-Ville distribution*

$$W_x(t,\nu) = \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right)e^{-i2\pi\nu\tau}d\tau \tag{4.3}$$

and $\Pi(t,\nu)$ some arbitrary parameterization function with the only constraint

$$\int\int_{-\infty}^{+\infty} \Pi(t,\nu)dtd\nu = 1. \tag{4.4}$$

The interest of Cohen's class is at least twofold. First, it provides a unification of different definitions which are simply characterized by different parameterization functions. Among the simplest examples, it is easily checked that the specification $\Pi(t,\nu) = W_h(t,\nu)$ is associated with

$$C_x(t,\nu;W_h) = \left|\int_{-\infty}^{+\infty} x(u)h^*(u-t)e^{-i2\pi\nu u}du\right|^2, \tag{4.5}$$

i.e. with the spectrogram. This latter appears therefore as only a special case within the more general framework of Cohen's class. Second, its parameterization provides a natural way of deriving specific definitions from requirements imposed *a priori*.

In this respect, it is known [9] that, for a number of theoretical reasons (marginal properties, localization in both time and frequency, estimation of instantaneous frequency,...), the Wigner-Ville distribution is to be preferred to the spectrogram and that, for more practical reasons, a useful approximation of it is the so-called *smoothed pseudo-Wigner-Ville distribution* whose parameterization function is *separable*, i.e. of the form $\Pi(t,\nu) = g(t)H(\nu)$, which leads to the formulation

$$C_x(t,\nu;gH) = \int\int_{-\infty}^{+\infty} h(\tau)g(u-t)x\left(u+\frac{\tau}{2}\right)x^*\left(u-\frac{\tau}{2}\right)e^{-i2\pi\nu\tau}dud\tau. \tag{4.6}$$

10

By analogy with the time-frequency case, it is expected that a companion situation could exist in the time-scale case. It has been shown that a natural counterpart of Cohen's class does exist [13] [26], the general class of time-scale energy distributions being

$$\Omega_x(t,a;\Pi) = \int\int_{-\infty}^{\infty} W_x(u,n)\Pi\left(\frac{u-t}{a}, an\right) dudn. \tag{4.7}$$

This general formulation is in fact deduced from the time-scale covariance requirement

$$x_{a\tau}(t) = \frac{1}{\sqrt{a}}x\left(\frac{t-\tau}{a}\right) \quad \Rightarrow \quad \Omega_{x_{a\tau}}(t,a;\Pi) = \Omega_x\left(\frac{t-\tau}{a}, \frac{a}{a}; \Pi\right) \tag{4.8}$$

imposed to bilinear forms [26].

Within this framework, the scalogram appears as only a special case associated with the choice $\Pi(t,\nu) = W_h(t,\nu)$, exactly as the spectrogram does within Cohen's class and, again, other choices can be preferred in some circumstances. This can be the case for the family of *affine smoothed Wigner-Ville distributions* [13] associated with the separable specification $\Pi(t,\nu) = g(t)H(\nu)$, and whose expression reads

$$\Omega_x(t,a;gH) = \frac{1}{a}\int\int_{-\infty}^{\infty} h\left(\frac{\tau}{a}\right) g\left(\frac{u-t}{a}\right) x\left(u+\frac{\tau}{2}\right) x^*\left(u-\frac{\tau}{2}\right) dud\tau. \tag{4.9}$$

4.2 Time-scale energy distributions and self-similar processes

The existence of a whole family of time-scale energy distributions offers a great versatility which does not necessarily reduce the problem of estimating scaling exponents to the use of scalograms. Loosely speaking, time-scale analysis can be thought of as a natural counterpart of time-frequency analysis, local scaling laws playing the role of instantaneous frequency. Precisely, we get [16], as a generalization of eq. (24),

$$\begin{gathered} \mathcal{E}\left([x(t+\tau)-x(t)]^2\right) \sim |\tau|^{2H(t)} \quad , \quad \tau \to 0 \\ \Rightarrow \\ \mathcal{E}\left(\Omega_x(t,a;\Pi)\right) \sim a^{2H(t)+1} \quad , \quad a \to 0 \end{gathered} \tag{4.10}$$

for all the distributions Ω_x such that their parameterization function Π has a 1D partial Fourier transform

$$\bar{\pi}(\xi,\nu) = \int_{-\infty}^{\infty} \Pi(t,\nu)e^{-i2\pi\xi t}dt \tag{4.11}$$

which satisfies

11

$$\tilde{\pi}\left(\xi,\frac{\xi}{2}\right)=0. \tag{4.12}$$

This holds in the case of the scalogram since, then,

$$\tilde{\pi}(\xi,\nu)=H\left(\nu+\frac{\xi}{2}\right)H^{*}\left(\nu-\frac{\xi}{2}\right) \quad \Rightarrow \quad \tilde{\pi}\left(\xi,\frac{\xi}{2}\right)=H(\xi)H^{*}(0)=0, \tag{4.13}$$

because of the admissibility condition on $h(t)$. This holds too for affine smoothed Wigner-Ville distributions such that

$$G(\xi)H\left(\frac{\xi}{2}\right)=0. \tag{4.14}$$

This latter case is of particular interest since, by controlling independently $g(t)$ and $h(t)$, we can come up, at will, with more or less smoothed versions (in time) of scalograms [13]. Therefore, given one recorded signal, affine smoothed Wigner-Ville distributions can be viewed as *estimators* of ensemble averaged scalograms for which the scaling property (24) holds. Within this interpretation, the above condition (39) ensures the estimator to be *unbiased*, whereas the time smoothing involved in the estimation reduces the variability of $\hat{H}(t)$. As expected, this modified procedure leads to effective improvements in the case of processes whose $H(t)$ is a piecewise constant or slowly-varying function [16] and its effectiveness for more general locally self-similar processes is currently under investigation.

5 Conclusion

Time-scale analysis is a powerful framework for characterizing self-similar processes. In the case of global self-similarity (e.g. fractional Brownian motion), orthonormal wavelet decompositions are particularly efficient whereas the study of local self-similarity has been preferably conducted, up to now, with the help of continuous wavelet transforms. In this respect, and with the idea that the estimation of local scaling exponents has to be performed locally, across scales, it has been argued that time-scale distributions more general than wavelet-based distributions can be used with possibly increased performance.

However, it has been recently shown [24] that *local* scaling properties of a (multifractal) self-similar process can be directly characterized by *global* scaling properties of some partition function built upon a wavelet transform. In this respect too, it is an interesting question to study how such an approach could be combined with the more general class of time-scale energy distributions discussed here, with the orthonormal framework (and its computational efficiency), or even with scale-based dynamical models [3].

12

References

1. A. Arnéodo, G. Grasseau and M. Holschneider, "Wavelet Transform Analysis of Invariant Measures of Some Dynamical Systems", in [5], pp. 182-196.

2. E. Bacry, A. Arnéodo, U. Frisch, Y. Gagne and E. Hopfinger, "Wavelet Analysis of Fully Developed Turbulence Data and Measurement of Scaling Exponents", in *Turbulence and Coherent Structures* (O. Métais and M. Lesieur, *eds.*), pp. 203-215, Kluwer, 1991.

3. K.C. Chou, S. Golden and A.S. Willsky, "Modeling and Estimation of Multiscale Stochastic Processes", IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-91, Toronto, pp. 1709-1712, 1991.

4. L. Cohen, "Time-Frequency Distribution - A Review", *Proc. IEEE*, Vol. 77, No. 7, pp. 941-981, 1989.

5. J.M. Combes, A. Grossmann and Ph. Tchamitchian (*eds.*), *Wavelets*, Springer-Verlag, New York, 1989.

6. I. Daubechies,"Orthonormal Bases of Compactly Supported Wavelets", *Con.m. Pure Appl. Math.*, Vol. XLI, No. 7, pp. 909-996, 1988.

7. K. Falconer, *Fractal Geometry*, J. Wiley and Sons, Chichester, 1990.

8. P. Flandrin, "On the Spectrum of Fractional Brownian Motions", *IEEE Trans. on Info. Theory*, Vol. IT-35, No. 1, pp. 197-199, 1989.

9. P. Flandrin, "Some Aspects of Nonstationary Signal Processing with Emphasis on Time-Frequency and Time-Scale Methods", in [5], pp. 68-98, 1989.

10. P. Flandrin, "Wavelets and Related Time-Scale Transforms", in *Advanced Signal Processing Algorithms, Architectures and Implementations*, (F.T. Luk, *ed.*), SPIE Proc., Vol. 1348, pp. 2-13, 1990.

11. P. Flandrin, "Fractional Brownian Motion and Wavelets", to appear in *Wavelets, Fractals and Fourier Transforms - New Developments and New Applications* (M. Farge, J.C.R. Hunt and J.C. Vassilicos, *eds.*), Oxford Univ. Press.

12. P. Flandrin, "Wavelet Analysis and Synthesis of Fractional Brownian Motion", *IEEE Trans. on Info. Theory*, April 1992.

13. P. Flandrin and O. Rioul, "Affine Smoothing of the Wigner-Ville Distribution", *IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-90*, Albuquerque (NM), pp. 2455-2458, 1990.

14. U. Frisch ang G. Parisi, "On the Singularity Structure of Fully Developed Turbulence", in *Turbulence and Predictability in Geophysical Fluid Dynamics* (M. Gil, R. Benzi and U. Frisch, *eds.*), pp. 84-88, North-Holland, 1985.

15. N. Gache, P. Flandrin and D. Garreau, "Fractal Dimension Estimators for Fractional Brownian Motions", IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-91, Toronto, pp. 3557-3560, 1991.

16. P. Gonçalves and P. Flandrin, "Scaling Exponents Estimation from Time-Scale Energy Distributions", *IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-92*, San Francisco (CA), 1992.

17. A. Grossmann and J. Morlet, "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape", *SIAM J. Math. Anal.*, Vol. 15, No. 4, pp. 723-736, 1984.

18. M. Kim and A.H. Tewfik, "Multiscale Signal Detection in Fractional Brownian Motion", in *Advanced Signal Processing Algorithms, Architectures and Implementations* (F.T. Luk, *ed.*), SPIE Vol. 1348, pp. 462-470, 1990.

19. S.G. Mallat, "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation", *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol. PAMI-11, No. 7, pp. 674-693, 1989.

20. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.

21. B.B. Mandelbrot and J.W. van Ness, "Fractional Brownian Motions, Fractional Noises and Applications", *SIAM Rev.*, Vol. 10, No. 4, pp. 422-437, 1968.

22. C. Meneveau and K.R. Sreenivasan, "The Multifractal Nature of Turbulent Energy Dissipation", *J. Fluid. Mech.*, Vol. 224, pp. 429-484, 1991.

23. Y. Meyer, "Orthonormal Wavelets", in [5], pp. 21-37, 1989.

24. J.F. Muzy, E. Bacry and A. Arnéodo, "Wavelets and Multifractal Formalism for Singular Signals: Application to Turbulence Data", *Phys. Rev. Lett.*, Vol. 67, No. 25, pp. 3515-3518, 1991.

25. J. Ramanathan and O. Zeitouni, "On the Wavelet Transform of Fractional Brownian Motion", *IEEE Trans. on Info. Theory*, Vol. IT-37, No. 4, pp. 1156-1158, 1991.

14

26. O. Rioul and P. Flandrin, "Time-Scale Energy Distributions - A General Class Extending Wavelet Transforms", *IEEE Trans. on Signal Proc.*, July 1992

27. A.H. Tewfik and M. Kim, "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motions", *IEEE Trans. on Info. Theory*, April 1992.

28. M. Vergassola and U. Frisch, "Wavelet Transforms of Self-Similar Processes", *Physica D*, Vol. 54, pp. 58-64, 1991.

29. G.W. Wornell, "A Karhunen-Loève-Like Expansion for $1/f$ Processes *via* Wavelets", *IEEE Trans. on Info. Theory*, Vol. IT-36, No. 4, pp. 859-861, 1990.

30. A.M. Yaglom, *Correlation Theory of Stationary and Related Random Functions*, Springer-Verlag, New York, 1986.

STRUCTURAL DECOMPOSITION OF SIGNALS

Stephane Mallat and Zhifeng Zhang

Courant Institute of Mathematical Sciences, New York University
251 Mercer Street, New York, NY 10012

Workshop on the Role of Wavelets in Signal Processing Applications

March 13th, 1992

1. Introduction

A fundamental problem in discrete signal processing is to find a numerical representation that is well adapted in order to perform processings such as compact coding, noise removal, feature enhancement, pattern detection or recognition. Most classical methods build signal representation with linear transforms, generally based on filtering technics. Some linear transforms such as the Karhunen-Loeve decomposition or wavelet-packets [1] can be adapted to the global signal properties. However, if the signal includes local patterns of very different types, such as edges and sinusoidal waves, the transform will not be adapted to at least one of these type of patterns. Fig. 3(a),4(a),5(a) show examples of such signals, composed of a sum of sinusoidal waves of different frequencies, plus a Dirac. Depending upon the relative energy of the sinusoidal waves and the Dirac, the best basis algorithm of Coifman and Wickerhauser [1] will choose either a Dirac representation (Fig. 3(b)) or a sinusoidal representation (Fig. 5(b)), or an intermediate representation (Fig. 4(b)) A remarkable property aspect of the wave-packet algorithm is that it chooses an orthogonal basis, however, the rigidity of this orthogonality can yield instabilities in the choice of the basis.

In this paper, we introduce a new non-linear operator, that defines local adaptive transforms. Although it is non-linear, this transform has an energy conservation law, as an orthogonal basis decomposition. The signal is decomposed into elementary structures, that match best its local patterns. These structures can well localized functions in the time/frequency plane, or

This research was supported by the NSF grant IRI-890331, AFOSR grant AFOSR-90-0040 and ONR grant N00014-91-J-1967.

other basic waveforms.

Section 2 introduces the basic theory of structure books. Section 3 describes in more details the particular examples of structure books built with local time/frequency transforms. Numerical examples and signal processing applications are given in the last section.

2. Structure Books

This section defines the decomposition of signals into structure books and the properties of this decomposition. The signal is decomposed into a sum of elementary waveforms that belong to a given set, called the dictionary. Depending upon the dictionary, the resulting transform can have very different properties. Let $\mathbf{H}$ be a Hilbert space. Let $D = (e_i)_{i \in I}$ be a family of normalized vectors that belong to $\mathbf{H}$. The family of vectors D is called a dictionary. The index set I might not be countable. We call C a choice function of D. Such a function associates to each nonempty subset E of D an element e_j that belongs to D, and we write $C(E) = e_j$. The axiom of choice guarantees that such a function exists.

The operator T subtracts to any vector f of $\mathbf{H}$ its projection on one particular vector e_j of D, that is chosen among the ones that have large inner products with f. Let α be a constant such that $0 < \alpha < 1$, and

$$S = \sup_{i \in I} \left\{ |\langle f , e_i \rangle| \right\} . \tag{1}$$

Let us also define

$$E = \left\{ e_i \in D \ / \ |\langle f , e_i \rangle| \geq \alpha S \right\} . \tag{2}$$

Let $e_j = C(E)$. The operator T is defined by

$$Tf = f - \langle f , e_j \rangle e_j . \tag{3}$$

It follows that

$$\| f \|^2 = \| Tf \|^2 + |\langle f , e_j \rangle|^2 . \tag{4}$$

To perform a full decomposition, we iterate on the operator T. We denote T^n the iteration n times of the operator T. For any $n > 0$ there exists $e_j^n \in D$ such that

$$T^{n+1} f = T^n f - \langle T^n f , e_j^n \rangle e_j^n . \tag{5}$$

The vector f can be decomposed into

$$f = \sum_{n=0}^{m-1} (T^n f - T^{n+1} f) + T^m f , \tag{6}$$

where $T^0 f = f$. As a consequence of equation (4), we obtain

$$f = \sum_{n=0}^{m-1} \langle T^n f , e_j^n \rangle e_j^n + T^m f \ . \tag{7}$$

Equation (5) yields the energy conservation equation

$$\| f \|^2 = \sum_{n=0}^{m-1} | \langle T^n f , e_j^n \rangle |^2 + \| T^m f \|^2 \ . \tag{8}$$

Let us denote by $\mathbf{V}$ closure of the space of vectors that are linear expansions of vectors in the dictionary D. Let $\mathbf{W}$ be the orthogonal complement of $\mathbf{V}$ in $\mathbf{H}$. Because of the energy conservation equation (8), one can prove that when m goes to $+\infty$, $T^m f$ converges weakly to the orthogonal projection of f on $\mathbf{W}$. We say that the dictionary is complete, if and only if $\mathbf{V} = \mathbf{H}$. In this case, $\mathbf{W} = \{ 0 \}$ and hence $T^m f$ converges weakly to 0. This means that the sum $\sum_{n=0}^{m-1} \langle T^n f , e_j^n \rangle e_j^n$ converges weakly to f. In general, the type convergence of $\sum_{n=0}^{m-1} \langle T^n f , e_j^n \rangle e_j^n$ depends upon the properties of the dictionary D. A detailed analysis of this convergence is under preparation [3]. In this short paper, we rather concentrate on the practical applications of this transform to discrete signal processing.

In signal processing, we manipulate finite discrete signals that can be considered as elements of a finite dimensional space $\mathbf{H}$. One easily can prove [3] the following theorem that guarantees that for finite discrete signal, a structuring transform provides a complete and stable signal representation.

Theorem

If the space $\mathbf{H}$ has a finite dimension, $T^n f$ converges in norm to the orthogonal projection of f on $\mathbf{W}$. If the dictionary D is complete, then

$$f = \sum_{n=0}^{+\infty} \langle T^n f , e_j^n \rangle e_j^n \ , \tag{9}$$

and

$$\| f \|^2 = \sum_{n=0}^{+\infty} | \langle T^n f , e_j^n \rangle |^2 \ . \tag{10}$$

We call a structure the information given by $(e_j^n , \langle T^n f , e_j^n \rangle)$ and structure book the sequence of structures $\left((e_j^n , \langle T^n f , e_j^n \rangle) \right)_{n \in \mathbb{N}}$. This theorem proves that a structure book is a complete and stable representation, of finite signals. In finite dimension, an example of choice function is implemented by selecting the first vector encountered in the dictionary data base, that satisfies

$$|\langle f , e_j \rangle| = \underset{i \in I}{Max} \left(|\langle f , e_i \rangle| \right) . \qquad (11)$$

3. Time/Frequency Dictionaries

The dictionary that is used to build the structure book must be adapted to the class of signals that is considered and to the particular applications. Local time/frequency dictionaries are often well adapted to characterize patterns of very different types. To create a signal representation that is invariant by translation, the dictionary must be composed of elementary waveforms that are translated on the signal grid. Let us suppose that the signals have P samples, and are periodized to solve border problems. Let $(e_k)_{1 \le k \le K}$ be a set of elementary discrete waveforms and $e_{k,p}$ by the translation of e_k by p:

$$e_{k,p}(n) = e_k(n-p) .$$

We build the dictionary $D = \left(e_{k,p} \right)_{1 \le p \le P, 1 \le k \le K}$. Suppose that a signal f is decomposed into a structure book given by

$$\left((e^n_{k,p} , \langle T^n f , e^n_{k,p} \rangle) \right)_{n \in \mathbb{N}} .$$

If the signal is translated by l samples, one can easily prove that the waveforms of the structure book are translated but not modified and are given by

$$\left((e^n_{k,p+l} , \langle T^n f , e^n_{k,p} \rangle) \right)_{n \in \mathbb{N}} .$$

In the following, we call shifting dictionary, any dictionary that is built by translating a set of elementary waveforms. One can easily prove that the dictionary is complete, if and only if there exists two constants $A > 0$ and $B > 0$, such that for all frequencies ω, their discrete Fourier transform $\hat{e}_k(\omega)$ of the signals e_k satisfy

$$A \le \sum_{k=1}^{K} |\hat{e}_k(\omega)|^2 \le B . \qquad (12)$$

The wavelet transform and multiscale window Fourier transform provide two particularly interesting local time/frequency dictionaries. If we want a representation that is invariant by scaling by any power of $s > 1$, we build a dictionary that is composed of wavelets dilated by s^m, $m \in \mathbb{Z}$, and translated on the signal grid. In this case, the structure book corresponds to a non-linear wavelet transform. The signal f is decomposed into a sum of wavelets translated and dilated. Although it is not an orthogonal wavelet transform, equation (10) proves that we have a conservation of energy.

To compute a discrete representation which is invariant by translation, dilation and modulation, we define a dictionary by dilating, translating and modulating a window function. The structure book defines a signal decomposition that is local in the phase plane. It regroups the waveforms of the time/frequency plane, that match best the signal patterns.

The general algorithm that computes a structure book is illustrated by the block diagram of Fig. 1. The algorithm requires to find a waveform in the dictionary, whose inner product (correlation) belongs to the set E defined in (2). For this purpose, we build a data base of correlation coefficients that carries the correlation of the signal with each waveform of the dictionary. The inner product of a signal with a translated waveform can be written as a convolution. For a shifting dictionary, this data base is thus computed by convolving the signal f with each elementary waveform, which can be done with fast Fourier transform. We then choose one correlation coefficient whose absolute value is maximum, subtract the corresponding waveform to the signal, as in equation (3) and update the correlation coefficients. This operation is repeated until the total energy of the selected structures is close to the signal energy.

Shifting dictionaries have the advantage of building signal representations that are shift invariant, but they require substantial amounts of computations. The wavepackets [1] define dictionaries, where the structure book can be computed efficiently. Orthogonal wavepackets are computed through a cascade of Quadrature Mirror Filter bank decompositions [4], along a tree, where each node has M sons, generally 2 for one-dimensional signals, as illustrated by Fig. 2. If we put a Dirac at the root of this tree, the discrete waveforms that are generated at the nodes of the tree are the wavepacket functions. The corresponding dictionary is a set of translated versions of these elementary waveforms. These wavepackets can be regrouped into a set of orthonormal bases. Coifman and Wickerhauser [1], showed that the wavepacket tree provides a local signal decomposition in the time/frequency plane. As opposed to the best basis algorithm, we extract from this tree a set of elementary waveforms that are not mutually orthogonals but that can match locally the signal patterns. For wavepackets, the update of correlation coefficients can be implemented efficiently with finite impulse response filters as the ones of Daubechies [2]. The selection of the correlation coefficients is done by using hash tables [3]. After processing the structure book, the reconstruction of a signal from a structure book is simply done by adding the waveforms of the structure book, as indicated by equation (9). The same type of computations can be performed for signals of any dimension. For wavepacket dictionaries, this algorithm can be efficiently implemented for images.

4. Numerical Experiments and Signal Processing Applications

This section presents numerical results for wavepackets dictionaries as well and signal processing applications. In order to give a better understanding of the structure book properties, we compare the structure book with the optimal wavepacket basis. Fig. 3,4 and 5 show each of these transforms for a Dirac embedded into a sum of sinusoidal waves. The wavepacket makes a global time/frequency localization choice that is controlled either by the Dirac of the sinusoids. On the contrary, the waveforms in the structure book are locally adapted to the signal properties. The structures therefore match the signal, independently from the relative global energy of the Dirac and the sinusoidal waves, as it can be seen in Fig. 3(c), 4(c) and 5(c). Each block shown in these figures corresponds to a structure. The darker the block, the larger the correlation coefficient.

The dictionary waveform of a signal structure can be viewed as a signal pattern and the correlation coefficient as the amplitude of this pattern in the signal. This signal representation has applications for pattern detection, recognition, noise removal, signal enhancement and compact coding. We illustrate the application to pattern detection and noise removal with several examples. Fig. 6 shows a simple example of signal separation. We know a priori that a Dirac yields highest correlation coefficients, for a wavelet time/frequency localization. Fig. 6(b) shows the structures of the whole structure book given in Fig. 5(c), that have a wavelet type time/frequency localization. Fig. 6(c) is the graph of the signal reconstructed from the selected structures shown in Fig. 6(b). The Dirac is clearly restored, with little remaining oscillatory component. Fig. 7(a) shows a Gabor function (Gaussian modulated by a sinusoidal wave). Fig. 7(b) is the sum of this Gabor function with a white noise. The total energy of the noise is much larger than the energy of the signal. Fig. 7(c) shows the best basis selection, which is completely driven by the white noise. Fig. 7(d) shows the structure book, where the Gabor function clear appears as a very dark box. This Gabor function is easily discriminated, because its energy is much more concentrated than the energy of the white noise. In Fig. 8(a), the input signal is a set of two Diracs, that can be viewed at radar impulses, for example. To these two Diracs is added a color noised shown and the sum is shown in Fig. 8(b). The Signal to Noise Ratio is -15.09 db. Fig. 8(d) shows the structure book of this signal. The structures of larger energy are noise components. However, one can recognize two elongated bars due to the Diracs. As previously mentioned, Diracs like any singularities, yield structures whose time/frequency localization are wavelet type time/frequency localization. To restore the impulse components of the signal, we thus select the structures of the structure book, that have a wavelet type time/frequency decomposition. These structures are shown in fig. 8(e). The signal reconstructed from these structures is shown in Fig. 8(c). The Diracs have been partly destroyed but are now clearly visible. Let us emphasize that this algorithm works well because the noise has a different type of time/frequency localization than the signal. After the noise removal, the SNR is 0.21 db, which represents a gain of -15.3 db.

The structure book representation can be used for applications such as speech recognition, where we know perfectly what is the information, but also to problems where we try find whether there is any "information" in a signal, and how to extract it. There are many such issues in medical signal processing, where new sensors such as high resolution blood pressure sensors, yield measurements that we do not know how to interpret.

References

1. Coifman, R., Meyer, Y., and Wickerhauser, V., "Size properties of wavelet-packets," in *Wavelets and their applications*, ed. Ruskai et. al., Jones and Bartlett, 1992.
2. Daubechies, I., "Orthonormal bases of compactly supported wavelets," *Communications in Pure and Applied Mathematics*, vol. 41, pp. 909-996, Nov. 1988.
3. Mallat, S. and Zhang, Z., "Structural analysis of signals," *NYU Techn. Report, Computer Science*, In preparation.
4. Rioul, O. and Vetterli, M., *Wavelets and Signal Processing*, p. IEEE Signal Processing Magazine, Oct. 1991.

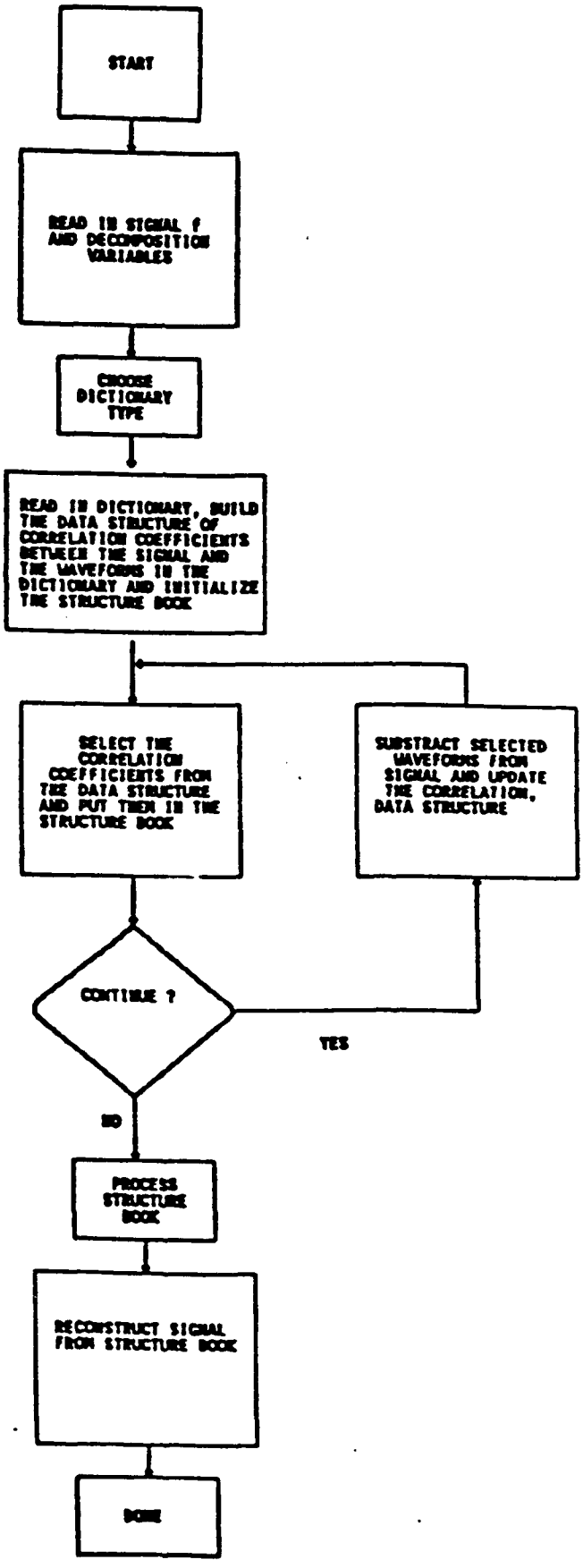


Fig.1 Algorithm to compute the structure book of a signal.

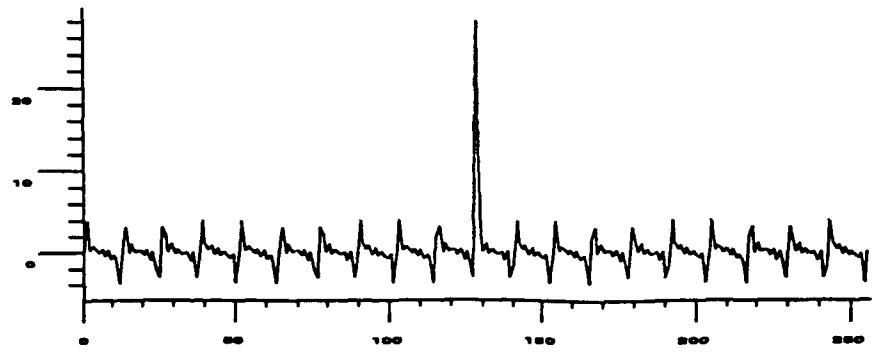


Fig.3(a) Dirac with sum of sunusoides. (dirac amplitude is 25)

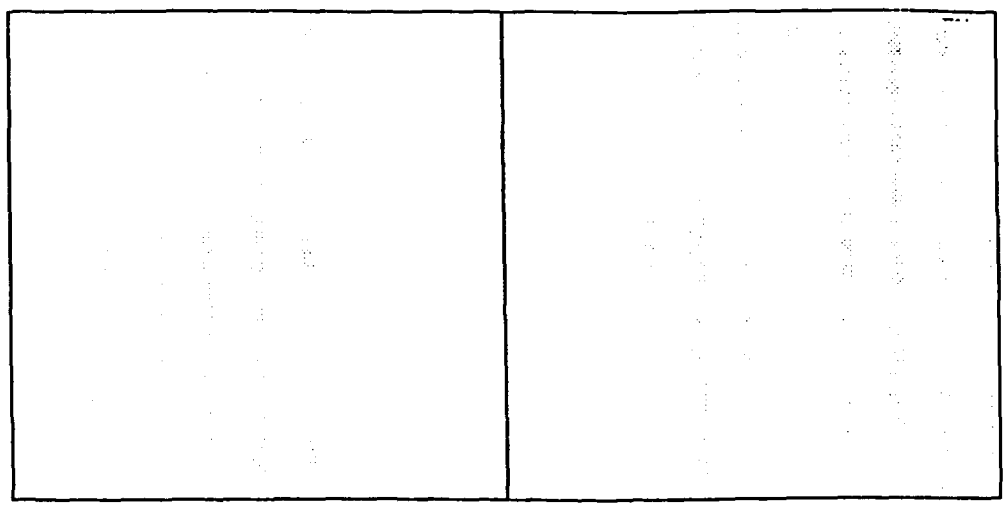


Fig.3(b) Time/frequency representation of the the best basis.

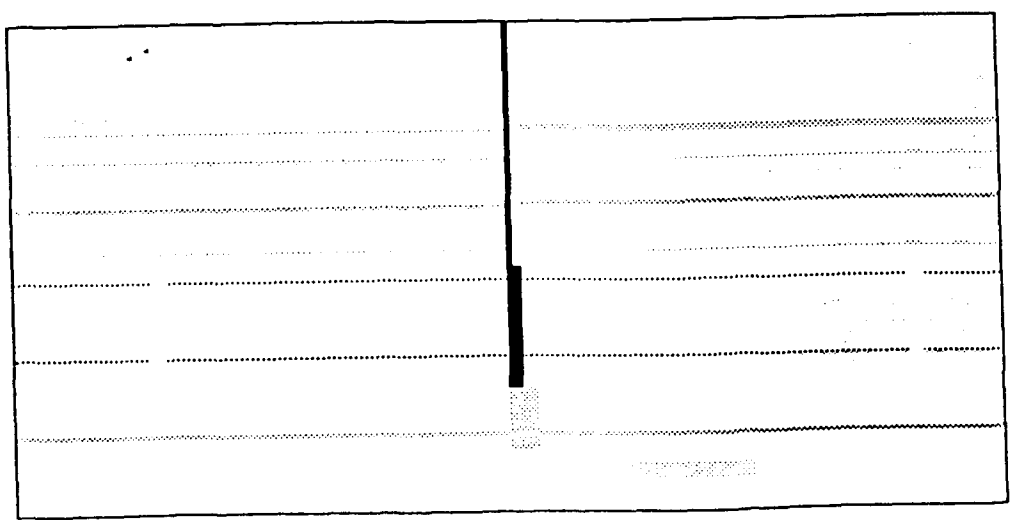


Fig.3(c) Time/frequency representation of the structure book. Each rectangle represents a particular structure which indexed by its position in the phase plane. The darker the rectangle, the larger the correlation coefficient.

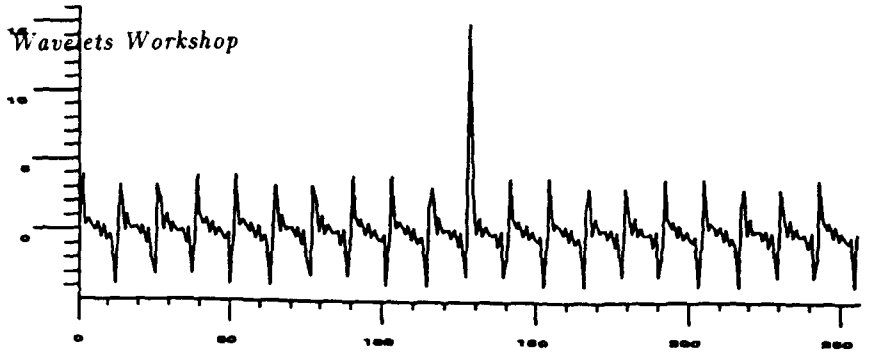


Fig.4(a) Dirac with sum of sunusoides. (dirac amplitude is 12)

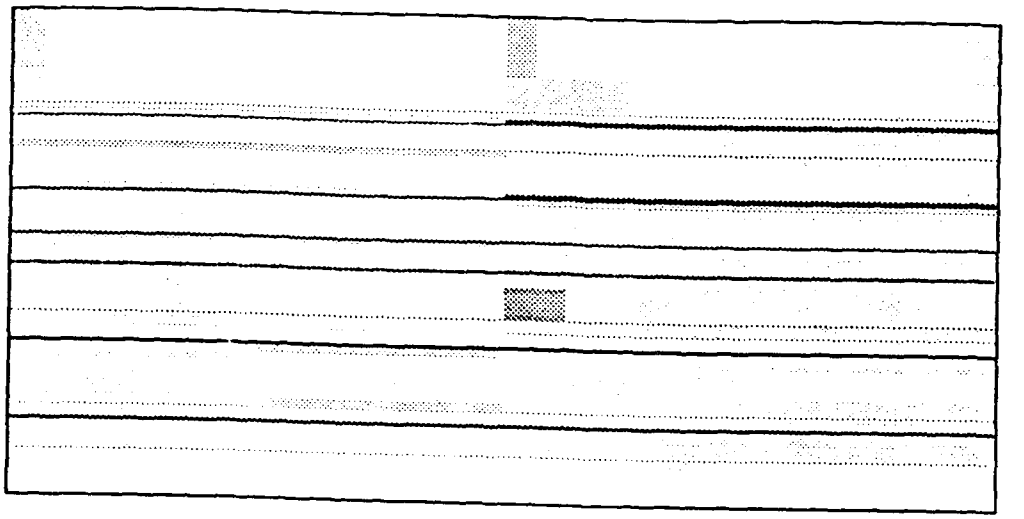


Fig.4(b) Time/frequency representation of the the best basis.

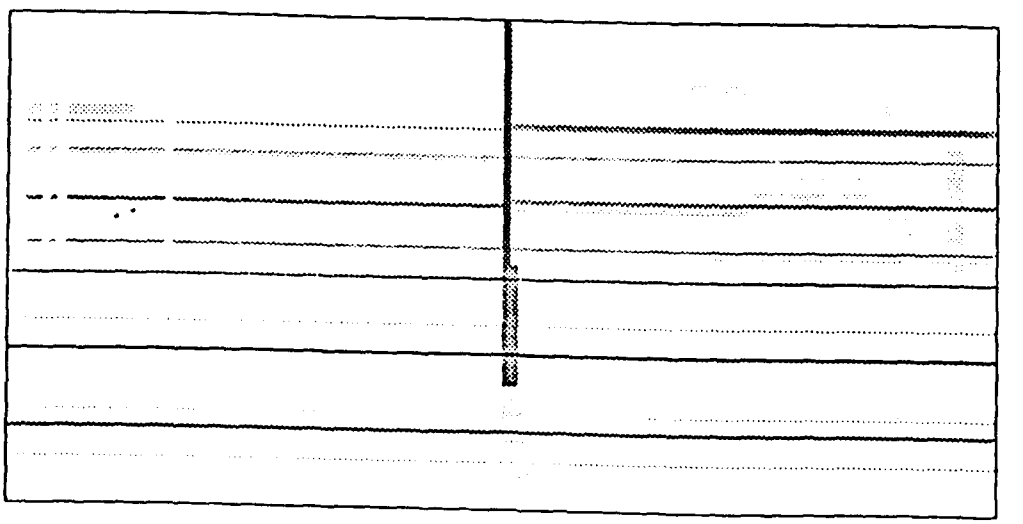


Fig.4(c) Time/frequency representation of the structure book.

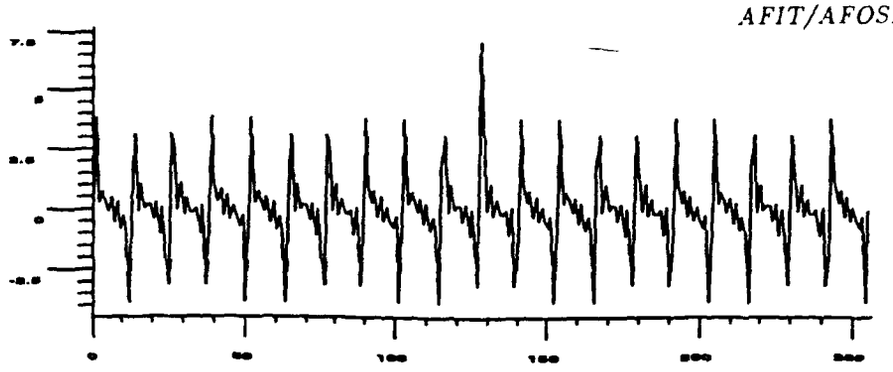


Fig.5(a) Dirac with sum of sunusoides. (dirac amplitude is 4)

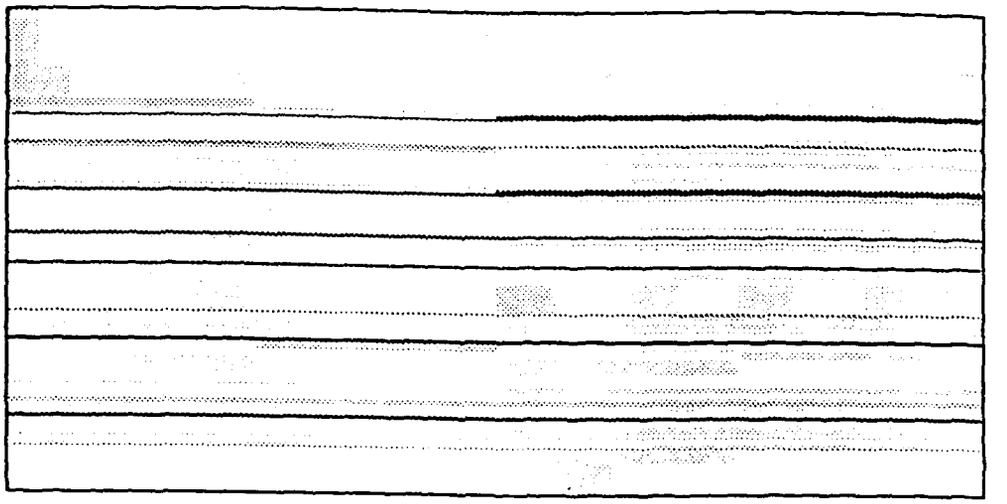


Fig.5(b) Time/frequency representation of the the best basis.

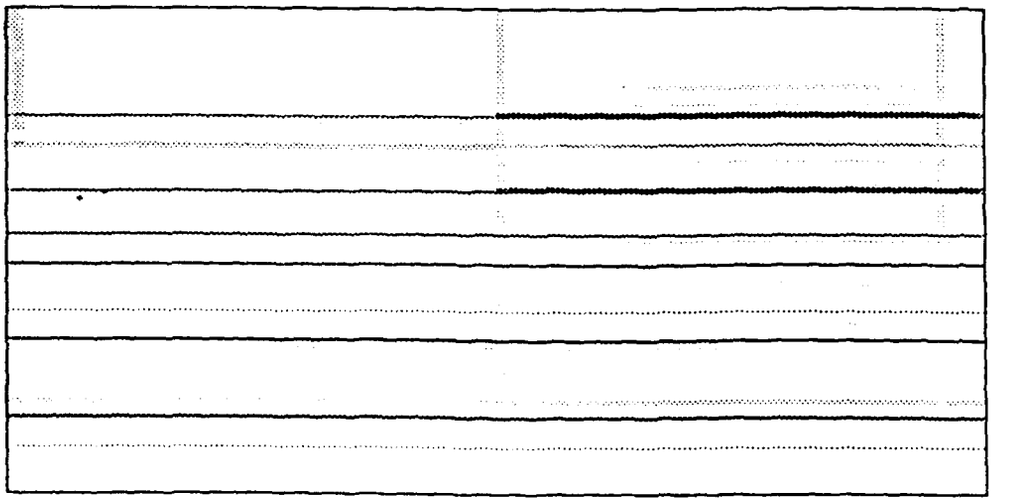


Fig.5(c) Time/frequency representation of the structure book.

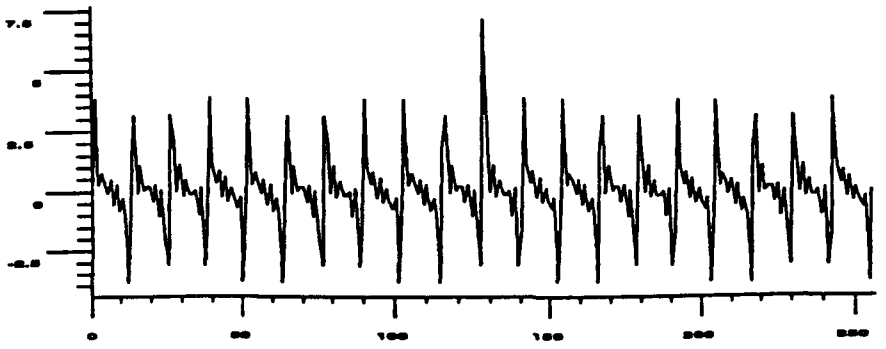


Fig.6(a) Dirac with sum of sinusoids (same as in Fig.5(a)).

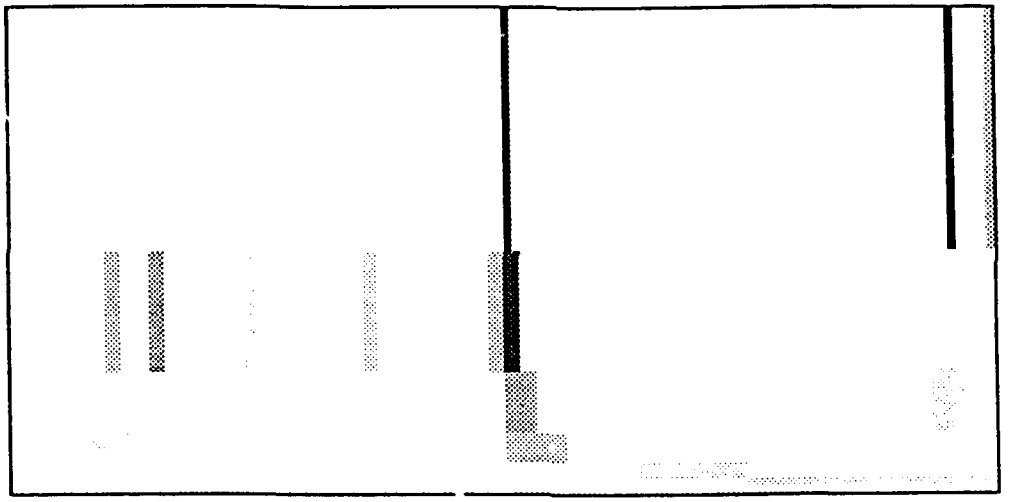


Fig.6(b) Selected structures from the structure book shown in Fig.5(c).

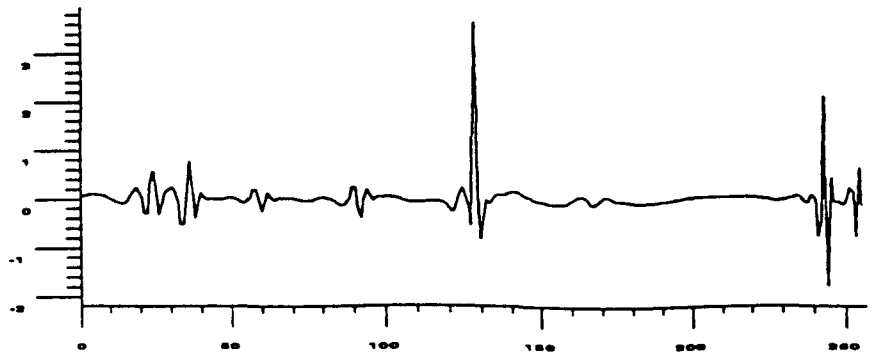


Fig.6(c) Reconstructed signals from selected structures.

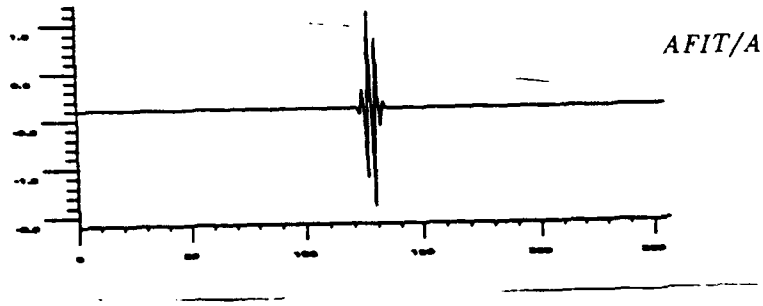


Fig.7(a) Gabor function (Gaussian modulated by a sinusoidal wave).

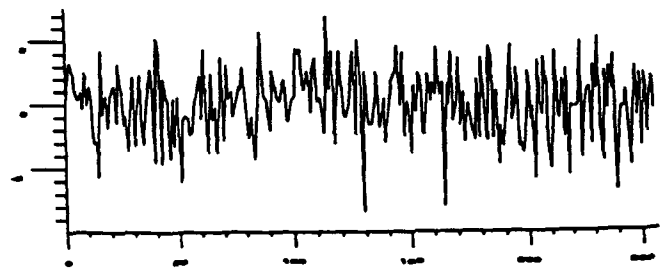


Fig.7(b) Gabor function plus white noise.

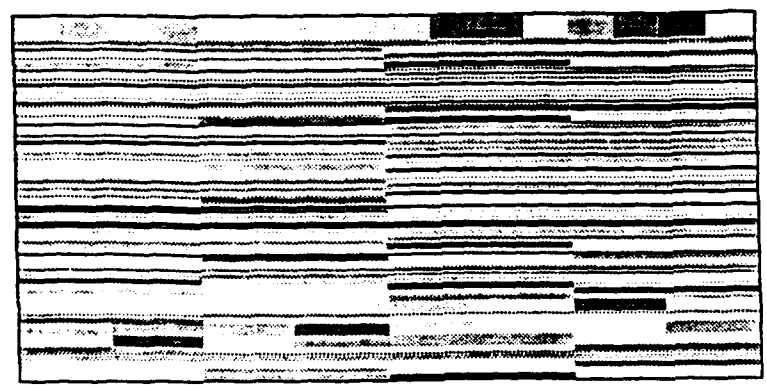


Fig.7(c) Time/frequency representation of the best basis.

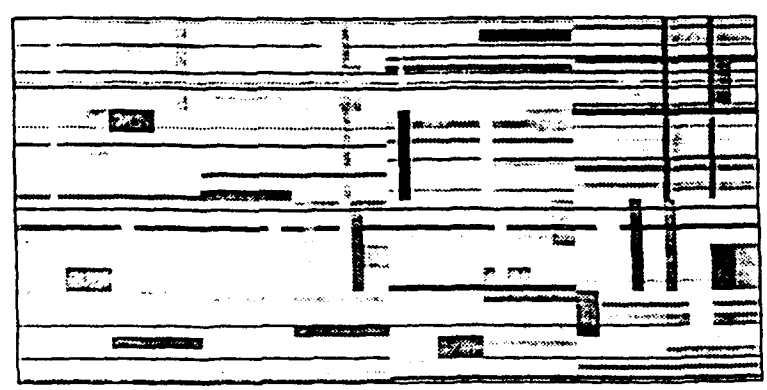


Fig.7(d) Time/frequency representation of the structure book.

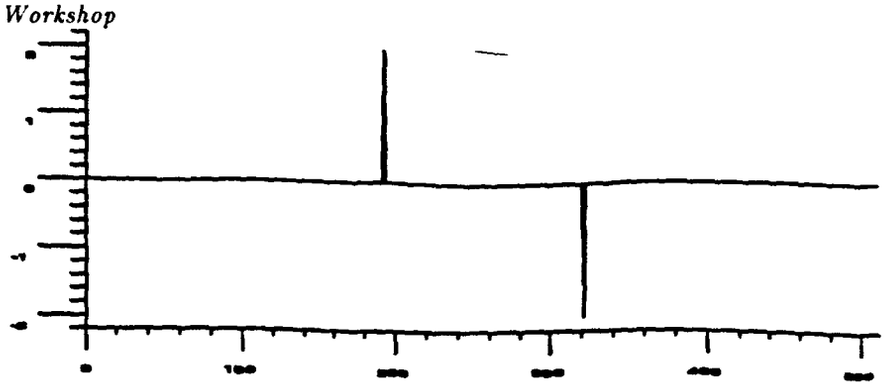


Fig.8(a) Signal diracs.

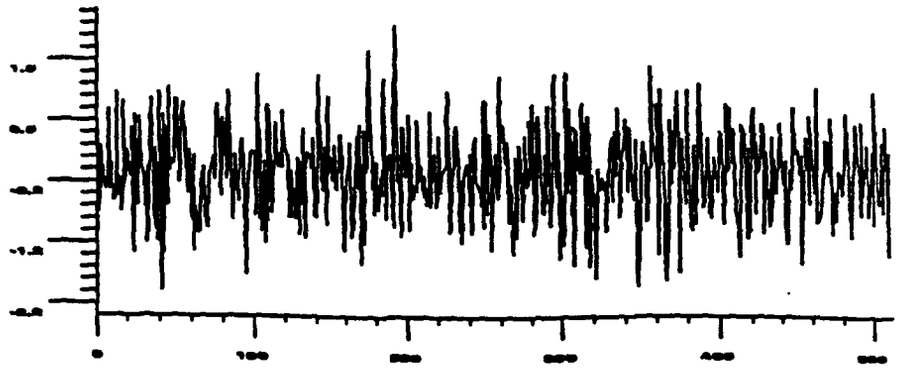


Fig.8(b) Diracs with colored noise. (SNR=-15.09 db)

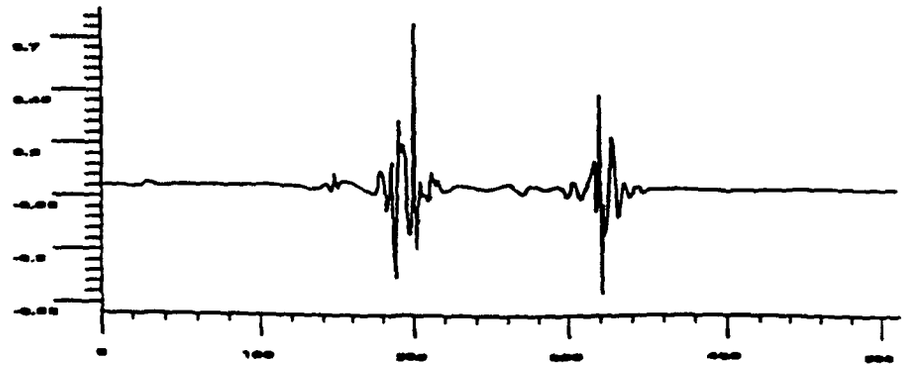


Fig.8(c) Reconstructed signals from the selected structures. (SNR=.21 db)

Copy available to DDC does not
permit fully legible reproduction

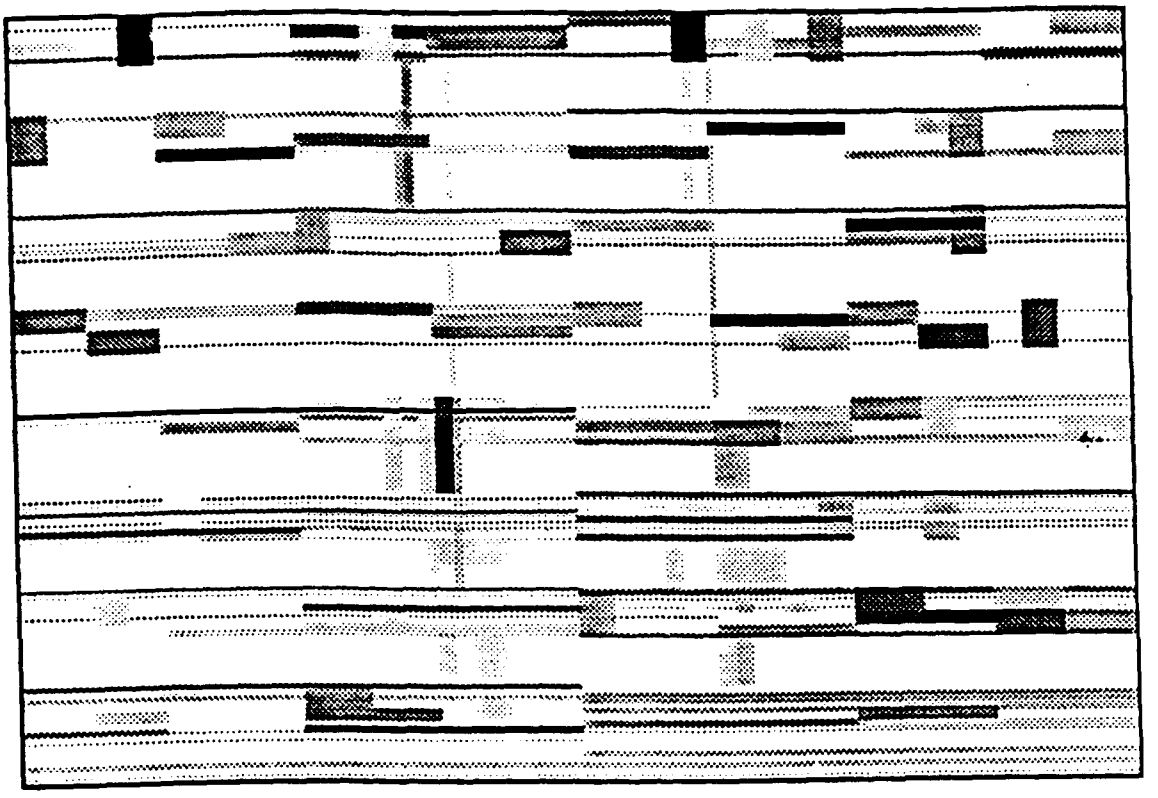


Fig.8(d) Time/frequency representation of the structure book.

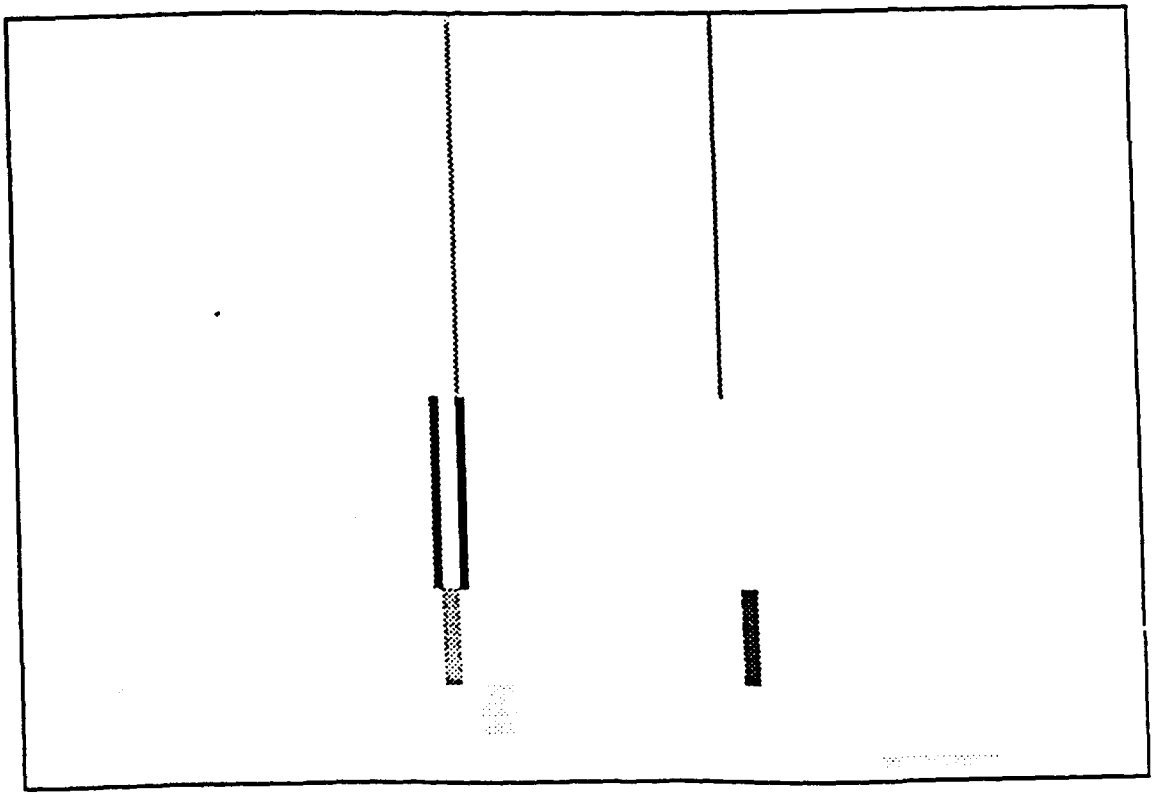


Fig.8(e) Time/frequency representation of the selected structures from

ON VARIABLE LENGTH WINDOWS AND WEIGHTED ORTHONORMAL FUNCTIONS

Bruce W. Suter and Mark E. Oxley

Air Force Institute of Technology

Wright-Patterson AFB, OH 45433

Abstract

A new formulation is presented for the analysis and synthesis of signals. This formulation is composed of a variable width window and a linear combination of weighted orthonormal functions. Tradeoffs in the specification of windows are examined. A sinusoidal example is considered, and a fast algorithm is provided for its evaluation.

This work was supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR-616-92-0019.

I. Introduction

In short time Fourier analysis, a signal is multiplied by a window and then the Fourier transform is computed (see, for example, Oppenheim and Schaefer [1]). The result of this transformation is not uniquely defined unless the window is specified. Towards this end, Harris [2] provides an encyclopedic presentation on windows.

These windows can be chosen to offer a great deal of flexibility for the user, but any windowing process inevitably limits the accuracy of real time spectral estimation. In an effort to overcome some of these limitations, Princen and Bradley [3] presented a technique that utilized a basis made from the product of a window and a sinusoidal function. The generality of their results was limited by the assumptions that the windows would be of constant length and would have fifty percent overlap with adjacent windows.

Using an approach which is conceptually similar to the earlier work of Princen and Bradley, Cassereau [4] introduced and Malver [5] further investigated a technique called the Lapped Orthogonal Transform (LOT). Although Malvar also constrained the windows to be of constant length and to have fifty percent overlap, he [6] was able to obtain perceptual improvements in the coding of speech through the "elimination of noise (extraneous tones)" that was associated with the edge effects of traditional windows. Recently, Akanasu and Wadas [7] applied Malvar's LOT to the coding of images and found that the energy compaction of LOT to be superior to block transforms for all cases considered. Generalizing the work of Malvar, Coifman and Meyer [8] provided conditions for windows of variable length.

Building on this body of knowledge, a more generalized formulation is presented for the analysis and synthesis of signals. This formulation involves a family of weighted

orthonormal bases of functions and a family of windows. Each window may vary in length, in peak amplitude, and in percent overlap with other windows. The following discussion will be limited to one dimensional functions, but the extension to multidimensional results can be achieved, in a straight-forward manner, by representing higher dimensional functions as a tensor product of unidimensional functions, as it is usually done in transform image coding [9].

II. A New Paradigm

We begin by partioning the Real line $\mathbf{R}$ with the strictly increasing sequence $\{a_j | j \in \mathbf{Z}\}$ so that $\mathbf{R} = \bigcup_{j\in\mathbf{Z}}[a_j, a_{j+1}]$. For each $j \in \mathbf{Z}$ let $\{f_{j,k} | k \in \mathbf{N}\}$ denote a real weighted orthonormal basis defined on the interval $I_j = [a_j, a_{j+1}]$ where orthogonality is measured with respect to the weight function $p_j(x)$, that is,

$$\int_{a_j}^{a_{j+1}} f_{j,k}(x) f_{j,l}(x) p_j(x) dx = \delta_{k,l}.$$

At each point a_j we center an interval, namely $[a_j - \epsilon_j, a_j + \epsilon_j]$ with $\epsilon_j > 0$. And to guarantee that this interval does not overlap with the interval centered at a_{j+1} we require $\epsilon_{j+1} + \epsilon_j \leq a_{j+1} - a_j$. Observe the redundancy with the overlapping intervals and $\mathbf{R} = \bigcup_{j\in\mathbf{Z}}[a_j - \epsilon_j, a_{j+1} + \epsilon_{j+1}]$. We define the extensions $\tilde{f}_{j,k}$ by constructing the odd extension of $f_{j,k}$ on $(a_j - \epsilon_j, a_j)$ and the even extension of $f_{j,k}$ on $(a_{j+1}, a_{j+1} + \epsilon_{j+1})$. Specifically, $\tilde{f}_{j,k}$ can

be expressed as

$$\tilde{f}_{j,k}(x) = \begin{cases} 0 & , \ -\infty < x \leq a_j - \epsilon_j \\ -f_{j,k}(2a_j - x) & , \ a_j - \epsilon_j < x < a_j \\ f_{j,k}(x) & , \ a_j \leq x \leq a_{j+1} \\ f_{j,k}(2a_{j+1} - x) & , \ a_{j+1} < x < a_{j+1} + \epsilon_{j+1} \\ 0 & , \ a_{j+1} + \epsilon_{j+1} \leq x < \infty. \end{cases}$$

Let $\tilde{p}_j$ denote the even extensions of p_j about both endpoints, specifically,

$$\tilde{p}_j(x) = \begin{cases} 0 & , \ -\infty < x \leq a_j - \epsilon_j \\ \tilde{p}_j(2a_j - x) & , \ a_j - \epsilon_j < x < a_j \\ p_j(x) & , \ a_j \leq x \leq a_{j+1} \\ p_j(2a_{j+1} - x) & , \ a_{j+1} < x < a_{j+1} + \epsilon_{j+1} \\ 0 & , \ a_{j+1} + \epsilon_{j+1} \leq x < \infty. \end{cases}$$

To simplify notation, let $g_{j,k}(x) = \sqrt{\tilde{p}_j(x)}\tilde{f}_{j,k}(x)$.

Let $w_j(x)$ denote the window function supported on the interval $(a_j - \epsilon_j, a_{j+1} + \epsilon_{j+1})$ having a peak amplitude of A_j. The amplitude-normalized window $\hat{w}_j(x)$ is given by $\hat{w}_j(x) = w_j(x)/A_j$. We choose amplitude-normalized windows $\hat{w}_j(x)$ with the following

properties (see Coifman and Meyer [8]):

$$
\begin{array}{lrcll}
(a) & \hat{w}_j(x) & = & 1 & \text{for} \quad x \in (a_j + \epsilon_j, a_{j+1} - \epsilon_{j+1}) \\
(b) & \hat{w}_j(x) & = & 0 & \text{for} \quad x \notin (a_j - \epsilon_j, a_{j+1} + \epsilon_{j+1}) \\
(c) & \hat{w}_j(a_j - s) & = & \hat{w}_{j-1}(a_j + s) & \text{for} \quad s \in [-\epsilon_j, \epsilon_j] \\
(d) & \hat{w}_j^2(x) + \hat{w}_{j-1}^2(x) & = & 1 & \text{for} \quad x \in [a_j - \epsilon_j, a_j + \epsilon_j]
\end{array}
$$

Now we form new functions $u_{j,k}(x)$ which are supported on the interval $(a_j - \epsilon_j, a_{j+1} + \epsilon_{j+1})$ for each $k \in \mathbf{N}$. Each function is the product of the amplitude-normalized window $\hat{w}_j(x)$ and the symmetric extension $g_{j,k}(x)$, that is, $u_{j,k}(x) = \hat{w}_j(x) g_{j,k}(x)$. A proof of the theorem that $\{u_{j,k} | j \in \mathbf{Z}, k \in \mathbf{N}\}$ is an orthonormal basis for $L^2(\mathbf{R})$ is given in Suter and Oxley [11].

III. Continuously Differentiable Orthonormal Functions

If $f_{j,k}$ is chosen to be continuously differentiable on (a_j, a_{j+1}) then the extension $\tilde{f}_{j,k}$ may not be continuously differentiable on $(a_j - \epsilon_j, a_{j+1} + \epsilon_{j+1})$. From the signal processing point of view, "noise" may be generated as a result of piecing together $\tilde{f}_{j,k}(x)$ with its odd extension at $x = a_j$ and even extension at $x = a_{j+1}$. To minimize this "noise" the following regularity conditions are imposed on $f_{j,k}$ at the endpoints to guarantee that $\tilde{f}_{j,k}$ is continuously differentiable on $(a_j - \epsilon_j, a_{j+1} + \epsilon_{j+1})$.

(a) $f_{j,k}(a_j) = 0$ and $f'_{j,k}(a_j)$ exists

(b) $f_{j,k}(a_{j+1})$ exists and $f'_{j,k}(a_{j+1}) = 0$.

As an example, assume $\tilde{f}_{j,k}(x)$ has the form of a sinusoidal function on $[a_j, a_{j+1}]$, in particular,

$$f_{j,k}(x) = A_{j,k} \sin\left(\omega_{j,k}\left[\frac{x - a_j}{a_{j+1} - a_j}\right]\right)$$

where $A_{j,k}$ and $\omega_{j,k}$ are to be determined. By construction $f_{j,k}(a_{j+1}) = 0$ and both $f_{j,k}(a_{j+1})$ and $f'_{j,k}(a_j)$ exist. The condition $f'_{j,k}(a_{j+1}) = 0$ yields

$$A_{j,k}\left(\frac{\omega_{j,k}}{a_{j+1} - a_j}\right)\cos\omega_{j,k} = 0$$

hence, choose $\omega_{j,k} = \pi(k + 1/2)$ for each $k \in \mathbf{N}$ and $j \in \mathbf{Z}$. We choose $A_{j,k}$ such that $f_{j,k}$ has unit energy norm, so

$$\int_{a_j}^{a_{j+1}} A_{j,k}^2 \sin^2\left(\pi(k + 1/2)\left[\frac{x - a_j}{a_{j+1} - a_j}\right]\right) dx = A_{j,k}^2\left(\frac{a_{j+1} - a_j}{2}\right) = 1$$

yields the orthornormal basis

$$f_{j,k}(x) = \sqrt{\frac{2}{a_{j+1} - a_j}} \sin\left(\pi(k + 1/2)\left[\frac{x - a_j}{a_{j+1} - a_j}\right]\right).$$

Coifman and Meyer presented this orthonormal basis in their recent paper [8]. Notice that the weight function $p_j(x) \equiv 1$ for $j \in \mathbf{Z}$ has a continuously differentiable extension $\tilde{p}_j(x)$.

IV. An Example of Windows

In section II, conditions were given for the amplitude-normalized window function $\hat{w}_j(x)$. In this section, we give an example of $\hat{w}_j(x)$. Assume the form of window in the interval

$[a_{j+1} - \epsilon_{j+1}, a_{j+1} + \epsilon_{j+1}]$ to be

$$\hat{w}_j(x) = \cos(B_j\{x - [a_{j+1} - \epsilon_{j+1}]\}) \text{ for } a_{j+1} - \epsilon_{j+1} \leq x \leq a_{j+1} + \epsilon_{j+1}$$

which, by construction, satisfies the condition $\hat{w}_j(a_{j+1} - \epsilon_{j+1}) = 1$. At the other endpoint, we require $\hat{w}_j(a_{j+1} + \epsilon_{j+1}) = 0$. Hence,

$$\cos(B_j 2\epsilon_{j+1}) = 0$$

implies the choice $B_j = \pi/4\epsilon_{j+1}$ for each $j \in \mathbf{Z}$. therefore

$$\hat{w}_j(x) = \cos\left(\frac{\pi}{4\epsilon_{j+1}}\{x - [a_{j+1} - \epsilon_{j+1}]\}\right) \text{ for } a_{j+1} - \epsilon_{j+1} \leq x \leq a_{j+1} + \epsilon_{j+1}.$$

Invoking symmetry, the window function becomes

$$\hat{w}_j(x) = \begin{cases} 0 & , \ -\infty < x \leq a_j - \epsilon_j \\ \sin\left(\frac{\pi}{4\epsilon_j}\{x - [a_j - \epsilon_j]\}\right) & , \ a_j - \epsilon_j \leq x \leq a_j + \epsilon_j \\ 1 & , \ a_j + \epsilon_j \leq x \leq a_{j+1} - \epsilon_{j+1} \\ \cos\left(\frac{\pi}{4\epsilon_{j+1}}\{x - [a_{j+1} - \epsilon_{j+1}]\}\right) & , \ a_{j+1} - \epsilon_{j+1} \leq x \leq a_{j,+1} + \epsilon_{j+1} \\ 0 & , \ a_{j+1} + \epsilon_{j+1} \leq x < \infty. \end{cases}$$

The advantage of this window is its simple implementation. The disadvantage of this window is that it is not differentiable at $x = a_j - \epsilon_j$ and $x = a_{j+1} + \epsilon_{j+1}$. Examples of windows which are differentiable at the endpoints are given in Suter and Oxley [11].

V. Coefficient Evaluation

Let $s(x)$ be a measured signal with finite energy, that is, $s \in L^2(\mathbf{R})$. Expanding $s(x)$ in terms of the orthonormal basis $\{u_{j,k} | j \in \mathbf{Z}, k \in \mathbf{N}\}$ yields

$$s(x) = \sum_{j \in \mathbf{Z}} \sum_{k \in \mathbf{N}} \alpha_{j,k} u_{j,k}(x)$$

where $\alpha_{j,k} = \langle s, u_{j,k} \rangle$. The coefficients $\alpha_{j,k}$ can be rewritten as (see [11] for the details)

$$\alpha_{j,k} = \int_{a_j}^{a_{j+1}} h_j(x) f_{j,k}(x) \sqrt{p_j(x)} dx,$$

where

$$h_j(x) = \begin{cases} s(x)\hat{w}_j(x) - s(2a_j - x)\hat{w}_j(2a_j - x) & , \ a_j \leq x \leq a_j + \epsilon_j \\ s(x)\hat{w}_j(x) & , \ a_j + \epsilon_j \leq x \leq a_{j+1} - \epsilon_{j+1} \\ s(x)\hat{w}_j(x) + s(2a_{j+1} - x)\hat{w}_j(2a_{j+1} - x) & , \ a_{j+1} - \epsilon_{j+1} \leq x \leq a_{j+1}. \end{cases}$$

VI. Coefficient Evaluation for Sinusoidal Functions

This section is a summary of the algorithm required to generate the coefficients of input data function using the sinusoidal basis example of section III. The derivation of the following algorithm is provided in [11].

Assume that the signal $s(x)$ is sampled at the rate $\delta > 0$. Let a_j be chosen at the sampled x's so that $a_{j+1} - a_j = \delta N_j$ where N_j is a positive integer. Thus, there are N_j samples taken in the interval $(a_j, a_{j+1}]$, and we denote $x_{j,l} = a_j + \delta l$ for $l = 0, 1, ..., N_j$. Choose $\epsilon_j = \delta M_j$ where M_j is also a postive integer. The nonoverlapping condition, $\epsilon_{j+1} + \epsilon_j \leq a_{j+1} - a_j$,

implies $M_{j+1} + M_j \leq N_j$. For each $j \in \mathbf{Z}$ we perform the following steps:

(1) Obtain data $s(x_{j,l})$ for $a_j - \epsilon_j \leq x_{j,l} \leq a_{j+1} + \epsilon_{j+1}$,

that is, $l = -M_j, ..., 0, 1, ..., N_j + M_{j+1}$.

(2) Multiply $s(x_{j,l})$ by window $\hat{w}_j(x_{j,l})$ for $l = -M_j, ..., 0, 1, ..., N_j + M_{j+1}$.

(3) Fold in the sequence by defining

$$H_{j,l} = \begin{cases} s(x_{j,l})\hat{w}(x_{j,l}) - s(2a_j - x_{j,l})\hat{w}_j(2a_j - x_{j,l}) & , \ a_j \leq x_{j,l} \leq a_j + \epsilon_j \\ s(x_{j,l})\hat{w}_j(x_{j,l}) & , \ a_j + \epsilon_j \leq x_{j,l} \leq a_{j+1} - \epsilon_{j+1} \\ s(x_{j,l})\hat{w}_j(x_{j,l}) + s(2a_{j+1} - x_{j,l})\hat{w}_j(2a_{j+1} - x_{j,l}) & , \ a_{j+1} - \epsilon_{j+1} \leq x_{j,l} \leq a_{j+1} \end{cases}$$

for $l = 0, 1, ..., N_j$.

(4) Define new array

$$\beta_{j,l} = H_{j,l} \sin\left(\frac{\pi l}{2N_j}\right) \quad \text{for} \quad l = 0, 1, \ldots, N_j \ .$$

(5) Define the even extension of $\beta_{j,l}$

$$\tilde{\beta}_{j,l} = \begin{cases} \beta_{j,l} & , \ 0 \leq l \leq N_j - 1 \\ \beta_{j,2N-l} & , \ N_j \leq l \leq 2N_j - 1. \end{cases}$$

(6) Perform an FFT of length $2N_j$ on $\tilde{\beta}_{j,l}$ (see for example,Ferguson [10]) and define

$$D_{j,k} = \mathcal{R}e\left(\frac{1}{2N_j} \sum_{l=0}^{2N_j-1} \tilde{\beta}_{j,l} e^{i2\pi kl/2N_j}\right).$$

(7) Interpret the results of the FFT

$$\begin{aligned} \alpha_{j,0}^T &= \sqrt{2\delta N_j} D_{j,0} \\ \alpha_{j,k}^T &= \alpha_{j,k-1}^T + 2\sqrt{2\delta N_j} D_{j,k} \end{aligned}$$

$\alpha_{j,k}^T$ approximates the coefficients $\alpha_{j,k}$.

VII. Reconstruction of Signal Using Sinusoidal Functions

In this section we give a summary of the algorithm to reconstruct the signal using the sinusoidal basis example of section III. The derivation of this algorithm is provided in [11].

Assume we have the set of coefficients $\{\alpha_{j,k} | j \in \mathbf{Z}, k \in \mathbf{N}\}$. We wish to reconstruct the signal $s(x)$ at the values of $x_{j,l} = a_j + \delta l$ for each $j \in \mathbf{Z}$ and $l = 0, 1, ..., N_j$. (See section VI.) For each $j \in \mathbf{Z}$ perform the following steps:

(1) Define odd extension of $\alpha_{j,k}$ for $k = 0, 1, 2, \ldots, N_j - 1$ by

$$\tilde{\alpha}_{j,k} = \begin{cases} \alpha_{j,k} & , \ 0 \le l \le N_j - 1 \\ -\alpha_{j,2N_j-k-1} & , \ N_j \le l \le 2N_j - 1 \end{cases}$$

(2) Perform a FFT of length $2N_j$ on $\tilde{\alpha}_{j,k}$ for $k = 0, 1, 2, \ldots, 2N_j - 1$ and produce the sequence

$$H_{j,l} = \sqrt{\delta 2N_j}\, \mathcal{I}m \left(e^{i\frac{\pi l}{2N_j}} \left[\frac{1}{2N_j} \sum_{k=0}^{2N_j-1} \tilde{\alpha}_{j,k} e^{i\frac{2\pi kl}{2N_j}} \right] \right)$$

(3) Reconstruct the signal on the interval $[a_j, a_{j+1}]$ at the data points $x_{j,l} = a_j + \delta l$ for $l = 0, 1, 2, \ldots, N_j$ by using

$$S_{j,l} = \begin{cases} \dfrac{H_{j-1,l}}{2\hat{w}_{j-1}(x_{j,l})} & , \quad x_{j,l} = a_j \\ \hat{w}_j(2a_j - x_{j,l})H_{j-1,N_{j-1}-l} - \hat{w}_j(x_{j,l})H_{j,l} & , \quad a_j < x_{j,l} < a_j + \epsilon_j \\ \dfrac{H_{j,l}}{\hat{w}_j(x_{j,l})} & , \quad a_j + \epsilon_j \leq x_{j,l} \leq a_{j+1} - \epsilon_{j+1} \\ \hat{w}_j(x_{j,l})H_{j,l} - \hat{w}_j(2a_{j+1} - x_{j,l})H_{j+1,N_j-l} & , \quad a_{j+1} - \epsilon_{j+1} < x_{j,l} < a_{j+1} \end{cases}$$

The values $S_{j,l}$ will approximate the signal evaluated at $X_{j,l}$, that is, $S_{j,l} \approx s(x_{j,l})$.

VIII. Conclusions

A formulation was presented for the analysis and synthesis of signals. This formulation permitted (a) a variable window length, (b) a variable percent overlap between windows, and (c) an arbitrary orthonormal basis inside the analysis interval. A sinusoidal example was examined and a fast algorithm were provided for both coefficient evaluation and signal reconstruction. It is important to note the complexity of both the spectrum generation and reconstruction are of the same complexity as the Fast Fourier Transform. Future planned work include (a) utilization of polynomial basis functions and (b) application of this general approach to the analysis of speech signals and the synthesis of computer graphics.

IX. References

[1] A.V. Oppenheim and R.W. Schaefer, Discrete Time Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[2] F.J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," Proceedings of IEEE, Vol. 66, p.51-83, January 1978.

[3] J.P. Princen, and A.B. Bradley, "Analysis/Synthesis Filter Band Design Based on Time Domain Cancellation," IEEE Transactions on Acoustics Speech, and Signal Processing, Vol. ASSP-34, p. 1153-1161, October 1986.

[4] P. Cassereau, "A new class of optimal unitary transforms for image processing," S.M. thesis, Dept. Elec. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, May 1985.

[5] H.S. Malvar and D.H. Staelin, "The LOT: Transform Coding Without Blocking Effects," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-37, p.553-559, April 1989.

[6] H.S. Malvar, "Lapped Transforms for Efficient Transform/Subband Coding," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-38, p.969-978, June 1990.

[7] A.N. Akanasu and F.E. Wadas, On Lapped Orthogonal Transforms, IEEE Transactions on Signal Processing, Vol. SP-40, No.2, p.439-443, February 1992.

[8] R. Coifman et Y. Meyer, "Remarques sur l'analyse de Fourier à fenêtre," C.R. Acad. Sci. Paris, t.312, Serie I, p.259-261, 1991.

[9] R.J. Clarke, Transform Coding of Images, Academic Press, London, 1985.

[10] W.E. Ferguson, "A simple derivation of Glassman's general N fast Fourier transform," Computers and Mathematics with Applications, Vol. 8, No.6, p.401-411, 1982.

[11] B.W. Suter and M.E. Oxley, "On Variable Overlapped Windows and Weighted Orthonormal Bases," IEEE Transactions on Signal Processing, submitted for publication April 1992.

ALPHATECH, Inc.

MULTI-RESOLUTION ESTIMATION FOR IMAGE PROCESSING AND FUSION

Robert. R. Tenney[1]

Alan S. Willsky[2]

This paper introduces a set of estimation algorithms based on multi-resolution models of random fields. The models of interest include statistical representations of terrain and other geophysical phenomena. Their structure employs a series of successively finer representations of the process. The estimation algorithms exploit this structure to combine information from different areas, perhaps at different resolutions, into updated estimates of the process variables. The algorithms extend recent work on estimation for stochastic tree processes; the descriptions of the images of interest require a somewhat more general structure than existing tree processes permit. We present a general theory of estimation on acyclic graphs, and specialize the results to one and two dimensional processes where the scale-to-scale relationships are midpoint deflection processes. Applications of this work will include anomoly detection, change detection, segmentation, and reconstruction with data from imaging sensors.

ACKNOWLEDGEMENT

Dr. Jon Sjogren of the US Air Force Office of Scientific Research sponsored the work reported in this paper, under contract F49620-91-C-0047.

OVERVIEW

Real-time video sensor technology has become quite mature, and therefore relatively inexpensive and widely used. Computing power continues to drop in cost, particularly when an algorithm can be structured to match massively parallel architectures. However, algorithm technology has not yet matured to the point where that computing power can be applied to real-time video processing for much more than simple image enhancement.

In addition, newer imaging techniques exploit very different physical phenomena which, while not providing the clarity expected from video sensors, can provide information about important phenomena that otherwise would be extremely difficult to observe. The lower signal to noise ratios of these sensors demand algorithms based on statistical and dynamic techniques, instead of the prevalent, largely deterministic approaches to feature extraction and identification.

The purpose of this effort was to develop efficient estimation algorithms to recover important information from real-time video imagery – information which is vital to subsequent monitoring or control applications. To achieve this goal, the algorithms must meet five criteria. First, they should have an *efficient computational structure*, measured in terms of the total number of numerical operations required. Second, they should have a *high degree of structural regularity* which can be mated with parallel computing architectures. Third, they should be *based on explicit models* of the physical processes which generate the imagery of interest. Fourth, those *models should be statistical in nature*, as Bayesian probability theory remains the most complete mathematical representation of uncertainty. Finally, the point of departure for this work is that those *models have a multi-scale structure* — where large structures find representation at coarse scales of the model, and local structures appear at the finer scales.

The mathematical approach taken here is more general than necessary, in order to remove any dependencies on the particular topology of the multi-scale model. The basic estimation theory derives from the properties of Bayesian networks, where a network representation highlights direct

[1] ALPHATECH, Inc., 15 Mall Road, Burlington, MA 01803 (617) 273-3388 x227

[2] Room 35-437, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139

statistical connections among an arbitrary set of variables. The application of the basic theory to image processing then becomes an exercise in representing the scale-to-scale dependencies of an image in terms of these networks.

The most evocative application of the class of models investigated here is in environmental or terrain reconstruction. The same models studied here have been widely used to generate synthetic, fractal landscapes in the training and entertainment industries. A slight change in a generating function permits synthesis of images with a granular or crystalline structure. This work addresses two questions: 1) *are multi-scale models able to represent real physical processes of interest?*, and 2) *do they lead to computationally efficient estimation algorithms?* The answer to both is "yes", and the major product of this work is *a foundation for algorithms which estimate the parameters which characterize the process.*

OBJECTIVES

Conventional image processing techniques operate directly on the spatial, pixel-level data of an image. Alternative techniques, inspired by the multi-scale functional bases of the affine wavelet transform, represent the image at successively finer levels of detail, and allocate processing power across scale as well as space. Potential advantages of these techniques include 1) the ability to *explicitly represent processes which truly are a product of multi-scale phenomena*, such as ocean waves; 2) the ability to *process data which comes from sources of different resolution* (scale), such as synthetic aperture radar and overhead infrared photography; and 3) the ability to *obtain computational advantages* from the allocation of processing power to appropriate scales and spatial areas, such as for image segmentation.

Model-based approaches to signal processing require a model of the process of interest, which includes (hypothesized) causal connections among the random elements of the signal. Multi-scale models range from the simple (where the random elements are coefficients on a set of dyadic, affine wavelet basis functions) to the complex (where successively finer scales of the signal representation are complicated functions of variables expressed at coarser scales).

Model-based techniques will only be successful if the resulting algorithms are *accurate* and *efficient*. Accuracy results from an inherent compatibility between the structure of the model and the structure of the process of interest. Efficiency is largely the result of regularity in the computational structure of the algorithms which are derived from the model. The chances of establishing regularity in an algorithm are greatly enhanced if the model has some degree of regularity as well.

Given this context, the purpose of this effort is to *investigate whether multi-scale models of image production could lead to accurate and efficient estimation algorithms for image processing*, with the long-term goal of embedding such algorithms into real-time image processing systems for image enhancement, object detection, or parameter estimation. Accuracy results from an approach which calculates the optimum estimate of each variable of interest, so that performance degradations result only from model mismatches and fundamental physical principles, not from approximations built into the solution algorithm. Efficiency results from the regular structure of a multi-scale model, where each refinement layer has the same structure as any other; layers differ only in numerical size and in scale-related parameters. To maintain accuracy, the mathematical foundation adopted for this work is Bayesian probability theory; to maintain efficiency, the multi-scale models have the same structure as both affine wavelet decompositions and mid-point deflection models of synthetic terrain, both of which lead to exceedingly fast computational techniques in other applications [2, 3, 5, 6, 18, 19, 27].

MODELING APPROACH

This section presents the rationale for investigating multi-scale models of multi-dimensional random fields as the basis for a new class of estimation algorithms. Subsequent sections will provide an overview of the mathematical elements of the models, key issues that drive estimation algorithm design, and examples of the processes that can be treated in this framework.

Models

Mathematical model building has been an essential element of scientific progress ever since the Renaissance. Building a model of a physical process that closely replicates observed behavior is substantial evidence that the process is in fact well understood. Models take many different forms, and it is difficult to generalize about them, but there appears to be an interesting shift in modeling perspective taking place at the end of the twentieth century.

Prior to Newton, models were largely static: a collection of relationships and numerical invariants. Newton and Leibnitz established a line of modeling techniques that is even more useful today than in the seventeenth century. Differential equations, and their stochastic counterparts, have led to magnificent successes in the analysis, estimation, and control of a myriad of physical processes. Their applicability to image processing problems, however, has been limited by two factors. First, prevalent digital computing techniques impose limitations, based on numerical effects, on the range of scale which can be employed in algorithms based on this class of model — the choice of step size in numerical integration being but a simple example of the tradeoffs imposed. Second, their extension to multi-dimensional processes, i.e. partial differential equations and their stochastic counterparts, has not yielded a class of models whose sample functions appear realistic for many applications contexts.

Fourier solidified an alternate view of physical processes based not on incremental change over time, but on composition of a large collection of elements which extend across all time. Representing a signal as a weighted superposition of basis elements immediately allows a wide variety of scales to be represented and, at least for one-dimensional signals, allows one to synthesize samples with a high degree of verisimilitude for many applications. Even better, it leads to extremely fast digital techniques for both synthesis and analysis. Again, there are limitations. The fundamental reliance on linearity restricts the class of phenomena which can be captured by these models. Also, one cannot use these models as synthesis tools for two-dimensional images — *the set of transform coefficients* which correspond to realistic images, at least in the optical spectrum, cannot be easily characterized.

Common to both perspectives is the fact that models built on differential or spectral structures have a great deal of difficulty synthesizing realistic images of common physical entities such as terrain. *If a model cannot reliably synthesize samples which are representative of the class of images of interest, the risk of model mismatch leading to ineffective algorithms is undoubtedly high.* Nonetheless, many of today's best image processing techniques do in fact rely on various combinations of differential or spectral techniques, although not always based on a common, clearly specified model.

What alternative is there? The emerging field of wavelets [6, 12, 13, 14, 22, 23] can be viewed as yet another superpositional model, offering an enlarged set of basis functions to represent non-stationary effects. This view is much too restrictive, however. The affine wavelet transform, and the multi-scale models it implies [17, 35], opens the door to an entirely new kind of causality based on neither incremental change nor superposition: *causality from scale to scale.* To fully exploit this class of models without inheriting the weaknesses of other spectral models requires some modification, however. One key modification is to introduce *explicit coupling* between the wavelet coefficients at successive scales, coupling that effectively reduces the enormous number of degrees of freedom presented by wavelet models.

The resulting multi-scale models lead to a fundamentally different perspective on a random process. They capture a notion of successive refinement: the fine features of the process may be dependent on the presence or absence of coarse features, but not the reverse. These models explicitly represent some form of *scale-to-scale causality*, and *they need not be linear.* Multi-scale techniques have proven extremely efficient in other applications, so empirical evidence suggests that estimation techniques inferred from a multi-scale model should also be extremely computationally efficient.

Example

Consider simple, one-dimensional Brownian motion. The Ito calculus provides a precise description of this stochastic process based on an incremental model. Weiner provided a model of the same process based on the superposition of sinusoids with random phase. What is a multi-scale model of the same process?

The simplest multi-scale model of Brownian motion employs the (non-orthonormal) wavelet basis illustrated in Figure 1. A series of approximations to the sample path are constructed by adding weighted translates of the basis function at successively finer scales. The basis functions are continuous, so each approximation must be continuous. If the weights are drawn from Gaussian distributions, then the joint distributions between all pairs of points in any approximation are Gaussian also. If the Gaussian distributions for the weights are shift-invariant, and have variances proportional to the length of the support set of the relevant basis function, then the statistics of the approximations approach those of Brownian motion. In fact, *they are identical to Brownian process statistics* at values of the independent variable s that map into the peaks of the basis functions at any scale n or coarser (since all basis functions at finer scales are zero at these points).

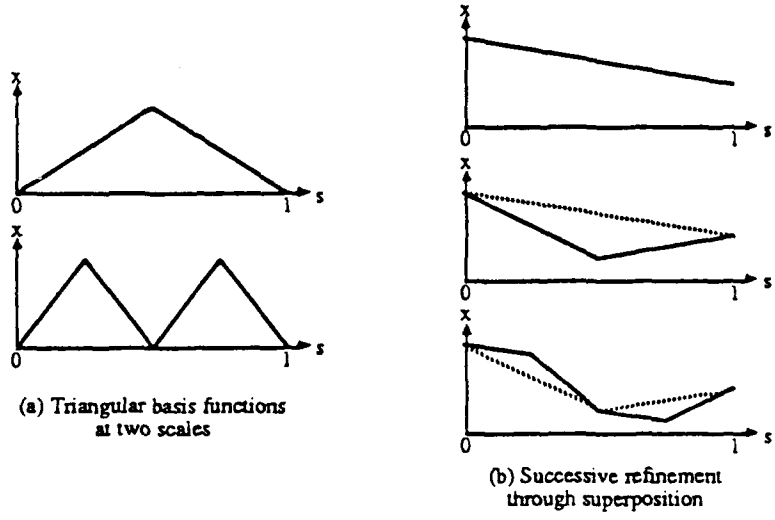


Figure 1: Wavelet Model of Brownian Motion. The independent variable, s, lies in the unit interval, and may be interpreted as time. The process value at s is x(s). (a) The "tent" functions of the basis set; the entire basis set consists of all dyadically scaled and integrally translated versions of the tent. (b) Construction of a sample path of Brownian motion by superposition of weighted basis functions. Weights for each basis function must be independent of one another and drawn from zero-mean Gaussian distributions which are stationary in s, and which have variances that decrease by a power of 2 at each successively finer scale.

This is a wavelet interpretation of a class of *midpoint deflection processes* which have been used to construct fractal signals [15, 24, 31, 33, 37, 38]. At each scale, the model constructs a value for the midpoint of the process over a set of equal intervals. At the next scale, it builds midpoints for each half-interval of the intervals at the preceding, coarser scale. The linear tails of the tent function simply interpolate values that have not yet been completely defined — the value of the process at a point that is not a midpoint of an interval is simply a linear combination of the process values at the neighboring points which have been defined.

Ignore the interpolation, and focus just on that finite set of points which have been completely defined by a finite number of scales. These points are at values of s which are integral multiples of a negative power of 2. The midpoint of an interval between two neighboring values of s at one scale can be directly constructed: average the values of x at the endpoints (interpolate), and

then add an amount drawn from a Gaussian distribution. Repeating this process leads to a *midpoint construction process*, where the stochastic process is defined over an expanding set of points which, in the limit, are countable but dense in the set of real numbers.

Figure 2 shows the data dependencies embedded in such a midpoint construction process. Assuming that x(0) and x(1) are given as boundary conditions, a sample of the process can be generated by successively subdividing intervals, and synthesizing a value for the midpoints of those intervals. For Brownian motion, it suffices to make the computation of each midpoint dependent only on the process values at its interval's endpoint, and on a draw from an independent, Gaussian random variable.

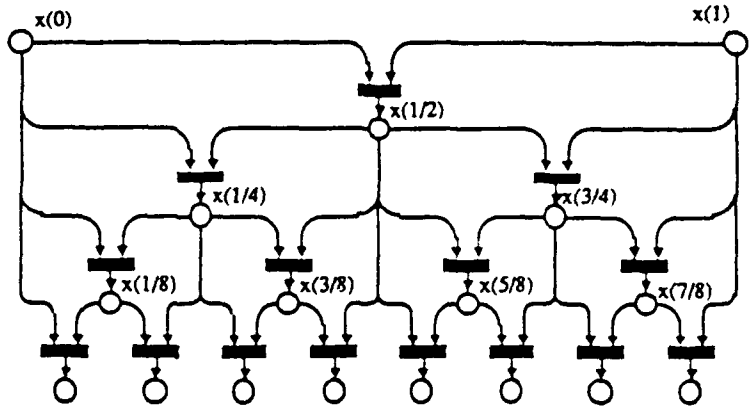


Figure 2: Data Dependencies in a Midpoint Construction Model. Circles denote values of the process at successive interval midpoints. Rectangles denote computations that construct the midpoints at the next level of refinement. Arcs indicate the process values on which these computations may depend. The output of each rectangle may also depend on a random value drawn from some specified distribution. For Brownian noise, these computations consist of linear averaging (with equal weights) to which a draw from a zero-mean Gaussian distribution is added. The variances of the distributions decrease by a factor of 2 from scale to scale.

Figure 3 shows the representation, at the 8 coarsest scales, of one sample path of a linear midpoint construction process. Viewed as a synthesis model, generation of such a sample path is extremely efficient due to the sparse nature of the dependencies. In fact, the dependency diagram of Figure 2 clearly depicts the *structural self-similarity* of the process, as the decomposition of any interval at one scale proceeds exactly as does the decomposition of any other interval at any other scale. Adding the requirement that the variances of the Gaussian distributions be proportional to the length of the interval being bisected assures that the process is *statistically self-similar* as well (and that the auto-correlation function of the process matches that of a Brownian process, at least for the values of s for which samples have been constructed).

Of course, *there is no reason for the midpoint computations to be restricted to a linear-Gaussian structure*. Nor need the values of the process at each value of s be restricted to a scalar. For synthesis, evaluation of discrete, nonlinear, or vector-valued functions is little more complicated than evaluation of a linear function. *This is the point at which our work departs from the limitations of multi-scale models supported by wavelet representations.*

Viewed as an analysis model, recovery of the representation of a sample path at coarser scales from fine scales is even more simple. The representation at scale n is simply decimated, by a factor of 2, to form the representation at the next coarser scale.

Viewed as the basis for estimation techniques, it is unclear whether models of the form shown in Figure 2 offer computational advantages — although similar tree-structured models have led to substantial speedups [4, 8, 9, 11, 32, 35]. (This is precisely the question which this work answers, in the affirmative). Estimation techniques will be simplest when the midpoint construction functions are linear-Gaussian, but even these can be used as the basis for linearized techniques when nonlinearities appear. (The results developed here deal with arbitrary probability

distributions, but only the linear-Gaussian case is known to lead to an exact, finite-dimensional implementation.)

The superficial similarity of samples of the two-dimensional analog of this model to terrain cross-sections has long been recognized [25, 28, 29, 30]. (It is not entirely appropriate as a terrain model, as the construction process does not prevent isolated depressions which, in real terrain, would be filled with lakes.) Nonetheless, this work is based on the supposition that *multi-scale models generate much more realistic representations of terrain than can be efficiently generated by either differential or superpositional approaches, particularly when the generating functions are allowed to be nonlinear*. Moreover, *the self-similar structure of the model which lies behind this sample impart a self-similar structure to the estimation algorithms derived from it*, leading to extremely efficient reconstruction techniques.

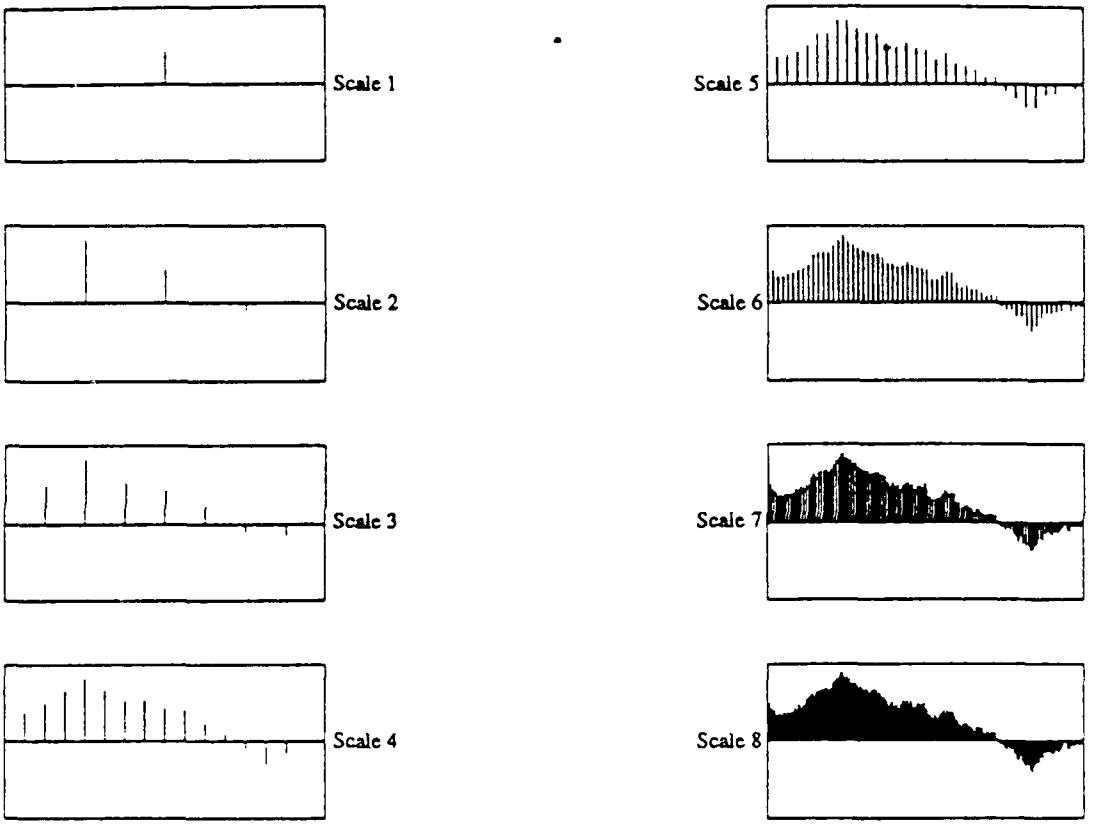


Figure 3: Refinement of A Sample Path from a Linear Midpoint Construction Process. Midpoints are constructed as linear averages of neighboring endpoints, to which noise is added. This sample uses triangular distributions for these increments in lieu of the computationally more expensive Gaussian distribution.

Thus the contribution of this work is *the extension of optimal estimation theory to multi-scale models* which, at the very least, can realistically represent terrain and other fractal textures.

MATHEMATICAL APPROACH

To avoid unnecessary dependencies on the specifics of a problem, the technical approach is based on models described as acyclic nets of random variables. One example of such a network is that of Figure 2 for one-dimensional processes; another appears later for square tessellations on a two-dimensional random field. Because the results apply to any acyclic net, a foundation exists for immediate transfer to other sensor topologies, such as triangular or hexagonal covers of an image.

Markovian Nets

The key assumptions that lie behind the results described in this work can be illustrated with Figure 4. As a synthesis tool for random processes, this network describes a partially ordered set of (pseudo)random computations. Some variables are given, or are drawn from initially specified distributions. These variables lie in the minimal places (with respect to the partial order defined by the directed arcs) of the acyclic graph.

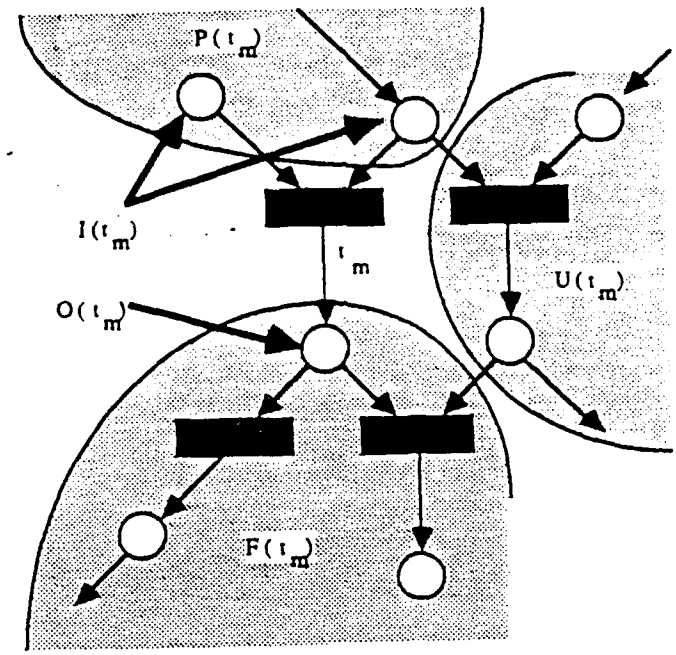


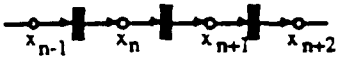
Figure 4: Parts of a Markovian Net. Circles denote *places* which hold the basic random variables of the process. Rectangles denote *transitions*, such as t_m, which represent stochastic operators that transform variables in input places to variables in a unique output place. Directed arcs specify the inputs $I(t_m)$, and output $O(t_m)$, of each transition. The bipartite graph must be acyclic, so the arcs also impose a partial order on the places and transitions. Those places which precede t_m in this partial order are in its past, $P(t_m)$; those which succeed it are in its future, $F(t_m)$; and every other place is in the unordered set $U(t_m)$.

All other places contain random variables derived from these initial variables. Each transition is characterized by a probability distribution on the random variables in its output places, conditioned on the values of the variables in the input places. In a synthesizer of a (pseudo) random process, this mechanism could be implemented either as a random draw from that conditional probability distribution, or as some deterministic operator applied to the values of the inputs and some exogenous random variable. In either case, a conditional probability distribution on the output variables, given the input variables, characterizes the statistical relations imposed by each transition.

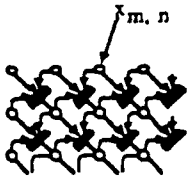
The key mathematical assumption required by this work is that the output variables of a transition depend only on the inputs and independent, exogenous randomness. Formally, this requires an independence assumption to the effect that the values of the outputs of a transition are equally predictable whether (1) only the inputs to the transition are known, or (2) each and every variable in the past and unordered sets of places (see Figure 4) are known. Clearly (1) is a (small) subset of (2). This independence assumption is analogous to the Markovian property required of a state-space description of a conventional Markov process; hence these graphs are called Markovian nets.

Examples

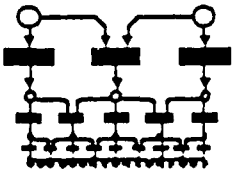
Figure 5 contains three examples of random processes depicted as Markovian nets. These serve as familiar cases for illustration, and for which estimation algorithms are known (at least for the first two examples). The algorithms developed in this work reduce to those known algorithms for the special cases of Figure 5(a) and (b).



a) Standard discrete time Markov process



b) Discrete index, quadrant causal Markov random field



c) Multiscale successive refinement process

Figure 5: Common Random Processes Expressed as Markovian Nets. a) Discrete time Markov process, where the random variable in each place is the state of the process at a given time index. b) Discrete two-dimensional Markov random field with a generator that specifies the field value in terms of three neighbors. The upper and leftmost field samples are taken as boundary conditions. c) Multi-scale model of a one-dimensional process. This structure is highly replicated, but is more general than that of Figure 2.

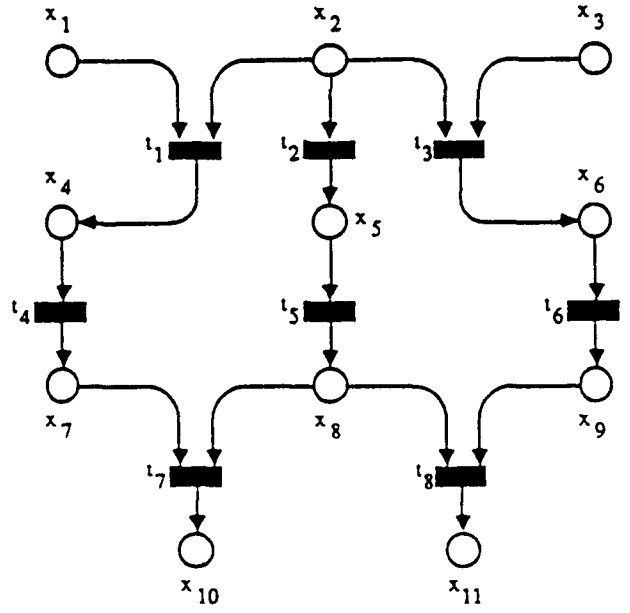


Figure 6: A Simple Markovian Net. Arcs denote explicit statistical dependences; many implicit dependencies must factor into estimation algorithm design as well. For example, x_{10} and x_{11} share common dependencies on x_2 and x_3. Special cases of this net useful for checking results include those where transitions simply copy all inputs to the output, or where outputs are independent of the inputs.

Figure 6 shows a small Markovian net useful for illustrating the fundamental issues to be addressed by estimation algorithms. It resembles a fragment of the net in Figure 5(c), with four "scales" of data. Of particular importance is the fact that x_{10} and x_{11} share statistical dependencies

at the coarsest scale (x_2), an intermediate scale (x_5), and a fine scale (x_8). When a measurement of x_{11} becomes available, the estimate of x_{10} must be somehow revised. If estimation algorithms are to exploit the regularity of multi-scale models, then the algorithm must strip out some of the fine level detail as an update to the estimate of x_8, pass some remaining information back to the intermediate scale to update the estimation of x_5, and then to the coarsest scale to update the estimation of x_2. Moreover, it must then construct an estimate for x_{10} which properly combines the revised x_2, x_5, and x_8 estimates. To make matters more complex, it must separate information about x_2 from information about x_3, both of which influence the measurement of x_{11}.

This example will carry through the following discussion of the general results of this effort; the algorithms described there will be transferred to the much more complex structures required for multi-scale models of one- and two- dimensional stochastic processes in a later section.

GENERAL RESULTS

The objective of an estimation algorithm is to compute estimates of imperfectly known random variables in response to a measurement taken of one of them. Since Markovian nets use a probabilistic description of uncertainty, these estimates ideally come from properly updated, conditional probability distributions. The algorithms which result from this work specify how to compute those distributions, *regardless of any specific structural assumptions* (e.g., self-similarity, linearity, or Gaussianness). The algorithms are basically straightforward applications of Chapman-Kolmogorov prediction, and Bayesian update, equations to a set of probability distributions.

The key research issue addressed in this work was not how to update distributions, but rather *what is the proper set of probability distributions to manipulate?* The unimaginative approach to estimation on Markovian nets works with the joint distribution on all of the random variables. This approach fails to exploit any simplifications made possible by the topology of the net itself — and it is well known that much simpler approaches are available for some of the network topologies shown in Figure 5.

Two important algorithms resulting from this work identify the proper set of distributions to manipulate. The first algorithm finds a set of distributions sufficient to allow one to construct estimates of all of the random variables given distributions on variables in the minimal (initial) places of the graph, and the transition probability distributions. This set is also adequate for updating distributions based on a measurement of one of the initial variables. The second algorithm augments the first set with some additional distributions required to backpropagate information from an intermediate or terminal place to other random variables. This section presents neither the graph theoretic algorithms which determine the requisite set of distributions, nor the estimation algorithms themselves; its objective to convey the types of manipulations performed by the algorithms and an intuitive justification for them.

Prediction

Given a Markovian net, what are the best estimates of the random variables that can be made prior to a measurement? This is the *prediction* problem, and is answered in the fullest by providing *a probability distribution for each of the random variables in the net.* Such a probability distribution is not given as part of the net; it must be derived from the distributions on the variables in the minimal places, and from the transition probability distributions.

The example net in Figure 6 shows why computation of these probability distributions is not immediate. Consider x_{11}. It is a function of both x_8 and x_9, and perhaps some exogenous noise. To compute the distribution on x_{11}, one must in general have the joint distribution on x_8, x_9, and the noise. If these are all independent, one can construct the joint distribution in a simple manner. In this case, they are not: x_8 and x_9 share a common dependence on x_2, so (except in degenerate cases) they are not independent. Hence one must compute the joint distribution on both x_8 and x_9 in order to arrive at a distribution for x_{11}.

Figure 7 shows the set of probability distributions which must be computed as intermediate steps towards finding prior distributions on each of the random variables in the net. It shows a sequence of computations of probability distributions which is sufficient to derive estimates of each random variable in the Markovian net. For example, the right half of the diagram shows the intermediate steps needed to compute a distribution on x_{11}. *The first major product of this effort is a graph theoretic algorithm to derive such a net for any given Markovian net*, along with the equations specifying each computation in terms of basic manipulations on arbitrary probability distributions.

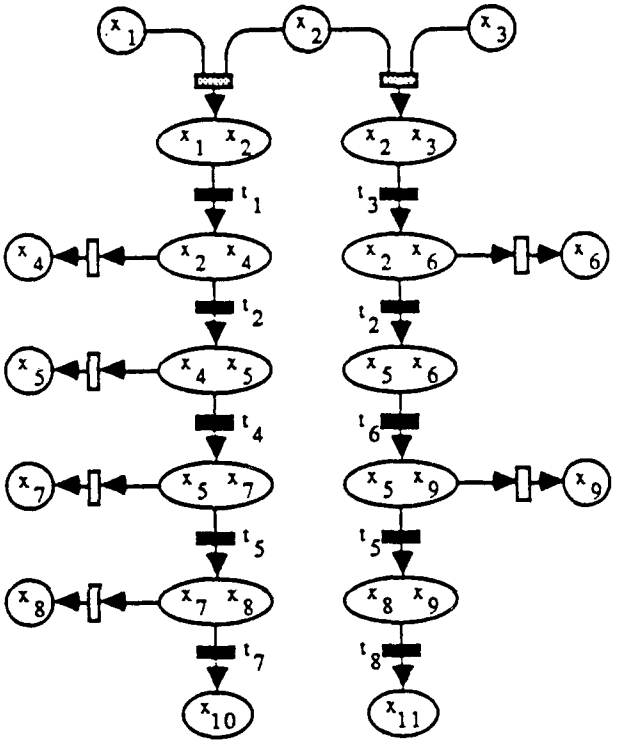


Figure 7: Sufficient Statistics for Prediction. Places enclose sets of random variables found in the underlying Markovian net of Figure 6. These are the sets on which probability distributions must be computed. Grey transitions compute the output distribution as the product of two marginal distributions. White transitions marginalize a joint distribution to obtain a distribution on one random variable. Placement of these is somewhat arbitrary. Black transitions transform an input distribution to an output distribution using a transition distribution from the underlying net. The labels on these transitions indicate the corresponding transition in Figure 6.

An important corollary captures the special features of linear-Gaussian nets. In these, the output of each transition (which may, of course, be a vector) is restricted to be a linear combination of the inputs, to which an exogenous Gaussian random variable is added. Also, the distributions on the variables in the initial places must be Gaussian. In this case, all of the distributions necessary for prediction remain Gaussian, so means and covariances of these distributions suffice.

The result of this part of the work is a specific algorithm to identify that part of a Markovian network's structure which can be exploited to simplify the prediction problem. The degree of simplification obtained depends, of course, on the topology of the net.

Update

Given a measurement of the random variable in an input place, the algorithm mentioned above can compute all of the posterior distributions. One only needs to replace the original

distribution at that place with the posterior distribution after a Bayesian update, and repeat the prediction process. If the measurement is taken elsewhere, however, things get more complicated.

To see that the distributions identified in Figure 7 are insufficient for the update process, consider a measurement of x_{11}. Figure 7 suggests that it can be used to update the distribution on x_{11}, and then the joint distribution on x_8 and x_9, and so on back to the joint distribution on x_2 and x_3. From the latter, one can construct an updated distribution on x_2 alone, which then can be propagated along the other half of the graph as if a direct measurement of x_2 had resulted in the update. The problem with this approach it that the update of the joint distribution on x_8 and x_9 resulted in an (implied) update to the distribution on x_8, but this does not get factored into the computation of the updated distribution on x_7 and x_8 in the prediction phase. Somehow, the update to the distribution on x_8 must be transferred between the two sides of the graph using some other mechanism.

That mechanism involves some additional statistics which augment those required for prediction alone. Called *crosslink statistics*, these preserve information on variables which appear in two or more unordered places in the prediction graph. *The second major product of this work is another graph theoretic algorithm which determines where crosslink statistics are needed*, along with two algorithms which respectively initialize and update them. The combination of the prediction statistics and the crosslink statistics are sufficient to update all of the distributions on individual random variables in the Markovian net.

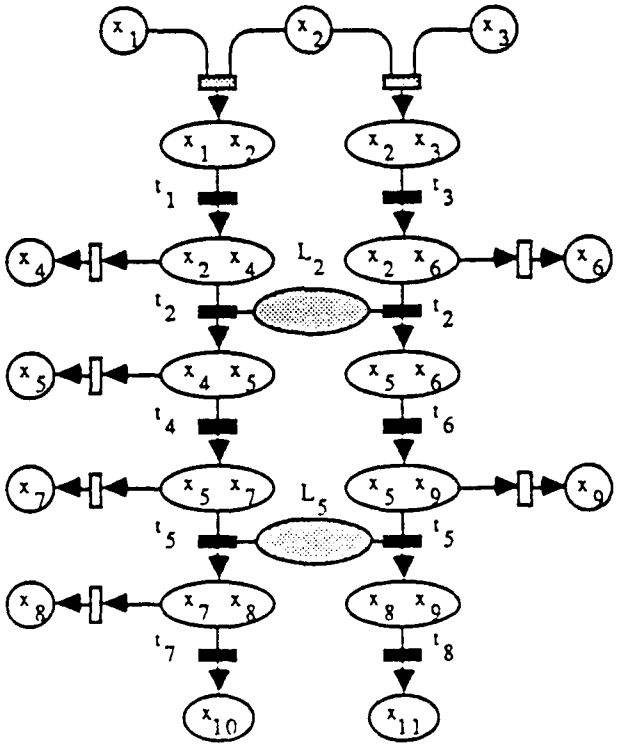


Figure 8: Sufficient Statistics for Updates. The two crosslink places, L_2 and L_5, correspond to the two transitions in the Markovian net which are referenced more than once in the prediction net. The statistics associated with these places can take one of several forms, but all preserve information on the random variables common to the input and output places of the connected transitions. In particular, L_5 statistics preserve information on x_8. When a measurement from a place is obtained, the directionality of some arcs must be reversed, and added to crosslink arcs, to determine the order of the update computations.

Figure 8 shows the crosslinks required for the example of Figure 6. These statistics are adequate regardless of the origin of a measurement, although the order in which the distributions

are updated does depend on the origin. Note that the crosslink statistics transfer information not only about x_2 and x_8, but also about x_5. Referring back to Figure 6, x_5 is also in the common past of x_{10} and x_{11}, so a measurement of x_{11} should update an estimate of x_5, which in turn should affect the estimate of x_{10}.

As with the prediction problem, Gaussianness is preserved under updates from linear-Gaussian measurements. In this case, the updated distributions remain Gaussian, and are completely determined from their means and covariances.

SPECIFIC RESULTS

With a general theory of estimation on Markovian nets available, estimation for multi-scale models becomes a special case. When applied to Markovian net descriptions of one- and two-dimensional random fields which have a high degree of structural regularity, it is not surprising that the resulting estimation algorithms are also highly regular.

One-Dimensional Processes

Figure 2 showed a Markovian network representation for a multi-scale model of Brownian motion. It also suggested a much richer class of stochastic processes, as the transition distributions need not be based on linear-Gaussian transformations. Since the general results in estimation on Markovian nets described above do not rely upon linearity or Gaussianness, sufficient statistics for estimation on arbitrary multi-scale models of one-dimensional processes may be identified.

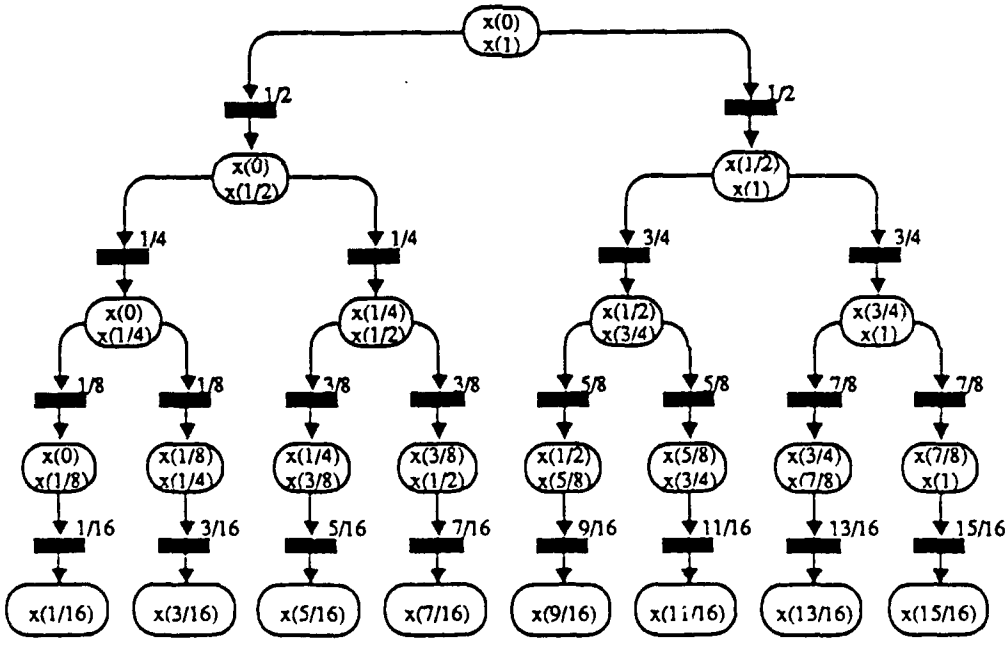


Figure 9: Sufficient Statistics For Prediction In a One-Dimensional Multi-Scale Model. Place labels indicate values of the process at specific points in time. Transition labels indicate the time to which a midpoint generated by that transition corresponds. Note that each sample in the interior of the unit interval appears in two places.

Figure 9 shows the network required for prediction in one-dimensional multi-scale processes. The sufficient statistics are simply joint distributions on process values at interval endpoints. They are connected in a tree structure, showing that the regularity of the lattice in Figure 2 imparts a regularity to the prediction process. Note that this diagram does not represent a Markovian net: the operation of transitions with identical labels is not independent, as each generates the value of the process at the midpoint of the interval associated with its input.

Figure 10 shows the crosslinks necessary for updates. The crosslinks connect places representing neighboring intervals. They account for the fact that a measurement of a process value is statistically related to the process values at the endpoints of that interval — and of all other intervals in which it is nested. Therefore, the update process propagates outward from the measurement point (through coarser scales) as estimates of endpoints over the enclosing intervals are revised, and then inwards (through finer scales) as estimates of values inside other intervals are updated to reflect revisions in the endpoint estimates.

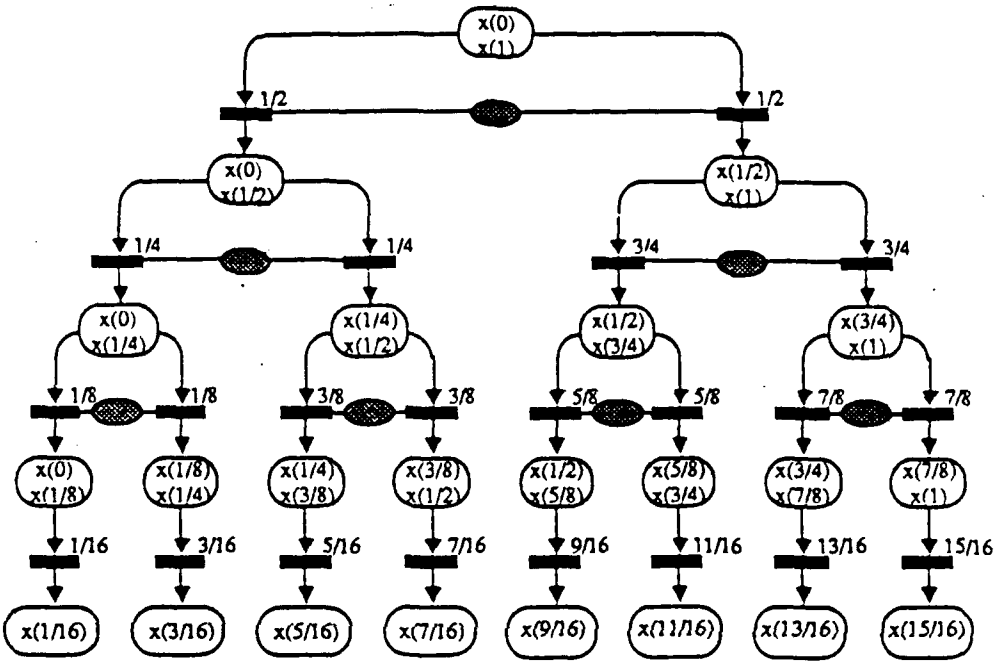


Figure 10: Sufficient Statistics for Update In a One-Dimensional Multi-Scale Model. Crosslinks connect places containing duplicate process values. Crosslink statistics may be merged with the prediction statistics at the immediately preceding places to reduce this structure to a tree. The complexity of the update statistics are independent of the number of scales included in the model.

The most important property of these multi-scale models is that *the complexity of the sufficient statistics does not depend on the number of scales included in the model.* This limits the computational effort required to perform an update, and the size of the data structures required to support that process. In fact, for processes that suit the representation of Figure 2, the complexity of the update process, for a single measurement, is *linear in the total number of sample points.*

What this work does not address is the question of parameterization of the distributions on the variables in the places in Figure 10. For linear-Gaussian models of the type illustrated in Figure 4, all required distributions are Gaussian. For other processes, finite parameterization of the conditional distributions at each place may not be possible.

Nonetheless, other forms of the transition probability distributions are important. For example, Figure 11 shows a highly nonlinear transition mechanism to construct values of the process at successive midpoints. This divides the unit interval into segments of constant process values. Segment boundaries can be determined through scale-to-scale causality by repeatedly assigning the midpoint of an interval either to one of the segments in which an endpoint is contained, or to a new segment. By restricting new segments to appear only when an interval's endpoints are in different segments, segments remain connected. The process value over each segment is determined by the value assigned to the first point that falls into it. Figure 12 shows a typical sample of one of these *midpoint selection* processes.

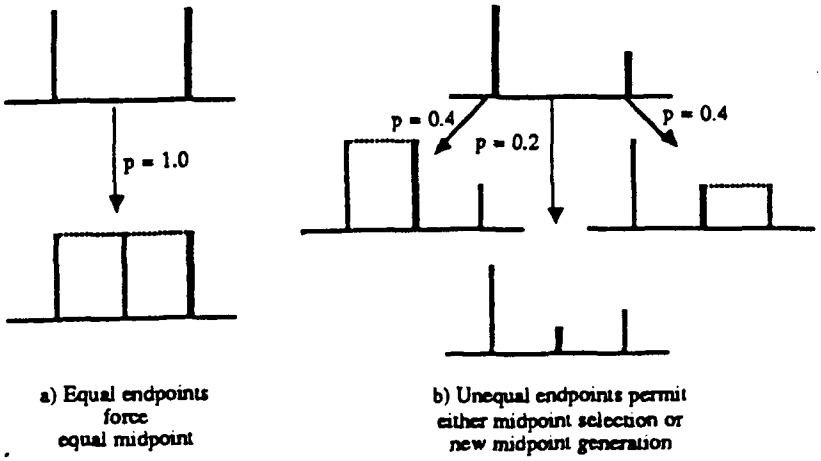


Figure 11: Example of Nonlinear Midpoint Construction. For each interval, the value of the process is determined as a function of the values at the endpoints of the interval. (a) If the endpoint values are equal, the midpoint value is set equal to them. (b) If the endpoint values differ, then with some specified probability the midpoint may take on a completely new value drawn from some distribution (e.g., uniform over the range of the process). If it does not, then one of the endpoints is selected, and its value copied to the midpoint.

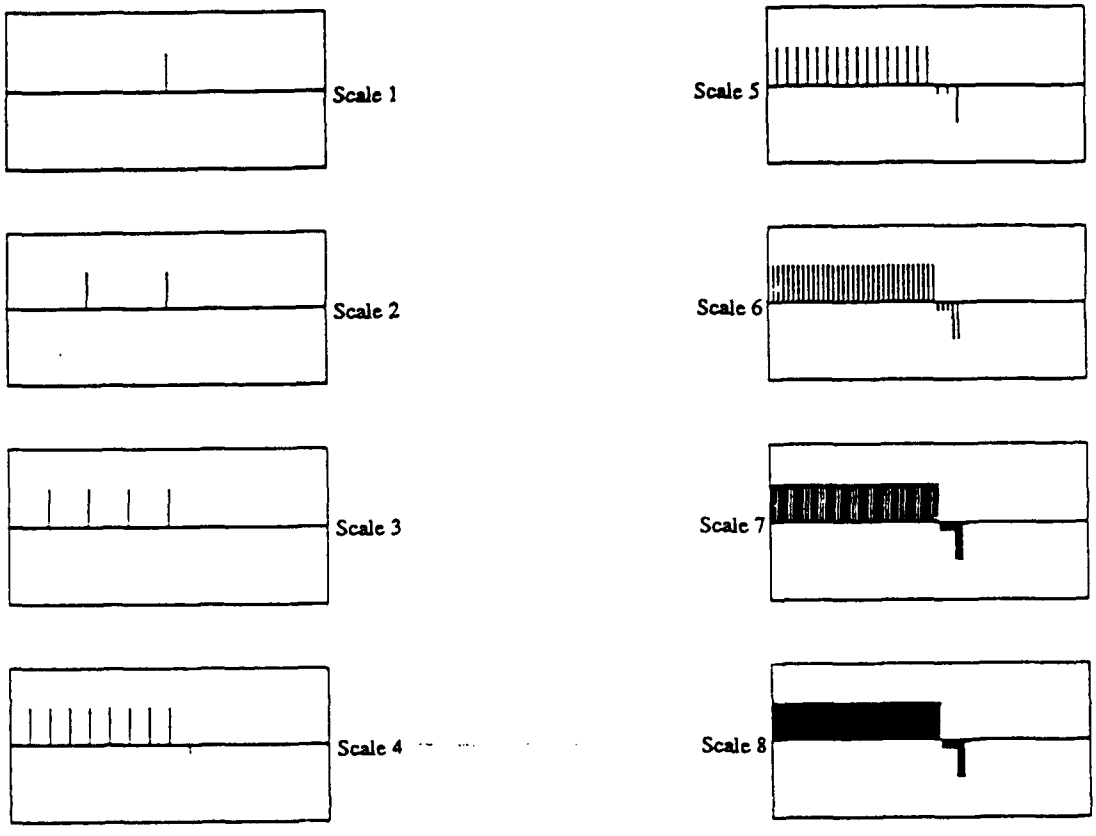


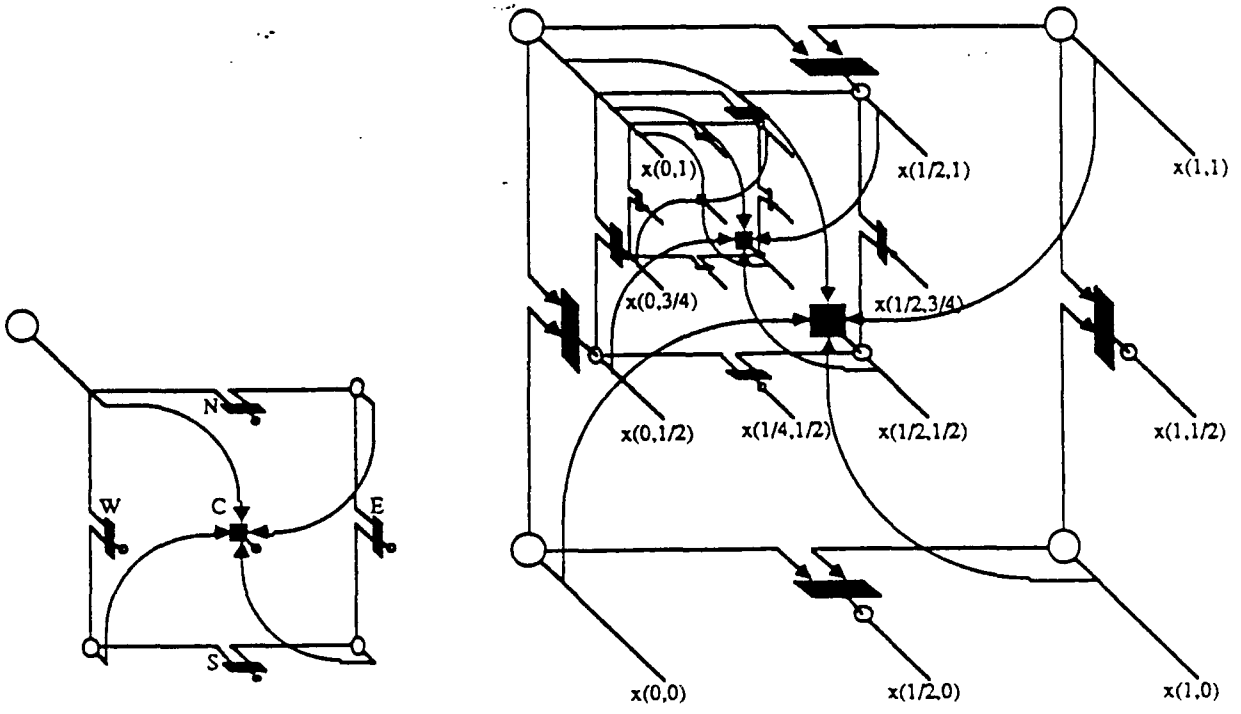
Figure 12: Sample Function of a Nonlinear Midpoint Construction Process. The eight coarsest scales of the evolution of a sample function from a midpoint selection process show how new segments are created when their first point is constructed, and extended by the assignment of midpoint values to one of the endpoint values. The endpoint values at scale 1 are set to 0.5 (left) and 0.0 (right) at the start of the construction.

The sufficient statistics for a midpoint selection process remain the same as for the midpoint deflection, or Brownian motion, processes: the joint distributions on process values at the endpoints of a nested set of intervals. Unlike the linear-Gaussian case, there is no obvious finite parameterization of these distributions. If, however, the number of levels which can be assigned to segments is small, then it is possible to store this joint distribution explicitly.

With the requisite distributions on hand, estimation algorithms for one-dimensional multi-scale processes become straightforward specializations of the general algorithms. For the processes discussed here, the algorithms are in fact equivalent to known algorithms for estimation on tree structures.

Two-Dimensional Processes

Image processing requires treatment of two-dimensional random fields, not just one-dimensional processes. Two-dimensional fields also admit a multi-scale Markovian net representation, although the topology of the Markovian nets becomes somewhat more complex.



a) Generator for one quadrant

b) Successive refinement of upper left quadrants

Figure 13: Markovian Net Representation of a Quadrant Decomposition for a Two Dimensional Random Field. a) Each quadrant inherits one corner value from a coarser scale, and three from the current scale. These are inputs into operators which determine process values at boundary midpoints and the center of the quadrant. b) Nesting quadrants permits refinement through an arbitrary number of scales. Note that each value of the process ultimately influences the construction of eight other process values.

Figure 13 illustrates one way to represent a two-dimensional field with a multi-scale Markovian net. The unit square is the domain of interest, and it can be divided into four equal quadrants. These quadrants may be recursively decomposed in a similar manner to give a quadtree structure [7]. However, the random variables of the Markovian net cannot be associated with just these nested quadrants, or the Markovian scale-to-scale independence assumption would preclude exact matches at boundaries between quadrants. Instead, each quadrant can be characterized by the

values of the random field at its four corners, and one-dimensional multi-scale processes inserted to describe the boundaries between quadrants.

As a synthesizer of samples of random fields, the model of Figure 13 operates as follows. Assume process values have already been determined at the four corners of a quadrant. Generate a value for the midpoint of each edge of the quadrant using the one-dimensional midpoint construction techniques presented earlier. Note that these midpoint values will be available to finer scale operations affecting both quadrants on either side of the boundary. Finally, generate a value for the process at the center point of the quadrant, and notice that all information necessary to enable the construction of the process in each sub-quadrant is now present.

Figure 14 shows the complete network for a three-scale model of a two-dimensional random field. It consists of 21 copies of the generating structure from Figure 13(a), nested in a quadtree structure. Unlike conventional quadtree representations, however, it includes *embedded one-dimensional processes along quadrant boundaries*. The samples generated by these processes affect the generation of samples in both quadrants at finer levels. This way, the sample fields described by the model may remain continuous, and *avoid the boundary artifacts of tree models.*

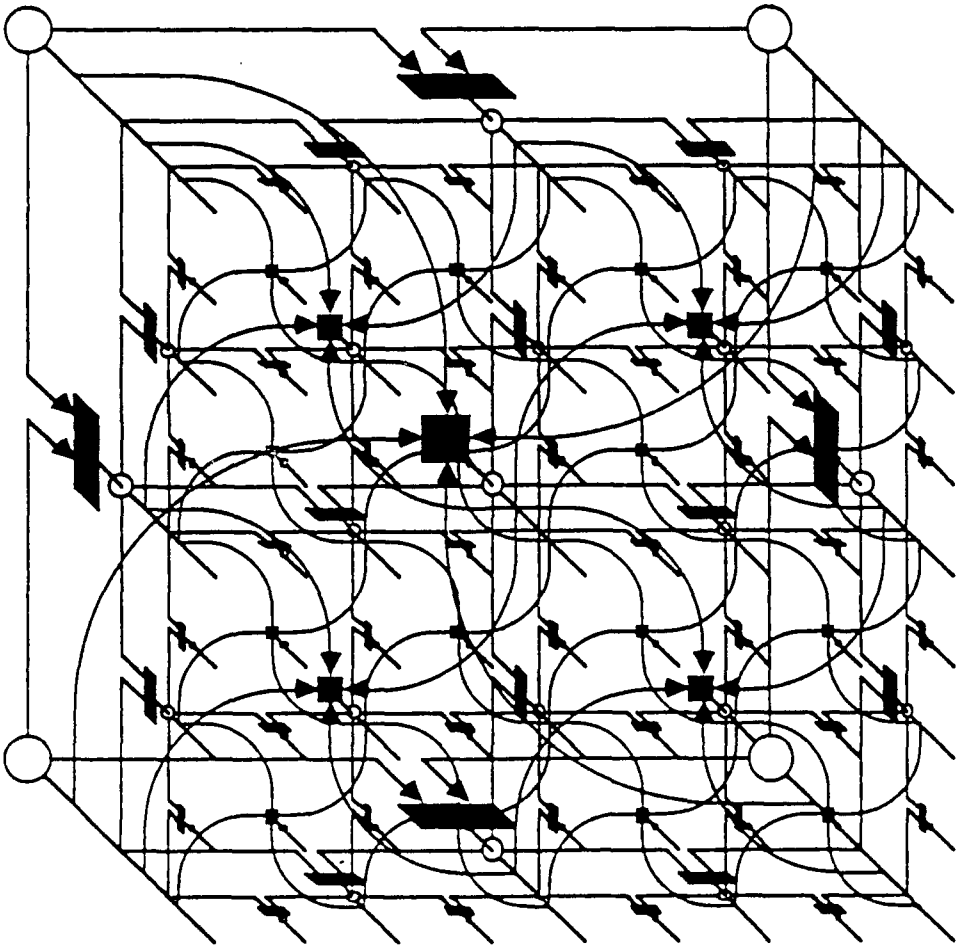


Figure 14: Top Three Scales of a Markovian Net Representation of a Two-Dimensional Field. The topology of the net is quite sparse and highly regular, although it cannot be reduced to an equivalent tree model. Note that each boundary of this net is just a one-dimensional multi-scale model, and that other one-dimensional models appear along internal boundaries.

This representation of a two-dimensional random field is clearly not a tree. Nor can it be reduced to a tree by some manipulation such as augmentation of the set of variables stored at each

place without making the number of variables stored at each place proportional to the depth of the model. Since the complexity of the statistics required by the optimal estimation algorithms is at least proportional to the complexity of the set of variables in the places, such a structural reduction would have unacceptable impacts on algorithm complexity. *This is the model structure which motivated the work on a general approach to estimation on Markovian nets described earlier.*

The graph theoretic algorithms which identify the structure of an estimator for a Markovian net can be applied directly to this network. Figure 15 illustrates one element of the structure required for prediction — an element corresponding to one quadrant. Several copies of this structure must be combined into a tree in order to prescribe the prediction structure for an entire model.

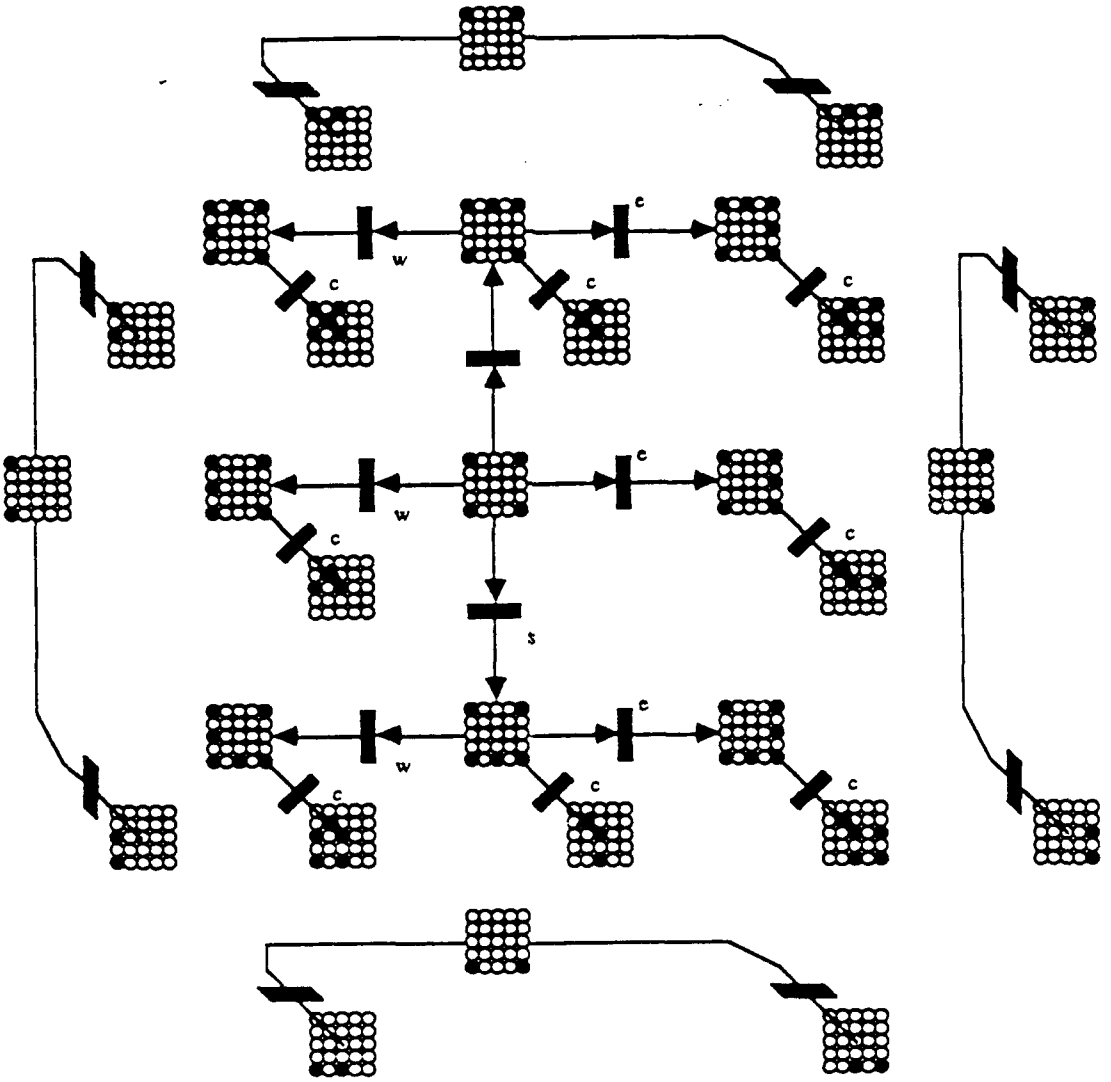


Figure 15: Prediction Statistics for One Quadrant's Decomposition. Each 5 x 5 array represents 25 samples of the random field at the finest scale shown for the generator of Figure 13(a). Darkened cells indicate which sample values are included in the set at each place. Transition labels refer to the transitions in the standard generator of Figure 13(a).

To understand the structure of these prediction statistics, first examine the four boundaries of the quadrant. Each of these is a one-dimensional process. The prediction statistics for one-dimensional processes appeared in Figure 9, and turned out to be distributions on process values at

the two endpoints of an interval. This is precisely what Figure 15 shows for the exterior boundaries: each place holds process values for the endpoints of a segment at the coarser level (at the top of each boundary tree fragment) or at the finer level (at the bottom of each fragment) corresponding to the two half-intervals.

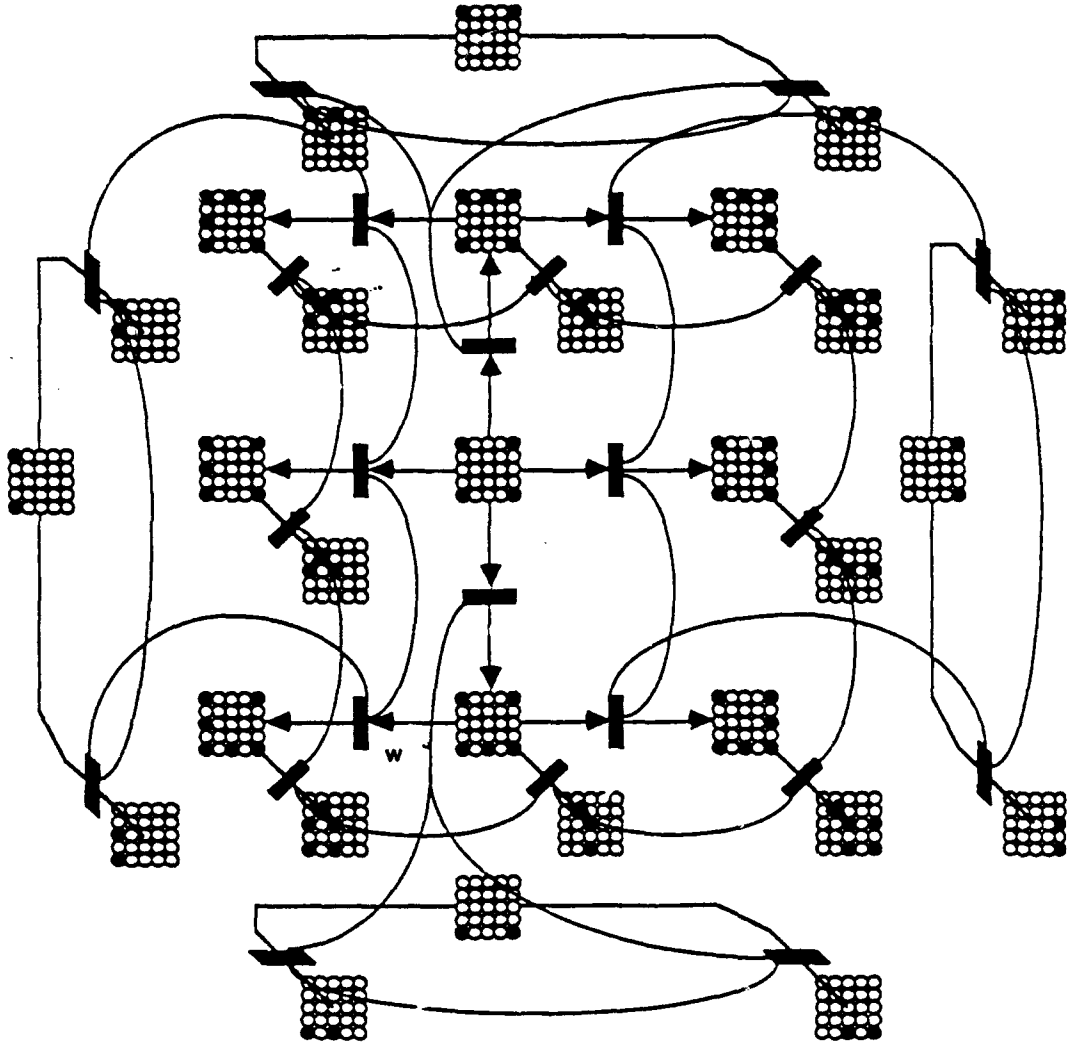


Figure 16: Update Structure for Two-Dimensional Processes. Crosslinks appear as curved arcs, without places. In this net, crosslinks connect more than two transitions due to the complexity of the connections in the basic Markovian net model. Connected transitions correspond to the same transition in Figure 13(a), and all introduce the same sample value into their output places.

The 17 places in the interior of the diagram also form a tree. The root of the tree contains sample values at the four corners of the quadrant. Four of the leaves contain the sample values at the four corners of the subquadrants, allowing these structures to be concatenated into a full tree. The remaining four leaves are the roots of another tree for the one-dimensional processes that form the boundaries between the subquadrants. The remaining 8 places are intermediate steps needed since each transition only introduces one new variable into the statistic set at a time.

To complete the treatment of two-dimensional fields, one must add the crosslinks which convey update information between the branches of the prediction tree. These appear in Figure 16. As with the one-dimensional case, the crosslink arcs have no directionality here; that is imposed by the source of a measurement. The purpose of the crosslinks is the same as discussed above:

when information from a measurement is propagated up the prediction tree, the part of it that pertains to the variables introduced at this scale must be preserved for the time when update information propagates back down the other branches of the prediction tree.

The most important feature of this update structure is that, once again, *the complexity of the statistics required for prediction and update are independent of the number of scales included in the model.* Thus *the computational load required to update all variables in the model from a measurement taken at one point is proportional to the number of nodes in the prediction tree, which in turn is proportional to the number of pixels* in the two-dimensional image.

Recall that the objective of this work was to develop accurate and efficient estimation algorithms based on multi-scale models of two-dimensional random processes. This objective has been met: *the algorithms introduced here are the most accurate possible, as they explicitly reconstruct the conditional probability distribution on each variable* in the process. As can now be seen, they are also extremely efficient, as *the computational effort required to process an update is simply proportional to the size of the process.*

APPLICATIONS

Work is not complete, however. Establishing the general structure of estimation algorithms for multi-scale models of two-dimensional random processes is an important, but not complete, step towards the ultimate goal of real-time image reconstruction, fusion, and identification. To see where additional work is necessary, consider some potential applications.

Problem Requirements

There are three broad classes of application for this work: static image analysis, dynamic image analysis, and image fusion. All but the first impose real-time computing constraints, and hence a close match between algorithm structure and available computing structures. Most importantly, however, one can now consider novel approaches to classical problems which can be posed as estimation problems within the structure investigated here, and rest assured that the solutions will inherit the computational regularity and simplicity of the general algorithms developed to date.

First, consider static image analysis. Important problems in this area include segmentation, texture identification and classification, and anomaly detection. All three problems can be posed in terms of multi-scale Markovian network models. Midpoint construction models provide a statistical characterization of a wide variety of irregular boundaries, such as shorelines. Different coefficients in linear-Gaussian transformations can lead to a wide range of spatially-invariant textures, such as pasture, woodland, or ocean surfaces. Combinations of the two provide models of landscapes whose sample fields are strikingly realistic. An estimator build for these models could not only reconstruct estimates of the field values at all sample points, but also estimate model parameters and segment boundaries. Such algorithms will require computationally realizable approximations to the exact algorithms derived here, analogous to the generalized likelihood ratio techniques for failure detection (as a solution to the segmentation problem), the extended Kalman filter (for model parameter identification as part of texture classification), and model mismatch evaluation (for anomaly detection).

One can view dynamic images as a three- or four-dimensional random field, with the time dimension having the property that information need only propagate in one direction along it. Too complex to illustrate, but easily represented in a digital computer, are the multi-scale Markovian net models for these higher dimensional processes. Temporal causality may allow substantial simplifications, just as it does in conventional Markov processes. For example, one can use the model of Figure 14 to describe the initial frame of an image sequence, and also (with different parameters) to describe the changes from frame to frame. Tracking moving boundaries and detecting changes in textures become two-dimensional analogs of tracking and detection algorithms for conventional time series.

Finally, image fusion algorithms must process image data at different scales, with different centerpoints and orientations, often involving different physical phenomenologies, into a composite representation of an area of interest. Representing different sensor scales is natural in this modeling environment; unfortunately, most phenomenological models have differential or spectral foundations instead of an explicit multi-scale structure. Therefore, a model identification algorithm, again based on linearization techniques similar to the extended Kalman filter, would be effective in estimating model parameters directly from imagery.

CONCLUSIONS

This work set out to investigate whether multi-scale models of random fields could lead to accurate and efficient estimation algorithms. It accomplished that objective, with affirmative answers. The estimation algorithms compute the exact posterior distributions on process variables given a measurement taken from any location, at any scale. They operate in time that is proportional to the size of the field of interest; this efficiency results from the essential fact that the sufficient statistics for estimation in a multi-scale model are of complexity that is independent of the number of scales employed.

With respect to practical implementations of the algorithms developed here, questions outnumber answers. The linear-Gaussian equations provide a finite-dimensional realization of an optimal estimator for one special case, but others may exist. Nothing is yet known about potential simplifications when one is presented with a set of measurements taken over a region of the image, instead of at a single point. Nor have linearized versions of the algorithms for use with mildly nonlinear problems been constructed. Nor, indeed, has duality been exploited to derive multi-scale optimization algorithms based on the same model structure.

Therefore, we conclude that *the domain of multi-scale stochastic process models is exceedingly rich.* We also conclude that *they lead to highly structured and efficient algorithms which solve a wide class of estimation problems* associated with that model. These algorithms offer the potential of vast simplifications for a variety of image analysis problems, as well as a basis for extensions into even more important applications such as real-time, video tracking.

REFERENCES

1. Barton, R.J. and V.H. Poor, "Signal Detection in Fractional Gaussian Noise," *IEEE Transactions on Information Theory*, Vol. IT-34, September 1988, pp. 943-959.
2. Basseville, M., A. Benveniste, and A.S. Willsky, "Multiscale Autoregressive Processes, Part I: Schur-Levinson Parametrizations," submitted to *IEEE Transactions on ASSP*.
3. Basseville, M., A. Benveniste, and A.S. Willsky, "Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling," submitted to *IEEE Transactions on ASSP*.
4. Basseville, M., A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and Estimation of MultiResolution Stochastic Processes," to appear in the Special Issue of the *IEEE Transactions on Information Theory* on *Wavelet Transforms and MultiResolution Signal Analysis*
5. Benveniste, A., R. Nikoukhah, and A. S. Willsky, "MultiScale System Theory," submitted to the *IEEE Transactions on Circuits and Systems*
6. Beylkin, G., R. Coifman, and V. Rokhlin, "Fast Wavelet Transforms and Numerical Algorithms I, to appear in *Comm. Pure and Appl. Math.*
7. Burt, P.J. and E.H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions Comm.*, Vol. COM-30, No. 4, April 1983, pp. 532-540.
8. Chou, K. C., A. S. Willsky, and A. Benveniste, "MultiScale Dynamic Models, Data Fusion, and Optimal Estimation," submitted to the *IEEE Transactions on Automatic Control*
9. Chou, K.C. A.S. Willsky, A. Benveniste, and M. Basseville, "Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, FL, December 1989.
10. Chou, K.C. and A.S. Willsky, "Multiscale Riccati Equations and a Two-Sweep Algorithm for the Optimal Fusion of Multiresolution Data," *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, December 1990.
11. Chou, K.C., S. Golden, and A.S. Willsky, "Modeling and Estimation of Multiscale Stochastic Processes," *Int'l. Conference on Acoustics, Speech, and Signal Processing*, Toronto, April 1991.
12. Coifman, R.R., Y. Meyer, S. Quake, and M.V. Wickehauser, "Signal Processing and Compression with Wave Packets," preprint, April 1990.
13. Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets," *Commun. Pure and Appl. Math*, Vol. 41, November 1988, pp. 909-996.
14. Daubechies, I., "The Wavelet Transform, Time-Frequency Localization, and Signal Analysis," submitted to the *IEEE Transactions on Information Theory*, Vol. 36, 1990, pp 901-1005
15. Flandrin, P. "On the Spectrum of Fractional Brownian Motions," *IEEE Transactions on Information Theory*, Vol. 35, January 1989, pp. 197-199.
16. Glowinski, R., W. Lawton, M. Ravachol, E. Tennenbakum, "Wavelet Solution of Linear and Nonlinear Elliptic, Parabolic, and Hyperbolic Problems in One Space Dimension," *9th Int'l. Conference on Comp. Methods in Appl. Sci. and Eng., SIAM, 1990.*
17. Golden, S. A., "Identifying MultiScale Statistical Models Using the Wavelet Transform," S.M. Thesis, MIT Dept. of EECS, May 1991
18. Hackbush, W., *Multigrid Methods and Applications*, Springer-Verlag, New York, 1985.
19. Kim, M. and A.H. Tewfik, " Fast Multiscale Detection in the Presence of Fractional Brownian Motions," *Proceedings of SPIE Conference on Advanced Algorithms and Architecture for Signal Processing V*, San Diego, CA, July 1990.

20. Lawton, W.M., "Wavelet Discretization Methods for Surface Estimation and Reconstruction," *SPIE/SPSE Symposium on Elec. Imaging Sci. and Tech.*, Santa Clara, CA, February 1990.
21. Lundahl, T., W.J. Ohley, S.M. Kay, and R. Siffert, "Fractional Brownian Motion: A Maximum Likelihood Estimator and its Application to Image Texture," *IEEE Transactions on Med. Imaging*, Vol. MI-5, September 1986, pp. 152-161.
22. Mallat, S.G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Anal. and Mach. Intel.*, Vol. PAMI-11, July 1989, pp. 674-693.
23. Mallat, S.G., "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Transactions on ASSP*, Vol. 37, December 1989, pp. 2091-2110.
24. Mandelbrot, B.B. and H.W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10, October 1968, pp. 422-436.
25. Mandelbrot, B.B., *The Fractal Geometry of Nature*, W.H. Freeman, San Francisco, 1983.
26. Pentland, A., "Finite Element and Regularization Solutions Using Wavelet Bases," MIT Media Lab Vision and Modeling Group TR-143, June 1990.
27. Pentland, A.J., "Fast Surface Estimation Using Wavelet Bases," MIT Media Lab Vision and Modeling Group TR-142, June 1990.
28. Pentland, A.P., "Fractal-Based Description of Natural Scenes," *IEEE Transactions on Patt. Anal. and Mach. Intel.*, Vol. PAMI-6, November 1989, pp. 661-674.
29. Szeliski, R., "Fast Parallel Surface Interpolation with Applications to Digital Cartography," SRI Tech. Note 470; June 1989.
30. Szeliski, R., "Fast Surface Interpolation Using Hierarchical Basis Functions," *IEEE Transactions on PAMI*, Vol. 12, No. 6, June 1990, pp. 513-528.
31. Taqqu, M.S., "Self-Similar Processes," in *Encyclopedia of Statistical Sciences*, Wiley, NY, 1985.
32. Terzopoulos, D., "Image Analysis Using Multigrid Relaxation Methods," *IEEE Transaction on PAMI*, Vol. PAMI-8, No. 2, March 1986, pp. 129-139.
33. Tewfik, A.H. and M. Kim, "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motions," submitted to *IEEE Transactions on Information Theory.*
34. Vetterli, M. and C. Herley, "Wavelets and Filter Banks: Relationships and New Results," *Proceedings of ICASSP*, Albuquerque, NM, 1990.
35. Willsky, A.S., K.C. Chou, A. Benveniste, and M. Basseville, "Wavelet Transforms, Multiresolution Dynamical Models, and Multigrid Estimation Algorithms," *1990 IFAC World Congress*, Tallinn, USSR, August 1990.
36. Witkin, A., D. Terzopoulos, and M. Kass, "Signal Matching Through Scale Space," *Int. J. Comp. Vision*, Vol. 1, 1987, pp. 133-144.
37. Wornell, G. W., "Synthesis, Analysis, and Processing of Fractal Signals," Ph.D. Thesis, MIT Dept. of EECS, May 1991
38. Wornell, G. W., and A. V. Oppenheim, "Deterministically Self-Similar Signals," submitted to the *IEEE Transactions on Information Theory*
39. Wornell, G.W. and A.V. Oppenheim, "Estimation of Fractal Signals from Noisy Measurements Using Wavelets," submitted to *IEEE Transactions on ASSP.*
40. Wornell, G.W., "A Karhunen-Loéve-Like Expansion for 1/f Processes via Wavelets," *IEEE Transactions on Information Theory*, Vol. 36, No. 9, July 1990, pp. 859-861.

8 Mar '92

PARAUNITARY AND ORTHONORMAL CONVOLVERS

P. P. Vaidyanathan, Fellow, IEEE
Dept. of Electrical Engineering, 116-81
California Institute of Technology
Pasadena, CA 91125, USA

Abstract. A maximally decimated filter bank system (with possibly unequal decimation ratios in the subbands) can be regarded as a generalization of the short time Fourier transformer. In fact, it is known that such a 'filter bank transformer' is closely related to the wavelet transformation. A natural question that arises when we conceptually pass from the traditional Fourier transformer to the filter bank transformer is: what happens to the convolution theorem, i.e., is there an analog of the convolution theorem in the world of 'filter bank transforms'? In this paper we address the question first for uniform decimation and then generalize it to the nonuniform case. The result takes a particularly simple and useful form for paraunitary (or orthonormal) filter banks. It shows how we can convolve two signals $x(n)$ and $g(n)$ by directly convolving the subband signals of a paraunitary filter bank and adding the results. The advantage of the method is that we can quantize in the subbands based on the signal variance and other perceptual considerations, as in traditional subband coding. As a result, for a fixed bit rate, the result of convolution is much more accurate than direct convolution. That is, we obtain a *coding gain* over direct convolution. We will derive expressions for optimal bit allocation and optimal coding gain for such paraunitary convolvers. As a special case, if we take one of the two signals to be the delta function (e.g., $g(n) = \delta(n)$), we can recover the well-known bit allocation and coding gain formulas of traditional subband coding.

Work supported in parts by National Science Foundation grants MIP 8919196 and matching funds from Rockwell Inc., and Tektronix, Inc.

1. INTRODUCTION

Fig. 1.1(a) shows the M channel maximally decimated digital filter bank, which has been studied by a number of authors in the past decade. Here $H_k(z)$, $F_k(z)$, $0 \leq k \leq M-1$ are the set of analysis and synthesis filters. The notations $\downarrow n_k$ and $\uparrow n_k$ denote the n_k-fold decimator and interpolator (unpsampler) as defined in several earlier references [1]–[5]. The boxes labeled Q_k denote quantizers which quantize the subband signals $x_k(n)$.

The relations between filter banks and wavelet transforms have been known for some time [6]–[12]. An excellent magazine article appeared recently [10], revealing this connection explicitly. It is well known that wavelet transforms provide more flexibility (in terms of time-frequency resolution) than the traditional Fourier transform. In this paper we deal only with discrete-time filter banks (both uniform and nonuniform decimators will be considered). It is known that discrete time filter banks can be considered as discrete time wavelet transformations. Here the analysis bank can be viewed as a transformation from 'time' to 'time-frequency'. We will simply refer to this as the filter bank transform, and the decimated subband signals $x_k(n)$ will be called the transform-domain signals. The synthesis bank is regarded as the inverse transformer (assuming perfect reconstruction, that is, $\hat{x}(n) = x(n)$).

Aim of the paper.

The advent of these transforms leads us to ask the question, "how do the standard properties of the Fourier transformation generalize to the case of 'filter-bank transforms'?" For example, what is the extension of the convolution theorem? To introduce the main topic of the paper, let $y(n)$ denote the convolution of two sequences $x(n)$ and $g(n)$, i.e., $y(n) = \sum_m x(m)g(n-m)$. According to Fourier transform theory, the transform of $y(n)$ is related to those of $x(n)$ and $g(n)$ as $Y(e^{j\omega}) = X(e^{j\omega})G(e^{j\omega})$, i.e., convolution becomes 'multiplication' in the transform domain. Now consider the 'filter bank transformer', with the decimated subband signals regarded as the 'transform domain'. What is the 'convolution theorem' in this case? To expand on this question, consider Fig. 1.1 where we show $x(n)$ and $g(n)$ as the inputs to two copies of the filter bank.

The transform domain 'coefficients' corresponding to $x(n)$ and $g(n)$ are the sequences $x_k(n)$ and $g_k(n)$, respectively. How should we combine $x_k(n), g_k(n)$, $0 \leq k \leq M-1$ so that the convolution $\sum_m x(m)g(n-m)$ can be obtained from these, assuming there are no subband quantizers?

In Sec. 2.1 we will derive this convolution theorem for the case of uniform filter banks (i.e., $n_k = M$ for all k). The result takes an exceptionally simple form in the case of *paraunitary* filter banks [2], [12]–[14]. More specifically, the convolution $x(n) * g(n)$ is reduced to computing the convolutions $x_k(n) * g_k(n)$ and adding. This will be stated more precisely in Theorem 2.1 (equal n_k) and Theorem 2.2 (unequal n_k). In Sec. 2.2, the result will be extended to the case of filter banks with nonuniform decimation ratios. Once again, it will be shown that when the synthesis filter coefficients form an orthonormal basis (this being the extension of the paraunitary concept), the 'convolution theorem' takes a special simple form. Even though the uniform filter bank is a special case, we have chosen to treat it separately first, because it is much simpler, while conveying most of the ideas well.

Usefulness.

The motivation for obtaining these 'convolution theorems' *does not* originate from a desire to obtain algorithms that are faster than the many well-known fast convolution technique. (Indeed, the 'state of art' for fast convolutions is already very advanced). The actual motivation comes from the fact that we can quantize in the subbands, and reduce the roundoff error (for fixed wordlength) by proper bit allocation schemes. Thus, instead of quantizing $x(n)$ and then convolving with $g(n)$, we can now quantize $x_k(n)$ and then convolve with $g_k(n)$ and add the results for all k. When performing this quantization in subbands, we can exploit the subband energy distribution and perform optimal bit allocation. In this way, we obtain increased accuracy for a given bit rate. That is, the system offers a coding gain. This idea is very similar in philosophy to subband coding [15] (e.g., see Chapter 11 of [16], and Chapter 1 of [17]). In a spirit similar to that described in the above references, we can define a coding gain for the paraunitary convolver. In Sec. 3 we will present a detailed study of this coding gain. We will obtain the optimal bit allocation formula, and study the

coding gain under optimal bit allocation. Unlike in usual subband coding, the paraunitary subband convolver can provide a coding gain > 1 even if $x(n)$ has a flat spectrum (i.e., is white).

It is important to notice that the computation of the subband signals $x_k(n)$ itself involves filtering. If this filtering complexity is comparable to the direct convolution of $x(n)$ and $g(n)$, then the above technique is clearly unworthy. It has potential applications only when $x(n)$ and $g(n)$ are very long sequences (in comparison with the lengths of the analysis filters $H_k(z)$). An useful special cases arises when the analysis filters have length $\leq M$ (which is analogous to transform coding). We will see that, even in this case, substantial coding gain can be exhibited.

It is meaningful to try to maximize the coding gain by optimization of the coefficients of the paraunitary filter bank (for a fixed order). Such an optimization is easier to formulate in the special case where the filter bank reduces to the transform coding system (to be described in Fig. 4.1 later, where T is a unitary matrix). This is a specical case of paraunitary filter banks (with *constant* polyphase matrices). We consider it separately, and address the problem of optimal choice of T (under the optimal bit allocation constraint). This, then is the generalization of the Karhunen-Leove transform (KLT) [16],[18],[19], for the case of unitary convolvers. We will formulate the problem in terms of two autocorrelation matrices, but unlike in the KLT problem, a closed form solution for T (e.g., in terms of the eigenvectors) is not possible. We will consider a numerical example based on speech signals, and show that the coding gain of the convolver is very close to the theoretical upper bound, if the matrix T is taken to be the DCT matrix. This observation parallels a similar well-known result in orthogonal transform coding of speech [16].

Outline. In Sec. 2.1 we derive the convolution theorem for paraunitary filter banks with uniform decimation. This is extended to the case of nonuniform filter banks (n_k not identical for all k) in Sec. 2.2. In this case the paraunitary property is replaced with a generalization (orthonormality). Section 3 presents a derivation of optimal subband bit allocation, as well as the corresponding coding gain expression for the paraunitary convolver. Once again, the uniform case will be considered first, and then generalized to the nonuniform case. Even though the former is

strictly a special case of the latter, we have chosen to treat them separately. This is because of the simplicity of the uniform case, which at the same time brings out many of the important features. In Sec. 3.4 we show how the well-known coding gain results for traditional subband systems can be obtained as special cases of the convolver coding gain expressions. Section 4 consider a further specicalization of the uniform paraunitary convolver, with analysis filter lengths constrained to be $\leq M$. This is, in principle, the extension of transform coding problem, to the case of convolution. It has the advantage that we can further maximize the coding gain by optimizing the transform matrix (generalization of the KLT). Section 5 presents several numerical examples, and provides a relative comparision of the coding gain, for different test conditions.

Notations and basics.

1. Bold faced quantities represent matrices and vectors. The notations $\mathbf{A}^T, \mathbf{A}^*$ and $\mathbf{A}^\dagger$ represent, respectively, the transpose, conjugate, and transpose-conjugate of $\mathbf{A}$. The accent 'tilde' as in $\tilde{\mathbf{H}}(z)$ stands for transposition, followed by conjugation of coefficients, followed by replacement of z with z^{-1}. On the unit circle $\tilde{\mathbf{H}}(z) = \mathbf{H}^\dagger(z)$.
2. The M-fold decimator $\downarrow M$ and interpolator $\uparrow M$ (or expander) are defined as in [1],[2]. Thus the input output relation for the decimator is $y(n) = x(Mn)$, and for the interpolator it is

$$y(n) = \begin{cases} x(n/M), & n = \text{integer mul. of } M \\ 0, & \text{otherwise.} \end{cases}$$

   In this paper, all decimation and interpolation ratios are positive integers. In equations, the notation $a(n)\big|_{\downarrow M}$ denotes the decimated sequence $a(Mn)$. (The vertical bar is omitted where it is unnecessary). With $A(z)$ denoting the z-transform of $a(n)$, the notation $A(z)\big|_{\downarrow M}$ denotes the z-transform of the decimated version $a(Mn)$. Let $A(z)$ and $B(z)$ be rational functions and let K and L be integers. The following identity can be easily verifed:

$$\Big(A(z^K)B(z)\Big)_{\downarrow KL} = \Big(A(z)(B(z)\big|_{\downarrow K})\Big)_{\downarrow L}. \tag{1.1}$$

3. $x(n) * g(n)$ denotes convolution of $x(n)$ with $g(n)$. The sequence $x(n) * g^*(-n)$ is the deterministic cross correlation between $x(n)$ and $g(n)$, and has z-transform $X(z)\tilde{G}(z)$.

Polyphase notation

For the case where $n_k = M$ for all k, the system of Fig. 1.1(a) can be redrawn as in Fig. 1.2 where $\mathbf{E}(z)$ and $\mathbf{R}(z)$ are $M \times M$ matrices. Defining the analysis and synthesis filter vectors as

$$\mathbf{h}(z) = [\, H_0(z) \quad H_1(z) \quad \ldots \quad H_{M-1}(z)]^T, \mathbf{f}(z) = [\, F_0(z) \quad F_1(z) \quad \ldots \quad F_{M-1}(z)]^T \tag{1.2}$$

we have

$$\mathbf{h}(z) = \mathbf{E}(z^M)\mathbf{e}(z), \quad \mathbf{f}^T(z) = \tilde{\mathbf{e}}(z)\mathbf{R}(z^M), \tag{1.3}$$

where $\mathbf{e}(z)$ is the delay chain vector, i.e.,

$$\mathbf{e}(z) = [\, 1 \quad z^{-1} \quad \ldots z^{-(M-1)}]^T. \tag{1.4}$$

$\mathbf{E}(z)$ and $\mathbf{R}(z)$ are, respectively, the polyphase matrices of the analysis and synthesis banks.

2. CONVOLUTION THEOREMS FOR ORTHONORMAL FILTER BANKS

2.1. Filter bank with equal decimation ratio in all branches

First consider Fig. 1.1, with $n_k = M$ for all k. The convolution theorem is obtained by analyzing this in absence of the quantizers Q_k. Assume that the set of filters $\{H_k(z), F_k(z)\}$ are chosen to satisfy the perfect reconstruction property, i.e.,

$$\hat{X}(z) = X(z), \quad \hat{G}(z) = G(z). \tag{2.1}$$

Using the fact that the M-fold upsamplers have outputs $X_k(z^M)$ and $G_k(z^M)$, we can express $\hat{X}(z)$ as $\sum_{k=0}^{M-1} X_k(z^M)F_k(z)$, and similarly for $\hat{G}(z)$. Using these together with (2.1) we obtain

$$X(z) = \sum_{k=0}^{M-1} X_k(z^M)F_k(z), \quad G(z) = \sum_{k=0}^{M-1} G_k(z^M)F_k(z). \tag{2.2}$$

Now consider the quantity $X(z)\tilde{G}(z)$ (with the 'tilde' notation as defined at the end of Sec. 1). We have

$$X(z)\tilde{G}(z) = \sum_{k=0}^{M-1}\sum_{m=0}^{M-1} X_k(z^M)\tilde{G}_m(z^M)F_k(z)\tilde{F}_m(z). \tag{2.3}$$

The inverse z-transform of $X(z)\tilde{G}(z)$ is equal to the convolution of $x(n)$ with $g^*(-n)$ (i.e., the determininstic cross correlation between $x(n)$ and $g(n)$). Similarly $X_k(z)\tilde{G}_m(z)$ represents the convolution of the subband signals $x_k(n)$ and $g_m^*(-n)$.

Paraunitary or orthonormal filter banks.

The above equation reduces to a much simpler form (the double summation reduces to a single summation) when the filter bank is paraunitary [12]–[14]. In this case the polyphase matrix $\mathbf{E}(z)$ satisfies

$$\tilde{\mathbf{E}}(z)\mathbf{E}(z) = \mathbf{I}, \tag{2.4}$$

and we choose $\mathbf{R}(z) = \tilde{\mathbf{E}}(z)$ for perfect reconstruction (so that $\mathbf{R}(z)$ is also paraunitary). In this case the analysis and synthesis filters are related as $F_k(z) = \tilde{H}_k(z)$, that is, $f_k(n) = h_k^*(-n)$. In order to ensure that $F_k(z)$ is stable, we assume that the analysis filters are FIR. Thus, $h_k(n)$ and $f_k(n)$ are FIR with same length. A paraunitary filter bank satisfies the following properties, regardless of the exact nature of $H_k(e^{j\omega})$ (i.e., regardless of filter quality) [12].

1. The energy of each analysis filter equals unity, that is $\int_0^{2\pi} |H_k(e^{j\omega})|^2 d\omega/2\pi = 1$.
2. The analysis filters are power complementary, that is, $\sum_k |H_k(e^{j\omega})|^2 = M$.
3. Since $f_k(n) = h_k^*(-n)$, we have $|F_k(e^{j\omega})| = |H_k(e^{j\omega})|$. So the above two properties hold for the synthesis filters as well.

(Notice, in particular, that in the case of idea brickwall filters, to be shown later in Fig. 3.3, the first two properties are evident.) The paraunitary property of $\mathbf{R}(z)$ is equivalent to the property that the synthesis filters satisfy an *orthonormality condition* [10]–[12], that is,

$$\sum_n f_k(n) f_m^*(n + Mi) = \delta(k - m)\delta(i). \tag{2.5}$$

In the z-domain this can be rewritten as

$$\left(F_k(z)\tilde{F}_m(z)\right)\Big|_{\downarrow M} = \delta(k - m). \tag{2.6}$$

Simplification of the convolution formula

Using the above orthonormality condition, Eq. (2.3) leads to

$$\left(X(z)\tilde{G}(z)\right)_{\downarrow M} = \sum_{k=0}^{M-1} X_k(z)\tilde{G}_k(z). \tag{2.7}$$

This can be rewritten in the time domain as

$$\left(x(n) * g^*(-n)\right)_{\downarrow M} = \sum_{k=0}^{M-1} x_k(n) * g_k^*(-n). \tag{2.8}$$

In the time domain, the left hand side represents the M-fold decimated version of the convolution of $x(n)$ with $g^*(-n)$. The kth term on the right hand side represents the convolution of the subband signal $x_k(n)$ with $g_k^*(-n)$. Summarizing, we have

Theorem 2.1. *Paraunitary convolution theorem.* Consider the two copies of a maximally decimated filter bank as in Fig. 1.1, with FIR analysis and synthesis filters, and $n_k = M$ for all k. Ignore the quantizers Q_k. Assume that the system has perfect reconstruction ($\hat{x}(n) = x(n)$ for any $x(n)$) and that the polyphase matrix $\mathbf{E}(z)$ (Fig. 1.2) is paraunitary (equivalently the synthesis filters are orthonormal, i.e., satisfy (2.5) or equivalently (2.6)). Then the M-fold decimated version of the convolution $x(n) * g^*(-n)$ can be computed by computing the convolutions $x_k(n) * g_k^*(-n)$, $0 \le k \le M-1$, and adding them. ◊

In order to obtain all the samples of the convolution $x(n) * g^*(-n)$, we have to repeat the above operation M times, by replacing $g(n)$ with $g(n-i)$, for $0 \le i \le M-1$. We can represent these operations mathematically as

$$\left(z^i X(z)\tilde{G}(z)\right)_{\downarrow M} = \sum_{k=0}^{M-1} X_k(z)\tilde{G}_k^{(i)}(z), \quad 0 \le i \le M-1. \tag{2.9}$$

where $G_k^{(i)}(z)$ is the subband signal obtained by replacing $g(n)$ with $g(n-i)$. Assuming that $x(n)$ is an input sequence and that $g(n)$ is a fixed filter, the quantities $G_k^{(i)}(z)$ are fixed (i.e., can be precomputed).

Application in decimation filtering. As a special situation, imagine that $g^*(-n)$ is a decimation filter for $x(n)$. This means that the result of convolution $x(n) * g^*(-n)$ is decimated

by some factor D. In this case, we do not have to repeat (2.9) for all values of i. For example if $D = M/2$, we only have to perform (2.9) for $i = 0$ and $i = M/2$.

Comments on complexity. Computational complexity is *not* the main advantage of the method of subband convolution. Assume for simplicity that $x(n)$ and $g(n)$ are N-point sequences. Then direct convolution of $x(n)$ and $g^*(-n)$ (without using standard fast techniques) requires N^2 multipliers. Assuming that N is much larger that the lengths of the subband filters $H_k(z)$ (so that the multiplications required to implement analysis filters are negligible) the signals $x_k(n)$ and $g_k(n)$ have lengths $\approx N/M$. Each subband convolution requires nearly $(N/M)^2$ multiplications, so that the total number of multiplications for all the M values of i in (2.9) is nearly N^2 again. It is true that we can employ the FFT, or even the fast 'short convolution algorithms' in the subbands, but again this is not the main point of the discussion.

The above reasoning does not hold if the analysis filters have length comparable to those of $x(n)$ and $g(n)$. In this case, the complexity of the analysis bank becomes comparable to the direct convolution of $x(n)$ with $g(n)$, and the method is not useful.

The *actual* advantage of the (paraunitary) subband convolver is that it allows us to allocate the computational accuracy (i.e., bits) among the subbands, resulting in a coding gain as elaborated in Sec. 3. In fact considerable coding gain can be obtained even in the special case where the analysis filters have small length (e.g., $\leq M$), as discussed in Sections 4 and 5.

2.2. Orthonormal filter bank with unequal decimation ratios

Now consider the case where the decimation ratios n_k in Fig. 1.1 are possibly unequal integers such that

$$\sum_{k=0}^{M-1} \frac{1}{n_k} = 1. \tag{2.10}$$

This condition implies that we have a maximally decimated system. The design of such systems has received attention recently [20], [21]. Such a system can be regarded as a discrete time wavelet decomposition system. The analysis bank is the 'wavelet transformer' and the synthesis bank the inverse transformer. Assuming perfect reconstruction (i.e., $\hat{x}(n) = x(n)$) we can express the signal

$x(n)$ in terms of the synthesis filters $F_k(z)$ and the wavelet coefficients $X_k(z)$ as follows:

$$X(z) = \sum_{k=0}^{M-1} F_k(z) X_k(z^{n_k}), \tag{2.11}$$

i.e., in the time domain,

$$x(n) = \sum_{k=0}^{M-1} \sum_{\ell} x_k(\ell) f_k(n - n_k \ell), \tag{2.12}$$

The doubly indexed set of sequences

$$\xi_{k,\ell}(n) \stackrel{\Delta}{=} f_k(n - n_k \ell) \tag{2.13}$$

are therefore the 'basis functions' for the expansion of $x(n)$.

Orthonormailty (nonuniform case). The above basis is said to be orthonormal if

$$\sum_n \xi_{k,\ell}(n) \xi^*_{m,i}(n) = \delta(k - m)\delta(\ell - i). \tag{2.14}$$

In terms of the the synthesis filters, the orthonormality property is

$$\sum_n f_k(n) f^*_m(n + n_k \ell - n_m i) = \delta(k - m)\delta(\ell - i). \tag{2.15}$$

This is a generalization of the orthonormality property (2.5) which followed earlier from paraunitariness. Let $n_{k,m}$ denote the greatest common divisor of n_k and n_m, i.e.,

$$n_{k,m} = \gcd\,(n_k, n_m). \tag{2.16}$$

We can then rewrite (2.15) as [22]

$$\sum_n f_k(n) f^*_m(n + n_{k,m} p) = \delta(k - m)\delta(p) \tag{2.17}$$

(see Appendix A). In the z-domain this is equivalent to

$$\left(F_k(z)\tilde{F}_m(z)\right)\Big|_{n_{k,m}} = \delta(k - m). \tag{2.18}$$

A simple example of a perfect reconstruction orthonormal filter bank with unequal n_k is obtained by use of a binary tree structure with paraunitary polyphase matrices at each level [10]–[12]. This results in filter responses that have an octave spacing.

Derivation of the convolution theorem (nonuniform case).

Assume that we have perfect reconstruction, i.e., $\hat{X}(z) = X(z)$ and $\hat{G}(z) = G(z)$. Using the expression (2.11) for $X(z)$ and similarly for $G(z)$, we have

$$X(z)\tilde{G}(z) = \sum_{k=0}^{M-1} \sum_{m=0}^{M-1} F_k(z)\tilde{F}_m(z)X_k(z^{n_k})\tilde{G}_m(z^{n_m}). \tag{2.19}$$

Let L be the least common multiple of the decimation ratios, i.e.,

$$L = \text{lcm } \{n_k\}. \tag{2.20}$$

For $0 \le k, m \le M-1$ we then have

$$L = n_k p_k, \quad L = n_{k,m} p_{k,m} \tag{2.21}$$

for some integers p_k and $p_{k,m}$. Consider now the L-fold decimated version of $X(z)\tilde{G}(z)$. Using the above decomposition of L and the identity (1.1), we can write

$$\Big(X(z)\tilde{G}(z)\Big)\Big|_{\downarrow L} = \sum_{k=0}^{M-1} \sum_{m=0}^{M-1} \left(\Big(F_k(z)\tilde{F}_m(z)\Big)_{\downarrow n_{k,m}} X_k(z^{n_k/n_{k,m}})\tilde{G}_m(z^{n_m/n_{k,m}}) \right)_{\downarrow p_{k,m}}. \tag{2.22}$$

Using the orthonormality property (2.18) this simplifies to

$$\Big(X(z)\tilde{G}(z)\Big)_{\downarrow L} = \sum_{k=0}^{M-1} \Big(X_k(z)\tilde{G}_k(z)\Big)_{\downarrow p_k}. \tag{2.23a}$$

Equivalently, in the time domain

$$\Big(x(n) * g^*(-n)\Big)\Big|_{\downarrow L} = \sum_{k=0}^{M-1} \Big(x_k(n) * g_k^*(-n)\Big)\Big|_{\downarrow p_k}. \tag{2.23b}$$

To obtain all the samples of the convolution $x(n) * g^*(-n)$, we have to repeat the above with the shifted versions $g(n-i), 0 \le i \le L-1$. *This result is summarized as follows.*

Theorem 2.2. *Convolution theorem for orthonormal nonuniform filter-banks.* Consider the maximally decimated filter bank of Fig. 1.1, and ignore the quantizers Q_k. Let

$$L = \text{lcm } \{n_i\}, \quad n_{k,m} = \text{gcd } (n_k, n_m), \quad p_k = L/n_k \quad \text{and} \quad p_{k,m} = L/n_{k,m}. \tag{2.24}$$

Assume that the system has perfect reconstruction ($\hat{x}(n) = x(n)$ for any $x(n)$) and that the synthesis filters are orthonormal, i.e., satisfy (2.17) or equivalently (2.18). Then the L-fold decimated version of the convolution $x(n) * g^*(-n)$ can be computed by computing the p_k-fold decimated versions of the convolutions $x_k(n) * g_k^*(-n)$, and adding them. We can obtain all samples of the convolution by repeating this for L successively shifted versions of $g(n)$. ◇

From a computational complexity view point, the comments following eqn. (2.9) continue to hold. It can be shown that the number of multiplications for a direct convolution $x(n) * g^*(-n)$ are nearly same as the total number of multiplications required to perform all the necessary subband convolutions. (This neglects the multiplications required to implement the analysis filters $H_k(z)$ and assumes that the lengths of $H_k(z)$ are much smaller than those of $x(n)$ and $g(n)$). The coding gain of the nonuniform orthonormal convolver will be derived in Sec. 3.3.

3. CODING GAIN OF PARAUNITARY CONVOLVERS

Fig. 1.1 shows the paraunitary convolver with quantizers inserted in the subbands of $x(n)$. We will first consider the uniform case ($n_k = M$ for all k). The nonuniform case will be addressed in Sec. 3.3. Assume that $g(n)$ is a fixed filter with no quantizers in its subbands (This assumption can be removed, but only with considerable loss of simplicity of mathematics). For simplicity of analysis we assume that $x(n), g(n)$ and the filter coefficients in $H_k(z)$ are real so that $x_k(n)$ and $g_k(n)$ are is real. This enables us to deal with quantizers that operate on real inputs.

Let b_k denote the number of bits per sample of $x_k(n)$, permitted by the quantizer Q_k. Thus the average bit rate is

$$b = \frac{1}{M} \sum_{k=0}^{M-1} b_k, \tag{3.1}$$

i.e., on the average, we have used b bits per sample of $x(n)$.

Because of the quantization in the subbands, the output of the paraunitary convolver is different from the ideal result $x(n) * g^*(-n)$. To analyze this error, we replace the quantizers Q_k with the noise sources $q_k(n)$ as shown in Fig. 3.1. Consider the paraunitary convolution formula (2.8). In

the presence of quantizers, we are actually computing

$$\sum_{k=0}^{M-1} \Big(x_k(n) + q_k(n) \Big) * g_k^*(-n). \tag{3.2}$$

(According to the realness assumption the conjugate sign is redundant, but we show it for consistency with previous sections). The quantization error is therefore

$$q(n) = \sum_{k=0}^{M-1} q_k(n) * g_k^*(-n). \tag{3.3}$$

The noise model

To perform a statistical analysis, we will make the following assumptions:

1. $x(n)$ is a zero-mean wide sense stationary random process so that the subband signals $x_k(n)$ are zero-mean WSS with some variance, say, $\sigma_{x_k}^2$. We consider $g(n)$ to be a determininistic sequence.
2. The quantization error $q_k(n)$ is zero-mean and white, with variance $\sigma_{q_k}^2$. Also $q_k(n)$ is uncorrelated to $q_m(i), k \neq m$, and to the input $x(n)$ (hence to the quantizer input $x_k(n)$).

It should be noticed that the above assumptions are reasonable as long as the bit rates b_k are moderate or high [23]. In any case, in the absence of such assumptions, it is not usually possible to find an expression for error variance.

3.1. Expression for the error variance

Let $\sigma_{x_k}^2$ denote the variance of $x_k(n)$, and $\sigma_{q_k}^2$ the variance of the quantizer error $q_k(n)$. In order to equalize the overflow probabilty across all the M subbands, these two should be related as

$$\sigma_{q_k}^2 = c\sigma_{x_k}^2 2^{-2b_k}. \tag{3.4}$$

(See, e.g., Chap. 4 of [16]). Here c is a constant, identical for all subbands (which is a valid assumption if all $x_k(n)$ have similar type of distribution, e.g., all Gaussian).

Using (3.3) and the noise model assumptions stated earlier, the variance of $q(n)$ can be expressed as

$$\begin{aligned}\sigma^2_{q(n)} &= \sum_{k=0}^{M-1} \sigma^2_{q_k} \sum_{\ell} |g_k(\ell)|^2 \\ &= c \sum_{k=0}^{M-1} 2^{-2b_k} \sigma^2_{x_k} \sum_{\ell} |g_k(\ell)|^2.\end{aligned} \tag{3.5}$$

This is for $i = 0$ in (2.9). For arbitrary i, the filter $g(n)$ is replaced with $g(n-i)$, and the above equation is modified to

$$\sigma^2_{q(n-i)} = c \sum_{k=0}^{M-1} 2^{-2b_k} \sigma^2_{x_k} \sum_{\ell} |g_k^{(i)}(\ell)|^2, \quad 0 \le i \le M-1, \tag{3.6}$$

where $g_k^{(i)}(n)$ is the kth subband output in response to $g(n-i)$. The dependence on i is removed by averaging over all i. The resulting average variance of $q(n)$ is given by

$$\sigma^2_{q,PU} = \frac{c}{M} \sum_{k=0}^{M-1} 2^{-2b_k} \sigma^2_{x_k} \alpha_k^2, \tag{3.7}$$

where

$$\alpha_k^2 \triangleq \sum_{i=0}^{M-1} \sum_{\ell} |g_k^{(i)}(\ell)|^2. \tag{3.8}$$

The inner summation above represents the energy in the kth subband in response to the input $g(n-i)$. The outer summation removes the dependence on i. Thus α_k^2 is proportional to the average energy of $g(n)$ in the kth subband. Using the paraunitary property, it can be shown that $\sum_k \alpha_k^2/M$ is the total energy in $g(n)$ (see eqn. (3.22) later).

The '*PU*' in the subscript in (3.7) is a reminder of 'paraunitary'. Equation (3.7) gives the average error variance (over a period of length M), and is independent of time.

3.2. Coding gain of the paraunitary convolver

Now consider direct convolution $x(n) * g^*(-n)$. Suppose $x(n)$ is directly quantized to b bits before convolution. Denoting $e(n)$ as the quantization error, the result of quantization is $[x(n) + e(n)] * g^*(-n)$ so that the error is $e(n) * g^*(-n)$. Under usual noise model assumptions, the variance of this error is

$$\sigma^2_{q,PCM} = \sigma_e^2 \sum_{n} |g(n)|^2, \tag{3.9}$$

where σ_e^2 is the variance of $e(n)$, which can be expressed, similar to (3.4), as $\sigma_e^2 = c\sigma_x^2 2^{-2b}$, where σ_x^2 is the variance of $x(n)$. Thus

$$\sigma_{q,PCM}^2 = c\sigma_x^2 2^{-2b} \sum_n |g(n)|^2 \tag{3.10}$$

The ratio

$$G_{PU}(M) = \frac{\sigma_{q,PCM}^2}{\sigma_{q,PU}^2} \tag{3.11}$$

is the coding gain of the paraunitary convolver. The argument M is a reminder that there are M subbands in the system. Substituting from (3.7) and (3.10), this becomes

$$G_{PU}(M) = \frac{2^{-2b}\sigma_x^2 \sum_n |g(n)|^2}{\frac{1}{M}\sum_{k=0}^{M-1} 2^{-2b_k}\sigma_{x_k}^2 \alpha_k^2}. \tag{3.12}$$

In this expression, $\sigma_{x_k}^2$ is the variance of the kth subband signal derived from the input $x(n)$, and $\alpha_k^2 \geq 0$ is related to the k subband of the *filter* $g(n)$. And b is the average bit rate (3.1). Notice that $\sigma_{x_k}^2$ and α_k^2 depend on the analysis filter response $H_k(e^{j\omega})$.

Optimum bit allocation

Under the average bit-rate constraint (3.1), we can maximize the coding gain by optimally allocating the bits b_k among subbands. The idea is very similar to the counterpart in subband coding [16]. For this we note that the numerator of (3.12) is independent of the bit-allocation. We only have to miminize the denominator. For this we invoke the arithmetic-geometric mean inequality [24] (AM-GM inequality) which says this: if $P_k, 0 \leq k \leq M-1$ is a set of nonnegative numbers, then

$$\frac{1}{M}\sum_{k=0}^{M-1} P_k \geq \Big(\prod_{k=0}^{M-1} P_k\Big)^{1/M}, \tag{3.13}$$

with equality if and only if $P_k = P$ for all k. Using this in conjunction with (3.1) we can show that

$$\frac{1}{M}\sum_{k=0}^{M-1} 2^{-2b_k}\sigma_{x_k}^2 \alpha_k^2 \geq 2^{-2b}\prod_{k=0}^{M-1}\left(\sigma_{x_k}^2 \alpha_k^2\right)^{1/M} \tag{3.14}$$

with equality if and only if all terms on the left side above are equal. Since the quantizer variances $\sigma_{q_k}^2$ are given by (3.4), we see that the above condition for equality implies

$$\sigma_{q_k}^2 = c\sigma_{x_k}^2 2^{-2b_k} = \frac{\text{constant}}{\alpha_k^2}$$

The output noise variance due to the kth quantizer (kth term in (3.7)) is therefore independent of k.

We obtain the formula for the optimal bit allocation by setting all the terms on the left side of (3.14) to be equal. The result is

$$b_k = C + \frac{1}{2}\log_2(\sigma_{x_k}^2 \alpha_k^2). \tag{3.15}$$

for some C. By using (3.1) we can eliminate C and obtain

$$b_k = b + 0.5\log_2(\sigma_{x_k}^2 \alpha_k^2) - \frac{0.5}{M}\sum_{k=0}^{M-1}\log_2\left(\sigma_{x_k}^2 \alpha_k^2\right). \tag{3.16}$$

This is very similar to the expressions which can be found in [16], [17] for traditional subband coding systems. The difference is that the product $\sigma_{x_k}^2 \alpha_k^2$ appears in the place of $\sigma_{x_k}^2$. Thus, the energy of the signal as well as the filter $g(n)$ in the kth subband determine the bits b_k. For high bit rate coding, the above expression is useful. As in subband coding, b_k might turn out to be non integral, and sometimes negative if b is not large enough.

Optimum coding gain. *The optimum coding gain is obtained when equality holds in* (3.14), i.e., when all the terms on the left side of (3.14) are equal. This optimum value is

$$G_{PU,optimal}(M) = \frac{\sigma_x^2}{\left(\prod_{k=0}^{M-1}\sigma_{x_k}^2\right)^{1/M}} \times \frac{\sum_n |g(n)|^2}{\left(\prod_{k=0}^{M-1}\alpha_k^2\right)^{1/M}} \tag{3.17}$$

Notice that the above analysis holds for any filter-bank convolver with paraunitary polyphase matrix, regardless of the the quality of the filter responses. The filter responses will in turn determine the values of $\sigma_{x_k}^2$ and α_k^2 for fixed $g(n)$ and $x(n)$.

Lemma 3.1. $G_{PU,optimal}(M) \geq 1$ regardless of the choice of paraunitary filters $H_k(z)$. ◊

Proof. We will rewrite the optimal coding gain (3.17) by expressing σ_x^2 in terms of $\sigma_{x_k}^2$, and $\sum_n |g(n)|^2$ in term of α_k^2.

The variance of the output of $H_k(z)$ is also the variance of the *decimated* subband signal $x_k(n)$ so that

$$\sigma_{x_k}^2 = \frac{1}{2\pi}\int_0^{2\pi} S_{xx}(e^{j\omega})|H_k(e^{j\omega})|^2 d\omega. \tag{3.18}$$

where $S_{xx}(e^{j\omega})$ is the power spectral density of $x(n)$. The paraunitary property $\tilde{\mathbf{E}}(z)\mathbf{E}(z) = \mathbf{I}$ implies $\sum_{k=0}^{M-1} |H_k(e^{j\omega})|^2 = M$. By computing $\sum_k \sigma_{x_k}^2$ from (3.18) we therefore obtain

$$\frac{1}{M}\sum_{k=0}^{M-1} \sigma_{x_k}^2 = \sigma_x^2. \tag{3.19}$$

Next consider the signals generated by the filter bank in response to $g(n-i)$ (Fig. 3.2). Define the vectors

$$\widehat{\mathbf{g}}^{(i)}(n) = \begin{bmatrix} g(Mn-i) \\ g(Mn-1-i) \\ \vdots \\ g(Mn-M+1-i) \end{bmatrix}, \quad \mathbf{g}^{(i)}(n) = \begin{bmatrix} g_0^{(i)}(n) \\ g_1^{(i)}(n) \\ \vdots \\ g_{M-1}^{(i)}(n) \end{bmatrix}, \tag{3.20}$$

for $0 \le i \le M-1$. The superscript i is a reminder that the input is $g(n-i)$. Using the paraunitary property, we conclude [12]

$$\sum_n [\widehat{\mathbf{g}}^{(i)}(n)]^\dagger \widehat{\mathbf{g}}^{(i)}(n) = \sum_n [\mathbf{g}^{(i)}(n)]^\dagger \mathbf{g}^{(i)}(n). \tag{3.21}$$

The left hand side is the energy $\sum_n |g(n)|^2$. Combining this with the definition (3.8) of α_k^2, we obtain

$$\sum_n |g(n)|^2 = \frac{1}{M}\sum_{k=0}^{M-1} \alpha_k^2. \tag{3.22}$$

Substituting (3.19) and (3.22) into (3.17), we arrive at

$$G_{PU,optimal}(M) = \frac{\frac{1}{M}\sum_{k=0}^{M-1} \sigma_{x_k}^2}{\left(\prod_{k=0}^{M-1} \sigma_{x_k}^2\right)^{1/M}} \times \frac{\frac{1}{M}\sum_{k=0}^{M-1} \alpha_k^2}{\left(\prod_{k=0}^{M-1} \alpha_k^2\right)^{1/M}}. \tag{3.23}$$

Using the arithmetic-geometric mean inequality we conclude that $G_{PU,optimal}(M) \ge 1$. $\nabla\nabla\nabla$

Notice that the above proof uses the paraunitary property. The property $G_{PU,optimal}(M) \ge 1$ cannot be claimed for a convolver based on an arbitrary filter bank (i.e., without paraunitary property). The appearance of the arithmetic-geometric mean ratio in the coding gain has been observed in other contexts in traditional subband coding applications. It has been formally proved for the case of ideal brickwall filters and for the case of orthogonal transform coding [16]. Such an expression has also been used for other types of (non ideal) filter banks [25]. The true justification for such use is based on the paraunitary property, as shown above and in [26].

In general the gain can exceed unity for two possible reasons. First the subband variances $\sigma_{x_k}^2$ could be different for different k. And second, the quantity α_k may not be the same in all subbands.

Special cases.

Paraunitary filter banks are special cases of a more general class of perfect reconstruction filter banks [2],[3]. However, they cover a wide range of practical filter banks. In fact, some of the approximate reconstruction systems (viz., the pseudo QMF banks [27]-[30]) are known to satisfy the paraunitary property 'approximately' (see [31]), even though these approximate systems were developed *before* paraunitary filter banks were reported.

1. A special case of paraunitary systems, primarily of theoretical interest, arises when the filters $H_k(e^{j\omega})$ are equispaced ideal brickwall filters as shown in Fig. 3.3. In this case

$$F_k(e^{j\omega}) = H_k(e^{j\omega}) = \begin{cases} \sqrt{M} & \text{if } \omega \in k\text{th passband} \\ 0 & \text{otherwise,} \end{cases} \tag{3.24}$$

and it can be shown that $\mathbf{E}(e^{j\omega})$ is paraunitary (see Sec. 6.2.2 of [12]). In this case, we have

$$\sigma_{x_k}^2 = M \int_{k\text{th band}} S_{xx}(e^{j\omega}) d\omega/2\pi \tag{3.25}$$

where $S_{xx}(e^{j\omega})$ is the power spectrum of $x(n)$. Thus, the coding gain is greater than unity as long as $x(n)$ does not have same 'variance' in all the consecutive frequency bands. The system (3.24) will be called the ideal SBC (subband coding) convolver.

2. A second special case of theoretical interest arises when $H_k(z) = z^{-k}$ for all k. In this case the above results still hold (since $\mathbf{E}(z) = \mathbf{I}$ which is paraunitary); and the coding gain can be verified to be unity.

3. *Case of white input.* Now consider the special case where $x(n)$ is zero-mean and white. Let the response of the filters $H_k(e^{j\omega})$ be arbitrary except for the FIR paraunitary property. Then $\sigma_{x_k}^2$ is identical for all k. This follows because paraunitariness implies in particular, that the energy $\int_0^{2\pi} |H_k(e^{j\omega})|^2 d\omega/2\pi$ is identical for all k (Appendix B). In this case, the coding gain can still exceed unity, because α_k^2 may not be identical for all k.

3.3. Coding gain for the nonuniform orthonormal convolver

In the nonuniform case, Eqn. (2.23) gives the L-fold decimated version of the convolution. To obtain all samples of the convolution, we repeat this operation with $g(n)$ replaced by $g(n-i)$, i.e., $g_k(n)$ replaced by $g_k^{(i)}(n)$ for $0 \le i \le L-1$.

With quantizers inserted as in Fig. 1.1(a), we can replace them with noise sources $q_k(n)$ as in Fig. 3.1. With $x(n)$ and $g(n-i)$ used as the filter bank inputs, the error in the computation of $x_k(n) * [g_k^{(i)}(-n)]^*$ is therefore $\sum_{k=0}^{M-1} q_k(n) * [g_k^{(i)}(-n)]^*$. Proceeding as before, we find the variance of this error to be

$$\sum_{k=0}^{M-1} \sigma_{q_k}^2 \sum_n |g_k^{(i)}(n)|^2. \tag{3.26}$$

Averaging over the L values of i, we obtain the average variance of the error $q(n)$ in the convolution as

$$\sigma_{q,\perp}^2 = \frac{1}{M} \sum_{k=0}^{M-1} \sigma_{q_k}^2 \alpha_k^2 = \frac{c}{M} \sum_{k=0}^{M-1} 2^{-2b_k} \sigma_{x_k}^2 \alpha_k^2. \tag{3.27}$$

The subscript $\perp$ stands for 'orthonormal' filter banks. This is the 'output error variance' of the convolver. Here we have used (3.4). Also, we have defined

$$\alpha_k^2 = (M/L) \sum_{i=0}^{L-1} \sum_n |g_k^{(i)}(n)|^2. \tag{3.28}$$

The bit rate for the kth subband is b_k/n_k. Assume that the total bit rate is constrained to be b. Then the bit rate constraint is

$$\sum_{k=0}^{M-1} \frac{b_k}{n_k} = b. \tag{3.29}$$

To obtain the optimal bit allocation, we can minimize $\sigma_{q,\perp}^2$ under the above constraint by use of the Lagrange multiplier method. That is, form the Lagrangian $\phi = \sigma_{q,\perp}^2 - \lambda(\sum_{k=0}^{M-1} \frac{b_k}{n_k} - b)$ and set $\partial\phi/\partial b_k = 0$. This results in the set of equations

$$2^{2b_k} = D\sigma_{x_k}^2 \alpha_k^2 n_k, \quad 0 \le k \le M-1. \tag{3.30}$$

where D is a constant independent of k. † Taking logarithm and using (3.29), we can evaluate the constant D to be

$$D = \frac{2^{2b}}{\prod_{i=0}^{M-1}\left(\sigma_{x_i}^2 \alpha_i^2 n_i\right)^{1/n_i}} \tag{3.31}$$

Substituting this into (3.30) and taking logarithm, we obtain the following formula:

$$b_k = b + 0.5\log_2(n_k \sigma_{x_k}^2 \alpha_k^2) - 0.5 \sum_{i=0}^{M-1} \frac{\log_2(n_i \sigma_{x_i}^2 \alpha_i^2)}{n_i}. \tag{3.32}$$

for optimal bit allocation. Under this condition, the variance of the kth quantizer noise is given by

$$\sigma_{q_k}^2 = c2^{-2b_k}\sigma_{x_k}^2 = \frac{c}{D\alpha_k^2 n_k} = \frac{c2^{-2b}\prod_{i=0}^{M-1}\left(\sigma_{x_i}^2 \alpha_i^2 n_i\right)^{1/n_i}}{\alpha_k^2 n_k} \tag{3.33}$$

which is proportional to $1/(\alpha_k^2 n_k)$. With optimum bit allocation, the output noise variance contributed by the kth quantizer (kth term in (3.27)) simplifies to $c/(DMn_k)$, and is proportional to $1/n_k$. The the total output noise variance is

$$\sigma_{q,\perp}^2 = \frac{c}{DM}\sum_{k=0}^{M-1} 1/n_k = \frac{c}{DM} = \frac{c2^{-2b}}{M}\prod_{i=0}^{M-1}\left(\sigma_{x_i}^2 \alpha_i^2 n_i\right)^{1/n_i} \tag{3.34}$$

The coding gain, defined as $G_\perp(M) = \sigma_{q,PCM}^2/\sigma_{q,\perp}^2$ can now be calculated. Thus, using (3.10) and (3.34) we obtain

$$G_\perp(M) = \frac{\sigma_x^2 \sum_n |g(n)|^2}{\frac{1}{M}\prod_{i=0}^{M-1}\left(n_i \sigma_{x_i}^2 \alpha_i^2\right)^{1/n_i}} \tag{3.35}$$

Notice that these results reduce to those in Sec. 3.2 if we set $n_k = M$ for all k. Another special case of interest in many applications (speech and image coding) is the filter bank with analysis filter responses as in Fig. 3.4. The responses have an octave spacing and correspondingly increasing bandwidths (constant Q filter bank). Such a system can be generated by use of a tree-structured system, where one of the two signals from the previous stage is further split into two in the next

† The fact that this represents a minimum rather than maximum can be verified in many ways. For example one can verify in this case that the Hessian of the Lagrangian [32] is a diagonal matrix with positive elements.

stage [10], and so forth. The orthonormality property can be satisfied in such a system by use of 2×2 paraunitary polyphase matrices at each level of the tree. The above theory can be applied for these systems, with

$$n_0 = n_1 = 2n_2 = 4n_3 = \ldots$$

3.4. The special case of traditional subband coding

The results derived above for the paraunitary convolvers (uniform as well as nonuniform) can be used to derive the optimal bit allocation and coding gain for orthonormal subband coding systems, i.e., systems of the form in Fig. 1.1(a). This is done by setting $g(n) = \delta(n)$. Under this condition, the quantity $g_k^{(i)}(n)$ is the decimated version of $h_k(n-i)$, where $h_k(n)$ is the impulse response of the analysis filter $H_k(z)$. Using the fact that the analysis filters have unit energy under the paraunitary constraint, one can verify that $\alpha_k^2 = M/n_k$. Substituting this we obtain the reconstruction error variance, i.e., variance of $x(n) - \hat{x}(n)$ in Fig. 1.1(a). This can be obtained from (3.27) as

$$\sigma_{q,\perp}^2 = \sum_{k=0}^{M-1} \frac{\sigma_{q_k}^2}{n_k} \tag{3.36}$$

The optimal bit allocation rule reduces to

$$b_k = b + 0.5\log_2(\sigma_{x_k}^2) - 0.5 \sum_{i=0}^{M-1} \frac{\log_2(\sigma_{x_i}^2)}{n_i}. \tag{3.37}$$

and the optimized coding gain becomes

$$G_\perp(M) = \frac{\sigma_x^2}{\prod_{i=0}^{M-1} \left(\sigma_{x_i}^2\right)^{1/n_i}} \tag{3.38}$$

Since $\alpha_k^2 = M/n_k$ in this case, we see from (3.33) that the variance $\sigma_{q_k}^2$ of the kth quantizer noise is independent of k, and is given by

$$\sigma_{q_k}^2 = c2^{-2b} \prod_{i=0}^{M-1} (\sigma_{x_i}^2)^{1/n_i} \tag{3.39}$$

The contribution to the output noise variance $\sigma_{q,\perp}^2$, coming from the kth quantizer (kth term in (3.36), is proportional to $1/n_k$.

Summarizing, the above expressions are applicable to any subband coder (possibly unequal decimation ratios, but maximally decimated) with orthonormal filters, under the noise model assumptions stated at the beginning of Sec. 3. The further special case where $n_k = M$ has been reported in many references in the past [16],[17],[25]. All these references assume ideal nonoverlapping subband filters, but that assumption is not necessary as the above analysis shows; orthonormality (paraunitariness in the uniform case) is really sufficient.

4. GENERAL ORTHOGONAL TRANSFORM CONVOLVER

The optimal coding gain (3.23) depends on the choice of the paraunitary matrix $\mathbf{E}(z)$. A natural problem of interest here is the choice of *optimal* paraunitary $\mathbf{E}(z)$ of a given degree J (for fixed number of channels M) which further maximizes the coding gain. In general this is a difficult problem, although some progress can be made in the special case where $J = 0$, i.e., $\mathbf{E}(z)$ is a constant unitary matrix $\mathbf{T}$. This is shown in Fig. 4.1(a). We will now consider the optimization problem for this special case. This special case is particularly attractive because the analysis filters $H_k(z)$ have length $\leq M$ (which could be much smaller that the lengths of $x(n)$ and $g(n)$). In this case the complexity of implementing the analysis and synthesis filters is negligible (compared to the complexity of the convolutions $x_k(n) * g_k^*(-n)$), and can therefore be disregarded. However, significant coding gain can still be achieved, as we will demonstrate later.

With $\mathbf{T}$ taken to be unitary, i.e., $\mathbf{T}^\dagger\mathbf{T} = \mathbf{I}$, the system is a paraunitary perfect reconstruction filter bank [2]. This is similar to the orthogonal transform coding system [16]. The convolution theorem (Theorem 2.1) clearly continues to hold in this case, and so do the coding gain expressions of the previous section. We will now address the problem of finding the optimal $\mathbf{T}$ that maximizes the coding gain (3.17) under optimal bit allocation. It will again be assumed that the signals $x(n), g(n)$ and the matrix $\mathbf{T}$ are real. We will first simplify the expression (3.17) by writing $\sigma_{x_k}^2$ and α_k^2 directly in terms of $\mathbf{T}$.

Expressions for $\sigma_{x_k}^2$ and α_k^2

First refer to Fig. 4.1(a). Define the vectors $\hat{\mathbf{x}}(n)$ and $\mathbf{x}(n)$ as

$$\hat{\mathbf{x}}(n) = \begin{bmatrix} x(Mn) \\ x(Mn-1) \\ \vdots \\ x(Mn-M+1) \end{bmatrix}, \quad \mathbf{x}(n) = \begin{bmatrix} x_0(n) \\ x_1(n) \\ \vdots \\ x_{M-1}(n) \end{bmatrix}. \tag{4.1}$$

Then $\mathbf{x}(n) = \mathbf{T}\hat{\mathbf{x}}(n)$. Assuming that $x(n)$ is WSS, the vector processes $\hat{\mathbf{x}}(n)$ and $\mathbf{x}(n)$ are WSS. Define the autocorrelations

$$\hat{\mathbf{R}}_{\mathbf{xx}} = E[\hat{\mathbf{x}}(n)\hat{\mathbf{x}}^{\dagger}(n)], \quad \text{and} \quad \mathbf{R}_{\mathbf{xx}} = E[\mathbf{x}(n)\mathbf{x}^{\dagger}(n)]. \tag{4.2}$$

Then

$$\mathbf{R}_{\mathbf{xx}} = \mathbf{T}\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{T}^{\dagger}. \tag{4.3}$$

The quantity $\sigma_{x_k^2}$ is the diagonal element $[\mathbf{R}_{\mathbf{xx}}]_{kk}$ so that the product of these (which appears in the denominator of (3.17)) is given by

$$\prod_{k=0}^{M-1} \sigma_{x_k}^2 = \prod_{k=0}^{M-1} \left(\mathbf{T}\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{T}^{\dagger}\right)_{kk} \tag{4.4}$$

Next refer to Fig. 4.1(b). Define the vectors $\hat{\mathbf{g}}^{(i)}(n)$ and $\mathbf{g}^{(i)}(n)$ as in (3.20). We then have $\mathbf{g}^{(i)}(n) = \mathbf{T}\hat{\mathbf{g}}^{(i)}(n)$. Thus

$$\begin{aligned} \alpha_k^2 &= \sum_{i=0}^{M-1} \sum_n \left(\mathbf{g}^{(i)}(n)[\mathbf{g}^{(i)}(n)]^{\dagger}\right)_{kk} \quad \text{(from the definition (3.8))} \\ &= \left(\mathbf{T} \sum_{i=0}^{M-1} \sum_n \hat{\mathbf{g}}^{(i)}(n)[\hat{\mathbf{g}}^{(i)}(n)]^{\dagger}\mathbf{T}^{\dagger}\right)_{kk} \\ &= \left(\mathbf{T}\hat{\mathbf{R}}_{\mathbf{gg}}\mathbf{T}^{\dagger}\right)_{kk} \end{aligned} \tag{4.5}$$

where

$$\hat{\mathbf{R}}_{\mathbf{gg}} = \sum_{i=0}^{M-1} \sum_n \hat{\mathbf{g}}^{(i)}(n)[\hat{\mathbf{g}}^{(i)}(n)]^{\dagger} \tag{4.6}$$

Summarizing, the coding gain (3.17) can be expressed as

$$G_{TC}(M) = \frac{\sigma_x^2 \sum_n |g(n)|^2}{\left(\prod_{k=0}^{M-1}\left(\mathbf{T}\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{T}^{\dagger}\right)_{kk} \prod_{k=0}^{M-1}\left(\mathbf{T}\hat{\mathbf{R}}_{\mathbf{gg}}\mathbf{T}^{\dagger}\right)_{kk}\right)^{1/M}} \tag{4.7}$$

The subscript TC stands for 'transform coding'. The expression (4.7) holds under the optimal bit allocation condition (3.16). The unitary matrix $\mathbf{T}$ should be chosen so as to minimize the product in the denominator.

Properties of the matrices $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\hat{\mathbf{R}}_{\mathbf{gg}}$. The $M \times M$ matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ is the autocorrelation matrix derived from a scalar WSS process $x(n)$, and is therefore Hermitian, Toeplitz, and positive semidefinite. It is also positive definite unless $x(n)$ is harmonic (i.e., the power spectrum is made of impulses $\delta(\omega - \omega_k)$). It can be shown that $\hat{\mathbf{R}}_{\mathbf{gg}}$ also has all these properties, i.e., Hermitian, Toeplitz and positive definite unless $G(e^{j\omega})$ is made of impulses. (See Appendix C). In fact it turns out that $[\hat{\mathbf{R}}_{\mathbf{gg}}]_{km} = \sum_n g(n)g^*(n + k - m))$ so that it is a determininstic autocorrelation matrix.

The problem of finding the optimal transformation $\mathbf{T}$ therefore reduces to the following: given the $M \times M$ Hermitian, Toeplitz and positive definite matrices $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\hat{\mathbf{R}}_{\mathbf{gg}}$, find a unitary matrix $\mathbf{T}$ such that

$$\prod_{k=0}^{M-1} \left(\mathbf{T}\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{T}^\dagger\right)_{kk} \prod_{k=0}^{M-1} \left(\mathbf{T}\hat{\mathbf{R}}_{\mathbf{gg}}\mathbf{T}^\dagger\right)_{kk} \tag{4.8}$$

is minimized.

Given a Hermitian positive definite matrix $\mathbf{P}$, consider the product $\prod_{k=0}^{M-1}\left(\mathbf{T}\mathbf{P}\mathbf{T}^\dagger\right)_{kk}$ where $\mathbf{T}$ is constrained to be unitary. It is known that this product is minimized if and only if the columns of $\mathbf{T}^\dagger$ are eigenvectors of $\mathbf{P}$. (This is how the traditional Karhunen-Loeve transform (KLT) is obtained [16]). Under this condition $\mathbf{T}\mathbf{P}\mathbf{T}^\dagger$ is diagonal. However, in our case, two positive definite matrices are involved. The problem of finding a single unitary matrix $\mathbf{T}$ that minimizes the product (4.8) does not appear to have a simple, known, solution.

If the matrices $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\hat{\mathbf{R}}_{\mathbf{gg}}$ are diagonalizable by the same unitary matrix $\mathbf{T}$, then this $\mathbf{T}$ maximizes the coding gain. This condition for simultaneous diagonalization is equivalent to either of the following two conditions [33]:

1. $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\hat{\mathbf{R}}_{\mathbf{gg}}$ commute, i.e., $\hat{\mathbf{R}}_{\mathbf{xx}}\hat{\mathbf{R}}_{\mathbf{gg}} = \hat{\mathbf{R}}_{\mathbf{gg}}\hat{\mathbf{R}}_{\mathbf{xx}}$.
2. $\hat{\mathbf{R}}_{\mathbf{xx}}\hat{\mathbf{R}}_{\mathbf{gg}}$ is Hermitian.

For the special case of 2×2 real matrices (i.e., $M = 2$, and $x(n), g(n)$ and $\mathbf{T}$ are real), the above

conditions are satisfied for the following reason: The matrices $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\hat{\mathbf{R}}_{\mathbf{gg}}$ are 2×2 symmetric Toeplitz, so that they are also *circulant*. But circulant matrices commute [34]. The two matrices are simultaneously diagonalizable by the unitary matrix

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{4.9}$$

With this choice of $\mathbf{T}$ the coding gain reduces to

$$G_{TC}(2) = \frac{1}{\sqrt{(1-\rho_x^2)(1-\rho_g^2)}} \tag{4.10}$$

where $\rho_x = E[x(n)x^*(n-1)]/\sigma_x^2$, and $\rho_g = \sum_n g(n)g^*(n-1)/\sum_n |g(n)|^2$. For example, if $\rho_x = \rho_g = 0.95$ then the coding gain is $G_{TC}(2) = 10.26$.

Bound on the coding gain.

For a Hermitian positive definite matrix $\mathbf{P}$, we have $[\mathbf{P}]_{ii} \geq \det \mathbf{P}$ with equality if and only if $\mathbf{P}$ is diagonal. Using this we see that the gain (4.7) is bounded as

$$G_{TC}(M) \leq \frac{\sigma_x^2 \sum_n |g(n)|^2}{\left([\det \hat{\mathbf{R}}_{\mathbf{xx}}][\det \hat{\mathbf{R}}_{gg}]\right)^{1/M}} = \frac{\sigma_x^2 \sum_n |g(n)|^2}{\left(\det [\hat{\mathbf{R}}_{\mathbf{xx}} \hat{\mathbf{R}}_{gg}]\right)^{1/M}} \tag{4.11}$$

5. NUMERICAL EXAMPLES

In the following examples, we will demonstrate the coding gains of the paraunitary convolvers. The signals $x(n), g(n)$, and the number of subbands M are chosen as follows:

1. Number of subbands $M = 6$ in all cases.
2. Many choices of $g(n)$ are used, but all of these are such that $G(e^{j\omega})$ is lowpass as demonstrated in Fig. 5.1. All choices have the same bandedges. To obtain different stopband attenuations, we change the length of $g(n)$, but retain the same band edges for $G(e^{j\omega})$.
3. The input signal $x(n)$ is taken to be an autoregressive process of order five [i.e., an $AR(5)$ process]. The autocorrelation coefficients $R(k)$, for $0 \leq k \leq 5$, are obtained from Table 2.2 of [16] (lowpass speech source). Where necessary, the power spectrum $S_{xx}(e^{j\omega})$ is computed as

$S_{xx}(e^{j\omega}) = \alpha/|1 + \sum_{n=1}^{5} a_n e^{-j\omega n}|^2$ where a_n are the autoregressive coefficients (obtainable by solving the optimal fifth order linear-prediction problem [16]).

Fig. 5.2 shows the coding gain of the paraunitary convolver (with optimal bit allocation) as a function of the stop band attenuation of $G(e^{j\omega})$, for three cases. The topmost curve corresponds to the ideal SBC convolver. In other words, the analysis and synthesis filters are as in Fig. 3.3 (ideal brickwall filters (Eq. (3.24)). The bottom curve is for the DCT convolver, that is an orthogonal transform convolver (Fig. 4.1) in which the matrix T is taken to be the 6 x 6 DCT matrix. (Four types of DCT matrix have been defined in the literature; we have used the one in Eq. (12.157) of [16].) † The middle curve shows the upper bound (4.11) for the orthogonal transform convolver. It is interesting to note that the DCT system is only about 0.5 dB worse than the bound. The ideal brickwall SBC convolver is about 2.5 dB better than the DCT convolver. The DCT convolver, however, is very simple to implement (less expensive than good filters approximating the ideal SBC filters). In all the above cases the coding gain improves with the attenuation of $G(e^{j\omega})$ because the AM/GM ratio in Eq. (3.17) improves.

In the above experiment suppose we take $g(n) = \delta(n)$. Then the coding gain of the convolver is equal to the coding gain of the traditional subband coding system. For the ideal SBC filters, this value is G = 6.72 dB, and for transform coding with DCT this is 5.3 dB (consistent with experiments on speech coding; for example, see page 542 of [16]). Thus, the additional gain seen in Fig. 5.2 is contributed by the filter $G(e^{j\omega})$ participating in the subband convolver.

We have not shown plots of the coding gain with respect to the number of channels M, as it does not reveal more insights than what is already known in subband coding practice [16],[36].

6. CONCLUDING REMARKS

In this paper we have introduced the convolution theorems for filter bank transfomers. Both

† The motivation for the use of the DCT is that, in traditional speech coding, it is known to be an excellent substitute for the optimal (KLT) transform.

uniform and nonuniform decimation ratios were considered, and the theorems simplified for the case of paraunitary and orthonormal convolvers. Expressions for optimal bit allocation and the optimized coding gain were derived, and numerically demonstrated. The contribution to coding gain comes partly from the nonuniformity of the signal spectrum $S_{xx}(e^{j\omega})$, and partly from nonuniformity of the filter spectrum $|G(e^{j\omega})|^2$. Thus, even if $x(n)$ is nearly white, the coding gain can be large. With $g(n)$ taken to be the unit pulse function $\delta(n)$, the coding gain expressions reduce to those for traditional subband and transform coding, many of which are well-known.

The paraunitary (more generally orthonormal) convolver has about the same computational complexity as a traditional convolver, if the analysis bank has small complexity compared to the convolution itself. Such, indeed, is the case in the special case of the orthogonal transform convolver (Fig. 4.1) where the analysis filter bank has filter lengths $\leq M$ (number of bands). In spite of this simplicity, the coding gain obtainable can already be quite significant. Even though there is no closed form expression for the optimal orthogonal convolver matrix T, we could derive an upper bound for this (for fixed M), and the DCT matrix offers a gain very close to this bound for the case of speech signals.

Appendix A

If $k = m$ we can rewrite (2.15) as $\sum_n f_k(n) f_k^*(n + n_k(\ell - i)) = \delta(\ell - i)$ and (2.17) as $\sum_n f_k(n) f_k^*(n + n_k p) = \delta(p)$. Evidently these imply each other.

Next let $k \neq m$. First assume that (2.15) holds. Recall $n_{k,m} = \gcd(n_k, n_m)$. Thus, there exist integers a and b such that $n_k a - n_m b = n_{k,m}$. Therefore, given any integer p there exists integers ℓ and i such that $n_k \ell - n_m i = n_{k,m} p$. Thus the left hand side in (2.17) can always be rewritten to resemble the left hand side of (2.15). Since $k \neq m$, this left hand side is indeed zero, so that the left hand side of (2.17) is zero as well. Conversely let (2.17) be true. Given a pair of integers ℓ, i we can always write $n_k \ell - n_m i = n_{k,m} p$ for some integer p. So the left side of (2.15) can be rewritten to resemble the left side of (2.17). Since $k \neq m$, (2.17) says that this is zero, so that the

same follows for (2.15).

Appendix B

If the filter bank in Fig. 1.1(a) (with $n_k = M$ for all k) is paraunitary, then the matrix $\mathbf{R}(z)$ (Fig. 1.2) is, in particular, paraunitary. This imples (2.5) which in turn means $\sum_n |f_k(n)|^2 = 1$. Since perfect reconstruction is achieved by choosing $f_k(n) = h_k^*(-n)$, we also have $\sum_n |h_k(n)|^2 = 1$. So all the analysis filters have the same energy.

Appendix C

Since $\hat{\mathbf{R}}_{\mathbf{xx}}$ is the autocorrelation matrix obtained from a scalar WSS process $x(n)$, it is positive semidefinite. It is therefore positive definite if and only if it is nonsingular. If this matrix is singular, then there exists $\mathbf{v} \neq 0$ such that $\mathbf{v}^\dagger \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{v} = 0$, i.e., $E[|\mathbf{v}^\dagger \hat{\mathbf{x}}(n)|^2] = 0$, i.e., $\mathbf{v}^\dagger \hat{\mathbf{x}}(n) = 0$. In other words, there exists an FIR filter $V(z) \triangleq v_0^* + v_1^* z^{-1} + \ldots + v_{M-1}^* z^{-(M-1)}$ such that the output in response to the WSS process $x(n)$ is zero. Thus if $S_{xx}(e^{j\omega})$ denotes the power spectrum of $x(n)$, then the power spectrum of the output is $S_{xx}(e^{j\omega})|V(e^{j\omega})|^2 = 0$. Since the FIR filter $V(z)$ can have at most $M-1$ zeros on the unit circle, this means that the power spectrum has the form $S_{xx}(e^{j\omega}) = \sum_{k=1}^{M-1} c_k \delta(\omega - \omega_k)$, i.e., $x(n)$ is a harmonic process. Thus, unless $x(n)$ is harmonic, $\hat{\mathbf{R}}_{\mathbf{xx}}$ is positive definite. This is a well-known fact [35], and is reviewed here only for completeness.

Next consider $\hat{\mathbf{R}}_{\mathbf{gg}}$ defined in (4.6). Using the definition of $\hat{\mathbf{g}}^{(i)}(n)$ in (3.20) we see that

$$\begin{aligned} [\hat{\mathbf{R}}_{\mathbf{gg}}]_{pq} &= \sum_{i=0}^{M-1} \sum_n g(Mn - i - p) g^*(Mn - i - q)] \\ &= \sum_\ell g(\ell - p) g^*(\ell - q) = R_{gg}(q - p) \end{aligned} \tag{C.1}$$

where $R_{gg}(k)$ is the deterministic autocorrelation of the sequence $g(n)$. Thus $\hat{\mathbf{R}}_{\mathbf{gg}}$ is a determininstic autocorrelation matrix and has all the properties of $\hat{\mathbf{R}}_{\mathbf{xx}}$. It can be written as

$$\hat{\mathbf{R}}_{\mathbf{gg}} = \sum_n \begin{bmatrix} g(n) \\ g(n-1) \\ \vdots \\ g(n-M+1) \end{bmatrix} [g^*(n) \quad g^*(n-1) \quad \ldots \quad g^*(n-M+1)] \tag{C.2}$$

If this is singular, then there exists a vector $\mathbf{c} \neq 0$ such that $\mathbf{c}^\dagger \hat{\mathbf{R}}_{\mathbf{gg}} \mathbf{c} = 0$. Thus, for each n in (C.2), we must have $c_0^* g(n) + c_1^* g(n-1) + \ldots c_{M-1}^* g(n-M+1) = 0$, where at least one c_i is nonzero.

Proceeding as in the previous paragraph, we see that this happens only if $g(n)$ is either zero or made of at most $M-1$ impulses.

ACKNOWLEDGEMENTS

I am thankful to See-May Phoong, graduate student, Dept. of Electrical Engineering, California Institute of Technology, for generating the numerical examples presented in Sec. 5.

References

[1] Crochiere, R. E., and Rabiner, L. R. *Multirate digital signal processing*, Englewood Cliffs, NJ: Prentice Hall, 1983.

[2] Vaidyanathan, P. P. "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," *Proc. of the IEEE*, vol. 78, pp. 56-93, Jan. 1990.

[3] Vetterli, M. "A theory of multirate filter banks," *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. ASSP-35, pp. 356-372, March 1987.

[4] Vetterli, M. "Running FIR and IIR filtering using multirate filter banks," *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. ASSP-36, pp. 730-738, May 1988.

[5] Smith, M. J. T., and Barnwell III, T. P. "A unifying framework for analysis/synth- esis systems based on maximally decimated filter banks," Proc. IEEE Int. Conf. Acoust. Speech, and Signal Proc., pp. 521-524, Tampa, FL, March 1985.

[6] Daubechies, I. "Orthonormal bases of compactly supported wavelets," *Comm. on Pure and Appl. Math.*, vol. 4, pp. 909-996, Nov. 1988.

[7] Daubechies, I. "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Info. Theory*, vol. IT-36, pp. 961-1005, Sept. 1990.

[8] Mallat, S. "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Pattern Anal., and Machine Intell.*, vol. 11, pp. 674-693, July 1989.

[9] Tewfik, A. H., Sinha, D., and Jorgensen, P. E. "On the optimal choice of a wavelet for signal representation," *IEEE Trans. Info. Theory*, to appear.

[10] Rioul, O., and Vetterli, M. "Wavelets and signal processing," *IEEE Signal Processing magazine*, pp. 14-38, Oct. 1991.

[11] Vetterli, M., and Herley, C. "Wavelets and filter banks," *IEEE Trans. on Signal Processing*, vol. SP-40, 1992.

[12] P. P. Vaidyanathan, *Multirate systems and filter banks*, Englewood Cliffs, NJ: Prentice Hall, 1992.

[13] Vaidyanathan, P. P. "Theory and design of M-channel maximally decimated quad-rature mirror filters with arbitrary M, having perfect reconstruction property," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-35, pp. 476-492, April 1987.

[14] Vaidyanathan, P. P. "Quadrature mirror filter banks, M-band extensions and perfect reconstruction techniques," *IEEE ASSP magazine*, vol. 4, pp. 4-20, July 1987.

[15] Crochiere, R. E. "Subband coding," *Bell System Tech. J.*, vol. 60, pp. 1633-1654, Sept. 1981.

[16] Jayant, N. S., and Noll, P. *Digital coding of waveforms*, Prentice Hall, Inc., Englewood Cliffs, 1984.

[17] Woods, J. W. *Subband image coding*, Kluwer Academic Publishers, Inc., 1991.

[18] Huang, Y., and Schultheiss, P. M. "Block quantization of correlated Gaussian random variables," *IEEE Trans. Comm. Syst.* pp. 289-296, Sept. 1963.

[19] Segall, A. "Bit allocation and encoding for vector sources," *IEEE Trans. on Info. Theory*, pp. 162-169, March 1976.

[20] Kovačević, J., and Vetterli, M. "Perfect reconstruction filter banks with rational sampling rate changes," Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc. pp. 1785-1788, Toronto, Canada, May 1991.

[21] Nayebi, K., Barnwell, III, T. P. and Smith, M. J. T. "The design of perfect reconstruction nonuniform band filter banks," Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc., pp. 1781-1784, Toronto, Canada, May 1991.

[22] Soman, A., and Vaidyanathan, P. P. "On orthonormal wavelets and paraunitary filter banks," IEEE Trans. Signal Processing, submitted.

[23] Barnes, C. W., Tran, B. N., and Leung, S. H. "On the statistics of fixed-point roundoff error," *IEEE Trans. on Acoust. Speech and Signal Processing*, pp. 595-606, vol. ASSP-33, June 1985.

[24] Beckenbach, E., and Bellman, R. *An introduction to inequalities.* Random House, 1961.

[25] Malvar, H. S. "Lapped transforms for efficient transform/subband coding," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-38, pp. 969-978, June 1990.

[26] Soman, A., and Vaidyanathan, P. P. "Coding gain in paraunitary analysis/synthesis syst" IEEE Trans. Signal Processing, submitted.

[27] Nussbaumer, H. J. "Pseudo QMF filter bank," *IBM Tech. disclosure Bulletin*, vol. 24, pp. 3081-3087, Nov. 1981.

[28] Rothweiler, J. H. "Polyphase quadrature filters, a new subband coding technique," Proc. of the IEEE Int. Conf. on ASSP, pp. 1980-1983, Boston, MA, April 1983.

[29] Cox, R. V. "The design of uniformly and nonuniformly spaced pseudo QMF," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-34, pp. 1090-1096, Oct. 1986.

[30] Masson, J., and Picel, Z. "Flexible design of computationally efficient nearly perfect QMF filter banks," Proc. of the IEEE Int. Conf. on ASSP, pp. 14.7.1-14.7.4, Tampa, FL, March, 1985.

[31] Koilpillai, R. D. and Vaidyanathan, P. P. "Cosine-modulated FIR filter banks satisfying perfect reconstruction," *IEEE Trans. on Signal Processing*, vol. SP-40, April 1992.

[32] Luenberger, D. G. *Introduction to linear and nonlinear programming*, Addison-Wesley, 1973.

[33] Horn, R. A., and Johnson, C. R. *Matrix analysis*, Cambridge Univ. Press, 1985.

[34] Davis, P. J. *Circulant matrices*, New York, Wiley, 1979.

[35] Kay, S. M., and Marple, S. L. "Spectrum analysis: a modern perspective," *Proc. of the IEEE*, vol. 69, pp. 1380-1419, Nov. 1981.

[36] Akansu, A. N., and Liu, Y. "On signal decomposition techniques," *Optical engr.*, vol. 30, pp. 912-920, July 1991.

LIST OF FIGURES

Fig. 1.1. The maximally decimated filter bank, (a) with input $x(n)$, and (b) with input $g(n)$.

Fig. 1.2. Polyphase representation of the filter bank with equal decimation ratios.

Fig. 3.1. The quantizer and its noise model.

Fig. 3.2. Response of the filter bank to a shifted input.

Fig. 3.3. Magnitude response of ideal brick-wall analysis filters. Synthesis filters for perfect reconstruction have the same magnitude responses.

Fig. 3.4. Magnitude responses of ideal analysis filters, for a well-known class of nonuniform filter banks.

Fig. 4.1. The orthogonal transform convolver. (a) $x(n)$ is input to the filter bank, and (b) shifted $g(n)$ is input to the filter bank.

Fig. 5.1. A typical magnitude response of the filter $g(n)$ used in the experiment.

Fig. 5.2. Demonstration of the coding gains of paraunitary convolvers.

FOOTNOTES

1. Manuscript Received _____

2. This work was supported by National Science Foundation Grant MIP 8919196, and funds from Tektronix, Inc., and Rockwell, International.

3. The author is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

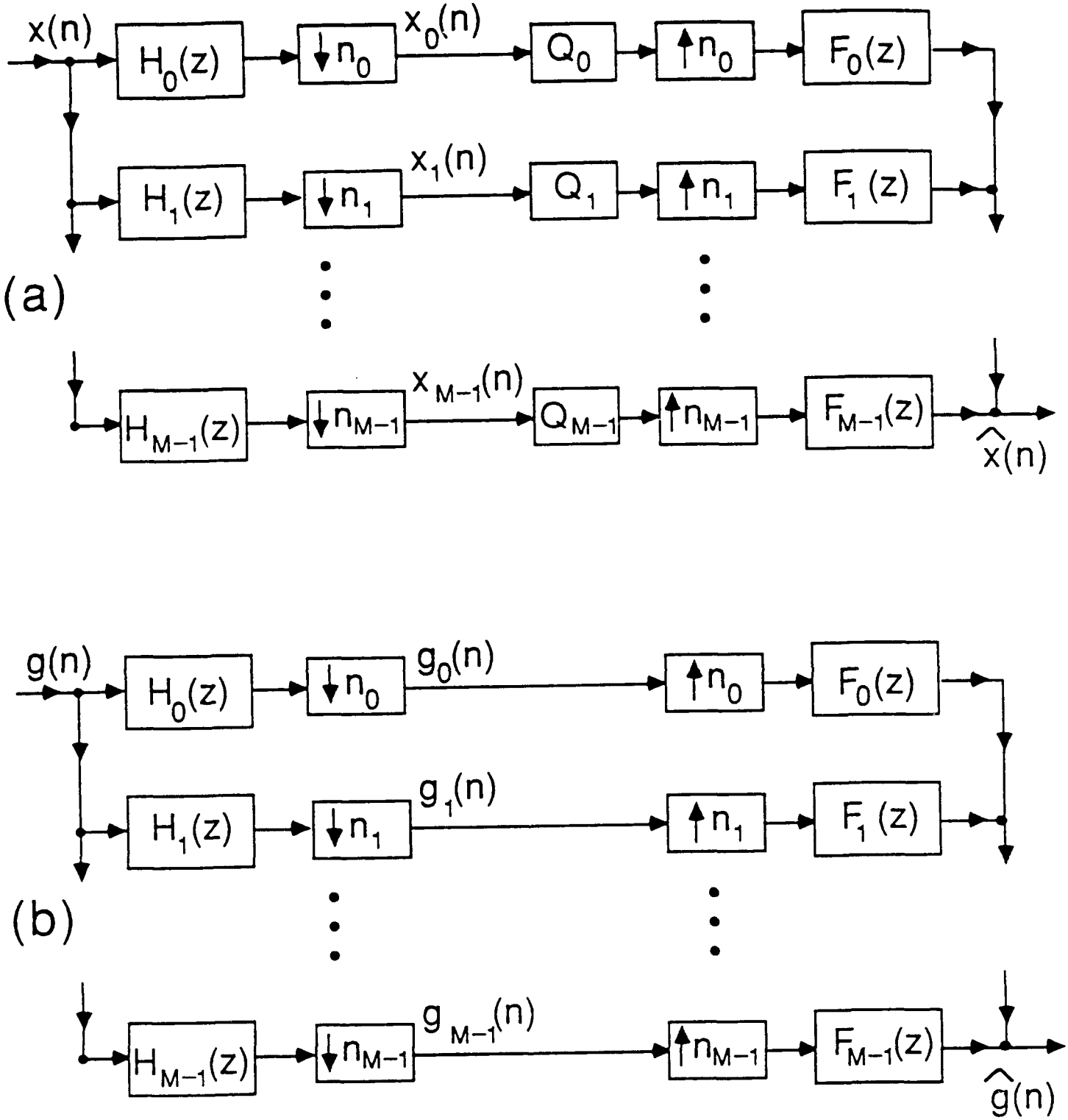


Fig. 1.1. The maximally decimated filter bank, (a) with input x(n), and (b) with input g(n).

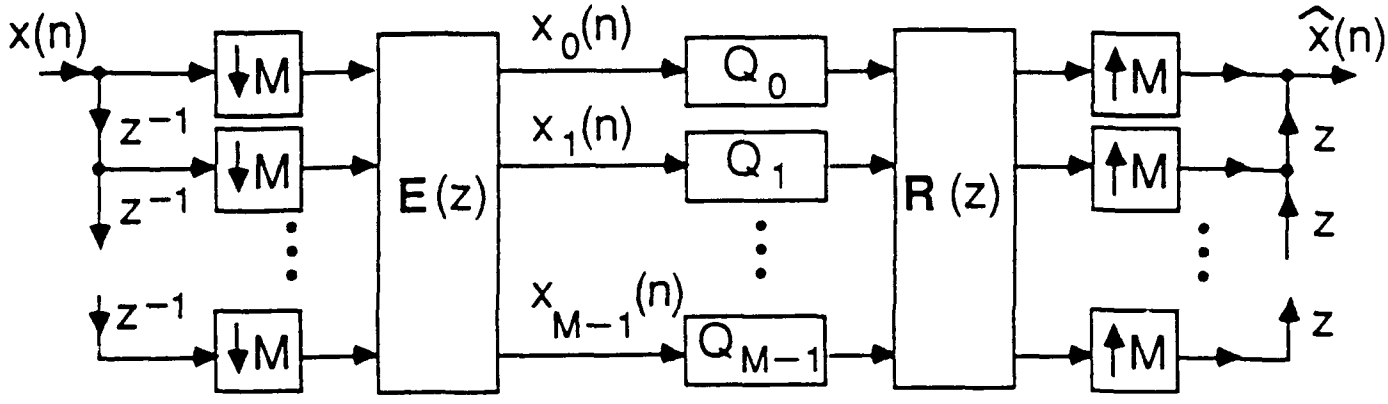


Fig. 1.2. Polyphase representation of the filter bank with equal decimation ratios.

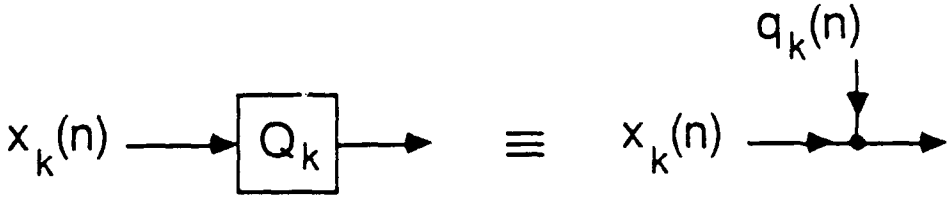


Fig. 3.1. The quantizer and its noise model.

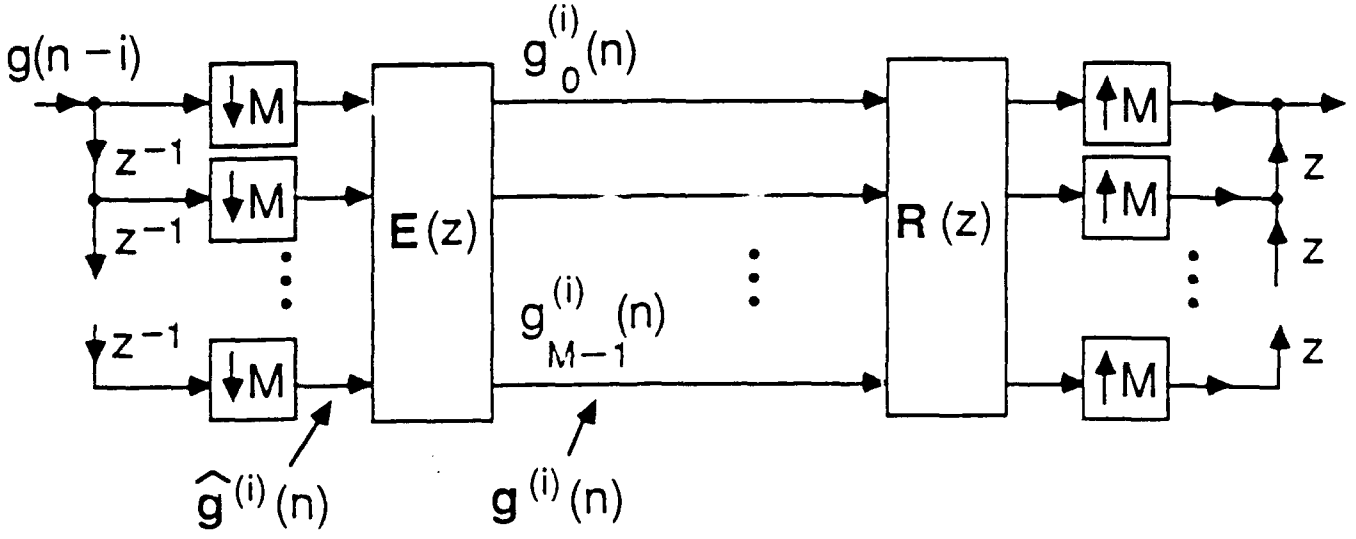


Fig. 3.2. Response of the filter bank to a shifted input.

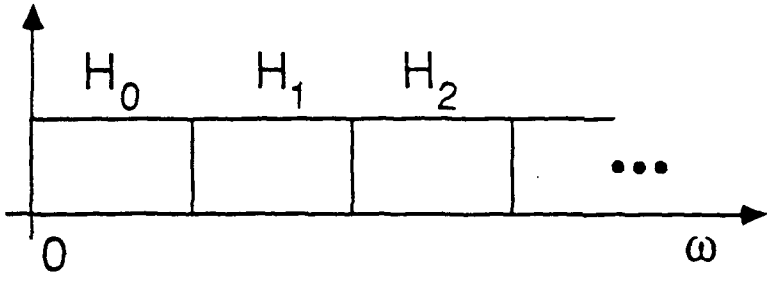


Fig. 3.3. Magnitude responses of ideal brick-wall analysis filters. Synthesis filters for perfect reconstruction have the same magnitude responses.

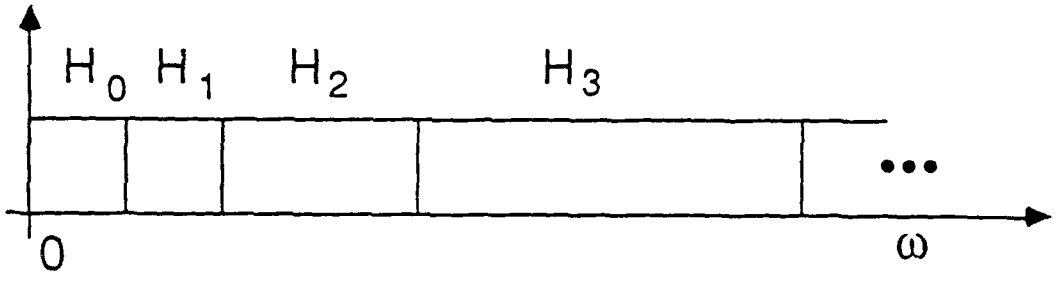


Fig. 3.4. Magnitude responses of ideal analysis filters, for a well-known class of nonuniform filter banks.

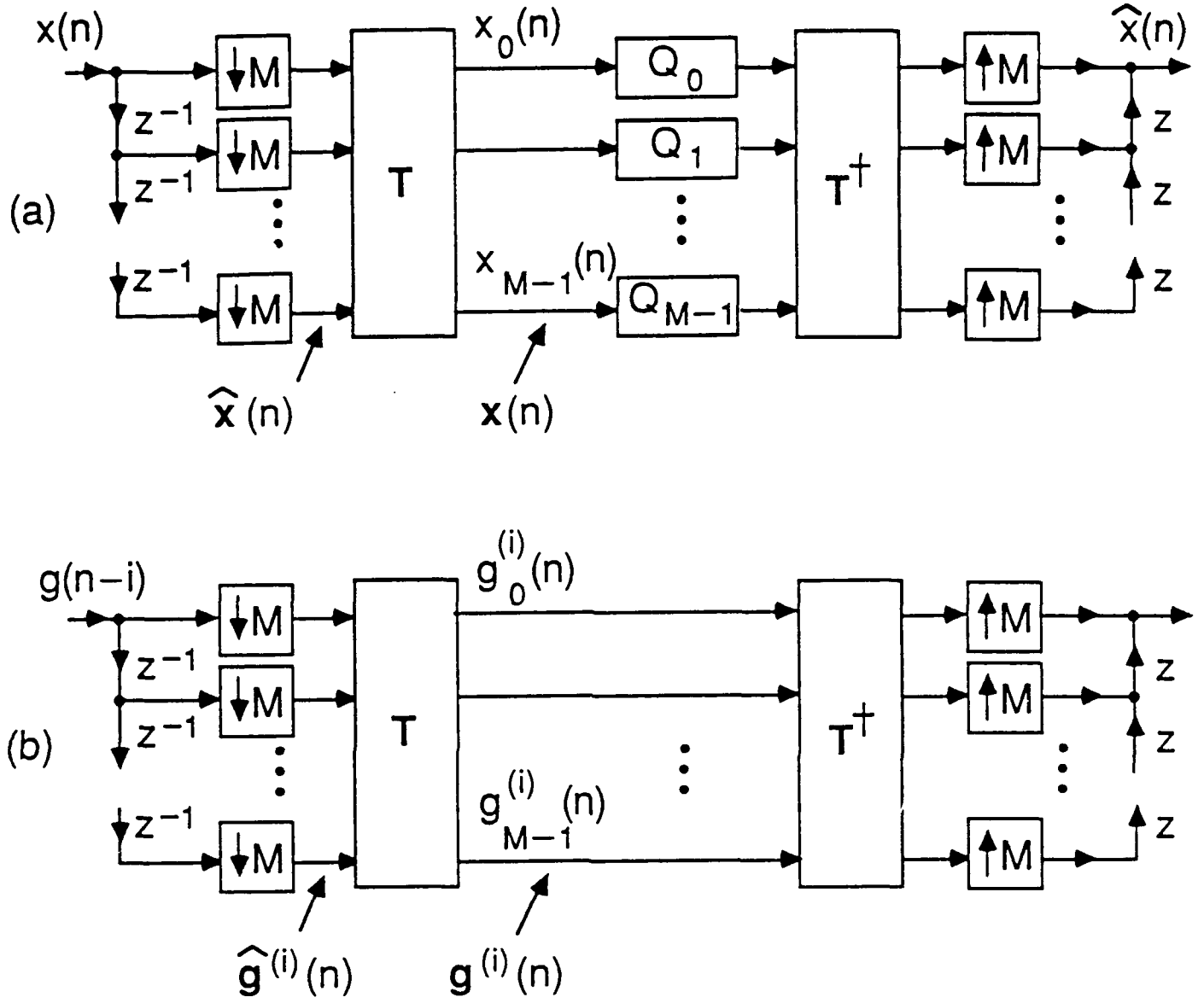


Fig. 4.1. The orthogonal transform convolver. (a) x(n) is input to the filter bank, and (b) shifted g(n) is input to the filter bank.

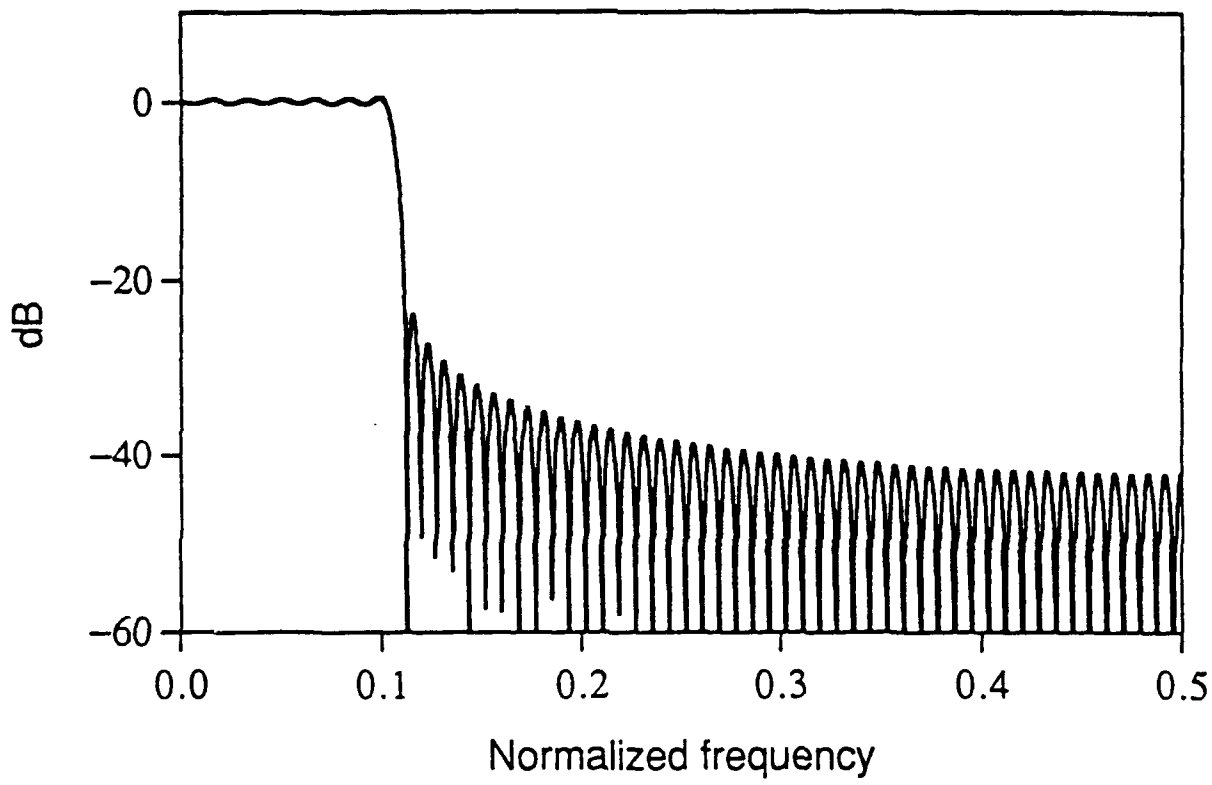


Fig. 5.1. A typical magnitude response of the filter g(n) used in the experiment.

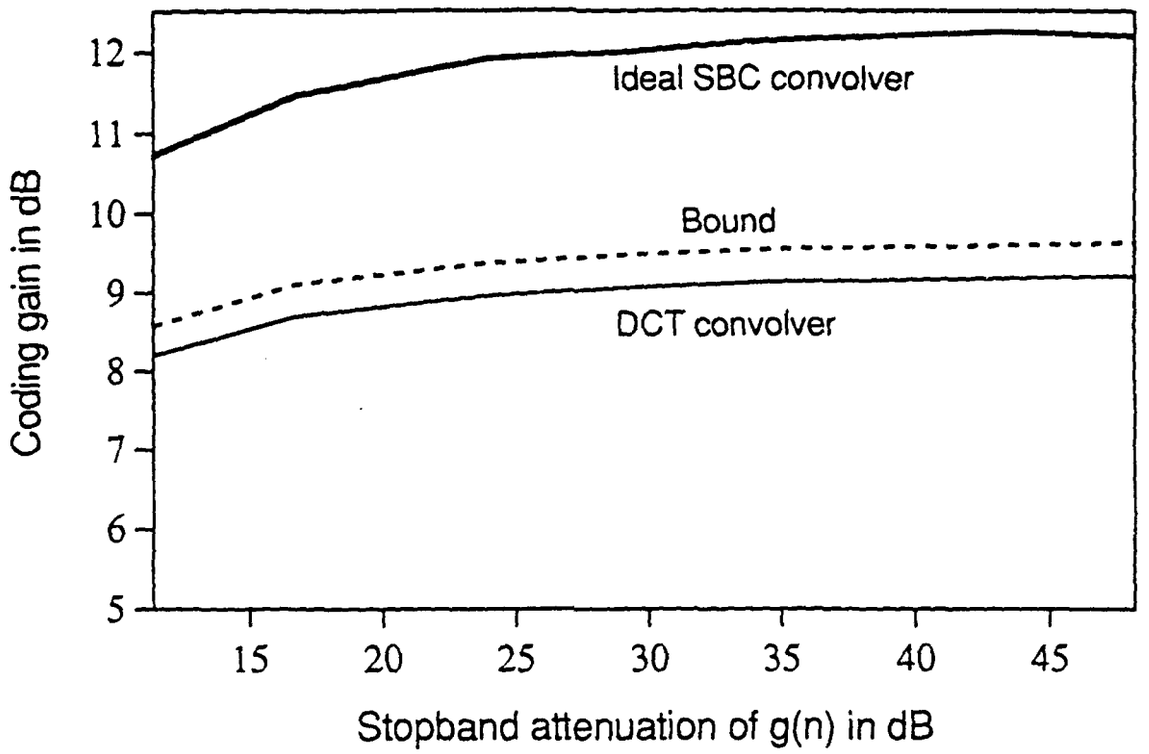


Fig. 5.2. Demonstration of the coding gains of paraunitary convolvers.

Submitted to IEEE Trans. on Image Processing, 1992

Best Wavelet Packet Bases in a Rate–Distortion Sense

Kannan Ramchandran and Martin Vetterli,
Department of Electrical Engineering
and Center for Telecommunications Research,
Columbia University,
New York, N.Y. 10026

January 22, 1992

Abstract

The use of an adaptive tree structure using wavelet packets as a generalized wavelet decomposition for signal compression was recently introduced by Coifman, Meyer, Quake, and Wickerhauser [1]. The idea is to decompose a discrete signal using all possible wavelet packet bases of a given wavelet kernel, and then to find the "best" wavelet packet basis. Unlike the work in [1], in this paper, we employ a framework that includes both rate and distortion. A fast algorithm is formulated to "prune" the complete tree, signifying the entire library of admissible wavelet packet bases, into that best basis subtree which minimizes the global distortion for a given coding bit budget. Arbitrary finite quantizer sets are assumed at each hierarchical level of the basis-family tree. Finally, a DCT "wavelet packet" basis quadtree segmentation is described as an image coding application in a JPEG environment, with good improvement shown over non-adaptive JPEG quantization.

1 Introduction

Source compression for stripping redundancy from typical highly correlated sources like speech and image waveforms has been studied extensively. Some popular techniques addressed in the literature include vector quantization (VQ), linear predictive coding, linear transform coding (like the KLT and the DCT), and subband coding as well as various hybrid combinations of these.

VQ is a popular and powerful scheme for compressing correlated discrete signal sets whose characteristics have been "trained" initially, but its complexity grows exponentially with vector dimensionality. Linear transformations like the DCT are less computationally demanding, but owing to their "fixed" non-adaptive nature, their compression potential relies heavily on the stationarity of the signal. For non-stationary sources, linear transforms or prediction techniques generally fail to exploit all of the source redundancy present. If one could combine the adaptability of VQ with the speed of linear transform coding, one could achieve a coding scheme which adapts to signal non-stationarities without sacrificing computational ease. Wavelet packets, introduced by Coifman, Meyer, Quake, and Wickerhauser (CMQW) [1, 2], to be described in section 2, permit exactly this combination, and offer a flexible yet computationally non-overwhelming framework in which to undertake efficient signal compression.

This paper is organized as follows: Section 2 provides a brief description of the background information on which the rest of the paper is founded, while also outlining the scope of applicability of this work and its relation to existing literature. Section 3 highlights the intuition and main idea of the algorithm. Section 4 states the problem formally, while section 5 undertakes a fast solution to the problem. Section 6 flowcharts the complete algorithm. Finally, section 7 provides an image coding application using quadtree segmentation, based on our fast algorithm.

2 Background and scope of this work

This section deals with a brief explanation of wavelet packets, a summary of bit allocation techniques based on operational rate-distortion theory, a brief citation of existing literature and the contribution of this paper, as well as

the scope of applicability of the algorithm described here.

2.1 Wavelet packets

Wavelet packets (WP) were introduced recently by CMQW as a family of orthonormal (ON) bases for discrete functions of $\mathbf{R}^N$, and include the well-known wavelet basis and the Short-Time-Fourier-Transform-like (STFT) basis as its members. While a brief description of wavelet packets, together with an intuitive feel for what they represent, will be provided here, the interested reader is referred to [1] and [2] for a detailed and mathematical treatment of the subject.

Wavelet packets represent a generalization of the method of multiresolution decomposition, and comprise the entire family of subband coded (tree) decompositions, from which the optimal decomposition subtree can be selected, to maximize compression by permitting the signal characteristics to be matched "on the fly." Thus, the potential of the CMQW wavelet packet decomposition scheme lies in its capacity to offer a rich menu of ON bases, from which the "best basis" can be chosen. If one represents the complete subband decomposition of a discrete signal set in $\mathbf{R}^N$ as a regular analysis tree of depth *log N* (see Figure 1), the CMQW approach permits a decomposition topology to be picked corresponding to any pruned subtree of the original tree, i.e. any subtree sharing the same root as the original tree. This is obviously isomorphic to all permissible subband topologies (see Figure 2), with the collection of terminal nodes (leaves) of every pruned subtree representing the entire library of permissible ON bases with which to decompose the original signal.

Thus, this decomposition might be used to code independent segments of a given non-stationary signal. It enables the coder to exhibit, for example, a STFT-like characteristic (regular tree) at one source instance, a wavelet characteristic (logarithmic tree) at another instance, or any intermediate characteristic (arbitrary WP subtree) at yet other instances, to best match the signal's non-stationary statistics. See Figure 2. In this powerful scenario, the popular wavelet and STFT decompositions are mere *special cases* of permissible WP structures. To emphasize, the CMQW ON decomposition enables each internal tree node (ON basis parent-member) to spawn off, as its replacement, branch nodes (ON basis child-members) that provide a complete, disjoint basis cover for the space spanned by their parent. This

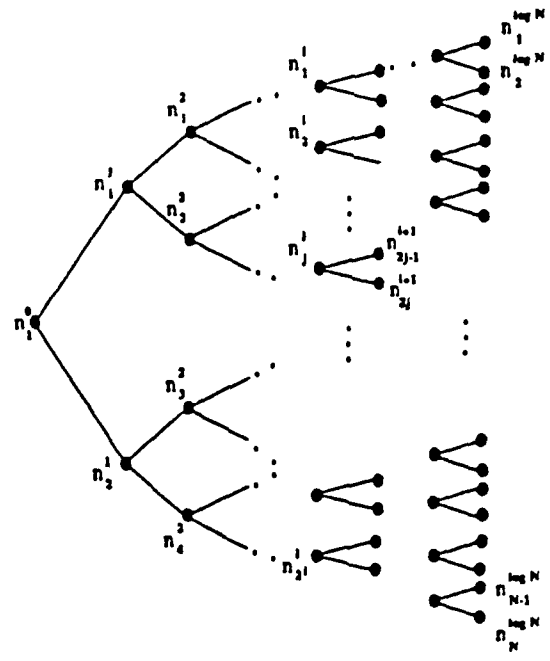


Figure 1: Complete wavelet packet tree of depth log N to code signal block of dimension N. Each node n_j^i contains the basis vector b_j^i with WP coefficient vector c_j^i. The complete set of all pruned subtrees represents the library of all admissible WP bases, or equivalently, all subband decomposition topologies.

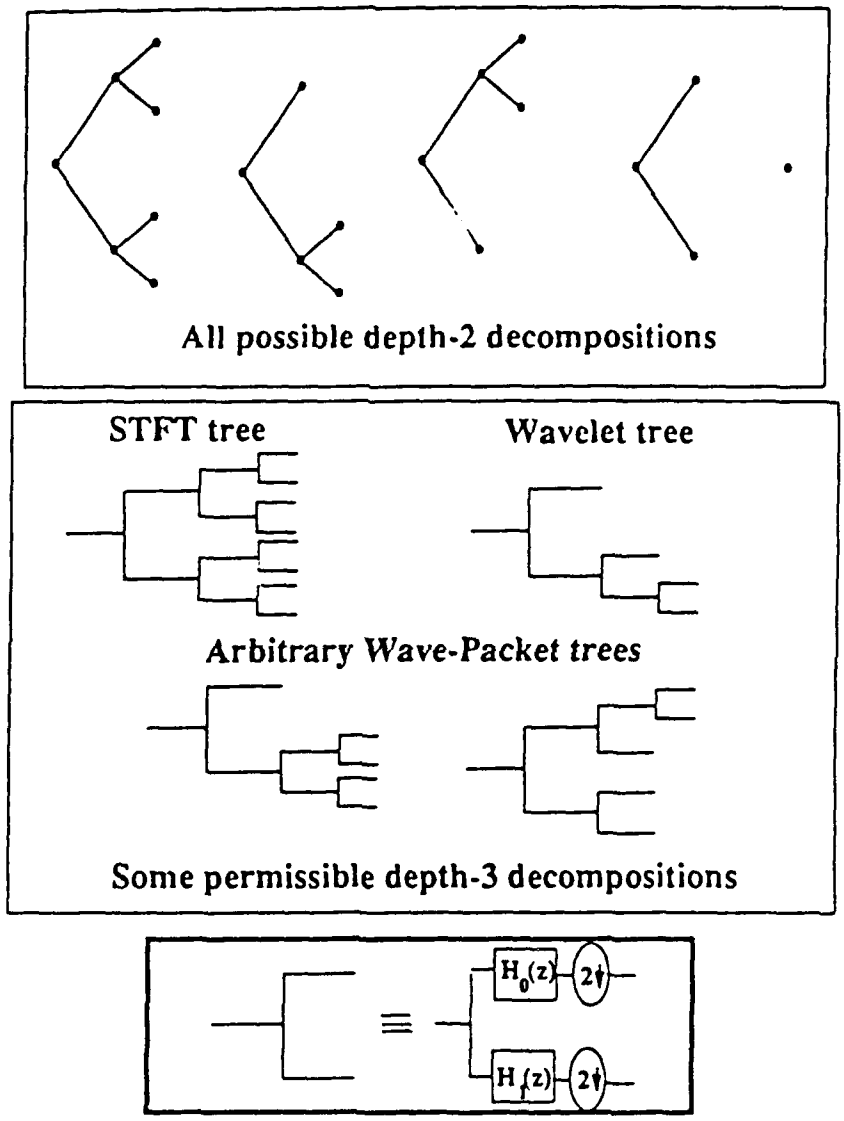


Figure 2: (a) All possible binary wavelet packet decompositions of depth 2. (b) Some typical depth-3 binary wavelet packet subtree decompositions. Note that H_1 and H_0 represent the "low pass" and "high pass" analysis filters.

property is vital to the development of the fast pruning algorithm, as it enables the coding rate and distortion corresponding to any node in the tree to be additive over the rates and distortions associated with the branches emanating from that node, with respect to an l_2 norm (like m.s.e or weighted m.s.e) criterion.

2.2 Quantization and bit allocation

The problem of bit allocation, where a given bit budget must be distributed efficiently among a set of given admissible quantization choices, is a classical problem in signal compression that has received exhaustive treatment in source coding literature [3, 4, 5, 6]. A classical framework for source coding is Shannon's rate-distortion theory, which deals with minimization of source distortion subject to a channel rate constraint, or the dual problem of minimization of channel rate subject to a distortion constraint. A practical coding environment involves a finite set of admissible quantizers, characterized by their (operational) rate-distortion functions, ranging from convex [3] to completely arbitrary [4]. These quantizers are used by the allocation algorithm to determine the best strategy to minimize the overall coding distortion subject to a total bitrate budget constraint. We use this framework to seek our best basis WP and best quantizer choices.

2.3 Related work and contribution of this paper

While the adaptivity and the speed best-basis search of [1] are unmistakable, the cost criterion and the coding (quantization) method used there to exploit this speed and flexibility are somewhat ad hoc. In this paper we formulate a fast algorithm, for a given total coding bitrate budget, to pick the optimal WP basis, together with the optimal quantizer choice for that optimal WP subtree, for each of the independent segments or "blocks" that the signal comprises. Optimality is with respect to a *global* distortion criterion that is additive over the signal blocks, e.g. m.s.e or weighted [1] m.s.e. We conduct our best basis hunt in a rate-distortion (R-D) framework that considers both aspects (i.e. rate and distortion) of the coding problem. This is a generalization of the treatment in [1, 2] where a one-sided "entropy" or m.s.e.

[1] In image processing applications, these weights may be designed, for example, to be commensurate with those of the Human Visual System.

distortion criterion is used. Our approach could be viewed, *in its quadtree application*, as an extension of the work by Shoham and Gersho [4] to provide a fast algorithm covering *hierarchies* of admissible quantizers. It may also be regarded, in its quadtree segmentation application, as a generalization of Chou et. al's (G-BFOS) algorithm [6] to the case where monotonicity constraints of rate and distortion with tree depth are removed. Figure 3 (a) gives an example of a rate-distortion characteristic that is constrained to be monotonic with tree depth [2], as identified by a single transition from the "merge" to "split" boundaries, a constraint that is necessary for the quadtree algorithm mentioned in [6]. Figure 3 (b) shows a non-restrictive case, where arbitrary transitions between the "split" and "merge" regions are permitted. A practical contribution of this paper involves the description of the results of a quadtree-based image compression application using a family of DCT "wavelet packet" bases. Our application is similar to that of independently done work by Sullivan and Baker [7], who performed efficient quadtree segmentation using VQ. Our example uses classified quantizers in a JPEG (DCT-based) [8] coding environment.

While bit allocation strategies for various coding environments have been formulated in the literature, the problem of using *arbitrary quantizers in a generalized multiresolution wavelet decomposition framework* has not, to the best of the authors' knowledge, been addressed.

2.4 Scope of applicability of our algorithm

It must be emphasized that the DCT-basis family tree of our application and the standard basis tree employed by [7] are not strictly WP trees, which are derived recursively using Quadrature Mirror Filters (QMF) filter banks or using multi-resolution wavelet analysis (whose equivalence has been established [12, 9]). The scope of applicability of our algorithm extends to all classes of structures which permit the construction of a hierarchy of basis covers for the input signal space. While this obviously includes structures like quadtrees and orthonormally transformed (e.g. DCT) quadtrees, to be described in detail in Section 7, other powerful structures such as the CMQW multiresolution decomposition wavelet packets and hierarchical subband coders are also applicable. As an example, our algorithm could be used to determine

[2] i.e. as the tree grows, the rate increases and the distortion decreases

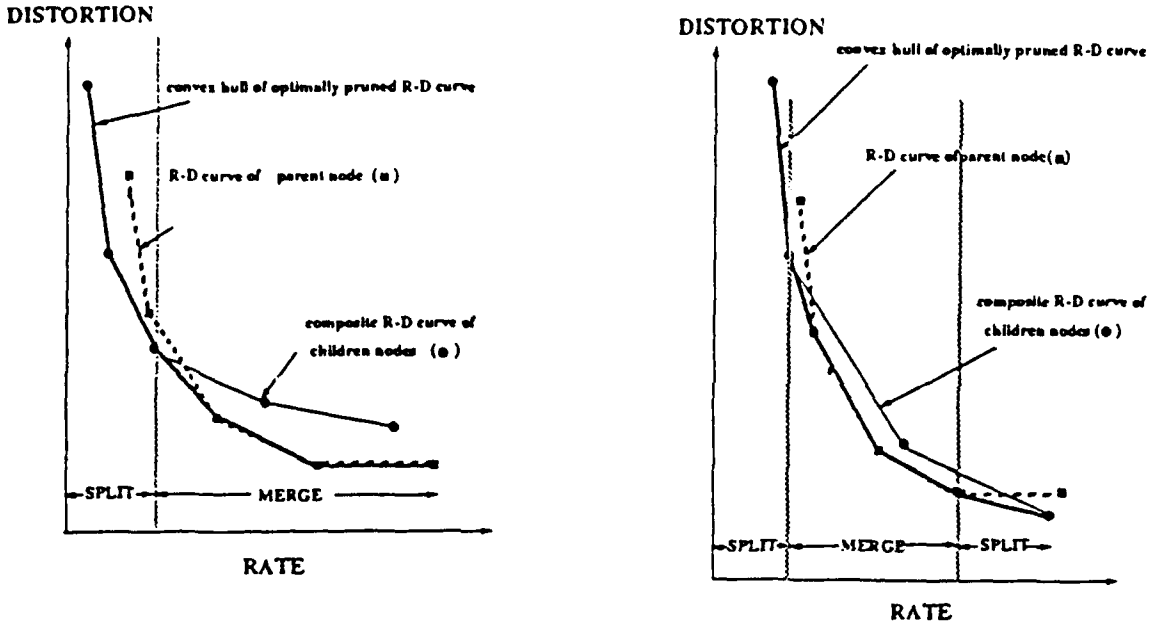


Figure 3: "Split/Merge" boundaries shown for (a) Monotonic case to which the G-BFOS algorithm is constrained, and (b) Non-restrictive case.

THIS PAGE IS BLANK

DUE TO A

PAGE NUMBERING ERROR

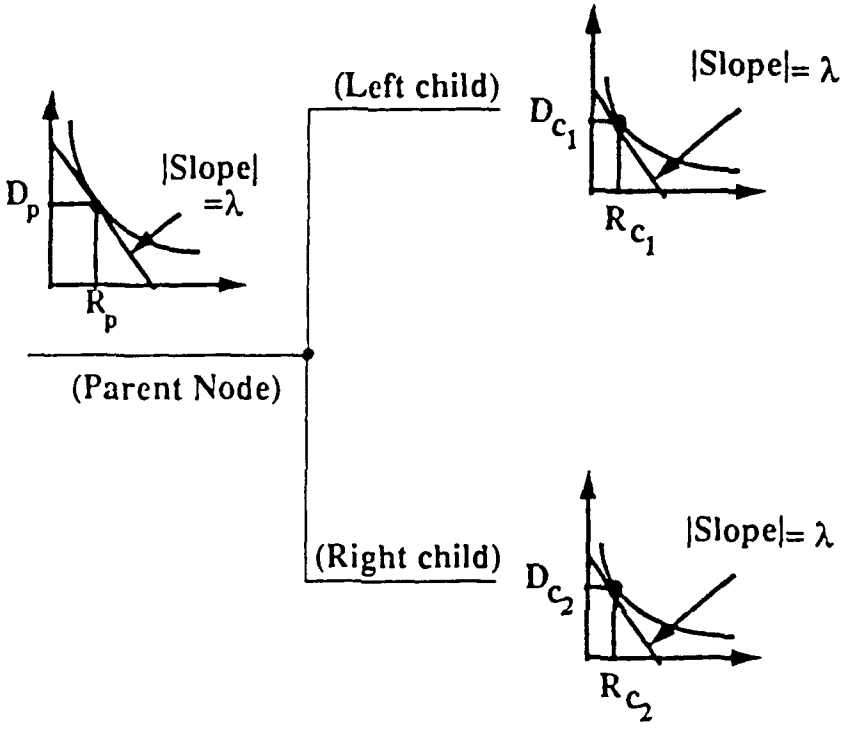
quantitatively, such important coder design considerations as the optimal decomposition depth for subband coding, or a performance comparison of filter banks of different kernels and topologies, or to determine an efficient DCT quadtree structure in a "hierarchical" JPEG application, as will be explained in Section 7.

3 Basic idea of the algorithm

We convert our budget-constrained search for the best wavelet packet basis (and best quantizer) sequence into an unconstrained one by minimizing the composite Lagrangian cost functional $J(\lambda) = D + \lambda R$, where λ^*, the optimal Lagrange multiplier for the given budget constraint, is found by a fast iterative scheme. The mathematical details and formal treatment are provided in Sections 4, 5, and 6. The translation into an unconstrained formulation makes it feasible to *deal with each signal block independently*. It also converts the problem into a fast iterative search for an operating point on the convex hull [3] of the composite operational rate-distortion curve. This is a function of the input signal characteristics, the wavelet kernel picked to generate the library of wavelet packet bases, and the set of admissible quantizers for each wavelet packet tree level.

As will be shown, at optimality, all nodes of all subtrees representing the sequence of signal blocks, must operate at "constant quality slope λ". See Figure 4. For a given $\lambda = \mid \Delta D/\Delta R \mid$, we populate each node of each tree block *independently* with the Lagrangian cost function associated with the best quantizer for that node. The best quantizer for a particular tree node is that one which "lies" at absolute slope λ on the convex hull of the operational R–D curve for that node, as shown in Figure 4. Then, by applying, *in parallel for each signal block*, the pruning criterion of Figure 4 recursively on every node, starting from the full-depth tree and proceeding towards the root, we find the sequence of best wavelet packet bases and associated best quantizers with which to code the signal. The recursive algorithm exploits Bellman's optimality principle by eliminating quickly a host of suboptimal subtrees from contention for the optimal solution, in a manner reminiscent of the popular Viterbi algorithm, and is similar to the pruning condition of [1], which, however, does not use our Lagrangian cost function. Finally, we

[3] i.e. the convex boundary of R-D points (see Fig. 3)



For "quality factor" λ, prune if

$$(D_{c_1}+D_{c_2}) + \lambda(R_{c_1}+R_{c_2}) > (D_p+\lambda R_p)$$

Figure 4: Lagrangian cost pruning criterion for "quality criterion" λ for each parent node of the wavelet packet tree. This condition is used recursively to do fast pruning from the complete tree depth towards the root to find the optimal subtree for a given λ.

show a fast way of iterating over the Lagrangian multiplier λ, in a convex search using Newton's method or bisection methods [10], to find the optimal λ^* satisfying the given budget constraint.

4 Formal Problem Definition

Without loss of generality, we will consider the problem of a binary wavelet packet decomposition tree of a discrete input signal (vector) of size N in $l^2(N)$ $\langle s_1, s_2, \ldots, s_N \rangle$. See Figure 1. Though omitted for convenience, each branch of the analysis tree consists of the appropriate filter: high-pass filter (HPF) H_0 for the upper child and low-pass filter (LPF) H_1 for the lower child. followed by a decimator by 2 (see Figure 2), with the corresponding synthesis tree consisting of an upsampler followed by the corresponding synthesis filters.

The analysis and synthesis filters of each branch satisfy the standard orthonormality conditions of paraunitary perfect reconstruction filter banks (PRFB's [9]). As is well known, iterating the orthonormal filter templates to the complete tree depth results in an equivalent generalized multiresolution decomposition tree (i.e. wavelet packet tree) whose nodes represent a family of orthonormal bases [1, 2]. As shown in Figure 1, we assume that there are M signal blocks to be coded independently, each of size N. To help provide a clear notation-free understanding of our algorithm, we introduce a "toy" example that we will invoke at various points in this paper. The example, shown in Figure 5, is that of coding a length-4 signal block, using the often-cited Haar basis (or sum and difference filters, using filter-bank jargon) as the wavelet packet kernel. Figure 2(a) shows the possible decompositions for the given signal.

Let us define the following terms to be used in the formulation:

- M : number of independently coded signal blocks.
- N : number of elements in each signal block (assumed to be one-dimensional, without loss of generality).
- T : complete wavelet packet tree (STFT tree), for each signal block, of depth $log\ N$ as shown in Fig. 1.
- t : any node of T, internal or terminal.

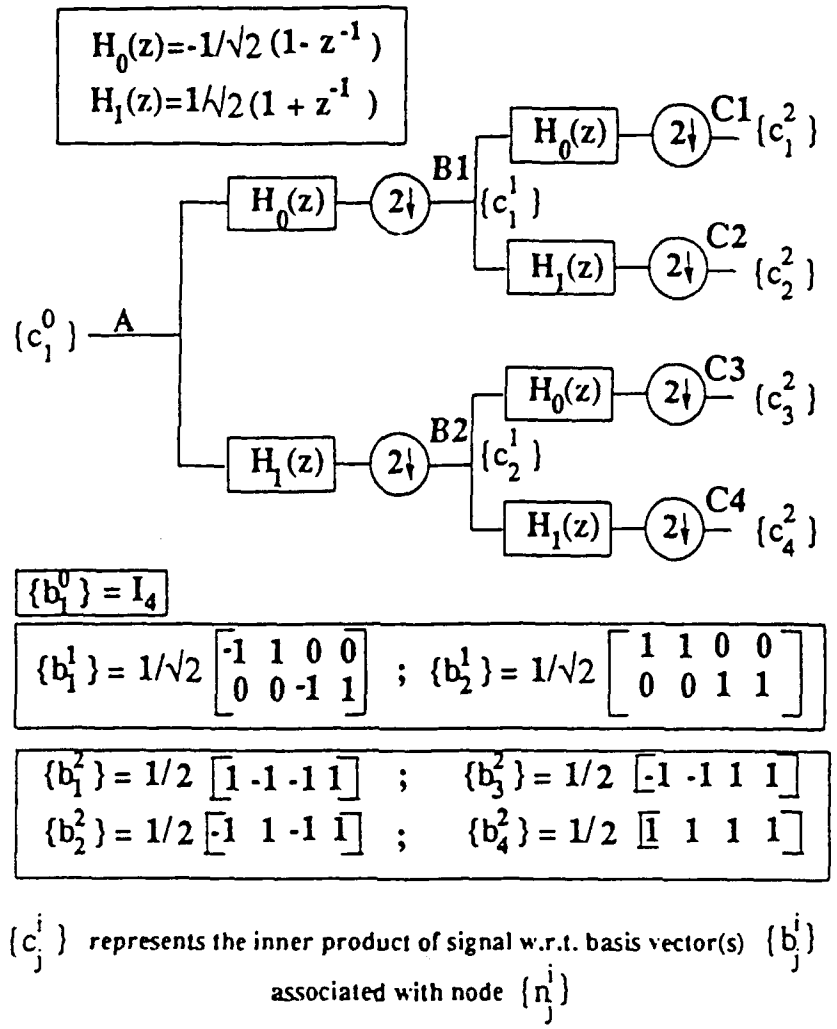


Figure 5: Toy example showing wavelet packet decomposition for a length-4 signal block using the popular Haar wavelet kernel.

- $S \preceq T$: pruned subtree of T, i.e. a (wavelet packet basis) subtree of T that shares its root; thus, S corresponds to any admissible wavelet packet basis.
- $\tilde{S}$: set of leaves or terminal nodes of subtree S.
- $\mathsf{q_R(t)}$: set of all admissible quantizers for node $t \in T$. The toy example at the end of this section presents some admissible quantizer choices for a particular case.
- $\mathsf{Q_R(S)}$: vector set of all admissible quantizers for the collection of individual leaf nodes of subtree S $=\{\mathsf{q_R}(t_1) \times \mathsf{q_R}(t_2) \times \ldots \times \mathsf{q_R}(t_L)\}$, where $\{t_1, t_2, \ldots t_L\} \in \tilde{S}$, is the complete set of leaves or terminal nodes of S.
- n_j^i, b_j^i, c_j^i : (see Figure 1) the jth (of the possible 2^i choices) node, basis, and coefficient vector respectively, at the ith tree-depth or "scale" (for $i = 1, 2, \ldots, \log N$). Note that b_j^i represents the $\mathbf{R}^N$-basis members associated with node n_j^i, while c_j^i represents the inner product of the signal with the basis vectors in b_j^i. See Figure 5 for a Haar basis kernel, where the length-4 signal is broken down into all possible wavelet packet coefficients for the complete depth-2 binary tree. Note also that to simplify notation, $\{t, b_t$ and $c_t\}$ will be invoked where convenient.
- $D_q(t), R_q(t)$: distortion and bitrate, respectively, associated with quantizing wavelet packet coefficient vector c_t of node t using quantizer $q \in \mathsf{q_R(t)}$.
- $D_Q(S), R_Q(S)$: distortion and rate, respectively, associated with coding subtree (or wavelet packet) S using quantizer $Q \in \mathsf{Q_R(S)}$. In our case, they are both linear tree functionals; i.e.: Total distortion $= D_Q(S) = \sum_{t \in \tilde{S}} D_q(t)$ and total rate $= R_Q(S) = \sum_{t \in \tilde{S}} R_q(t)$.

The problem to solve, then, is that of finding, given a total budget of R_{budget} to code M independent signal blocks, that sequence of (pruned) subtree best-bases $S_i^* \preceq T$ (for $i = 1, 2, \ldots, M$) together with their associated optimal quantizers $Q_i^* \in \mathsf{Q_R(S_i^*)}$ which minimize the global coding distortion. Stated mathematically, this boils down to determining $D_{min} = \sum_{i=1}^{M} D_{Q_i^*}(S_i^*)$, where

$$D_{Q_i^*}(S_i^*) = \min_{S_i \leq T} \left[\min_{Q_i \in Q_a(S_i)} D_{Q_i}(S_i) \right] \tag{1}$$

$$\text{such that } R_{total} = \sum_{i=1}^{M} R_{Q_i^*}(S_i^*) \leq R_{budget}, \tag{2}$$

where R_{budget} is the given bit budget constraint.

Toy Example

As an example, suppose we want to find the best wavelet packet basis corresponding to the Haar wavelet kernel for an input signal $s = [109, 23, -98, 13]$, with one block and dimension 4, i.e. N=4, M=1, for a coding budget of 21 bits, for the following classes of admissible quantizers for each tree level: Suppose we have three grades of uniform quantizers (coarse, medium, and fine) having step sizes of 16, 4, and 1 resp., or equivalently, a granularity of 16, 64, and 256 levels (using 4,6, and 8 bits) respectively, assuming a quantizer dynamic range from -128 to +128. As shown in Figure 5, for convenience, the tree scales are denoted by the labels A, B, and C. At full tree-depth C, the quantizers 1,2, and 3 denote the fine, medium, and coarse scalar quantizers for each of the 4 wavelet packet coefficients C1, C2, C3, and C4. At depth B, the quantizers 1,2, and 3 denote the [fine,fine], [medium,medium], and [coarse,coarse] combination of quantizers applied independently to each of the wavelet packet coefficients B1 and B2. At the tree root A, similar vectors of the three different grades of scalar quantizers are available to code the 4-D coefficient. Assume a m.s.e distortion criterion.

Note that the wavelet packet coefficients are the inner products of the input signal with the respective basis vectors (see Figure 5):

$$c_1^0 = [109, 23, -98, 13]$$
$$c_1^1 = [-60.81, 78.49] \; ; \; c_2^1 = [93.34, -60.1]$$
$$c_1^2 = [98.5]; c_2^2 = [12.5]; c_3^2 = [-108.5]; c_4^2 = [23.5]$$

Figure 6 shows the rate-distortion curves for all possible basis subtrees in our example, for the permissible quantization choices. Thus, for example, $c_1^0 = [109, 23, -98, 13]$ would be quantized to $[108, 24, -96, 12]$ (for a total

squared-error distortion of 7.0) with the medium grade (step-size 4) quantizer, and so on.

5 Fast solution

We solve the constrained problem of Equation (1) by converting it to an unconstrained problem using Lagrange multipliers. This section spells out the unconstrained approach, and explains how our problem is a hierarchical extension of that presented in [4]. A fast pruning algorithm is used to remove suboptimal subtrees that would not otherwise have been eliminated if we had resorted to a "flattened" version of our problem to emulate that solved in [4]. Solving the unconstrained problem for different positive values of the Lagrange multiplier results in the tracing out of convex hull points of the rate-distortion curve. The optimal convex hull point we solicit is that with the minimum distortion while not exceeding the given rate budget.

5.1 Unconstrained Optimization Approach

Instead of solving the constrained optimization problem (1), let us consider the following *unconstrained* formulation. Let us introduce the Lagrangian cost functional corresponding to the Lagrange multiplier $\lambda \geq 0$, for each signal block i, of basis subtree $S_i \preceq T$ and subtree quantizer set $Q_i \in \mathbf{Q_a}(S_i)$,

$$\begin{aligned} J_i(\lambda) &= J_\lambda(S_i, Q_i)) && (3)\\ &\triangleq D_{Q_i}(S_i) + \lambda R_{Q_i}(S_i) && (4)\\ &= \sum_{t \in \dot{S}_i} [D_{q_t}(t) + \lambda R_{q_t}(t)], && (5) \end{aligned}$$

where the last equation is written in terms of the leaf nodes of the subtree, as a result of the subtree rate and distortion functions being additive over its leaf nodes.

We now develop, by a simple extension of Theorem 1 in [4] to include the ensemble of wavelet packet bases $S \preceq T$ as well as their associated quantizers $Q(S) \in \mathbf{Q_a}(S)$, an equivalent unconstrained problem. This formulation is attractive because it decomposes the original problem into independent parallel

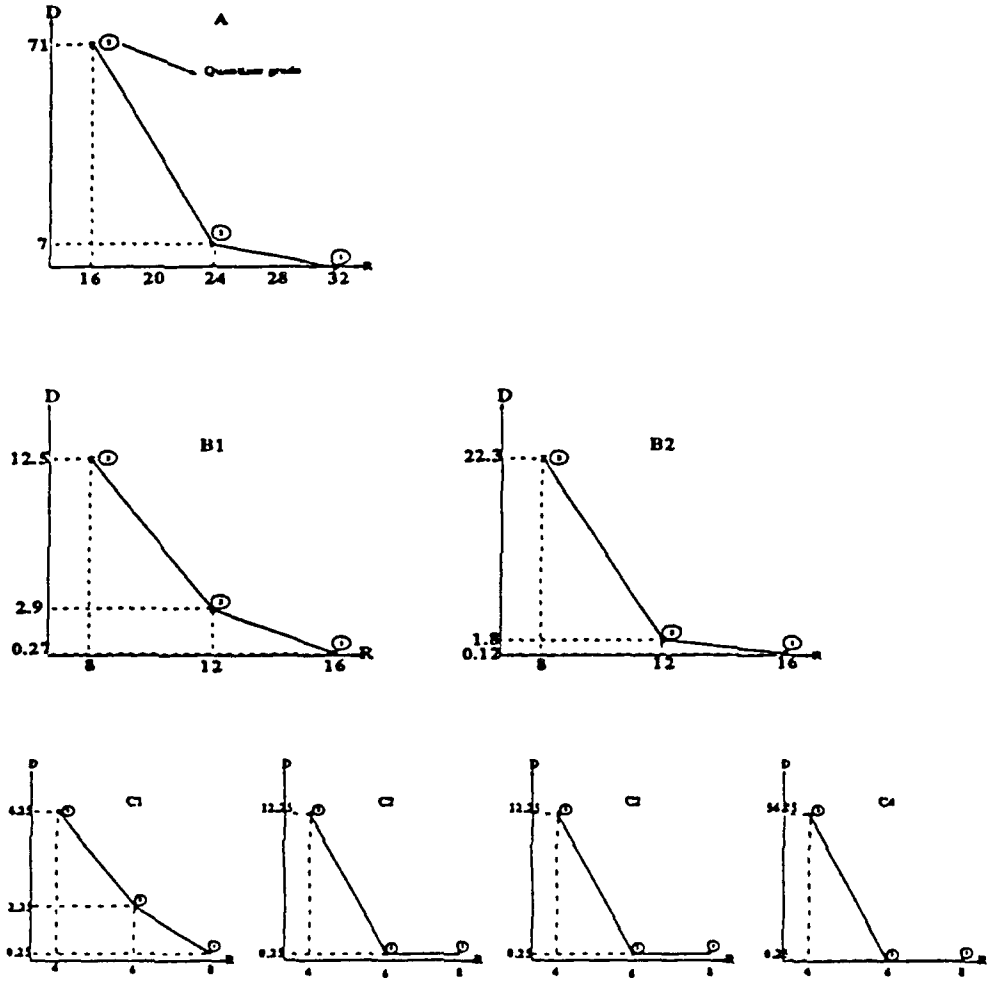


Figure 6: Toy example R-D curve at tree-depth 0 (A), depth 1 (B1,B2), and depth 2 (C1,C2,C3,C4). Note that B1 refers to the HPF output, and B2 to the LPF, and similarly (C1,C2) are the HPF/LPF outputs of the B1 input, and (C3,C4) of the B2 input. See Figure 5.

optimizations for each signal block $i = 1, 2, \ldots, M$. Mathematically stated, *for a fixed value of* λ, the unconstrained problem specified below is solved for S_i, Q_i for $i = 1, 2, \ldots, M$ (for the "correct" fixed value of λ , which is a function of the given budget constraint, and the hunt for which will be described later, we solve the original problem to within a convex hull approximation):

$$\begin{aligned} J_i^*(\lambda) &= J_\lambda(S_i^*, Q_i^*) && (6)\\ &= \min_{S_i \preceq T} \min_{Q_i \in \mathbf{Q_a}(S_i)} J_\lambda(S_i, Q_i) && (7)\\ &= \min_{S_i \preceq T} \min_{Q_i \in \mathbf{Q_a}(S_i)} [D_{Q_i}(S_i) + \lambda R_{Q_i}(S_i)] && (8)\\ &= \min_{S_i \preceq T} \sum_{t \in S_i} \min_{q \in \mathbf{Q_a}(t)} [D_q(t) + \lambda R_q(t)]. && (9) \end{aligned}$$

In order to make notation less cumbersome, if we consider the problem of a single wavlet packet tree (i.e. $M = 1$), with extention to $M > 1$ being trivial, due to the parallelization of the problem, the problem becomes determining:

$$D_{Q^*}(S^*) = \min_S \min_Q D_Q(S) \text{ s.t. } R_{Q^*}(S^*) \leq R_{budget}. \quad (10)$$

Thus, the unconstrained problem of Eq. (7) becomes finding:

$$J^*(\lambda) = J_\lambda(S^*, Q^*) = \min_{S \preceq T} \min_{Q \in \mathbf{Q_a}(S)} [D_Q(S) + \lambda R_Q(S)]. \quad (11)$$

The above approach identifies, for a fixed positive λ, an optimal operating point on the convex hull of the composite rate-distortion curve for the specified problem. If the original constrained problem happened to have a budget constraint that "hit" one of the convex-hull operating points, then, *the unconstrained and the constrained problems have identical solutions.* Mathematically stated, the equivalence is established in the following theorem, a direct hierarchical extension of Theorem 1 of Shoham and Gersho [4]:

Theorem 1 *If* (S_u^*, Q_u^*) *is the solution to the unconstrained problem of Eq. (11) corresponding to some fixed value of* λ, *then it is also the solution to the constrained problem of Eq. (10) for the particular case of* $R_{budget} = R_{Q_u^*}(S_u^*)$.

Proof:

$$\begin{aligned} J_\lambda(S_u^*, Q_u^*) &\leq J_\lambda(S, Q) && (12)\\ D_{Q_u^*}(S_u^*) + \lambda R_{Q_u^*}(S_u^*) &\leq D_Q(S) + \lambda R_Q(S) && (13)\\ D_{Q_u^*}(S_u^*) - D_Q(S) &\leq \lambda[R_Q(S) - R_{Q_u^*}(S_u^*)] && (14)\\ D_{Q_u^*}(S_u^*) - D_Q(S) &\leq \lambda[R_Q(S) - R_{budget}]. && (15) \end{aligned}$$

Since Equation (15) holds for all $S \preceq T$ and $Q \in \mathbf{Q}_a(S)$, it certainly holds for the subsets $\bar{S} \preceq T, \bar{\mathbf{Q}} \in \mathbf{Q}_a(\bar{S})$ which satisfy $R_Q(S) \leq R_{budget}$. That is,

$$R_Q(S) \leq R_{budget} \text{ for } S \in \bar{S}, Q \in \bar{\mathbf{Q}}. \quad (16)$$

Thus, from Eq. (15) and Eq. (16), since $\lambda \geq 0$, we have:

$$D_{Q_u^*}(S_u^*) - D_{Q_u}(S_u) \leq 0 \quad \forall S \in \bar{S}, Q \in \bar{\mathbf{Q}}, \quad (17)$$

i.e.(S_u^*, Q_u^*) also satisfies the original constrained optimization problem of Eq. (10) for the given budget constraint. □

Note again that the implication of the above result, when extended to the case of arbitrary M, is that if we solve the *unconstrained* problem of Eq. (8) for some $\lambda \geq 0$, and if R_{budget} of the *constrained* problem of Eq. (1) happens to be $\sum_{i=1}^{M} R_{Q_i^*}(S_i^*)$ of the unconstrained problem, then the solutions to both problems are identical. The unconstrained problem lends itself to a much easier solution than the constrained one, though of course, the latter may have an "inaccessible" solution, since the unconstrained approach maps out only convex hull points of the R-D curve. Figure 7 shows an example of an inaccessible convex-hull solution, where the budget constraint line does not pass through a convex-hull point. The excess $R_{budget} - \sum_{i=1}^{M} R_{Q_i^*}(S_i^*)$ bits, which, in practice, represent a negligible fraction of the original budget, can be allocated using "greedy" heuristics, or using a steepest descent algorithm similar to that of [4]. Bounds on the suboptimality of the unconstrained solution can be found in [7]. The power of the unconstrained approach obviously lies in its "parallelization" of the original problem into smaller independent optimization problems.

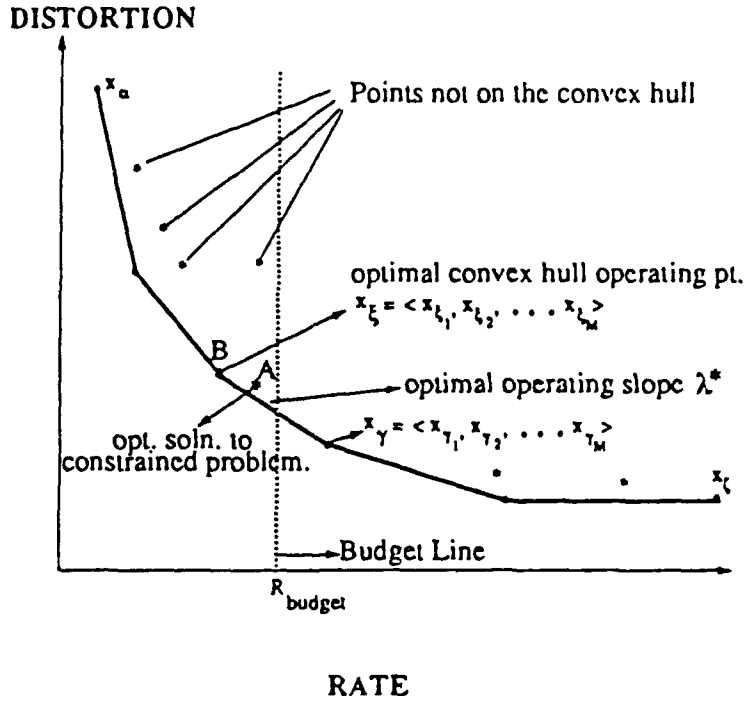


Figure 7: Composite R-D curve showing convex hull solution to an "inaccessible" problem.

5.2 "Flattening" the problem

If we define X_i for each independent block $i = 1, 2, \ldots, M$ to be the set of all admissible operating points for block i (i.e. the set comprising all combinations of subtrees and their associated quantizers, $\{S_i, Q_i\} \quad \forall S_i \preceq T, Q_i \in \mathcal{Q}_R(S_i)$), then we have a "flattened" version of our problem, similar to that solved in [4]. i.e. the constrained problem becomes:

$$\min_{x_i \in X_i} \sum_i D(x_i) \text{ subject to } \sum_i R(x_i) \leq R_{budget}. \tag{18}$$

Though the "flattened" version of our problem can be made unconstrained and solved as in [4], the hierarchical nature of our wavelet packet tree-structured problem lends itself to a fast algorithm. Thus, it would be computationally wasteful (for a tree of depth n, there are $O(2^n)$ subtrees!) to solve our problem by the exhaustive search method, which is what "flattening" of our problem would entail.

However, the flattened version does have some merits, e.g. it can be invoked to simplify notation, and it can serve as a base from which to inherit some important unchanged features of the unconstrained problem. Some of the key results inherited from [4], as they apply to our problem, are hence presented as a summary:

- The convex-hull optimally allocated rate and distortion values for each block, for a given λ, are monotonic non-increasing and non-decreasing step functions, respectively, of λ.
- The monotonic *step* function nature of $R_i^*(\lambda)$ and $D_i^*(\lambda)$ is caused due to the discrete nature of the problem. Thus λ could be interpreted as an index of operating quality as it varies from $0 \rightarrow \infty$.
- As λ is swept through all positive real numbers, all the convex hull points of the composite R-D curve are traced out. See Figure 7 for a typical composite R-D curve.

With $J_i(\lambda) = D(x_i) + \lambda R(x_i)$ representing the customary Lagrangian subcost for block i associated with operating point x_i for quality criterion λ, let us introduce the biased Lagrangian cost W as:

$$W'(\lambda) = W'(\lambda, \{x_i^*(\lambda)\}_{i=1}^M) \tag{19}$$

$$= W(\lambda, \mathbf{x}^*(\lambda)) \tag{20}$$

$$= (\sum_{i=1}^{M} J_i^*(\lambda)) - \lambda R_{budget} \tag{21}$$

$$= [(\sum_{i} \min_{x_i} [D(x_i) + \lambda R(x_i)]) - \lambda R_{budget}]. \tag{22}$$

Note that $\mathbf{x}$ above refers to a vector notation of the sequence $x_1, x_2, \ldots, x_M$. Then, following the optimization theory outlined in [13], we have the following result:

Lemma 1 *$W(\lambda)$ is a concave $\cap$ function of λ.*

Proof: See Appendix.

Now, if we find the maximum of $W(\lambda)$ over all positive λ,

$$W(\lambda^*) = W(\lambda^*, \mathbf{x}^*(\lambda^*)) = \max_{\lambda \geq 0} W(\lambda), \tag{23}$$

we have the following result for the unconstrained solution corresponding to the given budget constraint R_{budget}:

Theorem 2 *λ^* and $\mathbf{x}^*(\lambda^*)$ that maximize W in Eq. (23) are the optimal convex hull face slope and optimal convex hull operating point, respectively, for the unconstrained optimization problem of Eq. (8), for the given budget constraint R_{budget}.*

Proof: See Appendix.

Thus, the above result gives the condition on the desired operating quality slope which solves the flattened version of our original problem. By "unflattening" the above result, we now develop the unconstrained solution to our best wavelet packet basis and optimal quantization choice problem. The optimal slope λ^* is the solution to :

$$W(\lambda^*) = W(\lambda^*, \mathbf{x}^*(\lambda^*)) \tag{24}$$

$$= \max_{\lambda \geq 0} (\sum_{i=1}^{M} [\min_{S_i} \min_{Q_i} J_\lambda(S_i, Q_i)] - \lambda R_{budget}) \tag{25}$$

$$= \max_{\lambda \geq 0} (\sum_{i=1}^{M} [\min_{S_i} \{\sum_{t \in S_i} \min_{q} [D_q(t) + \lambda R_q(t)]\}] - \lambda R_{budget}). \tag{26}$$

This is then the unconstrained optimization formulation to our problem, which can be dissected into independent fast individual optimizations. Thus, Eq. (26) can be dissected into the following 3 optimizations:

- At first (innermost minimization), the best quantizer choice for every terminal node *of a fixed subtree S is found* (where the subtree cost is assumed to be additive over that of its terminal nodes) *for a fixed operating slope* λ, independently for every block.

- Then (outer minimization), the best basis subtree is determined for each block independently, *again for the fixed operating slope* λ, from amongst all permissible wavelet packet bases decompositions for the given wavelet kernel. A fast dynamic programming based pruning operation will be done to accomplish this (see Section 5.4).

- Finally (outermost maximization), the optimal slope λ^* that meets the given budget criterion R_{budget} is determined as the maximum of the $W(\lambda)$. Lemma 1 facilitates the use of fast search methods for finding the optimal λ^* in an iterative fashion (see Section 6.2).

5.3 Geometric interpretation

One insight to be made into the unconstrained optimization problem is that of a geometric approach. It can be shown that the optimal operating point on the R-D plane for each leaf node of the tree T for a given slope λ is that point in the *collection of R-D points* which is first "impinged upon" by a "plane-wave" of slope $-\lambda$ emanating from the fourth quadrant of the R-D plane towards the R-D curve in the first quadrant. This is because the Lagrangian cost J associated with any admissible operating point can be interpreted as the y-intercept of the straight line of slope $-\lambda$ passing through that point on the operational rate-distortion plane. See Figure 8. The minimum Lagrangian function (minimum y-intercept) is obviously achieved for that point which is "hit" first by the plane wave of absolute slope λ impinging on the rate-distortion curve. *Note also from Figure 8 that the biased Lagrangian function $W(\lambda)$ can be interpreted as the intercept, on the budget constraint line, of the straight line of slope -λ tangent to the convex hull of the R-D curve.* This geometric interpretation of the problem makes a lot of the properties itemized earlier, like monotonicity and existence of singular

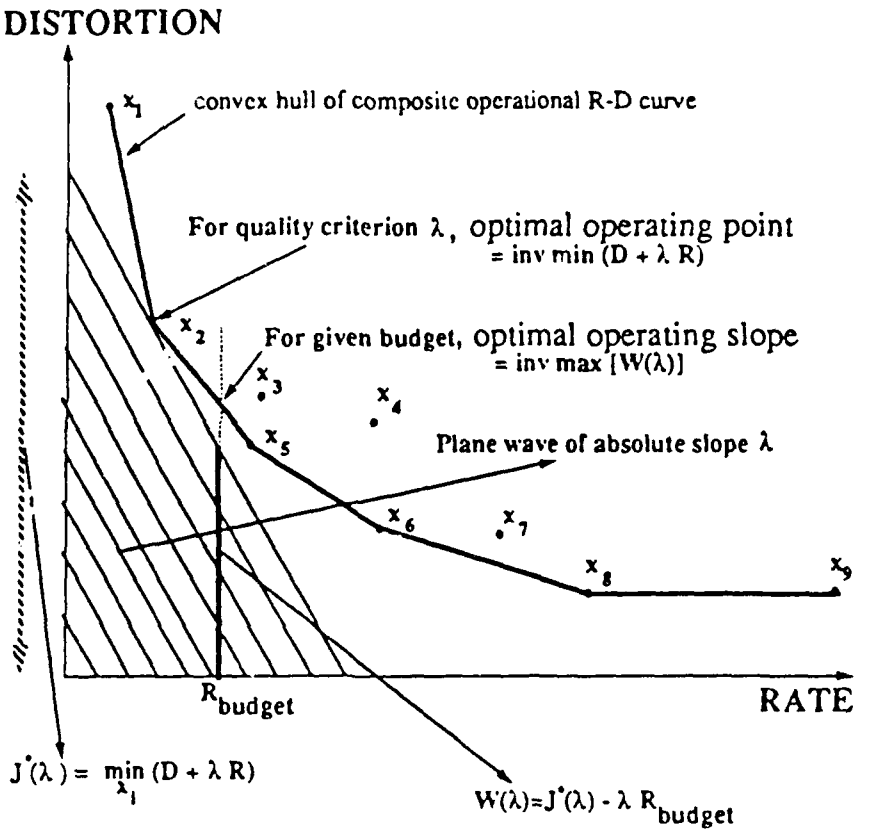


Figure 8: Geometric interpretation of the problem.

slope values, as well as the concavity of W, and the solution to the optimal convex-hull operating point, both intuitively pleasing and easy to show using straight-line geometry, though a more rigorous algebraic proof is provided in the appendix.

5.4 Finding the best basis subtree for each block

The difference between the flattened approach and the hierarchical approach is in the search for the best basis for each block of the signal. While a flattened approach entails an exhaustive search of the entire family of wavelet packets in a "brute force" manner, the hierarchical approach uses a fast "pruning" algorithm to determine the best basis.

While motivated by the "entropy" pruning criterion mentioned in [1], our formulation is in a dynamic programming framework using a Viterbi-like algorithm. Besides, it must be emphasized that the CMQW cost criterion used to populate the tree nodes prior to pruning uses a one-sided function (either rate or distortion), whereas we resort to a Lagrangian cost function, which is optimal in a rate-distortion sense.

A Viterbi-like fast dynamic programming technique is feasible due to the ON property of the WP basis family, that enables the signal space spanned by an arbitrary subtree rooted at internal node t of the tree to be identical to the space spanned by the twin subtrees rooted at the two branches emanating from node t. To be specific, let $t = n_j^i$, i.e. t is the jth node of the ith hierarchical level (or depth) of the tree T. Its two children are $t_1 = n_{2j-1}^{i+1}$ and $t_2 = n_{2j}^{i+1}$. See Figure 1. Then, because of the ON property, the subtrees rooted at t_1 and t_2 cover disjoint halves of the $\mathrm{R}^{N/2^i}$ signal space spanned by their parent node t.

This allows a direct *quantitative* one-to-one comparison between the $N/2^i$ basis coefficients $\{c_j^i\}$ associated with the basis subset $\{b_j^i\}$ of node t with the $(2 \times (N/(2^{(i+1)})))$ coefficients $\{\{c_{2j-1}^{i+1}\}, \{c_{2j}^{i+1}\}\}$ associated with the basis subsets $\{b_{2j-1}^{i+1}\}$ and $\{b_{2j}^{i+1}\}$ of nodes t_1 and t_2 respectively. The "split/merge" decision will be based on which option leads to a cheaper Lagrangian cost, as spelled out in Figure 4.

Assume known the optimal subtree from node $t = n_j^i$ "onwards" from node t "onwards" to the full tree-depth $\log N$. We could liken the subtrees to surviving paths in the Viterbi algorithm [14]. Then, by Bellman's optimality

principle [15], we know all surviving paths passing through node $t = n_j^i$ at depth i must invoke this same optimal "finishing" path. There are two contenders for the "surviving path" at every node of the tree, the parent and its children, with the winner having the lower Lagrangian cost.

Using this, we begin at the complete tree-depth $n = \log N$ and work our way towards the root t_0 of the tree, using the above cost criterion at each level i to determine whether to split or merge. This decision (or "path") is remembered and used to determine the best path when applying the same pruning criterion on the branches, which process is repeated till the root is encountered. At this point, the entire best path or best wavelet packet basis is known.

6 Complete Algorithm

The stage is now set to integrate the results of the previous two sections to formulate the optimal algorithm. This will be done in two phases. First, the optimal algorithm for a given operating slope λ will be flowcharted, followed by a description of the hunt for the optimal operating slope λ^*. Note that the algorithm is applied independently on each signal block to determine the best wavelet packet basis corresponding to that subblock.

6.1 Initialization

Prior to the actual pruning operation, a one-time fixed cost of gathering the statistics enlisted in Steps 1 and 2 below must be endured. Associated with every node n_j^i of T is a data structure of the form: $\{\hat{R}_j^i, \hat{D}_j^i, \hat{J}_j^i, split(n_j^i)\}$. The first three members refer to the rate, distortion, and Lagrangian cost associated with the optimal (for the given λ) subtree from n_j^i onwards, i.e. the optimal subtree rooted at n_j^i, while the last member of the data structure, $split(n_j^i)$, is a binary variable whose meaning (*yes* or *no*) reflects the decision of whether or not it is optimal to split the node into its children branches.

Step 1: Generate the coefficients $\{c_j^i\}$ for the entire WP family.

Step 2: Gather the given quantizer set dependent $(R_q(t), D_q(t))$ values for all the nodes $t \in T\ \forall q \in \mathrm{q_a(t)}$, to generate the R vs D points for each node.

Phase I: Optimality For A Given Operating Slope

Phase I of the algorithm is run for a given slope value λ, and could be considered a subroutine called by the Phase II, described later in the section, for the fixed budget allocation problem:

Step 3: For the λ of the current iteration, populate all the nodes t of the tree with the minimum Lagrangian cost function associated with that node ($J_t(\lambda)$, or equivalently, $J_j^i(\lambda)$ when referring to the j th node at scale i) where $J_t(\lambda) = \min_{q_t}[D_{q_t}(t) + \lambda R_{q_t}(t)]$

Step 4: Initialize $i \leftarrow n$, where $n = \log N$ is the maximum signal block tree-depth. For $t = n_j^n$, if q_t^* is the value of q_t that minimizes $J_t(\lambda)$ initialize:

$$\tilde{R}_j^n \leftarrow R_j^n \ \text{(where } R_j^n = R_{q_t^*}(t))$$
$$\tilde{D}_j^n \leftarrow D_j^n \ \text{(where } D_j^n = D_{q_t^*}(t)))$$
$$\tilde{J}_j^n \leftarrow J_j^n$$

Step 5: $i \leftarrow i - 1$. If $i < 0$, go to Step 8.

Step 6: $\forall j = 1, 2, \ldots, 2^i$ at the ith tree level:

if $J_j^i(\lambda) < \tilde{J}_{2j-1}^{i+1}(\lambda) + \tilde{J}_{2j}^{i+1}(\lambda)$,

then { $split(n_j^i) \leftarrow NO$; $\tilde{R}_j^i \leftarrow R_j^i$; $\tilde{D}_j^i \leftarrow D_j^i$; $\tilde{J}_j^i \leftarrow J_j^i$ }

else { $split(n_j^i) \leftarrow YES$; $\tilde{R}_j^i \leftarrow R_{2j-1}^{i+1} + R_{2j}^{i+1}$;

$\tilde{D}_j^i \leftarrow D_{2j-1}^{i+1} + D_{2j}^{i+1}$; $\tilde{J}_j^i \leftarrow J_{2j-1}^{i+1} + J_{2j}^{i+1}$ }

Step 7: Go to Step 5.

Step 8: Starting from the root t_0, and using, in a linked-list fashion, the node data-structure element $split$(node), selected optimally for all the nodes of T, carve out the optimal subtree $S^*(\lambda)$ and its associated optimal quantizer choice $Q^*(\lambda) \in \mathbf{Q_a}(S^*(\lambda))$. Also readily available at the data-structure for root node t_0 are $R_{Q^*}(S^*) = \tilde{R}_1^0$ and $D_{Q^*}(S^*) = \tilde{D}_1^0$, the rate and distortion of the optimal subtree $S^*(\lambda)$.

A point to be made is that it is possible to directly incorporate into

the pruning algorithm the cost of segmentation (in terms of overhead bits of the subtree map to be sent), if an *a priori* map-representation scheme is available. For example, if the subtree structure costs one bit per merge decision, this bit could be included in the Lagrangian cost comparison of the children nodes with the parent node in Step 6 of the Phase I algorithm outlined above. However, in our generalized algorithm such subtleties are not included (in practice, they make negligible difference anyway) as no *a priori* map-coding scheme is assumed. It is assumed that the R_{budget} criterion given for the problem is for pure coding expenditure without any overhead expenses, which may be minimized using entropy coding of the tree-map if necessary, or by coding only the locations of the non-zero coefficients, if that is cheaper. In our application to be described later, we found that the overhead represented a negligible proportion of the total budget.

6.2 Finding The Optimal Operating Slope

The problem of picking the optimal slope value for a given budget criterion R_{budget} will be the subject of discussion in Phase II of the algorithm; the iterative invocations of the Phase I subroutine in search of the optimal operating slope λ_{opt}, which satisfies the given budget constraint, will be described in this section.

As was shown in section 5, due to the concavity of $W(\lambda)$ in λ (see Figure 9), and since our optimal operating slope λ^* is $\max^{-1}(W)$, we can find our optimal operating point using a fast convex search algorithm like Newton's method or bisection methods [10]. Equivalently stated, we are interested in the zero-crossing operating slope of the derivative of W, $\partial W/\partial\lambda$. Recall that:

$$W(\lambda) = (\sum_i \min_{x_i \in \chi_i} [D(x_i) + \lambda R(x_i)]) - \lambda R_{budget}$$

which implies that, at non-singular values of λ,

$$\partial W/\partial\lambda = \sum_i R_i^*(\lambda) - R_{budget} \tag{27}$$

where $R_i^*(\lambda)$ is the rate associated with the optimal subtree/quantizer choice for block i. Due to the discrete nature of our problem, λ is singular at only a finite number of points (see Figure 9). Also, as was developed in Theorem 2 (see appendix), the optimal slope λ^* which maximizes W corresponds to

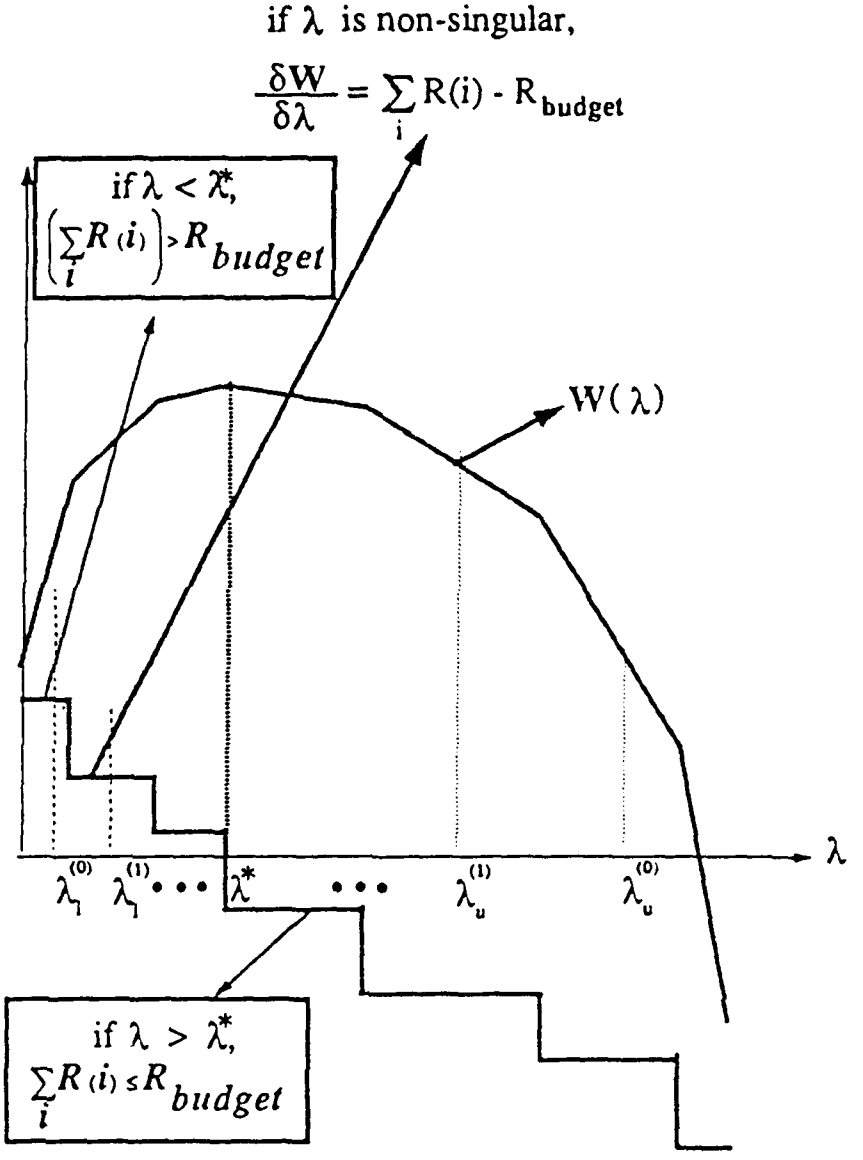


Figure 9: Concavity of the biased Lagrangian functional $W(\lambda)$ and mathematical justification for fast bisection algorithm.

a singular value. From Equation (27), at non-singular values of $\lambda < \lambda^*$, we have $\sum_i R_i^*(\lambda) > R_{budget}$, while for non-singular values of $\lambda > \lambda^*$, $\sum_i R_i^*(\lambda) \leq R_{budget}$. This then leads to the iterative fast convex search algorithm to be described.

As with most iterative solutions, the choice of a good initial operating point is the key to a fast convergence. Picking the initial λ could be a research topic all unto itself. Some heuristics are mentioned in [4]. In a predictive image coding environment, for example, a good guess would be λ_{opt} of the previous frame. Assume we have judiciously chosen two values of λ, λ_l and λ_u with $\lambda_l \leq \lambda_u$ which satisfy the relation:

$$\sum_i R_i^*(\lambda_u) \leq R_{budget} \leq \sum_i R_i^*(\lambda_l)$$

Note that failure to find any λ_l, λ_u which satisfy the above inequalities means that the given problem is unsolvable; i.e. the R_{budget} is inconsistent with the given sets of quantizers. A conservative choice for a solvable problem would be $\lambda_l = 0, \lambda_u = \infty$.

Phase II: Iterating towards the optimal operating point

Now the following "main" algorithm can be used to iteratively call the "subroutine" algorithm of the previous section:

Step 1: Pick $\lambda_l \leq \lambda_u$ such that

$$\sum_i R_i^*(\lambda_u) \leq R_{budget} \leq \sum_i R_i^*(\lambda_l)$$

If the inequality above is an equality for either slope value, stop. We have an exact solution. Otherwise, proceed to Step 2.

Step 2: $\lambda_{next} \leftarrow \left| \frac{\sum_i [D_i^*(\lambda_l) - D_i^*(\lambda_u)]}{\sum_i [R_i^*(\lambda_l) - R_i^*(\lambda_u)]} \right| + \epsilon$, where ϵ is a vanishingly small positive number picked to ensure that the lower rate point is picked if λ_{next} is a singular slope value.

Step 3: Run the Phase I optimal algorithm for λ_{next}.
$\Longrightarrow$ if $\{\sum_i R_i^*(\lambda_{next}) = \sum_i R_i^*(\lambda$ [illegible], then stop. $\lambda^* = \lambda_u$.

$\Longrightarrow$ else if $(\sum_i R_i^*(\lambda_{next}) > R_{budget}), \lambda_l \leftarrow \lambda_{next}$. Go to Step 2.

$\Longrightarrow$ else $\lambda_u \leftarrow \lambda_{next}$. Go to Step 2.

Toy Example

See Figure 10 for a plot of the convex hull to the operational rate–distortion curve for the given problem. Shown explicitly are the optimal quantizer and the best basis choice for each operating point, which corresponds to singular values of λ, whose sweep from 0 to ∞ results in the tracing out of all convex hull points. The budget constraint line of 21 bits is obviously an inaccessible convex hull solution, and one has to settle for the convex hull operating point using 20 bits. Note also the non-monotonic nature of the sequence of the depths of the best bases subtrees as one sweeps λ through all positive real numbers.

Let us first show an example of how the Phase I algorithm works for $\lambda = 10$ (to pick a nice number) and show how it leads to the lowest quality convex hull point of Figure 10, which is picked for all values of $\lambda > 5.43$:

i) Populate the tree with the minimum of all the Lagrangian cost functionals for $\lambda = 10$ as outlined in Phase I of the algorithm for each node A, B1,B2,C1,C2,C3, and C4 to get:

J_A=231 (achieved with quantizer Q3)
J_{B1}=92.5 (Q3); J_{B2}=102.3 (Q3)
J_{C1}=46.25 (Q3); J_{C2}=52.25 (Q3); J_{C3}=52.25 (Q3); J_{C4}=60.25 (Q2)

ii) Initialize i=2; $\tilde{J}_{C1}$=46.25; $\tilde{J}_{C2}$=52.25; $\tilde{J}_{C3}$=52.25; $\tilde{J}_{C4}$=60.25;

iii) i=1; Since $J_{B1} < \tilde{J}_{C1} + \tilde{J}_{C2}$, $split(B1) \leftarrow NO$;
$J_{B2} < \tilde{J}_{C3} + \tilde{J}_{C4}$, $split(B2) \leftarrow NO$;
$\tilde{J}_{B1} = J_{B1}; \tilde{J}_{B2} = J_{B2}$

iv) i=0; Since $J_A > \tilde{J}_{B1} + \tilde{J}_{B2}$, $split(A) \leftarrow YES$;
$\tilde{J}_A = \tilde{J}_{B1} + \tilde{J}_{B2}$;

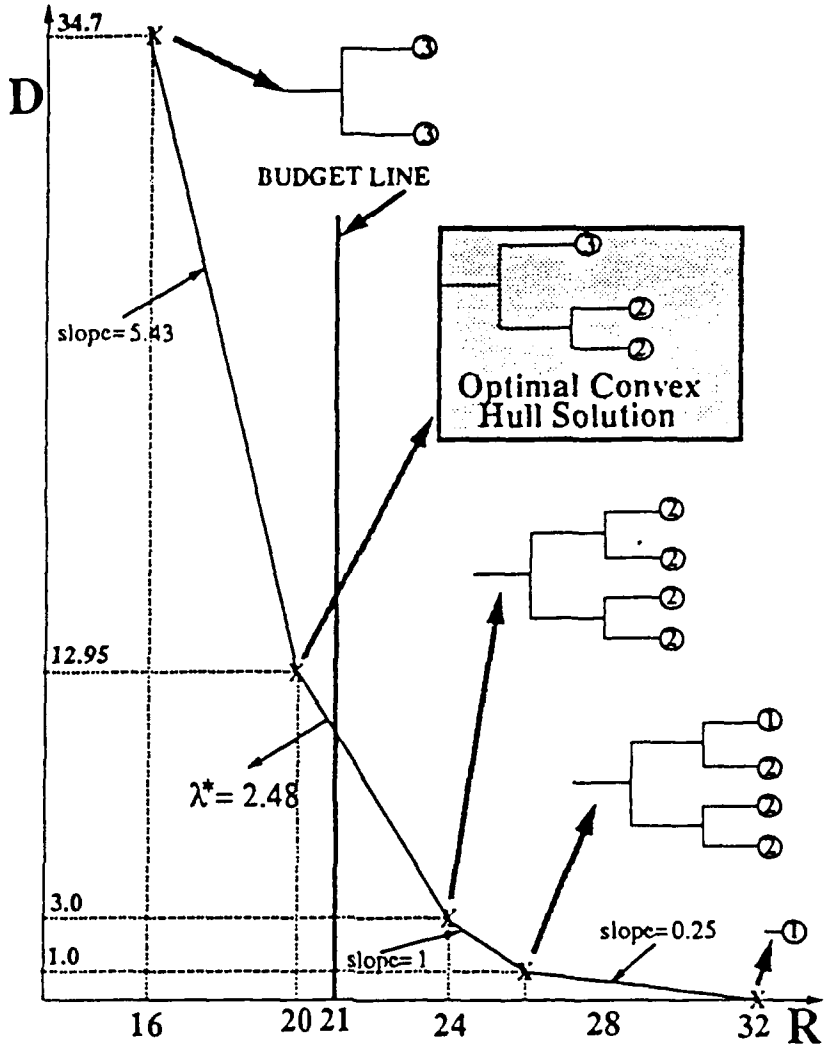


Figure 10: Composite R-D curve for toy example shown with best basis and best quantizer choices for all convex-hull points, and with optimal tree structure for the given budget constraint. Note the non-monotonicity of the R-D characteristics with tree depth, i.e. non-conformance with the Chou et. al.[6] assumptions.

We thus have our optimal basis subtree (with quantizer choice) for this value of λ, as shown as the lowest rate convex hull point of Figure 10. We now explain in detail the search for λ^* for the toy problem with a coding budget $R_{budget} = 21$ bits. Refer to Figure 10.

I) Initialize $\lambda_l^{(0)} = 0; \lambda_u^{(0)} = \infty$.
$R^*(\lambda_u) = 32; R^*(\lambda_l) = 16;$

II) $\lambda_{next} = \frac{34.7-0}{32-16} + \epsilon = 2.17 + \epsilon;$
$\Longrightarrow \lambda_l^{(1)} = 2.17; \lambda_u^{(1)} = \infty;$

III) $\lambda_{next} = \frac{34.7-3.0}{24-16} + \epsilon = 3.96 + \epsilon;$
$R^*(\lambda_{next}) = 20 < R_{budget} = 21;$
$\Longrightarrow \lambda_l^{(2)} = 2.17; \lambda_u^{(2)} = 3.96;$

IV) $\lambda_{next} = \frac{12.95-3.0}{24-20} + \epsilon = 2.48 + \epsilon;$
$R^*(\lambda_{next}) = R^*(\lambda_u) = 20;$
We have converged! $\Longrightarrow \lambda^* = 2.48;\ D(\lambda^*) = 12.95; R(\lambda^*) = 20;$

7 Image coding application using quadtree segmentation

We now describe an image processing application of the optimal pruning algorithm described earlier for the particular case where the tree structure is quadtree, and the basis family is the DCT transform. The coding environment is a modified version of the still-image coding standard, JPEG [8].

7.1 Adaptive quantization

One of the keys to achieving good signal compression is to have the quantization process adapt dynamically to the signal's non-stationarities. One way to accomplish this is to have classified quantizers. The concept of classifying quantizers for VQ applications was done in [16], and the use of classification in transform coding applications is also not new. The idea of classification is to have an assortment of classes with each class tuned to a particular signal

characteristic. For our image coding application, through empirical toil, we determined that a good tradeoff between overclassification with subsequent high side-information overhead cost and being "overstatic" by not adapting enough was to have 4 classes, each optimized for a particular image characteristic.

Our four quantizer classes were optimized for (1) a "typical" image subblock with low frequencies weighted much higher than the perceptually less sensitive higher frequencies, à la the JPEG suggested matrix; (2) horizontal edges; (3)vertical edges; and (4) image subblocks which are "white" in their frequency spectrum, with no discernible favoritism towards any specific orientation. A point of note when classifying transform coded images is that the DCT transform is completely symmetric with respect to phase reversals in the intensity gradients, thus requiring, for example, a single horizontal quantizer matrix, as opposed to one for a light-to-dark horizontal gradient, and another for a dark-to-light one. This cuts down the number of classification quantizers needed, as compared to the application of [16].

The admissible classes of quantizers described above were constructed for each of 3 hierarchical levels of the DCT basis tree: 4x4 blocks, 8x8 blocks, and 16x16 blocks for every 16x16 subblock of the original image. As mentioned earlier, this constraint is in keeping with the perceptual blockiness requirements [17], as well as the lack of usefulness of the DCT transform for sizes that are too small.

7.2 Coding description and simulation results

The DCT tree was grown to the three hierarchical levels described for every 16x16 subblock into which the image was divided. This is equivalent to a parallel pruning of the 16x16 subblock trees into which the original image was divided for independent coding of the subblocks. The optimal pruning algorithm as described earlier was invoked for each subblock tree. "Pseudo-JPEG" coding algorithms were followed for the non 8x8 blocks, [4] i.e. DCT transformation, quantization using classified quantizers, zigzag scanning, and RL coding of the zero runs, etc. Figure 11 shows comparisons of the adaptive DCT-quadtree coder versus the baseline JPEG coder plotting the PSNR (Peak Signal to Noise Ratio) defined as $10\log_{10}[255^2/(m.s.e)]$ versus bpp

[4]The standard JPEG algorithm is applicable only to 8x8 blocks

(bits per pixel) for some typical test images used in the image processing community. The results are compared with a typical JPEG coder deploying the suggested JPEG quantization matrix [8]. In formulating the pseudo-JPEG algorithm for the non-8x8 blocks, the same default Huffman coding table was used as outlined in the baseline JPEG specification for 8x8 blocks, and hence is suboptimal in general. This essentially places a lower bound on the performance of the adaptive coder, which can perform better if Huffman codes customized to 16x16 and 4x4 blocks are created. The hierarchy of classified quantizers was tried on both original images as well as difference images. The difference images were derived from a hybrid DCT-based pyramid coding scheme [18]. See Figure 12 for performance comparisons. As can be seen, for the hierarchy of classified quantizers picked and for the pseudo-JPEG coding scheme invoked, the adaptive DCT-based quadtree coder outperforms the standard static JPEG coder by about 1.5-2 dB at typical bitrates, or alternatively, by 15%-25% reduction in bitrate at typical PSNR values. For difficult images like "Barbara," even more impressive results were obtained. As can be seen from the plots, our adaptive scheme outperforms the static JPEG scheme for this image by about 2-3 dB at fixed bitrate, or equivalently about 25-35% compression advantage at fixed SNR, over an entire range of bit rates of interest.

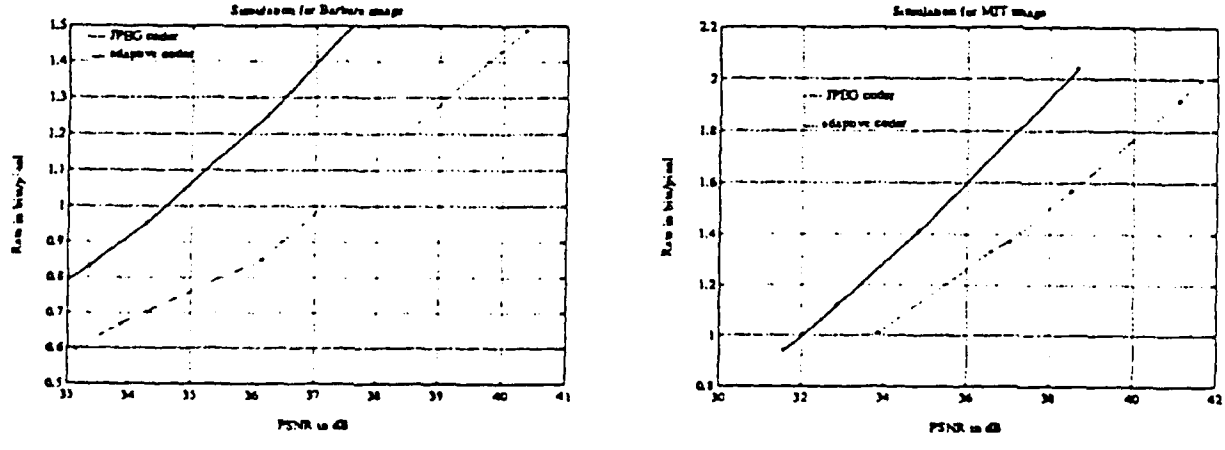


Figure 11: Comparison of adaptive depth-3 block DCT basis quadtree coding scheme with non-adaptive JPEG coding scheme for the "Barbara" and "mit" images.

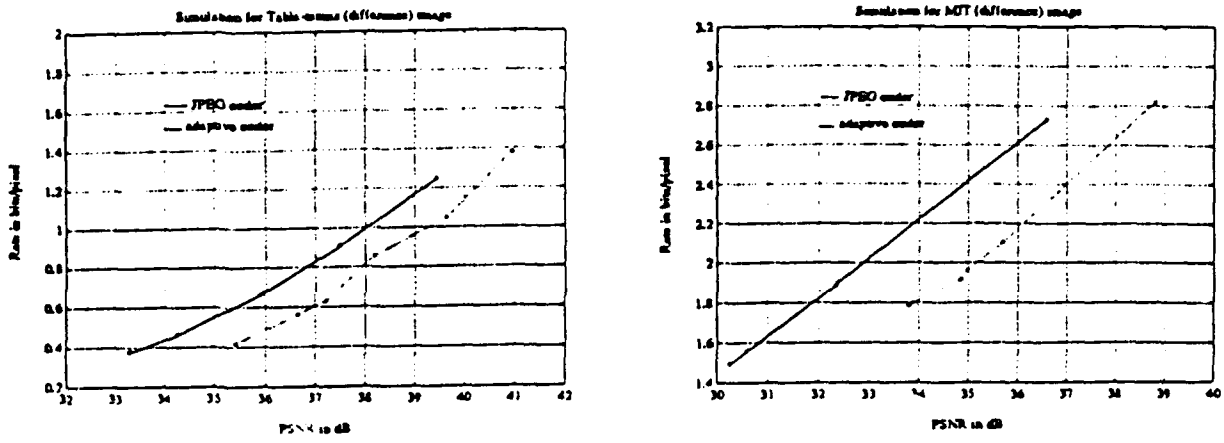


Figure 12: Comparison of adaptive depth-3 block DCT basis quadtree coding scheme with non-adaptive JPEG coding scheme for difference images derived from the "table-tennis" and "mit" images.

8 Conclusion

We have shown, for a given hierarchy of admissible quantizers, an efficient scheme for coding adaptive trees whose individual nodes spawn off descendents forming a disjoint and complete basis cover for the space spanned by their parent nodes. The scheme presented guarantees operation on the convex hull of the operational R-D curve for the admissible hierarchy of quantizers. Applications for this coding technique include the CMQW generalized multiresolution wavelet packet decomposition, iterative subband coders, and quadtree structures. An application to image processing involving quadtrees with a family of DCT bases has been demonstrated in a JPEG-like coding environment with good improvement shown over the static JPEG coding scheme.

Appendix

Proof of Lemma 1:

Proof: Denoting $\lambda_3 = \theta\lambda_1 + (1-\theta)\lambda_2$, where $0 \leq \theta \leq 1$, and recalling the vector notation $\mathbf{x}$ to represent the sequence $x_1, x_2, \ldots, x_M$, we have:

$$\begin{aligned} W(\lambda_3) &= W(\theta\lambda_1 + (1-\theta)\lambda_2) \\ &= \min_{\mathbf{x}\in\mathbf{X}^M} \{D(\mathbf{x}) + [\theta\lambda_1 + (1-\theta)\lambda_2]R(\mathbf{x}) - \lambda_3 R_b\} \\ &\geq \min_{\mathbf{x}\in\mathbf{X}^M} \theta[D(\mathbf{x}) + \lambda_1 R(\mathbf{x}) - \lambda_1 R_b] + \min_{\mathbf{x}\in\mathbf{X}^M} (1-\theta)[D(\mathbf{x}) + \lambda_2 R(\mathbf{x}) - \lambda_2 R_b] \\ &= \theta W(\lambda_1) + (1-\theta)W(\lambda_2) \quad \square \end{aligned}$$

Proof of Theorem 2:

Proof: Denote by λ' the slope of the convex hull face which "straddles" the budget constraint line on the R-D plane. See Figure 7. Let us consider the convex-hull face of slope λ' as a candidate for $\lambda^* = \max^{-1}(W(\lambda))$.

For $\lambda < \lambda'$, invoking the "lower rate" operating point on the convex hull of slope $\lambda', \mathbf{x}_\ell$, we have:

$$\begin{aligned} W(\lambda) - W(\lambda') &= \min_{\mathbf{x}\in\mathbf{X}^M} [D(\mathbf{x}) + \lambda R(\mathbf{x}) - \lambda R_b] - [D(\mathbf{x}_\ell) + \lambda' R(\mathbf{x}_\ell) - \lambda' R_b] \\ &\leq [D(\mathbf{x}_\ell) + \lambda R(\mathbf{x}_\ell) - \lambda R_b] - [D(\mathbf{x}_\ell) + \lambda' R(\mathbf{x}_\ell) - \lambda' R_b] \\ &\leq (\lambda - \lambda')(R(\mathbf{x}_\ell) - R_b) \\ &\leq 0 \end{aligned}$$

Similarly, for $\lambda > \lambda'$, invoking the "higher rate" operating point on the convex hull of slope $\lambda', \mathbf{x}_\gamma$, we have:

$$W(\lambda) - W(\lambda') \leq (\lambda - \lambda')(R(\mathbf{x}_\gamma) - R_b) \leq 0$$

Thus, for all positive values of $\lambda, W(\lambda) \leq W(\lambda^*)$.
But, by virtue of Lemma 1, we know that $W(\lambda)$, being concave $\cap$, has a

THIS PAGE IS BLANK

DUE TO A

PAGE NUMBERING ERROR

[10] W. K. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1988.

[11] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. XLI, pp. 909–996, 1988.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.

[13] M. Minoux, *Mathematical Programming: Theory and Algorithms.* Wiley, 1986.

[14] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding.* McGraw-Hill, 1979.

[15] R. Bellman, *Dynamic Programming.* Princeton University Press, 1957.

[16] B. Ramamurthi and A. Gersho, "Classified vector quantization of images," *IEEE Transactions on Communications*, vol. 34, pp. 1105–1115, Nov. 1986.

[17] D. J. Vaisey and A. Gersho, "Variable block-size image coding," *Proceedings of ICASSP*, pp. 1051–54, 1987.

[18] K. M. Uz, M. Vetterli, and D. LeGall, "Interpolative multiresolution coding of advanced television with compatible subchannels," *IEEE Transactions on CAS for Video Technology, Special Issue on Signal Processing for Advanced Television*, vol. 1, pp. 86–99, Mar. 1991.

unique maximum value which occurs at a singular slope. (If it were non-singular, then there exists an $\epsilon > 0$, no matter how small, for which $W(\lambda^* + \epsilon) - W(\lambda^*) = \lambda^* \epsilon > 0$, which contradicts the definition of $W(\lambda^*)$).

Thus, λ' is indeed this unique sigular maximum, with the optimal convex hull operating point obviously being $\mathbf{x}_\xi$. □

References

[1] R. Coifman, Y. Meyer, D. Quake, and V. Wickerhauser, "Accoustic signal compression with wave packets," *Wavelet Workshop, Marscille,*, Oct. 1990.

[2] M. V. Wickerhauser, "INRIA lectures on wavelet packet algorithms," tech. rep., Numerical Algorithms Research Group, Dept. of Mathematics, Yale University, Mar. 1991.

[3] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Transactions on Information Theory*, vol. IT-22, pp. 162–169, Mar. 1976.

[4] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on ASSP*, vol. 36, pp. 1445–1453, Sept. 1988.

[5] Y. Shoham and A. Gersho, "Efficient codebook allocation for an arbitrary set of vector quantizers," *Proceedings of ICASSP*, pp. 43.7.1–43.7.4, 1985.

[6] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Transactions on Information Theory*, vol. IT-35, pp. 299–315, Mar. 1986.

[7] G. J. Sullivan and R. L. Baker, "Efficient quadtree coding of images and video," *Proceedings of ICASSP*, vol. 4, pp. 2661–2664, May 1991.

[8] "JPEG technical specification: Revision (DRAFT), joint photographic experts group, ISO/IEC JTC1/SC2/WG8, CCITT SGVIII," Aug. 1990.

[9] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *To appear, IEEE Transactions on Signal Processing*, Sept. 1992.

Submitted to IEEE Trans. on Signal Processing, March 1992.

Wavelets and Recursive Filter Banks

Cormac Herley * and Martin Vetterli †
Department of Electrical Engineering
and Center for Telecommunications Research
Columbia University, New York, NY 10027-6699

March 2, 1992

Abstract

Recent work has shown that perfect reconstruction filter banks can be used to derive continuous-time bases of wavelets; the case of finite impulse response filters, which lead to compactly supported wavelets, has been examined in detail. In this paper we show that infinite impulse response filters lead to more general wavelets of infinite support. We give a complete constructive method which yields all orthogonal two channel filter banks, where the filters have rational transfer functions, and show how these can be used to generate orthonormal wavelet bases. A family of orthonormal wavelets, which shares with those of Daubechies the property of having a maximum number of disappearing moments, is shown to be generated by the halfband Butterworth filters. When there is an odd number of zeros at π we show that closed forms for the filters are available without need for factorization. A still larger class of orthonormal wavelet bases having the same moment properties is presented, and contains the Daubechies and Butterworth filters as the boundary cases. We then show that it is possible to have both linear phase and orthogonality in the infinite impulse response case, and give a constructive method. We show how compactly supported bases may be orthogonalized, and construct bases for the spline function spaces. These are alternatives to those of Battle and Lemarié, but have the advantage of being based on filter banks where the filters have rational transfer functions and are thus realizable. Design examples are presented throughout.

*Work supported in part by the National Science Foundation under grant ECD-88-11111.

†Work supported in part by the National Science Foundation under grants ECD-88-11111, and MIP-90-14189.

1 Introduction

The subject of wavelets has been studied by applied mathematicians for a number of years, as representing an alternative to traditional Fourier based analysis techniques. Considerable interest has been shown by the Signal Processing community more recently owing, in large measure, to the influence of pivotal papers by Mallat [1, 2] and Daubechies [3]. These demonstrate the strong link between the subject of Wavelets and that of multirate filter banks. Briefly put, multirate filter banks give the structures required to generate important cases of wavelets and the wavelet transform. Wavelets provides an elegant mathematical basis for multirate filter banks, and that mathematical machinery has led to new results [4].

Among the most celebrated wavelet bases are those of Meyer [5] and of Battle and Lemarié [6, 7], and these can be realized using orthogonal multirate filter banks; however the filters involved are not rational, and the corresponding wavelet cannot be computed exactly, so they are of limited worth from a Signal Processing point of view. More interesting are the compactly supported wavelets of Daubechies [3]. These are based on orthogonal Finite Impulse Response (FIR) filter banks, which have in fact been under study for some time [8, 9, 10].

Our principal interest in this paper is again orthogonal filter banks and their relation to wavelet bases. We consider Infinite Impulse Response (IIR) filter banks, which have been less studied, and which allow much greater freedom than their FIR counterparts.

The essential contribution of the paper consists of strong new results on orthogonal IIR filter banks which allow us to thoroughly examine the structure of possible solutions, and present new designs. The connection with Wavelets allows us to use these designs to get novel orthogonal wavelets, which are based on structures that are computable with finite complexity. Thus we present filters that are of interest in their own right, but which also allow us to generate wavelet bases which are in some senses comparable, and in others superior to those already published.

The summary of the paper is as follows. We present a succinct review of the relation between orthonormal wavelet bases and filter banks having orthogonality properties in section 2. We recall that designing a certain class of orthonormal wavelet bases is related to the simpler problem of designing orthogonal filter banks, provided that the filters satisfy certain regularity conditions. Since this material has been reviewed in a number of papers our treatment is limited to the essential points. Readers unfamiliar with the subject might consult [8, 11, 12] for additional coverage of filter banks, and [13, 3, 5, 4, 14] for treatments of the connection with wavelets.

In section 3 we present a constructive method to find all orthogonal filter banks, where the filters have rational transfer functions. In certain cases of considerable interest, we actually get closed form expressions for the filters; so that no factorization or approximation is necessary. This contrasts sharply with the FIR case, and seems to be the first closed form for a non-trivial implementable wavelet.

Section 4 demonstrates that wavelets with moment properties are derived from filter banks where the filter frequency responses are maximally flat. We construct the whole family of maximally flat filters for orthogonal filter banks, and show that the Butterworth halfband filters and the Daubechies solutions are included as special cases.

Section 5 illustrates that linear phase and orthogonality are not mutually exclusive properties for IIR filter banks, as they were in the FIR case. Filters of considerable interest can be designed that lead to orthogonal wavelets with symmetry.

We show in section 6 that if a compactly supported basis for one of the spaces in the multiresolution analysis structure exists, then we can always generate an orthogonal basis from realizable IIR filters. A special case is the N-th order spline function space; so we

construct bases which have the advantage over those of Battle and Lemarié that they are realizable.

Certain of the results have been presented in preliminary form in [15, 16, 17].

1.1 Notation

The set of real numbers will be represented by R, the set of integers by Z. The inner product over the space of square-summable sequences $l^2(Z)$ is:

$$< a(n), b(n) > = \sum_{n=-\infty}^{\infty} a^*(n)b(n),$$

where $a(n), b(n) \in l^2(Z)$, and superscript "$*$" denotes complex conjugation. Generally we shall deal with sequences and functions that are real. We define $\| a(n) \|_2^2 = < a(n), a(n) >$. The z-transform of a sequence is defined by $H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n}$. In an abuse of notation we shall use the same symbol for the Discrete Fourier Transfrom $H(e^{jw}) = H(z)|_{z=e^{jw}}$. Similarly over the space of square-integrable functions $L^2(R)$ we have the inner product:

$$< f(x), g(x) > = \int_{-\infty}^{\infty} f^*(x)g(x)dx,$$

where $f(x), g(x) \in L^2(R)$. The squared norm is given by $\| f(x) \|_2^2 = < f(x), f(x) >$. For continuous-time functions we will use subscripts to denote affine variable changes where the scales are powers of 2 as follows: $f_{jk}(x) = 2^{-j/2} \cdot f(2^{-j}x - k)$.

Our main interest is with filters that have z-transforms that can be written as $H(z) = z^k A(z)/B(z)$ for some $A(z)$ and $B(z)$ which are polynomials in z^{-1}. Since we deal with both causal and anticausal filters we shall often have positive and negative powers of z^{-1}. A function that has terms in both z and z^{-1} is not a polynomial, but we refer to it as an FIR function provided that it has a finite number of terms. The following shorthand notation for a causal FIR functions of length N is used: $\sum_{n=0}^{N-1} a_n z^{-n} = (a_0, a_1, a_2, \cdots a_{N-1})$.

In the following we shall refer to any symmetric or antisymmetric filter that has a central term as having whole sample symmetry (WSS) or whole sample antisymmetry (WSA), and one that does not have a central term as having half sample symmetry or antisymmetry (HSS or HSA). In the case of FIR filters WSS and WSA correspond to filters of odd length, and are often referred to Type I and Type III filters respectively [18]; whereas HSS and HSA imply filters of even length of Type II and Type IV respectively. Some of the basic properties of symmetric sequences that we will have need of are reviewed in appendix A.1.

2 Wavelets and filter banks

2.1 Multiresolution signal processing

The material of this section can also be found in [3, 1, 4, 19]. Two texts give very comprehensive treatments [5, 20]; a more tutorial approach is given in [21].

2.1.1 Continuous-time bases for multiresolution analysis

The axiomatic description of a multiresolution analysis scheme, as introduced by Mallat and Meyer [22, 2, 5] is that we should have:

(i) A succession of spaces:

$$\cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \cdots, \tag{1}$$

where the union of all the V_j's is $L^2(R)$, and the intersection of all of the spaces contains only the origin,

(ii) $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j-1}$,

(iii) $\exists\ \phi(x) \in V_0$ such that the set $\phi(x-n), n \in Z$ constitutes an orthonormal basis for V_0[1].

It follows that the set $\{\phi_{jk}(x) = 2^{-j/2} \cdot \phi(2^{-j}x - k),\ k \in Z\}$ is an orthonormal basis for V_j.

Next, let W_j be the orthogonal complement of V_j in V_{j-1}; that is $x \in V_j$, $y \in W_j \Rightarrow < x, y > \ = \ 0$ and $V_{j-1} = V_j \oplus W_j$. Obviously $V_j \subset V_{j-1}$ and $W_j \subset V_{j-1}$, so that the basis functions of V_j and W_j ($\phi_{jk}(x)$ and $\psi_{jk}(x)$ respectively) can be written as a linear combination of the basis functions of V_{j-1} ($\phi_{j-1k}(x)$). This gives the relations:

$$\phi(x) \ = \ 2^{1/2} \cdot \sum_{n=-\infty}^{\infty} h_0(n) \cdot \phi(2x - n), \tag{2}$$

$$\psi(x) \ = \ 2^{1/2} \cdot \sum_{n=-\infty}^{\infty} h_1(n) \cdot \phi(2x - n), \tag{3}$$

where infinitely many of the $h_0(n)$ and $h_1(n)$ may differ from zero. Since (2) relates $\phi(x)$ and $\phi(2x)$ it is called a two-scale difference equation. Note that $\phi(x)$ and $\psi(x)$ are called the scaling function and wavelet respectively.

Because V_j and W_j are orthogonal we find:

$$< \phi(x), \psi(x-k) > = 0 \ = < h_0(n), h_1(n-2k) > . \tag{4}$$

So the two sequences $h_0(n)$ and $h_1(n)$ must be orthogonal with respect to shifts by two.

Equally, by imposing the constraint that the bases for V_j and W_j be orthogonal:

$$< \phi(x), \phi(x-k) > = \ \delta_k \ = < \psi(x), \psi(x-k) >,$$

we find that:

$$< h_0(n), h_0(n-2k) > = \ \delta_k \ = < h_1(n), h_1(n-2k) > . \tag{5}$$

2.2 Wavelets derived from filter banks

We have seen above that the multiresolution analysis scheme with orthogonal basis functions satisfying (2) and (3) implies certain restrictions on the related sequences $h_0(n)$ and $h_1(n)$; that is (4) and (5) must hold. Also, since $< \phi(2x-k), \phi(x) > \ = \ 2^{-1/2} \cdot h_0(k)$, and $< \psi(2x-k), \phi(x) > \ = \ 2^{-1/2} h_1(k)$, it is obvious that once the basis functions $\phi(x)$ and $\psi(x)$ are known the related filters are easily found. However it is not yet obvious how functions satisfying the desired constraints may be found.

[1] Actually it is sufficient to have a basis, which can then be orthogonalized.

2.2.1 Limit functions of filter banks

A way to construct $\phi(x)$ and $\psi(x)$ from the associated discrete sequences was first shown by Daubechies in [3]. Essentially it entails considering the limit of a sequence of functions $f^{(i)}(x)$ which are piecewise constant on intervals of length $1/2^i$. The value of the constant is equal to the coefficient of an filter found by cascading i copies of the filter $H_0(z)$ followed by a subsampler [4].

Assuming for the moment that the limit exists, and that the filters $h_0(n)$, $h_1(n)$ satisfy the orthogonality constraints (4) and (5), we have as $i \rightarrow \infty$

$$f^{(\infty)}(x) = 2^{1/2} \cdot \sum_{m=-\infty}^{\infty} h_0(m) f^{(\infty)}(2x - m). \tag{6}$$

Taking the Fourier transform:

$$F^{(\infty)}(w) = 2^{-1/2} \cdot H_0(e^{jw/2}) \cdot F^{(\infty)}(w/2). \tag{7}$$

Now define $M_0(e^{jw}) = 2^{-1/2} H_0(e^{jw})$, so that:

$$\begin{aligned} F^{(\infty)}(w) &= M_0(e^{jw/2}) F^{(\infty)}(w/2) = M_0(e^{jw/2}) M_0(e^{jw/4}) F^{(\infty)}(w/4) \\ &= \prod_{l=1}^{\infty} M_0(e^{jw/2^l}) \cdot F^{(\infty)}(0). \end{aligned} \tag{8}$$

Now also consider the related function:

$$G^{(\infty)}(w) = 2^{-1/2} \cdot H_1(e^{jw/2}) \cdot F^{(\infty)}(w/2),$$

so that:

$$g^{(\infty)}(x) = 2^{1/2} \cdot \sum_{m=-\infty}^{\infty} h_1(m) f^{(\infty)}(2x - m). \tag{9}$$

On comparing (6) and (2), (9) and (3) we now see that $f^{(\infty)}(x)$, and $g^{(\infty)}(x)$ satisfy the two scale difference equations required of the orthonormal wavelet construction. It can be verified that the orthogonality relations of the functions $f^{(\infty)}(x)$, and $g^{(\infty)}(x)$ follow from the orthogonality of the filters [3, 4].

Thus the problem of finding the basis functions for the wavelet scheme is reduced to one of finding appropriate pairs of sequences $h_0(n)$ and $h_1(n)$. For much of the rest of the paper, we will concern ourselves with this.

2.2.2 Regularity

That the infinite product (8) converges as $i \rightarrow \infty$ cannot be taken for granted. Cases where convergence fails altogether, or where the product converges to a discontinuous function are easily found. We would like some guarantees about the convergence of (8) and the continuity of the functions $\phi(x)$ and $\psi(x)$, when they exist. Exactly such a criterion is derived in [3], and is reviewed below.

First factor $M_0(z)$ into its roots at $z = -1$ (if there is not at least one then the infinite product cannot converge [23, 24]) and a remainder function $K(z)$, in the following way:

$$M_0(z) = [(1 + z^{-1})/2]^N K(z).$$

Note that it can be shown that $K(1) = 1$ from the definitions; *i.e.* (5) gives $H_0^2(1) = 2$, so that $M_0(1) = 1$. Now call B the supremum of $|K(z)|$ on the unit circle: $B = \sup_{\omega \in [0,2\pi]} |K(e^{j\omega})|$. Then the following sufficient, but not necessary, test from [3] can be used:

Proposition 2.1 (Daubechies 1988) *If $B < 2^{N-1}$, then the piecewise constant function $f^{(i)}(x)$ defined in (8) converges pointwise to a continuous function $f^{(\infty)}(x)$.*

2.3 Filter banks

We now put the connection between filter banks and wavelets to work. Our interest in this paper is orthogonal wavelet bases, hence we restrict our attention to orthogonal filter banks. More general perfect reconstruction filter banks give rise to biorthogonal systems just as in the FIR case [4]. We assume some familiarity with the basic properties of multirate operations; these are detailed for example in [8, 11, 12].

2.3.1 Perfect reconstruction

The structure shown in Figure 1 is a maximally decimated two channel multirate filter bank. If $\hat{X}(z) = X(z)$ the filter bank has the perfect reconstruction property, and we refer to it as a PRFB. We now make the following choice for the synthesis filters:

$$[G_0(z),\ G_1(z)] \ = \ [H_0(z^{-1}),\ -H_1(z^{-1})]. \tag{10}$$

and choose $H_1(z) = z^{2k-1} H_0(-z^{-1})$. It is easily shown that the output $\hat{X}(z)$ of the overall analysis/synthesis system is then given by:

$$\hat{X}(z) \ = \ \frac{1}{2}[G_0(z)\, G_1(z)] \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \begin{bmatrix} X(z) \\ X(-z) \end{bmatrix} \tag{11}$$

$$= \ \frac{1}{2}[H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1})] \cdot X(z). \tag{12}$$

So for this arrangement of the filters it is clear that we get perfect reconstruction provided:

$$H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 2.$$

The importance of this construction is established by the next lemma.

We first introduce aditional notation that we will need. The 2×2 matrix in (11) is called $\mathbf{H}_m(z)$, the modulation matrix of the system. The following polyphase notation for the filters is standard [8, 11, 12].

$$H_i(z) = H_{i0}(z^2) + z^{-1} H_{i1}(z^2),$$

that is $h_{i0}(n)$ contains the even-indexed coefficients of the filter $h_i(n)$, while $h_{i1}(n)$ contains the odd ones. Thus:

$$\begin{bmatrix} H_{00}(z^2) & H_{01}(z^2) \\ H_{10}(z^2) & H_{11}(z^2) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix}. \tag{13}$$

The matrix on the left hand side is called the polyphase matrix $\mathbf{H}_p(z^2)$.

Lemma 2.2 *The following are equivalent:*
(a) $[\mathbf{H}_m(z^{-1})]^T \cdot \mathbf{H}_m(z) = 2 \cdot \mathbf{I}$,
(b) $[\mathbf{H}_p(z^{-1})]^T \cdot \mathbf{H}_p(z) = \mathbf{I}$,
(c) $H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 2$ *and* $H_1(z) = z^{2k-1}H_0(-z^{-1})A(z^2)$, *where* $A(z)$ *is allpass,*
(d) $< h_i(n), h_j(n-2k) > = \delta_k$ *and* $< h_0(n), h_1(n-2k) > = 0 \quad \forall\, k \in Z$.

A proof can be found in [25]. It is also proved that the choice of synthesis filters (10) is unique for the orthogonal construction.
Because of the impulse response relations in *(d)* we shall refer to any filter bank satisfying the conditions of Lemma 2.2 as *orthogonal*; in the filter bank literature the terms orthogonal, paraunitary and lossless are often used interchangeably [26]. Observe that in Lemma 2.2(c) we use functions of the form $H_i(z)H_i(z^{-1})$, which are called *autocorrelation* or *positive real* functions. It deserves mention that the study of lossless systems and positive real functions has a long history in both circuit theory and signal processing [27, 28, 29, 26]. When we wish to impose orthogonality on the filter bank to be used, we shall use whichever of the equivalent conditions of Lemma 2.2 is most convenient.

Note also that if we define $P(z) = H_0(z)H_0(z^{-1})$ then *(c)* requires that in addition to being an autocorrelation, $P(z)$ satisfies

$$P(z) + P(-z) = 2. \tag{14}$$

Since this condition plays an important role in what follows, we will refer to any function having this property as *valid*. Much of the focus of the paper will be in designing autocorrelation functions that are valid. We shall be interested only in valid functions that are rational, so that they can be factored into rational filters $H_0(z)$ and $H_0(z^{-1})$. These filters can then be implemented using recursive difference equations [18], whereas filters that do not have rational transfer functions have no finite complexity physical implementation.

3 Orthogonal IIR filter banks

We have already seen that constructing an orthogonal filter bank can be reduced to the task of finding a function $P(z)$ which is a valid autocorrelation; that is a function that satisfies (14) and can be factored as $P(z) = H(z)H(z^{-1})$. We first establish an important preliminary result on the form of valid rational functions.

Lemma 3.1 *If a valid rational function* $P(z)$ *has no common factors between the numerator and denominator, then the denominator is one of the two upsampled polyphase components of the numerator.*

Proof: We can write:

$$P(z) = \sum_{n=-\infty}^{\infty} [p(2n) + p(2n+1)z^{-1}]z^{-2n},$$

so the constraint gives:

$$P(z) + P(-z) = 2 \cdot \sum_{n=-\infty}^{\infty} p(2n)z^{-2n} = 2 \qquad \Rightarrow p(2n) = \delta_n. \tag{15}$$

Clearly:

$$P(z) = 1 + \sum_{n=-\infty}^{\infty} p(2n+1)z^{-2n-1} = 1 + z^{-1}F(z^2). \tag{16}$$

If $F(z^2)$ has no common factors between its numerator and denominator, then they must each be functions of z^2, possibly multiplied by some delay z^k. That is $F(z^2) = z^k N(z^2)/(z^k D(z^2))$. So we have:

$$P(z) = \frac{z^{-k}D(z^2) + z^{-(k+1)}N(z^2)}{z^{-k}D(z^2)}.$$

Thus the denominator is the first polyphase component if k is even, and the second if k is odd. The numerator and denominator of $P(z)$ are coprime if and only if $N(z)$ and $D(z)$ are. □

3.1 Structure of the solutions

Clearly Lemma 3.1 gives a simple method to design a rational function $P(z)$ which is valid. Lemma 2.2 then shows that this can be used to give an orthogonal filter bank if this function is an autocorrelation, that is it can be factored as $P(z) = H_0(z)H_0(z^{-1})$, since the essential requirement of Lemma 2.2(c) is that $H_0(z)H_0(z^{-1})$ be valid. The next theorem puts these parts together and shows how to design valid autocorrelation functions, and hence orthogonal filter banks. It's utility is that it is constructive and complete.

Theorem 3.2 *All orthogonal rational two channel filter banks can be formed as follows:*
(i) Choosing an arbitrary polynomial $R(z)$, form:

$$P(z) = \frac{2 \cdot R(z)R(z^{-1})}{R(z)R(z^{-1}) + R(-z)R(-z^{-1})}, \tag{17}$$

(ii) Factor as $P(z) = H(z)H(z^{-1})$,
(iii) Form the filter $H_0(z) = A_0(z)H(z)$, where $A_0(z)$ is an arbitrary allpass,
(iv) Choose $H_1(z) = z^{2k-1}H_0(-z^{-1})A_1(z^2)$, where $A_1(z)$ is again an arbitrary allpass.
(v) Choose $G_0(z) = H_0(z^{-1})$, and $G_1(z) = -H_1(z^{-1})$.

Proof: From Lemma 2.2(c) it is necessary and sufficient to find a valid rational autocorrelation function $P(z)$; since once this is factored as $P(z) = H_0(z)H_0(z^{-1})$ then $H_1(z)$ is specified by Lemma 2.2(c), and $G_0(z)$ and $G_1(z)$ by (10).

We show first that (17) always gives a valid, rational autocorrelation. It is valid, since:

$$\begin{aligned} P(z) + P(-z) &= \frac{2 \cdot R(z)R(z^{-1}) + R(-z)R(-z^{-1})}{R(z)R(z^{-1}) + R(-z)R(-z^{-1})}, \\ &= 2. \end{aligned}$$

It is clearly rational, $R(z)$ being a polynomial. The numerator of (17) is an autocorrelation; so is the denominator, since it is the sum of two autocorrelations $R(z)R(z^{-1})$ and $R(-z)R(-z^{-1})$. Hence $P(z)$ itself is an autocorrelation and can be factored:

$$P(z) = H(z)H(z^{-1}) = H_0(z)H_0(z^{-1}) \cdot A_0(z)A_0(z^{-1}),$$

for some $H(z)$ and an arbitrary rational allpass $A_0(z)$.

Next we show that any valid rational autocorrelation can be written as in (17) for some polynomial $R(z)$.

First, any common factors between the numerator and denominator of the given function can be cancelled; the result is clearly still a valid rational autocorrelation. So it can be written

$$P(z) = \frac{R(z)R(z^{-1})}{B(z)B(z^{-1})},$$

for some polynomials $R(z)$ and $B(z)$. Now we can use **Lemma 3.1** to get that the denominator, $B(z)B(z^{-1})$, is one of the upsampled polyphase components of the numerator:

$$D_0(z^2) = [R(z)R(z^{-1}) + R(-z)R(-z^{-1})]/2,$$

or

$$D_1(z^2) = [R(z)R(z^{-1}) - R(-z)R(-z^{-1})]/2.$$

Note that $R(z)R(z^{-1})$ is always of odd length and is symmetric. It follows that one of its upsampled polyphase components, $D_0(z^2)$, is whole sample symmetric (WSS), while $D_1(z^2)$ is half sample symmetric (HSS). Since half sample symmetric polynomials always have at least one zero at $z = -1$ (see appendix A.1), $D_1(z^2)$ is not a suitable choice for the denominator, as we wish to avoid poles on the unit circle. We therefore have that:

$$P(z) = \frac{2 \cdot R(z)R(z^{-1})}{R(z)R(z^{-1}) + R(-z)R(-z^{-1})}. \square \tag{18}$$

Note: The introduction of the allpass factors $A_0(z)$ and $A_1(z)$ affect only the phase of the filters to be implemented, and not their magnitudes. Equally, in the factorization required by step (ii) there is considerable choice for the phase of the filters; $H(z)$ could be minimum phase or maximum phase or mixed phase. The magnitude of course does not change. Irrational orthogonal factorizations of a rational $P(z)$ function are also possible. We give an example in section 4.4.

The theorem shows that if $R(z)$ ranges over the polynomials then (18) is complete for rational $P(z)$ functions. If $R(z)$ is chosen to be any function, rational or not, it is clear by inspection that (18) will still be a valid autocorrelation, but not in general rational. Completeness is less obvious in this case.

If $R(z)R(z^{-1})$ is itself valid, that is $R(z)R(z^{-1}) + R(-z)R(-z^{-1}) = 2$, and $A_0(z)$ and $A_1(z)$ are both chosen to be delays, then all of the filters specified by Theorem 3.2 are FIR. The synthesis filters are always time reversed versions of the analysis filters, just as in the orthogonal FIR case [4]. All of the FIR orthogonal filter banks can be implemented in a paraunitary lattice structure [10]; a similar result is true for IIR orthogonal filter banks [30], so an efficient and numerically robust implementation is always available.

3.2 Closed form factorization

Theorem 3.2 establishes the importance of valid rational functions which are autocorrelations. Numerical factorization poses certain difficulties however. This is certainly a problem in the FIR case; for example even when $P(z)$ is known exactly, the accuracy with which the

coefficients of $H_0(z)$ can be determined is dependent on the numerical robustness of the root extraction procedure.

We now show that in the special case where $R(z)$ is symmetric and of even length, a closed form factorization is available. The requirement that $R(z)$ be symmetric is very reasonable, since the numerator has to control the stopband of the filter $H(z)$ and typically has all of it's zeros on the unit circle; if this is so, then $R(z)$ is symmetric provided that it is real. For example all of the digital Butterworth, Chebyshev and elliptic filters have symmetric numerators.

Consider a causal symmetric FIR function $R(z)$ of even length $N+1$. Using the relationship between the polyphase components given in fact A.1 in the appendix: $R_1(z) = R_0(z^{-1})z^{-(N-1)/2}$ we can simplify:

$$R(z) = R_0(z^2) + z^{-1}R_1(z^2) = R_0(z^2) + z^{-N}R_0(z^{-2}). \tag{19}$$

This gives:

$$\begin{aligned} R(z)R(z^{-1}) &= [R_0(z^2) + z^{-N}R_0(z^{-2})] \cdot [R_0(z^{-2}) + z^{N}R_0(z^{2})] \\ &= 2R_0(z^2)R_0(z^{-2}) + [z^{-N}R_0(z^{-2})R_0(z^{-2}) + z^{N}R_0(z^{2})R_0(z^{2})]. \end{aligned}$$

Clearly, since N is odd:

$$D_0(z^2) = [R(z)R(z^{-1}) + R(-z)R(-z^{-1})]/2 = 2R_0(z^2)R_0(z^{-2}).$$

And hence

$$P(z) = \frac{R(z)R(z^{-1})}{2R_0(z^2)R_0(z^{-2})}.$$

It is now obvious that one possible choice for factorizing $P(z)$ is:

$$H(z) = \frac{R(z)}{\sqrt{2}R_0(z^2)}. \tag{20}$$

Since $R(z)$ and $R_0(z^2)$ are known exactly, this is a closed form; so $H(z)$ is directly available. Example 4.1 below illustrates this. The importance of this result can be seen by noting that the coefficients of the wavelet expansion can be obtained exactly, since they do not depend on any numerical procedure to find the transfer functions $H_0(z)$ and $H_1(z)$. This appears to be the first closed form for the filters used to generate a non-trivial realizable wavelet.

Observe that $H(z)$ can be rewritten:

$$H(z) = 2^{-1/2} \cdot (1 + z^{-N}A(z^2)), \tag{21}$$

where $A(z) = R_0(z^{-1})/R_0(z)$ is an $(N-1)/2$-th order allpass. The other analysis and synthesis filters have similar expressions, and thus can be implemented very efficiently. It is worth pointing out that the filters in this particular case are themselves valid.

4 Wavelets with moment properties

According to Proposition 2.1 the limits of iterated orthogonal digital filter banks can be used to derive wavelet bases. The sufficient condition to guarantee continuity of the wavelets was that the iterated lowpass filter, that is $H_0(z)$, should contain an adequate number of zeros at $z = -1$. It is for this reason that in the design of compactly supported wavelet bases [3, 31, 4] the emphasis was placed on using filters that have a maximum number of zeros at $z = -1$. In addition, a zero of order N at $z = -1$ in $H_0(z)$ implies N vanishing moments for the wavelet [3]:

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0 \quad k = 0, 1, \cdots N-1. \tag{22}$$

It can be shown that having a maximum number of zeros at $z = -1$, implies a maximally flat characteristic for the filters involved [3, 25, 32]. This implies that both the wavelet and the filter spectrum have considerable smoothness, which may be advantageous in certain contexts.

Our procedure to design orthogonal filters amounts then to the following:

(i) Choosing $B_{2N}(z) = (1 + z^{-1})^N (1 + z)^N$ for some N,

(ii) Finding least degree positive real $F(z) = F_N(z)/F_D(z)$ such that

$$P(z) = B_{2N}(z)F(z) = \frac{(1 + z^{-1})^N (1 + z)^N F_N(z)}{F_D(z)},$$

is valid, and

(iii) Factoring $P(z) = H_0(z)H_0(z^{-1})$.

Of course in [3] only FIR solutions were of interest; so the solutions had $F_D(z) = 1$. In other words the multiplicative factor $F(z)$ required to make $B_{2N}(z)F(z)$ valid had only zeros. In the next subsection we examine the opposite extreme, where $F(z)$ is all-pole, *i.e.* $F_N(z) = 1$. These in fact give rise to the Butterworth halfband filters.

In section 4.2 we examine solutions intermediate between the Daubechies ($F(z)$ all zero), and Butterworth ($F(z)$ all-pole); that is where $F(z)$ is still of minimal degree, but has some combination of poles and zeros.

4.1 Butterworth wavelets

Using Theorem 3.2, constructing regular IIR filter banks that lead to infinitely supported wavelets is very simple. Following Daubechies and the FIR case, if we again place a maximum number of zeros at $z = -1$ then we simply choose $R(z) = (1 + z^{-1})^N$. This gives:

$$P(z) = \frac{(1 + z^{-1})^N (1 + z)^N}{(z^{-1} + 2 + z)^N + (-z^{-1} + 2 - z)^N} = H_0(z)H_0(z^{-1}). \tag{23}$$

These filters are the IIR counterparts of the FIR filters given in [3] in that they generate wavelets with regularity that increases linearly with the degree N of the zero at $z = -1$.

These are in fact the N-th order halfband digital Butterworth filters [18]. That these particular filters satisfy the conditions for orthogonality was also pointed out in [33], and

their use for the construction of wavelets in [34, 35]. The Butterworth filters are known to be the maximally flat IIR filters of a given order.

We propose these Butterworth wavelets as alternatives to the compactly supported examples of [3]; they enjoy exactly the same moment properties, but achieve much better filtering action for the same complexity, and are considerably smoother. An additional advantage is that since $R(z)$ is symmetric we can make use of the closed form factorization of section 3.2 if we choose N to be odd. So in this case we can explicitly write:

$$H_0(z) = \frac{\sum_{k=0}^{N} \binom{N}{k} z^{-k}}{\sqrt{2} \cdot \sum_{l=0}^{(N-1)/2} \binom{N}{2l} z^{-2l}},$$

and the other filters follow from Theorem 3.2.

Example 4.1 *Take $R(z) = (1 + z^{-1})^N$ as above and $N = 7$, so that we can use the closed form factorization, hence:*

$$\begin{aligned} P(z) &= \frac{(1, 14, 91, 364, 1001, 2002, 3003, 3432, 3003, 2002, 1001, 364, 91, 14, 1) \cdot z^7}{14z^6 + 364z^4 + 2002z^2 + 3432 + 2002z^{-2} + 364z^{-4} + 14z^{-6}} \\ &= \frac{E(z)E(z^{-1})}{F(z)F(z^{-1})}, \end{aligned}$$

where

$$\frac{E(z)}{F(z)} = \frac{(1 + 7z^{-1} + 21z^{-2} + 35z^{-3} + 35z^{-4} + 21z^{-5} + 7z^{-6} + z^{-7})}{\sqrt{2} \cdot (1 + 21z^{-2} + 35z^{-4} + 7z^{-6})}.$$

So using the description of the filters in Theorem 3.2, with the simplest case $A_0(z) = A_1(z) = 1$ and $k = 0$ we find:

$$\begin{aligned} H_0(z) &= \frac{(1 + 7z^{-1} + 21z^{-2} + 35z^{-3} + 35z^{-4} + 21z^{-5} + 7z^{-6} + z^{-7})}{\sqrt{2} \cdot (1 + 21z^{-2} + 35z^{-4} + 7z^{-6})} \\ H_1(z) &= z^{-1} \frac{(1 - 7z^{1} + 21z^{2} - 35z^{3} + 35z^{4} - 21z^{5} + 7z^{6} - z^{7})}{\sqrt{2} \cdot (1 + 21z^{2} + 35z^{4} + 7z^{6})} \\ G_0(z) &= H_0(z^{-1}) \qquad\qquad G_1(z) = -H_1(z^{-1}). \end{aligned}$$

The wavelet, scaling function and their spectra are shown in Figure 2.

4.2 Intermediate solutions

At the beginning of the section we pointed out that in the construction of wavelets with a certain number of vanishing moments, the essence of the design was finding a minimal degree $F(z) = F_N(z)/F_D(z)$, such that $P(z) = B_{2N}(z)F(z)$ was valid. We now explore examples between the extremes of the Daubechies ($F_D(z) = 1$) and the Butterworth ($F_N(z) = 1$) cases.

First note that when $P(z)$ is a rational autocorrelation, both numerator and denominator will be of odd length, and symmetric. As pointed out in the proof of Theorem 3.2 the denominator is in fact the upsampled whole sample symmetric (WSS) polyphase component of the numerator. There are two cases:

- A symmetric FIR function of length $4k+1$ has an upsampled WSS component of length $4k+1$.

- A symmetric FIR function of length $4k+3$ has an upsampled WSS component of length $4k+1$.

To find a solution where $P(z)$ has less poles than in the Butterworth case we must find a function $F_N(z)/F_D(z)$ where $F_D(z)$ is of length $4(k-p)+1$ for some $0 < p < k$ and $F_N(z)$ is of minimal degree such that:

$$P(z) = \frac{(1+z^{-1})^N(1+z)^N F_N(z)}{F_D(z)},$$

is valid. In the Daubechies case we fixed $F_D(z) = 1$ and found the minimal degree $F_N(z)$, and in the Butterworth we fixed $F_N(z) = 1$ and found the minimal degree $F_D(z)$. For the intermediate cases we fix the length of $F_D(z)$ as $4(k-p)+1$ for some $0 < p < k$ and then find the minimal degree $F_N(z)$. For a given binomial factor $(1+z^{-1})^N(1+z)^N$ the total number of poles and zeros of $F(z)$ will not necessarily be the same for the Daubechies, intermediate and Butterworth solutions, although, in fact, it will never vary by more than two.

Note that $F_D(z)$ is the WSS component of $(1+z^{-1})^N(1+z)^N F_N(z)$, but is to be of lower degree than the WSS component of $(1+z^{-1})^N(1+z)^N$. Thus it is apparent that some of the terms of $(1+z^{-1})^N(1+z)^N F_N(z)$ must be zero, and the WSS component must not contain the endterms. This last condition implies that we must have that $(1+z^{-1})^N(1+z)^N F_N(z)$ is of length $4k+3$; since otherwise, if it is of length $4k+1$, the WSS component is also of length $4k+1$, and contains the endterms. It is convenient to treat separately the two cases for N even and odd.

N = 2k + 1 odd: The length of the denominator is $4(k-p)+1$. If we try $F_N(z)$ of length $4p+1$ then the length of the numerator, $(1+z^{-1})^N(1+z)^N F_N(z)$, is $4(k+p)+3$, and that of its WSS component $4(k+p)+1$. The difference between the length of the WSS component of $(1+z^{-1})^N(1+z)^N F_N(z)$, and the length of $F_D(z)$ is hence $4 \cdot 2p$. Since the WSS component of the numerator is symmetric, and a functions of z^2, setting one pair of its endterms to zero in fact decreases its length by 4. If this can be done $2p$ times then the WSS component of the numerator, and the denominator will be of the same length. Note that $2p$ is also the number of independent elements in $F_N(z)$. In fact the solution is found by solving a $2p \times 2p$ system of linear equations.

N = 2k even: The length of the denominator again is $4(k-p)+1$. Now if we try $F_N(z)$ of length $4(p-1)+3$, the length of the numerator is $4(k+p-1)+3$, and that of its WSS component $4(k+p-1)+1$. The difference between the length of the WSS component of the numerator, and the length of $F_D(z)$ is hence $4 \cdot (2p-1)$. Again $2p-1$ is the number of independent elements in $F_N(z)$. In this case the solution is found by solving a set of $(2p-1) \times (2p-1)$ linear equations.

Example 4.2 $N = 7$. *Note that* $N = 2k+1$ *where* $k = 3$. *There are thus two intermediate solutions for* $p = 1, 2$. *Taking the* $p = 1$ *case first, note that* $F_N(z)$ *is of length* 5, *and we wish to set* $2p = 2$ *pairs of endterms of the WSS component of the numerator to zero. The*

situation is illustrated below.

	z^9	z^8	z^7	z^6	z^5	z^4	z^3	z^2	z	z^0	
$z^2B_{14}(z)$	1	14	91	364	1001	2002	3003	3432	3003	2002	$\cdots$
$azB_{14}(z)$		1	14	91	364	1001	2002	3003	3432	3003	$\cdots$
$bB_{14}(z)$			1	14	91	364	1001	2002	3003	3432	$\cdots$
$az^{-1}B_{14}(z)$				1	14	91	364	1001	2002	3003	$\cdots$
$z^{-2}B_{14}(z)$					1	14	91	364	1001	2002	$\cdots$
$F_N(z)$	d	x	d	x	d	x	d	x	d	x	$\cdots$
		↑		↑							

We have used "d" to indicate elements of the HSS component of the numerator, and "x" for elements of the WSS component. Clearly if the indicated endterms of the WSS component equal zero, then the denominator will be of length 9. It is easily seen, that the conditions to set the endterms to zero are

$$\begin{aligned} 14 + a &= 0 \\ 364 + 92a + 14b &= 0, \end{aligned}$$

$\Rightarrow (a, b) = (-14, 66)$. *Thus* $F_N(z) = (z - 12 + z^{-1})$. *The wavelet, scaling function and spectra are shown in Figure 3.*
The second intermediate solution is for $p = 2$, *so that* $F_N(z)$ *is of length* 9, *and we want the WSS component of the numerator to have* $2p = 4$ *pairs of endterms set to zero.*

	z^{11}	z^{10}	z^9	z^8	z^7	z^6	z^5	z^4	z^3	z^2
$z^4B_{14}(z)$	1	14	91	364	1001	2002	3003	3432	3003	$\cdots$
$az^3B_{14}(z)$		1	14	91	364	1001	2002	3003	3432	$\cdots$
$bz^2B_{14}(z)$			1	14	91	364	1001	2002	3003	$\cdots$
$cz^1B_{14}(z)$				1	14	91	364	1001	2002	$\cdots$
$dzB_{14}(z)$					1	14	91	364	1001	$\cdots$
$cz^{-1}B_{14}(z)$						1	14	91	364	$\cdots$
$bz^{-2}B_{14}(z)$							1	14	91	$\cdots$
$az^{-3}B_{14}(z)$								1	14	$\cdots$
$z^{-4}B_{14}(z)$									1	$\cdots$
	d	x	d	x	d	x	d	x	d	$\cdots$
		↑		↑		↑		↑		$\cdots$

The conditions to set the indicated endterms to zero now are:

$$\begin{aligned} 14 + a &= 0 \\ 364 + 91a + 14b + c &= 0 \\ 2002 + 1001a + 364b + 92c + 14d &= 0 \\ 3432 + 3004a + 2016b + 1092c + 364d &= 0. \end{aligned}$$

The solution is $(a, b, c, d) = (-14, 592/7, -274, 3218/7)$. *The wavelet, scaling function and spectra are shown in Figure 4.*

Note: The denominator is of length $4(k-p)+1$ in these intermediate solutions, with $0 < p < k$. For $p = 0$ we would get the Butterworth solution, and for $p = k$ the Daubechies'. For $k = 0, 1$, that is $N = 1, 2, 3$, there are obviously no intermediate solutions.

4.3 Tabulating the $P(z)$ functions

A point that we would wish to emphasize is that in all of the design techniques discussed above it was the construction of $P(z)$ that was central. This was the case for the Daubechies' designs of [3], and the Butterworth and intermediate designs of sections 4.1 and 4.2. Once $P(z)$ is determined the magnitude spectrum of $H_0(z)$ is fixed irrespective of the allpass factors $A_0(z)$ and $A_1(z)$ of Theorem 3.2, and the factorization chosen.

If we desire filters that are maximally flat, or equivalently, wavelets that have a maximum number of vanishing moments then we design a $P(z)$ with the maximum number of zeros at $z = -1$. Those minimum degree $P(z)$'s with this property are easily listed, and this has been done in Table 1 for the cases $N = 1, 2, \cdots 7$. The table exhausts the minimal degree maximally flat $P(z)$ autocorrelation functions for these orders. A crude estimate of the regularity of the wavelets associated with the each function is given.

For comparison purposes the graph of the $N = 7$ Daubechies wavelet and scaling function are given in Figure 5.

4.4 Irrational Factorizations

Theorem 3.2 demonstrates how to calculate all valid rational autocorrelation functions. For implementation reasons we have been interested only in orthogonal rational factorizations. It is nonetheless possible to take an irrational factorization of a rational $P(z)$ function and use it to derive an orthonormal wavelet basis. For example if we take $P(z) = H_0(z)H_0(z^{-1})$ where $H_0(z) = \sqrt{P(z)}$ we end up with linear phase filters. That $H_0(z)$ is necessarily irrational, where one of the $P(z)$ functions designed in this section is used, is guaranteed by Lemma 5.1 below. For example if we use the Butterworth $N = 7$ case, as in example 4.1, we get the wavelet and scaling function shown in Figure 6. The magnitude spectra plots are of course identical to those in Figure 2, since these are independent of the factorization chosen. It is worth pointing out that the wavelet is very similar to an orthonormal wavelet constructed by Meyer, but based on irrational filters [5].

Clearly Theorem 3.2 generates all orthogonal filter banks where $P(z)$ is rational, even if the filters themselves are not so.

5 Linear phase orthogonal IIR solutions

In [3, 31, 4] it was pointed out that it is not possible to generate a nontrivial basis of real finite length wavelets which are orthonormal and symmetric. In fact the only solution is the Haar basis, which is not continuous. If we were prepared to consider complex FIR filters it would be possible [36], but filters with complex coefficients are not generally of interest.

We have not thus far addressed the possibility of achieving linear phase with orthogonal rational IIR filters. We first consider the possibility that one of the maximally flat $P(z)$ functions already derived might factor $P(z) = H(z)H(z^{-1})$, where $H(z)$ is a rational linear

phase filter. The next lemma proves that this is never possible for the Daubechies, intermediate or Butterworth $P(z)$ functions of any order. In other words if we desire linear phase filters the solutions presented so far will not serve.

After considering once more the structure of orthogonal IIR solutions however, we see how the linear phase condition can be structurally imposed, and use this to generate designs. While the filters never have as many zeros at $z = -1$ as those of section 4, they give wavelets that are very smooth. This result was presented in preliminary form in [15, 16].

5.1 Structure of linear phase orthogonal solutions

We first show that none of the particular orthogonal solutions presented so far can be used if rational filters are required.

Lemma 5.1 *The Daubechies, intermediate and Butterworth solutions to the equation:*

$$P(z) + P(-z) = 2,$$

can never be factored $P(z) = H(z)H(z^{-1})$ where $H(z)$ is a rational linear phase filter.

The proof is in appendix A.2.

The above result is not unexpected: all of these designs were found by merely ensuring that Lemma 2.2(c) was satisfied. If we wish in addition to guarantee linear phase we shall have to impose this structurally before we begin the design. We find it more convenient to work with the equivalent condition Lemma 2.2(b). We first recall an important preliminary result on the structure of orthogonal polyphase matrices [3, 26].

Lemma 5.2 *An orthogonal polyphase matrix is necessarily of the form:*

$$\mathbf{H}_p(z) = \begin{bmatrix} H_{00}(z) & H_{01}(z) \\ -H_{01}(z^{-1})\Delta_p(z) & H_{00}(z^{-1})\Delta_p(z) \end{bmatrix}, \tag{24}$$

where:

$$H_{00}(z)H_{00}(z^{-1}) + H_{01}(z)H_{01}(z^{-1}) = 1 = \Delta_p(z)\Delta_p(z^{-1}), \tag{25}$$

and $\Delta_p(z) = \det \mathbf{H}_p(z)$ is an allpass function.

Proof: Lemma 2.2(b) gives immediately:

$$\begin{bmatrix} H_{00}(z^{-1}) & H_{10}(z^{-1}) \\ H_{01}(z^{-1}) & H_{11}(z^{-1}) \end{bmatrix} = \frac{1}{\Delta_p(z)} \begin{bmatrix} H_{11}(z) & -H_{01}(z) \\ -H_{10}(z) & H_{00}(z) \end{bmatrix}$$

which leads to:

$$H_{11}(z) = H_{00}(z^{-1})\Delta_p(z) = H_{00}(z^{-1})/\Delta_p(z^{-1})$$

from which follows that $\Delta_p(z^{-1}) = [\Delta_p(z)]^{-1}$, that is, $\Delta_p(z)$ is an allpass filter [26]. Also:

$$H_{10}(z) = -H_{01}(z^{-1})\Delta_p(z). \square$$

We have seen before that linear phase filters are of two types, those that have half sample symmetry or antisymmetry (HSS or HSA) and those that have whole sample symmetry or antisymmetry (WSS or WSA); again we find it convenient to treat them separately.

5.2 Half sample symmetric case

If linear phase filters are half sample symmetric or antisymmetric then the polyphase components are related as in fact A.1. We can use this to force linear phase on the polyphase filter matrix (24).

Lemma 5.3 *In an orthogonal filter bank, where the filters are half sample symmetric, it is necessary and sufficient that the polyphase matrix be of the form:*

$$\mathbf{H}_p(z) = \begin{bmatrix} A(z) & z^{-l}A(z^{-1}) \\ -z^{l-n}A(z) & z^{-n}A(z^{-1}) \end{bmatrix}, \tag{26}$$

where $A(z)A(z^{-1}) = 1$.

Proof: One of the filters must be HSS while the other is HSA, since these always have at least one zero each at $z = -1$ and $z = 1$ respectively, and, because of (25), the filters must have no zeros in common.

Hence if $H_0(z) = H_{00}(z^2) + z^{-1}H_{01}(z^2)$ is HSS then $H_{00}(z) = z^l H_{01}(z^{-1})$ for some l. Similarly $H_{10}(z) = -z^m H_{11}(z^{-1})$ for some m. The HSS polyphase matrix is

$$\mathbf{H}_p(z) = \begin{bmatrix} H_{00}(z) & z^{-l}H_{00}(z^{-1}) \\ H_{10}(z) & -z^{-m}H_{10}(z^{-1}) \end{bmatrix}. \tag{27}$$

On equating (24) and (27) we get $H_{01}(z) = z^{-l}H_{00}(z)$, $H_{10}(z) = -H_{00}(z)\Delta_p(z)z^l$, and:

$$-z^{-m}H_{10}(z^{-1}) = z^{-m-l}H_{00}(z^{-1})\Delta_p(z^{-1}) = H_{00}(z^{-1})\Delta_p(z).$$

Now the fact that $\Delta_p(z) = 1/\Delta_p(z^{-1})$ gives $\Delta_p^2(z) = z^{-m-l}$, so that $\Delta_p(z)$ is a delay z^{-n}, and $2n = m + l$. This is the desired result. □

For example choosing $l = n = 0$, we get:

$$H_0(z) = A(z^2) + z^{-1}A(z^{-2}) \tag{28}$$
$$H_1(z) = -A(z^2) + z^{-1}A(z^{-2}) \tag{29}$$

In order to force some regularity we might wish to design $H_0(z)$ to have again the maximum possible number of zeros at $z = -1$. This can be done by solving a fairly simple set of nonlinear equations. Taking the filters in (28) and (29) and the simple allpass section:

$$A(z) = \frac{1 + az^{-1} + bz^{-2}}{b + az^{-1} + z^{-2}}$$

with $a = 6$, $b = 15/7$ we get that $H_0(z)$ contains five zeros at $z = -1$, has a reasonable lowpass response and gives a wavelet that is very smooth. The wavelet and its spectrum are shown in Figure 7.

5.3 Whole sample symmetric case

Next suppose $H_0(z)$ is to be whole sample symmetric (WSS). In this case one of the polyphase components must be half sample symmetric, the other whole sample symmetric, and both must be either symmetric or antisymmetric. Since antisymmetric filters always have a zero at $z = 1$ the latter case can never satisfy (25).

It is also implied by (25) that the denominators of $H_{00}(z)$ and $H_{01}(z)$ are equal, so we must solve:

$$N_{00}(z)N_{00}(z^{-1}) + N_{01}(z)N_{01}(z^{-1}) = D(z)D(z^{-1}),$$

where $N_{00}(z)$ and $N_{01}(z)$ are the numerators and $D(z)$ is the common denominator.

Since a rational IIR filter is symmetric if and only if both numerator and denominator are, we need consider only the symmetry of $N_{00}(z)$, $N_{01}(z)$ and $D(z)$. There are four cases that give that $H_{00}(z)$ and $H_{01}(z)$ have the whole/half sample symmetries described above. One can verify that these are that $D(z)$, $N_{00}(z)$ and $N_{01}(z)$ are all symmetric and have lengths that are respectively (odd, odd, even), (odd, even, odd), (even, even, odd) and (even, odd, even). The last two cases, where $D(z)$ has even length, are immediately ruled out, since a symmetric even length FIR function implies at least one zero on the unit circle.

For example for the (odd, odd, even) case $H_{00}(z)$ has whole sample symmetry, $H_{01}(z)$ has half sample symmetry, and the polyphase matrix is lossless and gives filters that have whole sample symmetry.

Finding good solutions is not as easy as in ths HSS case, since the method is not constructive. However examples can be constructed by solving a set of nonlinear equations. Consider the small example: $N_{00}(z) = a + bz^{-1} + az^{-2}$, $N_{01}(z) = c + cz^{-1}$, and $D(z) = a + dz^{-1} + az^{-2}$. The values $(a, b, c, d) = ((5 + 4\sqrt{2})/14, 1, (12 + 4\sqrt{2})/14, (21 + 24\sqrt{2} + 16 \cdot 2^{3/2})/49)$ gives a solution such that the lowpass filter $H_0(z)$ has two zeros at $z = -1$.An estimate of it's regularity gives $r > 0.5$.

6 Orthogonalization of wavelet bases

One of the interesting wavelet bases is that derived by Battle and Lemarié [6, 7], which has the property of being a basis for the spline function spaces.

The B-spline functions obviously form a basis for this space, but are not orthogonal with respect to integer shifts; in the language of section 2.1.1 we have a basis for V_0, but not an orthogonal one. The condition for orthogonality can also be written in the Fourier domain using the Poisson summation [20, 25]:

$$< \phi(x), \phi(x-n) > = \delta_n \quad \Leftrightarrow \quad \sum_{k=-\infty}^{\infty} |\Phi(w + 2\pi k)|^2 = 1. \tag{30}$$

Now assume that we have a non-orthogonal basis for a multiresolution analysis, given by a function $g(x)$ and it's integer translates. Then it is easy to see one way that the orthogonalization of the non-orthogonal basis $g(x-n)$ may be performed in the Fourier domain:

$$\Phi(w) = \frac{G(w)}{\sqrt{\sum_{k=-\infty}^{\infty} |G(w + 2\pi k)|^2}}. \tag{31}$$

Clearly $\Phi(w)$ satisfies the Fourier domain orthogonality condition (30), and the rest of the multiresolution analysis machinery follows; this is precisely the procedure followed by Battle

and Lemarié [6, 7]. The sequence $h_0(n)$ associated with the two scale difference equation (2) for $\phi(x)$ is not given by a rational function however. Hence the filter bank implementation is not realizable, by which we mean that there is no finite complexity recursive implementation of the filters. Often for such non-realizable filters a truncated version of the infinite impulse response is taken, so that an approximate FIR implementation is used; see for example [1].

6.1 Orthogonalizing continuous-time bases with recursive filters

Of course there are many different orthonormal bases that span the same space; the ones derived by Battle and Lemarié are by no means the only ones for the spline spaces. We next show that if there is a compactly supported wavelet basis for V_0, then it is always possible to find an orthonormal basis, which is infinitely supported, but for which the filters involved are rational and thus realizable. As a special case we shall construct realizable bases for the spline spaces, which are alternatives to those of Battle and Lemarié.

Theorem 6.1 *If the set $\{g(x-k), k \in Z\}$ forms a non-orthogonal basis for V_0, obeys a two-scale difference equation, and $g(x)$ is compactly supported, then it is always possible to find an orthonormal basis $\{\phi(x-k), k \in Z\}$, where:*

$$\Phi(w) = \prod_{i=1}^{\infty} H_0(e^{jw/2^i}), \tag{32}$$

and where $H_0(e^{jw})$ is a rational function of e^{jw}.

Proof: The proof is constructive. The normalizing function used in (31) (*i.e.* the denominator of the right hand side) is 2π-periodic, and can be written as a discrete-time Fourier transform:

$$\sum_{k=-\infty}^{\infty} |G(w+2\pi k)|^2 = \sum_n c_n e^{-jwn} = C(e^{jw}). \tag{33}$$

It can be shown [20, 25] that the Fourier coefficients are obtained from:

$$c_n = \int_{-\infty}^{\infty} g(x)g(x-n)dx. \tag{34}$$

Since $g(x)$ is compactly supported, it is obvious that only finitely many of the c_n are non-zero. Equally, since the c_n are the Fourier coefficients of a positive real function, it is clear that we can always factor:

$$C(z) = E(z)E(z^{-1}). \tag{35}$$

Note that $C(z)$ cannot have zeros on the unit circle, because of the fact that $g(x-n)$'s form a basis [20].

The choice

$$\Phi(w) = \frac{G(w)}{E(e^{jw})}, \tag{36}$$

clearly satisfies (30); so that the $\phi(x-k)$ are orthogonal. Since $E(e^{jw})$ is 2π-periodic we get:

$$\phi(x) = \sum_k f(k)g(x-k),$$

where $F(e^{jw}) = 1/E(e^{jw})$. That is $\phi(x)$ is a linear combination of shifted versions of $g(x)$; hence the span of $\{\phi(x-k)k \in Z\}$ is also V_0.

So we now have that both the sets $\{g(x-k), k \in Z\}$ and $\{\phi(x-k), k \in Z\}$ form bases for V_0. But $g(x)$ obeys the two scale difference equation (2); so for some $l(n)$:

$$g(x) = 2^{1/2} \cdot \sum_{n=-\infty}^{\infty} l(n) \cdot g(2x-n) \qquad \Rightarrow G(w) = L(e^{jw/2}) \cdot G(e^{jw/2}). \tag{37}$$

However, since the two sets span the same space, we can always write the function $\phi(x)$ as a linear combination of the functions $g(x-k)$:

$$\phi(x) = 2^{1/2} \cdot \sum_{n=-\infty}^{\infty} \alpha(n) \cdot g(x-n), \tag{38}$$

so that by substituting in the expression for $g(x)$ from (37) we get that for some sequence $h_0(n)$:

$$\phi(x) = 2^{1/2} \cdot \sum_{n=-\infty}^{\infty} h_0(n) \cdot \phi(2x-n) \qquad \Rightarrow \Phi(w) = H_0(e^{jw/2}) \cdot \Phi(w/2). \tag{39}$$

Thus $\phi(x)$ satisfies a two scale difference equation also.

Substituting (36) into the Fourier version of (39) we get:

$$\frac{G(w)}{E(e^{jw})} = \frac{H_0(e^{jw/2}) \cdot G(w/2)}{E(e^{jw/2})}.$$

Comparing this with (37) gives the relation:

$$H_0(e^{jw}) = \frac{L(e^{jw}) \cdot E(e^{jw})}{E(e^{j2w})}. \tag{40}$$

Note that $L(e^{jw})$ is an FIR function, since when it is iterated in (37) it gives $g(x)$, which is compactly supported. Equally $E(e^{jw})$ is FIR, since it is one of the factors of $C(e^{jw})$. Hence $H_0(e^{jw})$ is a rational function of e^{jw}, and corresponds to a filter that can be implemented recursively.□

Since $\phi(x)$ gives an orthogonal basis for V_0 we see from section 2.1.1 that $h_0(n)$ and $h_1(n)$, (given by $H_1(z) = z^{2k-1}H_0(-z^{-1})$), satisfy (5) and (4). In other words the conditions of Lemma 2.2(d) hold and we have an orthogonal filter bank, with rational filters, that generates our basis for V_0. It follows from theorem 3.2 that the function ,

$$P(z) = \frac{L(z)L(z^{-1}) \cdot E(z)E(z^{-1})}{E(z^2)E(z^{-2})},$$

is valid.

Note that we do not have to separately consider the convergence of the infinite product implied in (39). Because of (40) successive numerators and denominators of the product cancel:

$$\Phi(w) = \prod_{i=1}^{\infty} H_0(w/2^i) = \prod_{i=1}^{\infty} L(w/2^i) \cdot \prod_{i=1}^{\infty} \frac{E(e^{jw/2^i})}{E(e^{j2w/2^i})}$$
$$= G(w) \cdot \frac{E(e^{jw/2})}{E(e^{jw})} \cdot \frac{E(e^{jw/4})}{E(e^{jw/2})} \cdots = G(w) \cdot \frac{E(e^{j0})}{E(e^{jw})}. \quad (41)$$

So the infinite product for $H_0(e^{jw})$ converges since that for $L(e^{jw})$ does. This also means that we do not have to separately make regularity estimates for $\Phi(w)$ if the regularity of $G(w)$ is known, since $\phi(x)$ is a linear combination of integer shifts of the function $g(x)$, and thus has the same regularity.

6.2 Bases for the spline spaces using recursive filter banks

An application of the above result is to find bases for the spline spaces. First note that the N-th order B-spline function, which is defined by: $g(x) = s(x) * s(x) \cdots s(x)$, where there are N convolutions, and $s(x)$ is the characteristic function of the interval $[0, 1)$, is compactly supported. Further the set $\{g(x-k), k \in Z\}$ is a basis for the N-th order spline function space. To get an orthogonal basis from this we apply theorem 6.1.

Note that the Fourier transform of the B-spline $g(x)$ can be written [20]:

$$G(w) = \prod_{i=1}^{\infty} (1 + e^{-jw2^{-i}})^N.$$

In other words $L(e^{jw}) = (1 + e^{-jw})^N$.

The coefficients of $E(z)E(z^{-1})$ are found from (34), that is by evaluating $< g(x), g(x-n) >$ [20, 37]. Those of $E(z)$ are then obtained by spectral factorization (35). So we end up with:

$$H_0(z) = (1 + z^{-1})^N E(z)/E(z^2). \quad (42)$$

Successive terms in the infinite product cancel, as in (41), and we get

$$\Phi(w) = G(e^{jw}) \cdot \frac{E(e^{j0})}{E(e^{jw})}.$$

Hence:

$$\phi(x) = \sum_{k=-\infty}^{\infty} f(k) g(x-k),$$

where $F(e^{jw}) = 1/E(e^{jw})$ is an all-pole filter, that is $\phi(x)$ is a linear combination of splines.

Finding polynomial solutions $E(z)$ such that $H_0(z)$ in (42) gives an orthogonal filter bank was also done by Strömberg [38]. This solution was also noted by Unser and Aldroubi [39].

That the wavelet and scaling function are indeed splines is most easily seen for the $N = 2$ case where they are piecewise linear. The wavelet and scaling function are shown in Figure 8.

The relation between the wavelet basis proposed here, and those of Battle and Lemarié is readily seen if we consider the associated function $P(z)$:

$$P(z) = \frac{(1+z^{-1})^N(1+z)^N \cdot E(z)E(z^{-1})}{E(z^2)E(z^{-2})}. \tag{43}$$

Observe that if we factor it as $P(z) = \sqrt{P(z)} \cdot \sqrt{P(z)}$ and use $H_0(z) = \sqrt{P(z)}$ in (32) the cancellation property between successive numerators and denominators still holds, and we end up with:

$$\Phi_{BL}(w) = G(w) \cdot \sqrt{\frac{E^2(e^{j0})}{E(e^{jw})E(e^{-jw})}},$$

which is the same as the form in (31) when $E(e^{j0}) = 1$.

In words: the different orthonormal bases here correspond to different factorizations of $P(z)$. Note however that it is not in general true that different orthogonal factorizations of $P(z)$ give rise to wavelet bases that span the same space. For example the choice $H_0(z) = (1+z^{-1})^N E(z)/E(z^{-2})$ gives an orthogonal basis, but we do not get the cancellations in the infinite product, and the wavelets do not span the spline spaces.

It is clear that the filters that generate the Battle-Lemarié wavelets have linear phase; however they are not rational for any order, as is proved by the next lemma.

Lemma 6.2 *The Spline solutions to the equation*

$$P(z) + P(-z) = 2,$$

can never be factored $P(z) = H(z)H(z^{-1})$ where $H(z)$ is a rational linear phase stable filter.

The proof is in Appendix A.2. Note that in this case unstable solutions are possible, *i.e.* where $H(z)$ has poles on the unit circle, whereas no solutions at all were possible for the cases covered in Lemma 5.1.

7 Conclusion

We have examined in detail the structure of orthogonal two channel filter banks, and their relation with orthonormal bases of wavelets. We placed particular stress on filters that have a maximum number of zeros at π; since these maximally flat filters give rise to wavelets that have a large number of disappearing moments and are very smooth. The Daubechies, Butterworth and intermediate solutions were of this form. The filters that were used to realize bases for the spline spaces also had a large, but not maximum, number of zeros at π.

It should also be pointed out that while in this paper we have been interested exclusively with orthogonal filter banks it is of course possible to factor any of the $P(z)$ functions we have presented in a non-orthogonal fashion. This was essentially the procedure followed in [31, 4], where linear phase factorizations of the Daubechies' $P(z)$ were taken. As noted in [40, 4] however it can be difficult in the FIR case to get filters with flat spectra when linear phase is desired. We observe that the problem becomes even worse when IIR filters are involved. In other words it is very difficult to factor any of the IIR $P(z)$'s listed in tables 1 and 2 to obtain linear phase rational filters which still have acceptable response. Of course

it is always possible, as we saw for the Butterworth case in section 4.4 and Battle-Lemarié case in section 6.2, to factor any of these $P(z)$'s in a linear phase orthogonal fashion, but where the filters involved are irrational.

An important consideration that is often encountered in the design of wavelets, or of the filter banks that generate them, is the necessity of satisfying competing design constraints. This makes it necessary to clearly understand whether desired properties are mutually exclusive. For example in designing nontrivial linear phase wavelets it is found necessary to abandon orthogonality [3, 4], or to use filters with complex rather than real coefficients [36], or to abandon compact support and use rational filters (section 5 above) or irrational filters [6, 7, 5]. Table 3 attempts to clarify some of the conflicts by tabulating which of the properties orthogonality, linear phase, FIR, real coefficients and rational transfer function are simultaneously attainable and commenting on the solutions.

A Appendix

A.1 Filters with symmetry

Fact A.1 *For symmetric discrete sequence* $R(z) = R_0(z^2)+z^{-1}R_1(z^2)$ *the following relations between the polyphase components hold:*
(i) $R(z)$ *WSS:* $R_0(z^{-1}) = R_0(z), \quad z^2R_1(z^{-1}) = R_1(z),$
(ii) $R(z)$ *HSS:* $R_1(z^{-1}) = R_0(z),$
(iii) $R(z)$ *WSA:* $R_0(z^{-1}) = -R_0(z), \quad z^2R_1(z^{-1}) = -R_1(z),$
(iv) $R(z)$ *WSS:* $R_1(z^{-1}) = -R_0(z).$

Proof: *(i)*: WSS $\Rightarrow R(z) = R(z^{-1})$. So: $R_0(z^2)+z^{-1}R_1(z^2) = R_0(z^{-2})+zR_1(z^{-2})$. Equating even and odd powers of z^{-1} we find: $R_0(z^{-1}) = R_0(z), \quad z^2R_1(z^{-1}) = R_1(z)$.
(ii): HSS $\Rightarrow R(z) = z^{-1}R(z^{-1})$. So $R_0(z^2) + z^{-1}R_1(z^2) = z^{-1}R_0(z^{-2}) + R_1(z^{-2})$. Equating even and odd powers of z gives $R_1(z^{-1}) = R_0(z)$.
The other properties follow by similar analysis. □

It follows immediately that an HSS filter always has a zero at $z = -1$, and a HSA filter always has one at $z = 1$.

Fact A.2 *For a rational IIR filter that has linear phase, let* N_1 *be the length of the numerator, and* N_2 *the length of the denominator, then if* $N_1 - N_2$ *is odd the filter is WSS or WSA, and if* $N_1 - N_2$ *is even it is HSS or HSA.*

Proof: If $H(z) = N(z)/D(z)$ then:

$$H(e^{jw}) = \left|\frac{N(e^{jw})}{D(e^{jw})}\right| \cdot e^{j(\phi_N(w)-\phi_D(w))},$$

where $\phi_N(w)$ and $\phi_D(w)$ are the phases of the numerator and denominator respectively. Clearly $H(z)$ will have linear phase if and only if both numerator and denominator do.

Since $N(z)$ and $D(z)$ are linear phase FIR functions we have: $N(z) = z^l N(z^{-1})$ where l is even if the N_1 is odd, and l is odd if N_2 is even. Also $D(z) = z^m D(z^{-1})$, with similar constraints for m. Hence

$$H(z) = \frac{N(z)}{D(z)} = z^{l-m} \cdot \frac{N(z^{-1})}{D(z^{-1})}.$$

Now $l-m$ is even if N_1 and N_2 are both even or both odd (*i.e.* N_1-N_2 is even), and is odd otherwise (*i.e.* N_1-N_2 is odd). Using fact A.1 $l-m$ even implies that $H(z)$ is WSS or WSA, and $l-m$ odd implies that it's HSS or HSA. □

A.2 Proof of Lemmas 5.1 and 6.2

Proof: If $H(z)$ is linear phase then:

$$H(z) = \frac{z^k C(z)C(z^{-1})}{D(z^2)D(z^{-2})},$$

for some integer delay z^k, and:

$$P(z) = H(z)H(z^{-1}) = \left[\frac{C(z)C(z^{-1})}{D(z^2)D(z^{-2})}\right]^2.$$

Hence every pole and every zero must be double.
Butterworth case: In the Butterworth case $P(z)$ can be written:

$$P(z) = \frac{(1+z^{-1})^N(1+z)^N}{(z^{-1}+2+z)^N + (-z^{-1}+2-z)^N} = \frac{(1+z^{-1})^{2N}}{(1+z^{-1})^{2N} + (1-z^{-1})^{2N}(-1)^N}.$$

Note that the denominator, $W(z)$, is a polyphase component of $(1+z^{-1})^{2N}$ following Lemma 3.1. If all poles of $P(z)$ are to double we must have:

$$W(z_0) = 0 \Rightarrow \left.\frac{dW(z)}{d(z^{-1})}\right|_{z_0} = 0.$$

But

$$\frac{dW(z)}{d(z^{-1})}| = 2N\cdot(1+z^{-1})^{2N-1} - 2N\cdot(-1)^N\cdot(1-z^{-1})^{2N-1},$$

is a polyphase component of $(1+z^{-1})^{2N-1} = B_0(z^2) + z^{-1}B_1(z^2)$. So the polyphase components of two successive binomials must share a zero. Consider:

$$\begin{aligned}(1+z^{-1})^{2N} &= (1+z^{-1})\cdot(B_0(z^2)+z^{-1}B_1(z^2)) \\ &= (B_0(z^2)+z^{-2}B_1(z^2)) + z^{-1}(B_0(z^2)+B_1(z^2)).\end{aligned}$$

If the polyphase component of $(1+z^{-1})^{2N}$, which is $(B_0(z^2)+z^{-2}B_1(z^2))$, and that of $(1+z^{-1})^{2N-1}$ share a zero, then clearly $B_1(z^2)$ must contain this zero also. This would imply that $B_0(z)$ and $B_1(z)$ are not coprime; this is a contradiction however, the polyphase components of $(1+z^{-1})^N$ are known to be coprime for all N [4].
Intermediate cases: Here we shall make use of fact A.2 to show that the solutions have to be half sample symmetric or antisymmetric if they are to have linear phase, and then show that they do not satisfy the form of Lemma 5.3.
(a) Consider the $N=2k+1$ case: since the numerator and denominator of $P(z)$ have length $4(k+p)+3$ and $4(k-p)+1$ respectively, the numerator and denominator of $H(z)$ should

have lengths $N_1 = 2(k+p)+2$ and $N_2 = 2(k-p)+1$. Hence $N_1 - N_2 = 4p+1$, which is odd, and $H(z)$ must be HSS or HSA by fact A.2. But by Lemma 5.3 for $H(z)$ to be HSS its polyphase components must be allpass filters. Each of the polyphase components have numerator and denominator of lengths $2(k+p)+1$ and $2(k-p)+1$ respectively; hence they cannot be allpass if $p \neq 0$.
(b) Consider $N = 2k$: here $N_1 = 2(k-p+1)+2$ and $N_2 = 2(k-p)+1$, so $N_1 - N_2 = 2(2p-1)+1$ which is again odd. So again $H(z)$ must be HSS or HSA, and the polyphase components must be allpasses. As before examing the lengths of the numerator and denominator of H_{00} and H_{01} rules this out. The lengths are $2(k+p-1)+1$ and $2(k-p)+1$ respectively.
Daubechies case: the filters are always of even length, and hence either HSS or HSA if linear phase. Hence their polyphase components must be allpasses; but since the only FIR allpass is a delay the only solutions are those of length two. This was already proved in [3]. □

Proof of Lemma 6.2 Again all poles and zeros of $P(z)$ must be double if the filters have linear phase. Recall that in this case we require the $P(z)$ with the form given in (43) to be valid. Suppose indeed that every pole and zero were double, then we could write, for some $D(z)$:

$$P(z) = \frac{(1+z^{-1})^{2N} z^N \cdot (D(z)D(z^{-1}))^2}{(D(z^2)D(z^{-2}))^2}. \tag{44}$$

Since the numerator is a polyphase component of the denominator:

$$(1+z^{-1})^{2N} z^N \cdot (D(z)D(z^{-1}))^2 + (1-z^{-1})^{2N}(-z)^N \cdot (D(-z)D(-z^{-1}))^2 = (D(z^2)D(z^{-2}))^2.$$

Evaluate at $z = 1$ to get:

$$2^{2N}(D(1)D(1))^2 = (D(1)D(1))^2,$$

which is clearly a contradiction unless $D(1) = 0$. $D(1) = 0$ however implies poles on the unit circle; for rational solutions to exist they must be unstable. □

Note rational linear phase solutions which are unstable do exist in the spline case. For example when $N = 2$:

$$\begin{aligned} P(z) &= \frac{(1,4,6,4,1)(1,-4,6,-4,1)}{(1,0,-4,0,6,0,-4,0,1)} \\ &= \left[\frac{(1,2,1)(1,-2,1)}{(1,0,-2,0,1)}\right]^2. \end{aligned}$$

The denominator in this case has double roots at $z = 1$ and $z = -1$, as does the numerator.

References

[1] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Recognition and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[2] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 37, pp. 2091–2110, Dec. 1989.

[3] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. XLI, pp. 909–996, 1988.

[4] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. on Signal Proc.*, Sept. 1992. To appear.

[5] Y. Meyer, *Ondelettes*, vol. 1 of *Ondelettes et Opérateurs*. Paris: Hermann, 1990.

[6] G. Battle, "A block spin construction of ondelettes. Part I: Lemarié functions," *Commun. Math. Phys.*, vol. 110, pp. 601–615, 1987.

[7] P. G. Lemarié, "Ondelettes à localisation exponentielle," *J. Math. pures et appl.*, vol. 67, pp. 227–236, 1988.

[8] M. J. T. Smith and T. P. Barnwell III, "Exact reconstruction for tree-structured subband coders," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 34, pp. 434–441, June 1986.

[9] F. Mintzer, "Filters for distortion-free two-band multirate filter banks," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 33, pp. 626–630, June 1985.

[10] P. P. Vaidyanathan and P.-Q. Hoang, "Lattice structures for optimal design and robust implementation of two-band perfect reconstruction QMF banks," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 36, pp. 81–94, Jan. 1988.

[11] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," *Proc. IEEE*, vol. 78, pp. 56–93, Jan. 1990.

[12] M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Proc.*, vol. 10, no. 3, pp. 219–244, 1986.

[13] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$," *Trans. of American Math. Soc.*, vol. 315, pp. 69–87, Sept. 1989.

[14] O. Rioul, "A discrete-time multiresolution theory unifying octave-band filter banks, pyramid and wavelet transforms," *IEEE Trans. Acoust., Speech, Signal Proc.*, 1990. Submitted.

[15] C. Herley and M. Vetterli, "Linear phase wavelets: theory and design," in *Proc. IEEE Int. Conf. ASSP*, (Toronto), pp. 2017–2020, May 1991.

[16] C. Herley and M. Vetterli, "Biorthogonal bases of symmetric compactly supported wavelets," in *Wavelets, Fractals and Fourier Transforms* (M. Farge *et al*, ed.), Oxford University Press, 1991. To appear.

[17] C. Herley and M. Vetterli, "Wavelets generated by IIR filter banks," in *Proc. IEEE Int. Conf. ASSP*, (San Francisco), March 1992. To appear.

[18] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[19] O. Rioul, "A unifying multiresolution theory for the discrete wavelet transform, regular filter banks and pyramid transforms." IEEE Trans. Acoust., Speech, Signal Proc. Submitted 1990.

[20] I. Daubechies, *Ten Lectures on Wavelets.* SIAM, 1992. To appear.

[21] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Proc. Mag.*, Oct. 1991.

[22] S. Mallat, "Multiresolution approximation and wavelets," tech. rep., Dept. Computer and Information Science, University of Pennsylvania, Philadelphia, PA, Sept. 1987.

[23] N. Dyn and D. Levin, "Interpolating subdivision schemes for the generation of curves and surfaces." School of Mathematical Sciences, Tel Aviv University, preprint, 1990.

[24] O. Rioul, "Dyadic up-scaling schemes: simple criteria for regularity," *SIAM J. Math. Anal.*, 1991. Submitted.

[25] C. Herley and M. Vetterli, "Recursive filter banks and wavelets," tech. rep., Columbia University, 1992.

[26] P. P. Vaidyanathan and Z. Doğanata, "The role of lossless systems in modern digital signal processing," *IEEE Trans. Education*, vol. 32, pp. 181–197, Aug. 1989. Special issue on Circuits and Systems.

[27] V. Belevitch, *Classical Network Synthesis.* San Francisco, CA: Holden Day, 1968.

[28] E. Deprettere and P. Dewilde, "Orthogonal cascade realization of real multiport digital filters," *Circuit theory and appl.*, vol. 8, 1980.

[29] A. Fettweis, "Wave digital filters: theory and practice," *Proc. IEEE*, vol. 74, pp. 270–316, Feb 1986.

[30] Z. Doğanata and P. P. Vaidyanathan, "Minimal structures for the implementation of digital rational lossles systems," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 38, pp. 2058–2074, Dec. 1990.

[31] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets." Submitted, 1990.

[32] M. J. Shensa, "The discrete wavelet transform: wedding the à trous and Mallat algorithms," *IEEE Trans. on Signal Proc.*, Oct. 1992. To appear.

[33] M. J. T. Smith, "IIR analysis/synthesis systems," in *Subband Coding of Images* (J. W. Woods, ed.), (Norwell, MA), Kluwer Academic, 1991.

[34] P. G. Lemarié and G. Malgouyres, "Ondelettes sans peine." Unpublished manuscript.

[35] G. Evangelista, "Wavelet transforms and wave digital filters," in *Wavelets and Applications* (Y. Meyer, ed.), pp. 396–412, Springer-Verlag, 1992. Proc. of conference in Marseille, 1989.

[36] W. Lawton, 1990. Personal Communication.

[37] C. K. Chui and J. Z. Wang, "A cardinal spline approach to wavelets," *Proc. Amer. Math. Soc.*, 1991. To appear.

[38] J. O. Strömberg, "A modified Franklin system and higher order spline systems on R^N as unconditional bases for Hardy spaces," in *Proc. of Conf. in honour of A. Zygmund* (W. Beckner *et al*, ed.), pp. 475–493, Wadsworth Mathematics series, 1982.

[39] M. Unser and A. Aldroubi, "Polynomial splines and wavelets: a signal processing perspective," in *Wavelets: a Tutorial* (C. K. Chui, ed.), (San Diego, CA), Academic Press, 1992. To appear.

[40] M. Vetterli and D. Le Gall, "Perfect reconstruction FIR filter banks: some properties and factorizations," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 37, pp. 1057–1071, July 1989.

B Figures

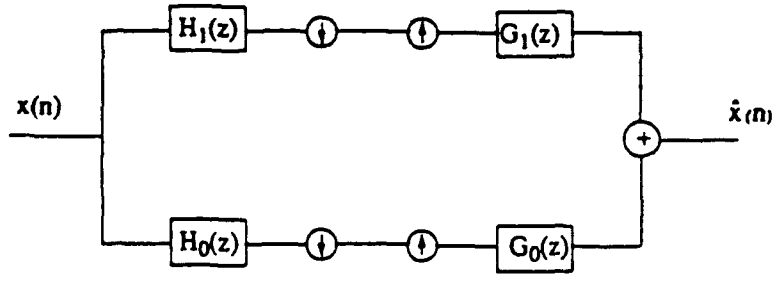


Figure 1: Maximally decimated two channel multirate filter bank.

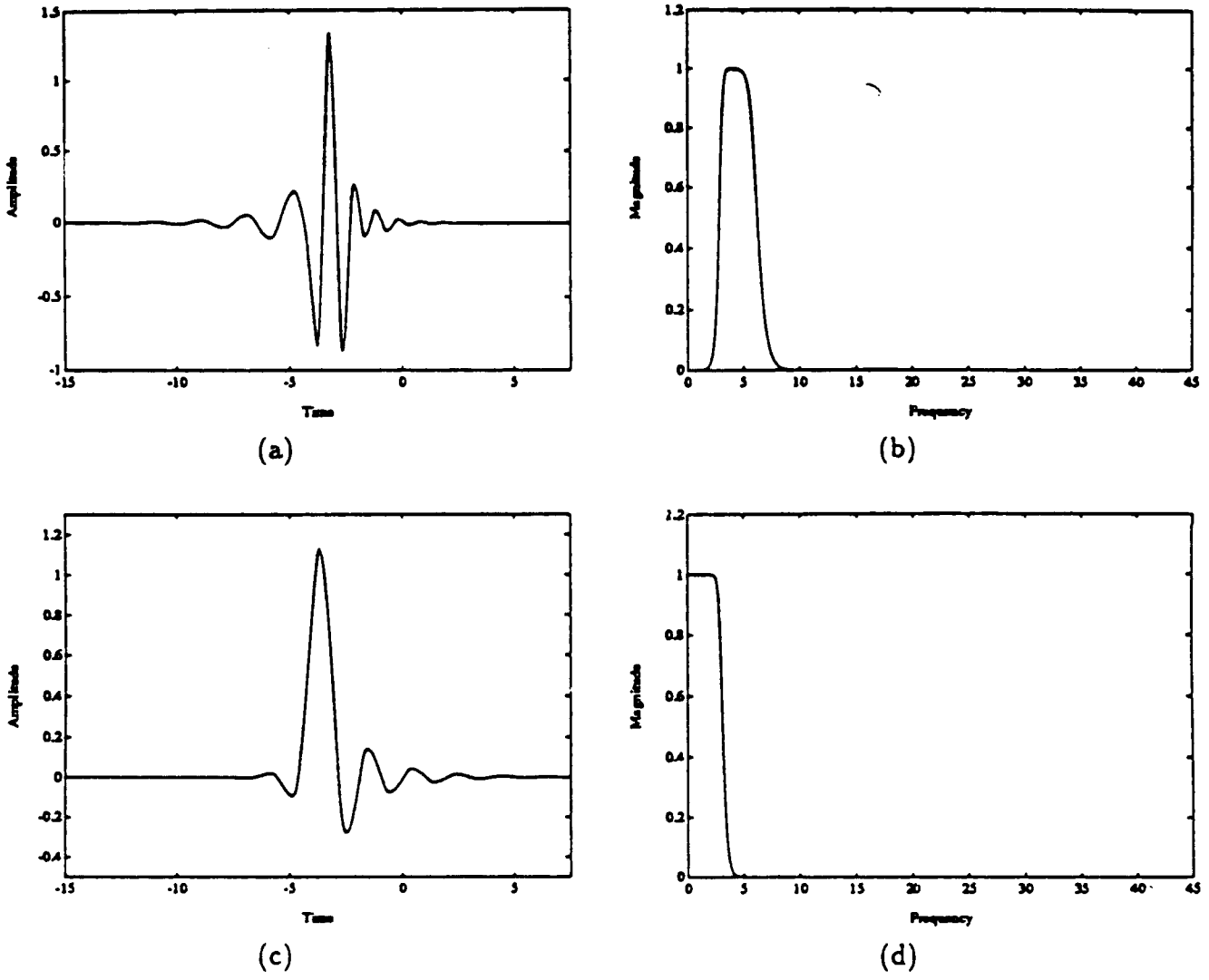


Figure 2: Example of Butterworth orthogonal wavelet; here $N = 7$, and the closed form factorization has been used. (a) The wavelet. (b) Spectrum of the wavelet. (c) Scaling function. (d) Spectrum of the scaling function.

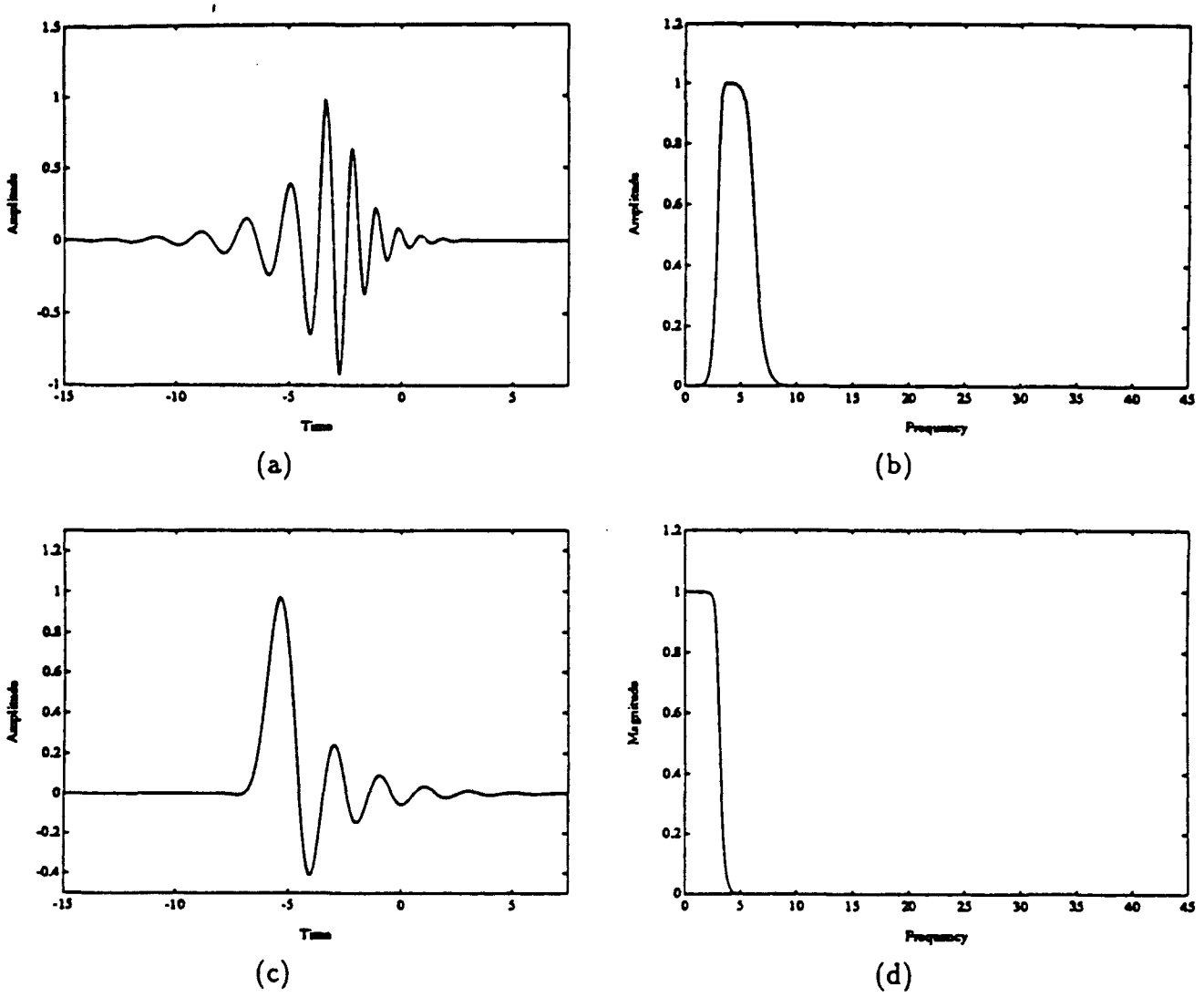


Figure 3: Example of intermediate solution orthogonal wavelet; this is the $N = 7$, $p = 1$ solution. (a) The wavelet. (b) Spectrum of the wavelet. (c) Scaling function. (d) Spectrum of the scaling function.

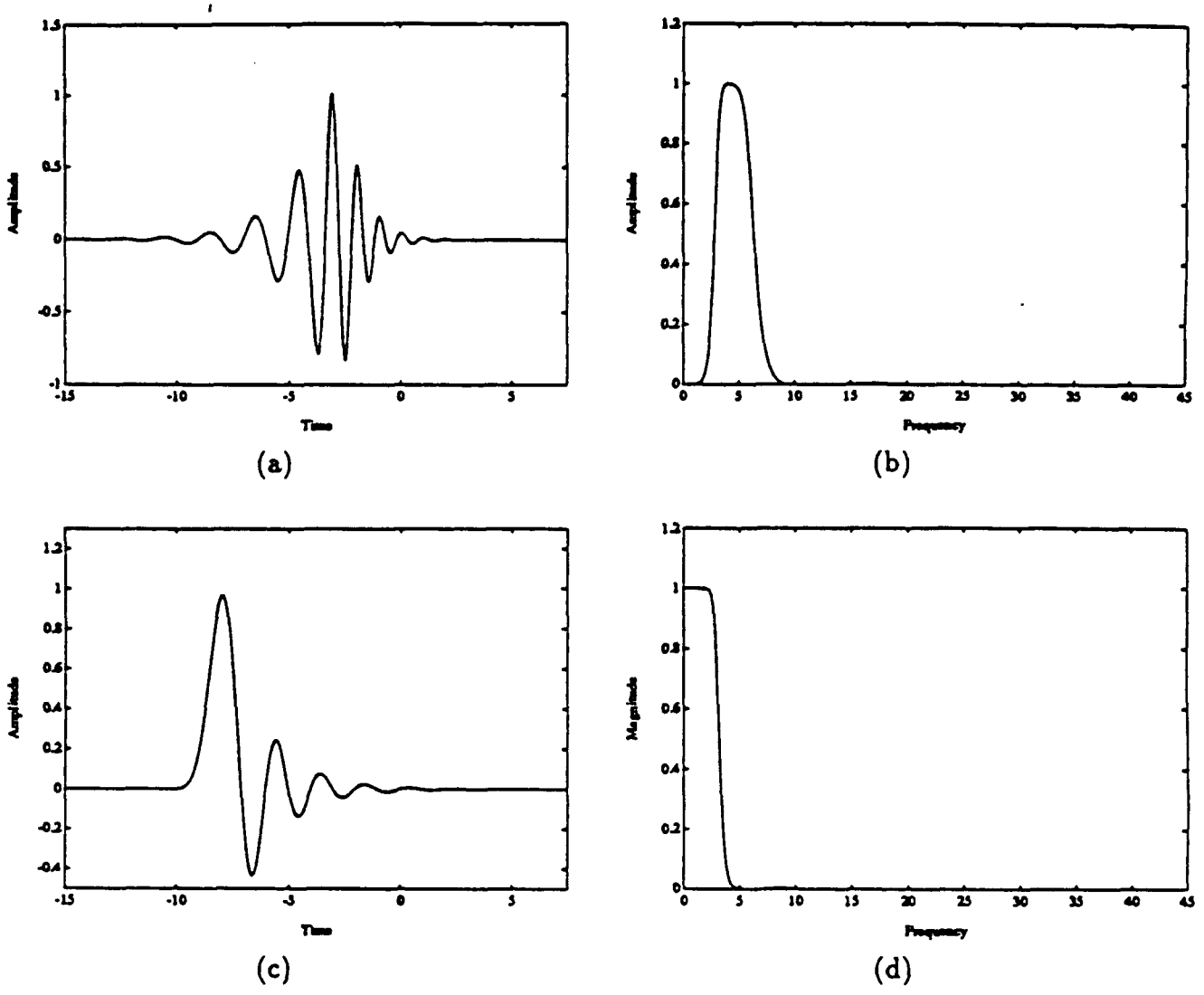


Figure 4: Example of intermediate solution orthogonal wavelet; this is the $N = 7$, $p = 2$ solution. (a) The wavelet. (b) Spectrum of the wavelet. (c) Scaling function. (d) Spectrum of the scaling function.

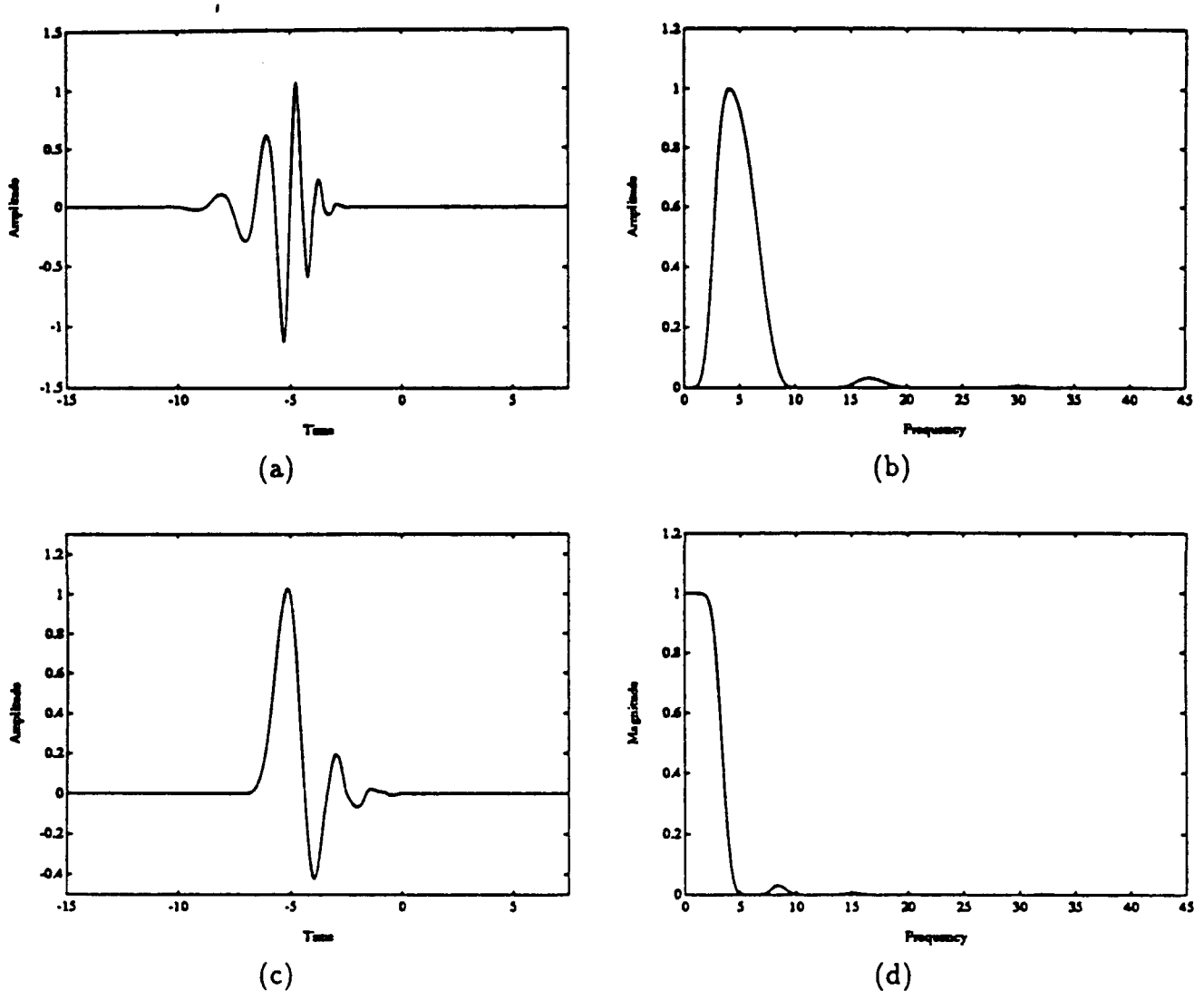
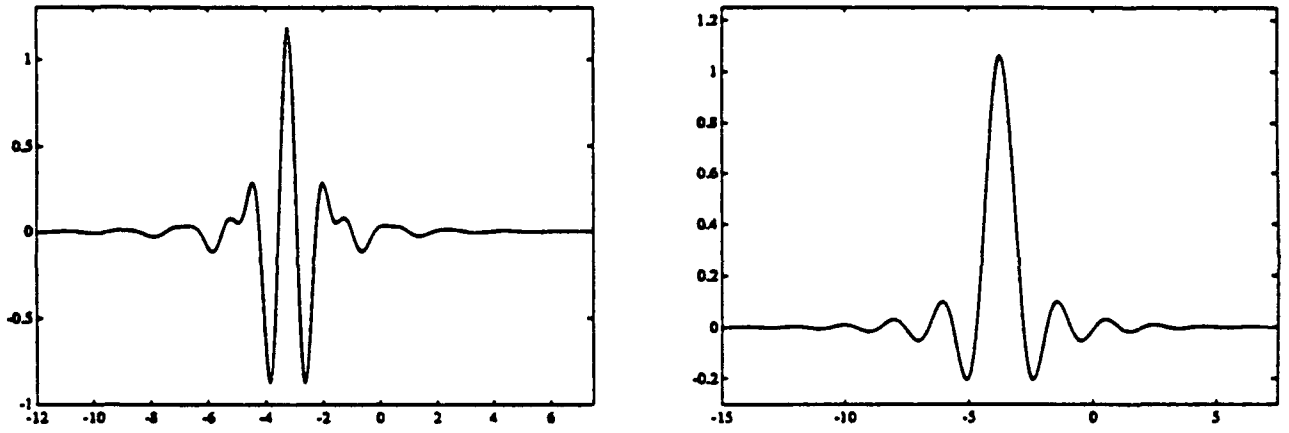


Figure 5: Example of Daubechies orthogonal wavelet; this is the $N = 7$ case. (a) The wavelet. (b) Spectrum of the wavelet. (c) Scaling function. (d) Spectrum of the scaling function.



(a) (b)

Figure 6: Orthogonal basis from irrational factorization of Butterworth case $N = 7$. (a) The wavelet. (b) The scaling function.

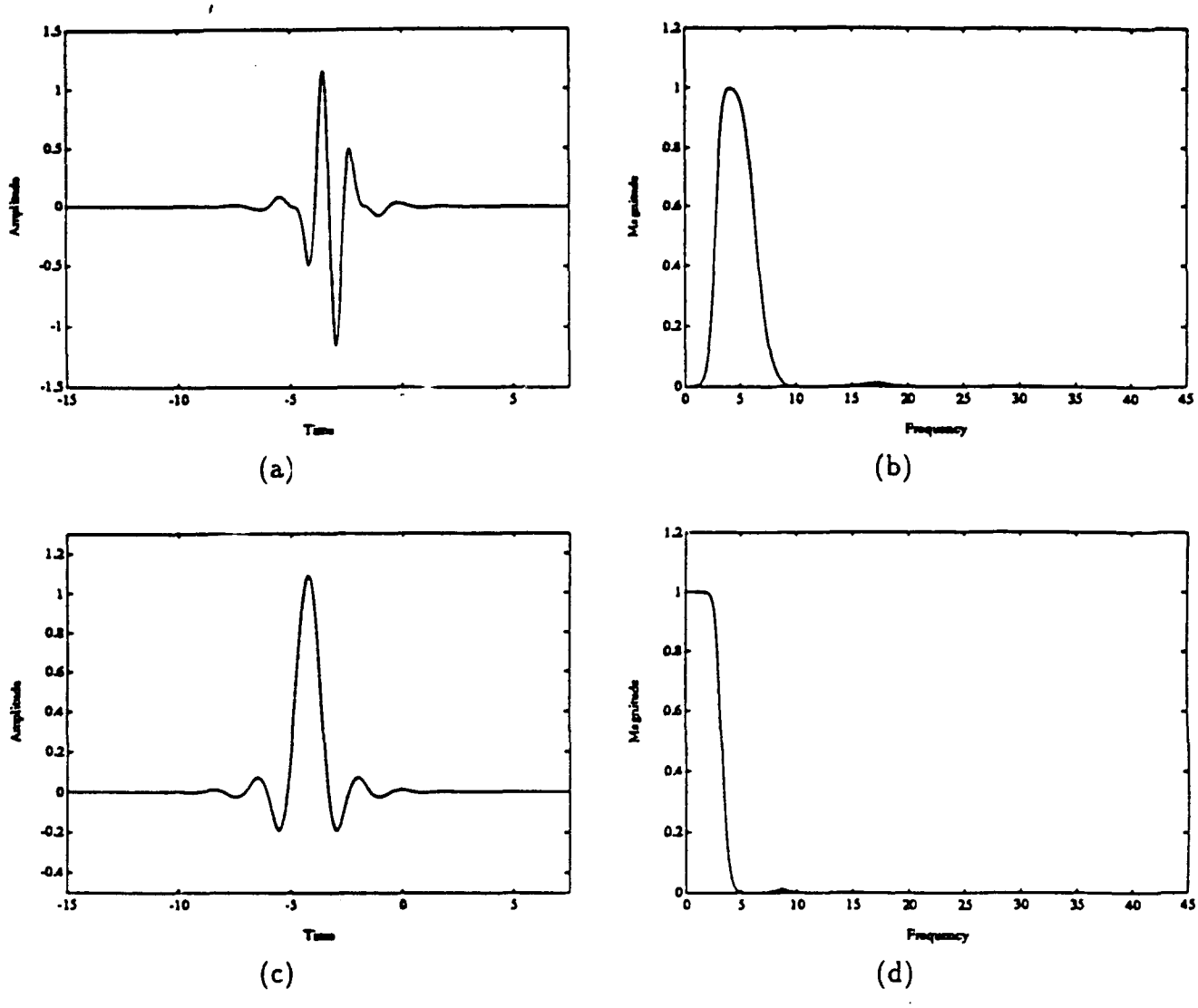
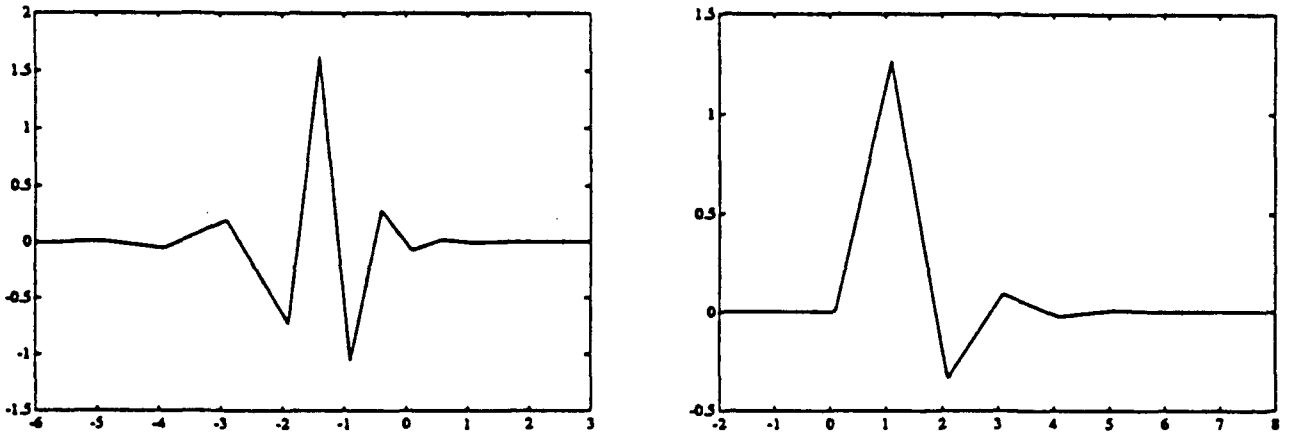


Figure 7: Example of linear phase orthogonal wavelet. (a) The wavelet (antisymmetric). (b) Spectrum of the wavelet. (c) The scaling function (symmetric). (d) Spectrum of the scaling function.



(a) (b)

Figure 8: Orthogonal basis for the piecewise linear spline space constructed from realizable filters. (a) The wavelet. (b) The scaling function.

Solution	$P(z)$	Regularity
$N = 1$		
Haar	$(1+z)(1+z^{-1}) \cdot 2^{-1}$	$r = 0$
$N = 2$		
Daubechies	$(1+z)^2(1+z^{-1})^2 \cdot (-1,4,-1) \cdot z2^{-4}$	$r > 0.5 - \epsilon$
Butterworth	$\frac{(1+z)^2(1+z^{-1})^2 z^{-2}}{(1,0.6,0.1)}$	$r > 0.5$
$N = 3$		
Daubechies	$(1+z)^3(1+z^{-1})^3 \cdot (3,-18,38,-18,3) \cdot z^2 2^{-6}$	$r > 0.9150$
Butterworth	$\frac{(1+z)^3(1+z^{-1})^3 \cdot z^{-2}}{(6,0,20,0,6)}$	$r > 1.0$
$N = 4$		
Daubechies	$(1+z)^4(1+z^{-1})^4 \cdot (-5,40,-131,208,-131,40,-5) \cdot z^3 2^{-12}$	$r > 1.2750$
Intermediate	$\frac{(1+z)^4(1+z^{-1})^4 \cdot (1,-8,1) \cdot z^{-1}}{(160,0,448,0,160)}$	$r > 1.497$
Butterworth	$\frac{(1+z)^4(1+z^{-1})^4 z^{-4}}{(1,0.28,0.70,0.28,0,1)}$	$r > 1.5$
$N = 5$		
Daubechies	$(1+z)^5(1+z^{-1})^5 \cdot$ $(35,-350,1520,-3650,5018,-3650,1520,-350,35) \cdot z^4 2^{-16}$	$r > 1.5960$
Intermediate	$\frac{(1+z)^5(1+z^{-1})^5 \cdot (1,-10,34,-10,1)}{(1792,0,4608,0,1792)}$	$r > 1.9991$
Butterworth	$\frac{(1+z)^5(1+z^{-1})^5 z^{-4}}{(10,0,120,0,252,0,120,0,10)}$	$r > 2.0$
$N = 6$		
Daubechies	$(1+z)^6(1+z^{-1})^6 \cdot z^5 2^{-20} \cdot$ $(-63,756,-4067,12768,-25374,32216,-25374,12768,-4067,756,-63)$	$r > 1.8880$
Intermediate I	$\frac{(1+z)^6(1+z^{-1})^6 \cdot (1,-12,1) \cdot z^{-3}}{(560,0,4928,0,9504,0,4928,0,560)}$	$r > 2.5$
Intermediate II	$\frac{(1+z)^6(1+z^{-1})^6 \cdot (1,-12,58.2,-126.4,58.2,-12,1) \cdot z}{(147456,0,360448,0,147456)}$	$r > 2.476$
Butterworth	$\frac{(1+z)^6(1+z^{-1})^6 z^{-6}}{(1,0.66,0.495,0.924,0.495,0.66,0.1)}$	$r > 2.5$
$N = 7$		
Daubechies	$(1+z)^7(1+z^{-1})^7 \cdot z^6 2^{-24} \cdot$ $(231,-3234,20706,-79674,203161,-356132,430908,$ $-356132,203161,-79674,20706,-3234,231)$	$r > 2.158$
Intermediate I	$\frac{(1+z)^7(1+z^{-1})^7 \cdot (1,-14.66,-14,1) \cdot z}{(10752,0,79872,0,10752)}$	$r > 2.998$
Intermediate II	$\frac{(1+z)^7(1+z^{-1})^7 \cdot (1,-14.592/7,-274.3218/7,-274.592/7,-14,1) \cdot z}{(720896/7,0,1703936/7,0,720896/7)}$	$r > 2.998$
Butterworth	$\frac{(1+z)^7(1+z^{-1})^7 z^{-6}}{(14,0.364,0.2002,0.3432,0.2002,0.364,0.14)}$	$r > 3.0$

Table 1: The various $P(z)$ solutions for a given number of zeros at $z = -1$. Daubechies, intermediate and Butterworth solutions for $N = 1, \cdots 7$ are shown.

Solution	$P(z)$	Regularity
$N = 1$	$(1 + z)(1 + z^{-1}) \cdot 2^{-1}$	$r = 0$
$N = 2$	$\frac{(1+z)^2(1+z^{-1})^2 \cdot (1,4,1) \cdot z^{-1} 2^3}{(1,0,4,0,1)}$	$r = 1.0$
$N = 3$	$\frac{(1+z)^3(1+z^{-1})^3 \cdot (1,26,66,26,1) \cdot z^{-2} 2^5}{(1,0,26,0,66,0,26,0,1)}$	$r = 2.0$
$N = 4$	$\frac{(1+z)^4(1+z^{-1})^4 \cdot (1,120,1191,2416,1191,120,1) \cdot z^{-3} 2^7}{(1,0,120,0,1191,0,2416,0,1191,0,120,0,1)}$	$r = 3.0$
$N = 5$	$\frac{(1+z)^5(1+z^{-1})^5 \cdot (1,502,14608,88234,156190,88234,14608,502,1) \cdot z^{-4} 2^9}{(1,0,502,0,14608,0,88234,0,156190,0,88234,0,14608,0,502,0,1)}$	$r = 4.0$

Table 2: The various $P(z)$ solutions for a given number of zeros at $z = -1$. Spline solutions for $N = 1, \cdots 5$ are shown.

Orthog.	Lin. phase	FIR	Real	Rational	Solutions
1	1	1	1	1	Haar Basis (1910)
0	1	1	1	1	Biorthogonal solutions [31, 4]
1	0	1	1	1	Daubechies [3], FIR paraunitary lattice [10]
1	1	1	0	1	Complex factorizations [36]
1	1	0	1	1	Linear Phase IIR solutions section 5
1	1	0	1	0	Battle-Lemarié bases, Meyer bases
1	0	0	1	1	Characterized by theorem 3.2

Table 3: Properties which are simultaneously achievable for two-channel filter banks, and comments on the solutions. A "1" in a particular box indicates that the solution necessarily has the corresponding property.

Multiresolution Broadcast for Digital HDTV Using Joint Source-Channel Coding

K. Ramchandran*, A. Ortega †, K. M. Uz‡, and M. Vetterli§
Department of Electrical Engineering
and Center for Telecommunications Research
Columbia University, New York, NY 10027-6699

December 17, 1991

Abstract

The use of multiresolution joint source–channel coding in the context of digital terrestrial broadcasting of High Definition Television (HDTV) is shown to be an efficient alternative to traditional single resolution techniques. While the single resolution schemes suffer from a sharp threshold effect in the fringes of the broadcast area, we show how a matched multiresolution approach to both source and channel coding can provide a stepwise graceful degradation. Furthermore, this multiresolution approach improves the behavior, in terms of coverage and robustness of the transmission scheme, over systems that are not specifically designed for broadcast situations. This paper examines the alternatives available for multiresolution transmission, through embedded modulation, possibly trellis-coded to increase coverage range, and error correction codes. We present coding results and simulations of noisy transmission. From a systems point a view, we also discuss the trade-offs involved in the choice of coverage areas for the low and high resolution, as well as the comparative costs and complexities of the different multiresolution transmission alternatives.

*Work supported in part by the New York State Science and Technology Foundation's CAT.

†Work supported in part by the Fulbright Commission and the Ministry of Education of Spain.

‡Work supported by the National Science Foundation under grants ECD-88-11111. K.M.Uz is now with David Sarnoff Research Center in Princeton, NJ 08543

§Work supported in part by the National Science Foundation under grants ECD-88-11111 and MIP-90-14189.

1 Introduction

General discussion of the problem

Recent advances in video compression techniques have spurred interest in the idea of digital HDTV. Even the most demanding delivery mechanism, namely terrestrial broadcast, might turn digital. Digital broadcast differs from digital point–to–point transmission in that different receivers have different channel capacities, i.e. channel capacity decreases with distance from the emitter. Furthermore, in a digital environment, the transition from reliable to unreliable reception is very abrupt, creating the so-called threshold effect. Hence, if digital broadcast is tackled as a single resolution (SR) problem, one would in effect be designing for the fringes of the coverage area, thus reducing the spectral efficiency in areas close to the emitter, as pointed out in [1]. In light of the current interest in digital terrestrial broadcast of HDTV in the U.S., the concern for *spectral efficiency* becomes even more pressing, especially given the conditions set by the FCC in terms of bandwidth allocation.

The approach of designing for the fringes is known from information theory to be suboptimal: when dealing with different channels, one can do better than to transmit only for the worst one, or to perform "naive" time or frequency multiplexing between the different channels! Cover [2] showed that one could trade capacity from the poor channels for more capacity in the better ones, and that the trade-off can in theory be worthwhile. These ideas point out the efficiency of using a multiresolution (MR) approach to digital broadcast. However, to the best of the authors' knowledge, no real end–to–end system has been designed using these results.

We approach this problem as one of joint source and channel coding in a *multiresolution (MR) framework*, extending our work of [3] (see Figure 1 for an example of a two–resolution system). In the two–resolution case, the source is split into "base" information, the coarse channel, and "refinement" information, the fine channel[1]. As in Figure 1, the idea is to match the different resolution levels to different channel capacities, thus creating a *MR channel coding scheme*, so that the receiver closer to the emitter can decode the full quality signal, while the distant receiver has access to the lower resolution quality, providing a stepwise graceful degradation. Furthermore, we show that the use of error concealment in the source decoder of a MR system (see Figure 1) improves the robustness of the full resolution signal, thus increasing the coverage of "indistinguishable quality" delivery over SR schemes.

[1]Note that, throughout this paper, we use "coarse" synonymously with the lower resolution channel and "detail" or "fine" with the refinement or augmentation channel of the two-resolution hierarchy

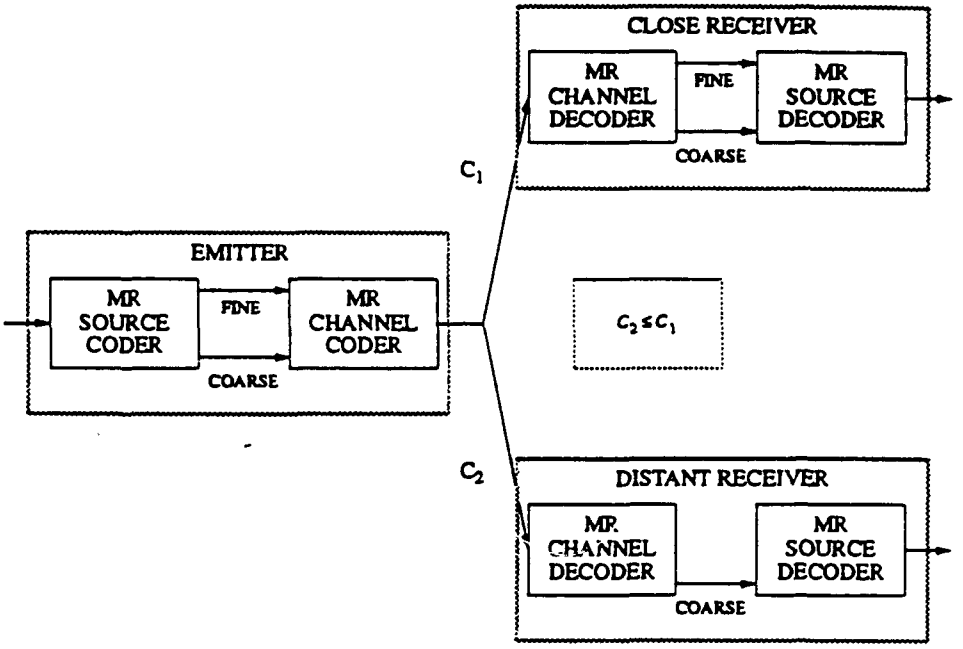


Figure 1: Block diagram of a multiresolution digital broadcasting scheme shown for two receivers with channel capacities C_1 and C_2 with $C_2 \leq C_1$.

We explore the available alternatives to an embedded transmission design and show how MR modulation schemes, combined with trellis coded modulation (TCM) techniques, can be used for this purpose, while pointing out the relative difficulty of designing efficient Error Correction Codes (ECC) to solve the same problem. We consider, in our experiments, a specific high quality MR HDTV coder[4] whose coarse to refinement channel bit rates are in the ratio of 1:2. We assume a spectral efficiency of 6 bits/symbol for our specific example, though, depending on the available broadcast bandwidth, other scenarios may use 3-4 bits/symbol. We evaluate the performance of the system in terms of both coverage area and subjective quality.

Past and current work

Most proposals to the FCC for digital terrestrial broadcast in the U.S. initially approached the problem as one of point–to–point transmission. The idea of graceful degradation, previously proposed as a natural advantage of multiresolution systems [5], has been recently included in the AT&T/Zenith proposal[6], a change from their single resolution scheme advocated earlier[7]. The Sarnoff/NBC/Philips/Thomson[8] proposal includes prioritization in its coding scheme, but does not possess the "embedded" MR transmission to be described in this paper. The idea of efficient multiplexing of the different resolutions of a MR transmission scheme has been studied, using multidimensional constellations, in [9], although a joint source and channel coding design is not addressed. Schreiber has pointed out [1, 10] the problem of spectral efficiency for broadcast, and has proposed a hybrid analog–over–digital scheme, which, though multiresolution in nature, does not fully exploit recent advances in digital compression technology. Note that though several works [11, 12, 13] have considered, in different contexts, the problem of joint source channel coding of images, none has tackled the problem in a broadcast scenario.

Outline of paper

The outline of the paper is as follows. Section 2 presents the digital broadcast problem and suggests a multiresolution formulation. Section 3 reviews MR video coding [14] and summarizes the specific scheme [4] that is used in this paper for the HDTV source coding. Section 4 discusses the idea of MR transmission for broadcast channels. It reviews the classic idea of embedding [2] and shows how it can be applied to digital broadcast. We introduce the concept of *embedded constellations* and show, through a series of examples, how these, possibly combined with Trellis Coded Modulation (TCM) and ECC's, can provide an efficient solution. Section 5 discusses the alternatives and proposes a recipe for the broadcast problem as posed in Section 2. Finally, Section 6 verifies the benefits of using an embedded multiresolution design and illustrates the robustness achievable by

using efficient *error concealment* techniques in a MR coding environment.

2 The digital broadcast problem: a MR formulation

While Shannon [15] established the theoretical optimality of the separation of source coding, or removal of redundancy from a source, from channel coding, or insertion of redundancy to combat a noisy channel, his results hold only in the limit of infinitely complex and long codes, and, more important, for a single channel or point-to-point communication system. For the broadcast or multichannel environment, where a source communicates with a multitude of receivers of varying strengths, as will be explained in detail in section 4, Cover [2] established that optimal broadcast scenarios are multiresolution or embedded in character. This justifies the choice of a multiresolution (MR) source coding scheme to represent a source compactly in a hierarchy of resolutions, to which a "matched" MR transmission can be designed in order to produce an efficient end-to-end design.

2.1 Matched MR source channel coding

While the problem of joint source and channel coding has been addressed previously in various coding contexts, as stated in Section 1, in this paper, we propose the idea of designing an end-to-end joint MR system, i.e. *one which includes a MR channel coding scheme (an analog MR constellation, possibly using a MR Trellis Coded Modulation (TCM), and/or a digital MR ECC)* that is matched to the MR source coding scheme outlined in section 3.

Figure 2 outlines the importance of employing a joint design. For the different receiver Carrier-to-Noise Ratios (CNR's) throughout the broadcast area, the MR digital transmission system (see Figure 2(a)) can reliably deliver different user bit rates.

The idea is to design the MR source and channel coders so that their delivered rates are efficiently matched. The channel rates correspond to the MR modulation scheme, while the source rates refer to the different resolutions of the source coder, whose characteristics are shown in Figure 2(d) [2], resulting in the broadcast characteristics of Figure 2(c).

This paper suggests an efficient way to do this matching. We explain how the MR channel coder curve, which we attempt to match to the MR source coder, can be designed using the concept of embedded transmission, using a modulation parameter λ. Note that

[2]Note that while we use SNR as a source quality measure in this discussion, we do so with the usual disclaimer that while perceptual measures are more meaningful, they are difficult to quantify. Besides, any meaningful measure can be used in place of SNR without changing the nature of the joint source channel coding philosophy we outline here. Also note that SNR is a source quality measure, and CNR is a channel quality measure.

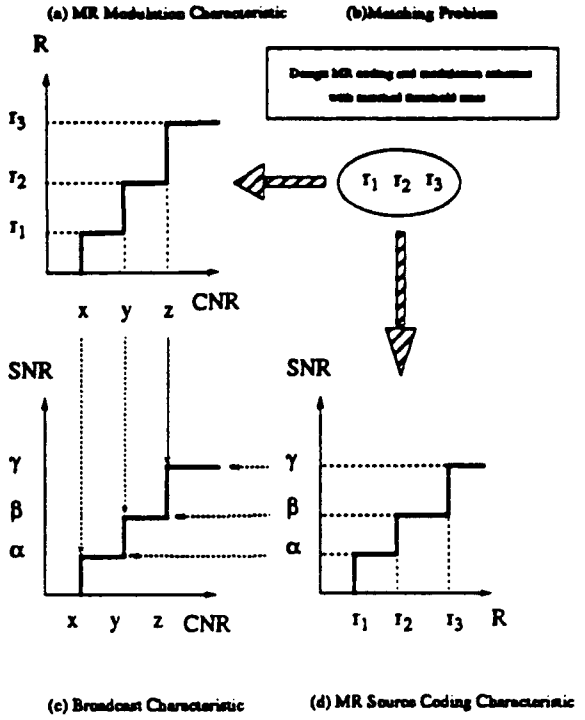


Figure 2: Matching of MR source and channel coders for desired broadcast characteristics. (a) MR channel coder characteristics (Rate vs. CNR). (b) Matching of threshold rates of channel and source coders to achieve desired broadcast characteristics. (c) Achieved broadcast characteristics. (d) MR source coder characteristics (SNR vs. Rate).

while embedded transmission for broadcast is efficient even for a single resolution source, it is even more natural to invoke when the source coder is hierarchical in nature, as is our case, to be described in Section 3.

To solve the matching problem of Figure 2(b), i.e. design the source and channel coders with matched rates, one needs a broadcast performance criterion over which to optimize the parameters. We now address this problem and suggest a tractable formulation.

2.2 The problem of choosing a cost function

The main difficulty in assessing the performance of a digital broadcast system is that of defining a cost function. In other words, one would like to have some way of measuring the performance of a system in terms of, say, the coverage area and the delivered quality, for a given set of resources to be used, such as bandwidth, power, etc. When studying a digital broadcast problem, this measure is not simple: the threshold effect mentioned earlier, simplistically stated, boils down to a trade-off between coverage area and quality

of reception in the case of a single resolution scheme. A multiresolution scheme will face the same trade-off but in a more complex way. For example, in the two resolution case, one can trade-off high quality (full resolution) coverage area for a lower quality (lower resolution) coverage area, as well as the quality of each resolution for a larger coverage area without affecting the area of the other resolution. Would it be better to cover a wide area with relatively low quality or a small area with high quality? The answer is not obvious and points to the lack of a clear cost function for this problem. However, making some assumptions about both the system and the requirements helps us set the system parameters without resorting to a cost analysis.

2.3 Setting the objectives for the system

Assume a two-resolution system. It is reasonable to expect the system to provide the two possible grades of service (full resolution closer to the emitter and a reduced but still acceptable quality further away) for the respective areas defined by distances of d_c and d_f from the emitter ($d_c \geq d_f$). The crucial point is to define what those distances represent in terms of quality. Since different systems will deliver different qualities, it is convenient to define those distances as the maximum distances at which each channel is received reliably (see Figure 3). We can, for instance, equate reliability with the delivered error rate being below a desired threshold. In summary, the system requirements can be set up in terms of providing full-resolution and lower-resolution quality at certain specified distances from the emitter. Now, the source and channel coding have to be chosen so as to guarantee that the required areas are covered, while maximizing the received quality.

Before we address a way of dealing with the stated problem, we analyze the system components, namely the source and channel coders.

3 Multiresolution source coding

Many popular and efficient source coding schemes are either directly or indirectly MR in nature. Methods like subband and wavelet coding have a natural multiresolution interpretation, while others, like DCT based techniques, which represent a common theme in all the digital HDTV proposals to the FCC, have an "acquired" MR interpretation. For a comprehensive review of multiresolution digital coding techniques, the reader is referred to [14].

Multiresolution (MR) source coding schemes can be seen as successive approximation methods. While they can be slightly suboptimal in terms of compression over a single resolution (SR) scheme that achieves the same full resolution quality *for point to*

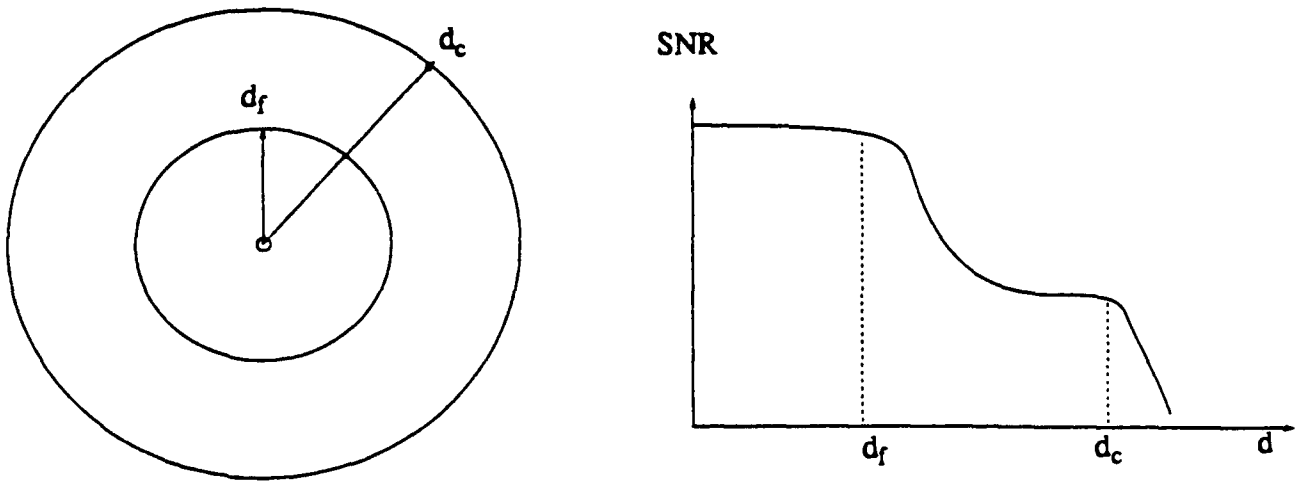


Figure 3: (a) Definition of the coverage area. (b) Distances as a function of the delivered quality

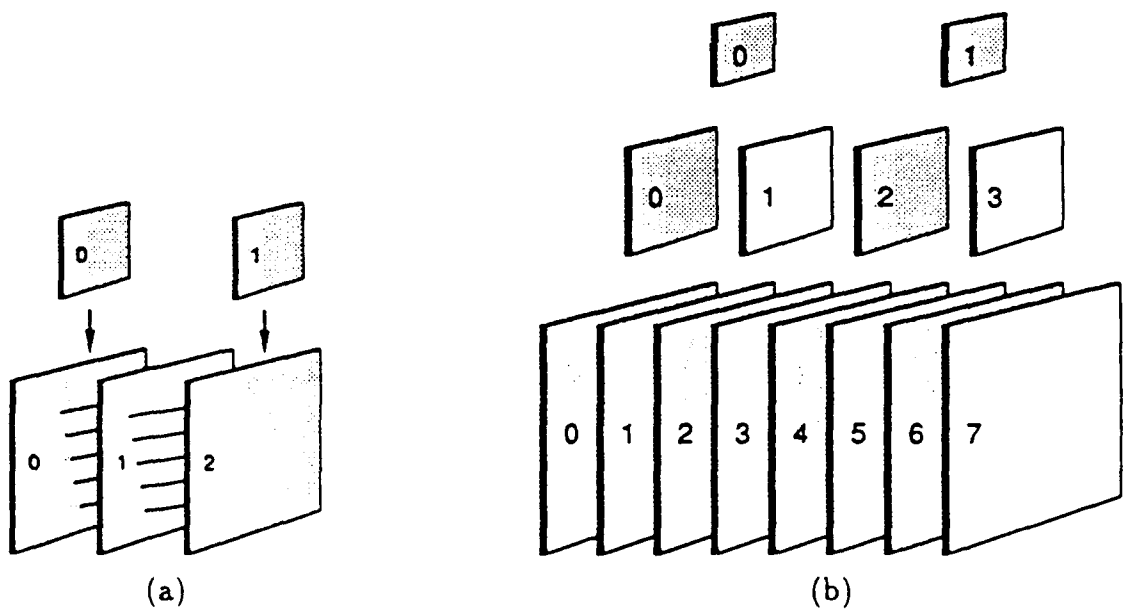


Figure 4: Reconstruction of the pyramid (a) One step of coarse-to-fine scale change (b) The reconstructed pyramid. Note that approximately one half of the frames in the structure (shown as shaded) are spatially coded/interpolated.

point communications, they can be superior in a ***broadcast*** situation, which is a multiuser communications problem. Even for point to point communications, it has been shown theoretically [16, 17] that MR source coding, using a successive refinement approach, can be theoretically optimal under certain conditions, and recently, an efficient practical MR source coder has been suggested [18], that compares very favorably with single resolution approaches that achieve the same full-resolution quality. The advantage of MR based schemes over SR schemes in a broadcast environment comes from the presence, in the former, of a coarse channel (which comes as a "by-product") that, combined with error concealment techniques used at the source decoder, can be used to increase robustness. A more detailed analysis of this robustness issue will be made in section 6. A note of interest, especially in the wake of the ongoing standards and compatibility debates, is that a MR decomposition affords a hierarchy of resolutions that are both natural and useful for the compatibility and broadcast problems.

A specific MR source coding scheme for HDTV

We now review the MR source coder that is an integral part of the joint MR source and channel coding method we undertake in this paper. Our MR video coder [4] is a three-dimensional pyramidal decomposition, based on spatiotemporal interpolation, forming a hierarchy of video signals at increasing temporal and spatial resolutions (see Figure 4 (b)). The structure is formed in a bottom–up manner, starting from the finest resolution, and obtaining a hierarchy of lower resolution versions. Spatially, images are subsampled after anti–aliasing filtering. Temporally, the reduction is achieved by simple frame skipping, since temporal filtering would be inadequate when motion is present (it would produce double images).

The encoding is done in a stepwise fashion, starting at the top layer and working down the pyramid in a series of successive refinement steps. The coarse–to–fine scale change step is illustrated in Figure 4 (a). At each step, first the spatial resolution is increased by linear interpolation, then the temporal motion based interpolation is done based on these new frames at the finer scale. We describe the interpolation procedure only briefly, and refer the reader to [4] for details.

The unshaded frames shown in Figure 4 (b) are interpolated in time. For these frames, the encoder computes a set of motion vectors that are transmitted along with the residual, i.e. the difference between the original and the interpolated frame. The motion vectors are computed in a MR fashion, using a hierarchical blockmatching algorithm [4]. For each block in the interpolated frame, three different motion vector candidates for the following interpolation modes are considered: *backward interpolation*: the motion vector that yields the best replacement from the previous frame; *forward interpolation*: the motion vector

that yields the best replacement from the next frame; ***motion averaged interpolation***: the motion vector **d** that yields the best replacement by averaging the block displaced by **d** in the previous frame and displaced by **-d** in the next frame. The mode that results in the best interpolated block (in the MSE sense) is selected, and the mode selection information is also encoded and transmitted to the receiver.

A discrete cosine transform (DCT) based coder is used to encode the top layer and the subsequent bandpass difference images. Quantizer steps, and consequently bit allocation at different levels in the hierarchy are determined to obtain good perceptual quality. Another major consideration in the bit allocation scheme is in "matching" the subsequent channel coding, to be described later in the paper.

It is important to note that, for the MR source coder we consider in our system, if one resorts to a two-resolution hierarchy comprising the two coarsest layers of the spatio-temporal pyramid in the coarse resolution source channel, and the difference layer in the detail channel, then the bit ratio of coarse to detail information is roughly 1:2 at high perceptual quality for typical sequences. This ratio is more accurate if one considers that the "vital" overhead associated with motion-vectors and synchronization would have to be carried in the lower resolution channel as well. This 1:2 ratio is a key parameter in the development of our joint MR source channel coding system.

4 Multiresolution transmission: embedding

The problem of efficient communication of digital information from a single source to multiple receivers with various Carrier-to-Noise Ratios (CNR's) is key to digital broadcast of HDTV. While the theory of digital broadcast has received attention in early information-theoretical literature [2, 16, 19], there is no evidence of the application of the theoretical maxims proffered in [2] to the design of practical digital broadcast channels. An efficient end-to-end broadcast system should have its *transmission constellation matched to its source coding scheme,* and this is the crux of our work, which we undertake in a *multiresolution environment.*

4.1 Efficiency of using embedding for digital broadcast

Figure 5(a) depicts a typical broadcast environment, with a source wishing to convey information $\{r, s_1\}$ to a stronger receiver and $\{r, s_2\}$ to a weaker one. Note that r represents the common message to be conveyed to both receivers. In [2], Cover establishes the efficiency of superimposing information, i.e. broadcasting in a multiresolution embedded fashion, where the *detailed information meant for the stronger receiver necessarily includes the coarse information meant for the noisier receiver.* The efficiency of embedded

broadcast, in terms of theoretically deliverable bitrate, compared to independent sharing of the broadcast channel resources in time or frequency among the receivers is depicted in Figure 5(b), where the superior curve is obtained by superimposing the detail information within the coarse information. That is, the superior receiver 1, in an optimal scenario, necessarily has access to the information $\{r, s_2\}$ meant for the weaker receiver 2. Note that the plot portrays the potentially deliverable bitrates which are upper bounded by the Shannon capacities of the channels, and has the same drawback of providing no more than existential knowledge, as in Shannon's classical results on channel coding [15]. In this work, we show a practical way of realizing this embedding gain.

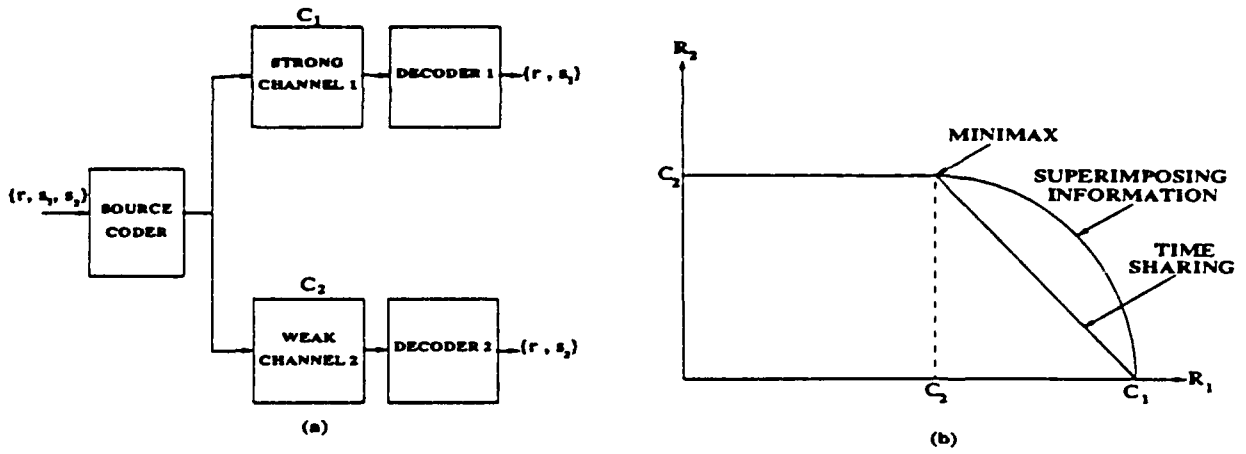


Figure 5: Typical broadcast environment. (a) Single source broadcasting to receivers of channel capacities C_1 and C_2 (b) Set of achievable broadcast bitrates for receivers 1,2.

4.2 Embedding in the modulation domain

Cover's concept of embedding the coarse information within the detailed information is generic in scope, and places no restrictions on the domain in which the embedding should be performed. To describe the effect of an analog domain embedded modulation, we refer to Figure 6 to point out some typical MR embedded modulation constellations. The basic idea is that each constellation consists of "clouds" of mini-constellations or "satellites," where the detail information is represented in the satellites, while the coarse information is carried in the clouds. Thus, the loss of coarse information is associated with the receiver's inability to decipher correctly which cloud was transmitted, while the loss of

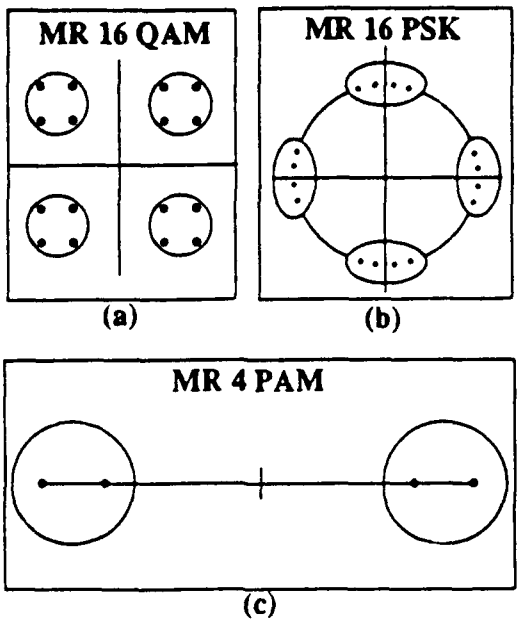


Figure 6: Some multiresolution constellations: (a) MR 16 QAM, (b) MR 16 PSK , (c) MR 4 PAM.

refinement information occurs from the receiver's confusing one intracloud signal point for another. The decoder first decodes the likeliest cloud (coarse information), "subtracts" the decoded cloud value from the received point, and then decodes the likeliest satellite within the cloud (detail information). Thus, the MR 16 QAM constellation of Figure 6(a) has 4 bits/symbol, of which 2 bits are coarse (4 clouds) and 2 bits are detail (4 satellites/cloud). Similarly, the MR 16 PSK scheme has 2 coarse bits/symbol and 2 detail bits/symbol, while the 4 PAM constellation has 1 bit/symbol of each. For our specific source coder, we consider the MR 64 QAM constellation of Figure 7.

While we present a two resolution hierarchy, the principles hold for any number of hierarchical levels desired, and would result in a "fractal" modulation constellation, although at increased complexity and decreased practicality. We point out later how one can combine an embedded ECC scheme with an embedded modulation scheme to increase the number of broadcast resolutions in a practical manner without sacrificing efficiency in the information-theoretical sense.

4.2.1 MR 64-QAM

Consider as Example A the constellation of Figure 7. For every 6 composite bits per channel symbol emitted by the 1:2 source (see Section 3), 2 coarse bits select one of the 4 clouds, while the remaining 4 detail bits select one of the 16 satellites within the

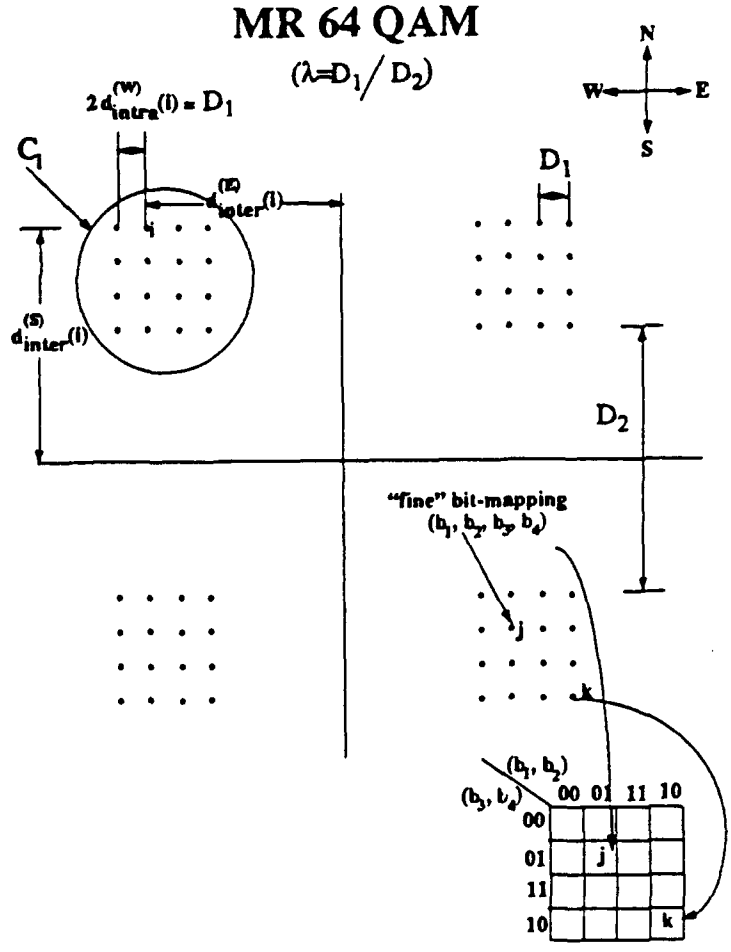


Figure 7: Example A: MR 64 QAM system constellation with definitions of λ, C_i, $d^k_{intra}(i)$, $d^k_{inter}(i)$, and the "fine" bit mapping of the constellation signal points according to the well-known Karnaugh-map partitioning. Note that λ=0 represents uniform 4 QAM, while λ=1 denotes uniform 64 QAM.

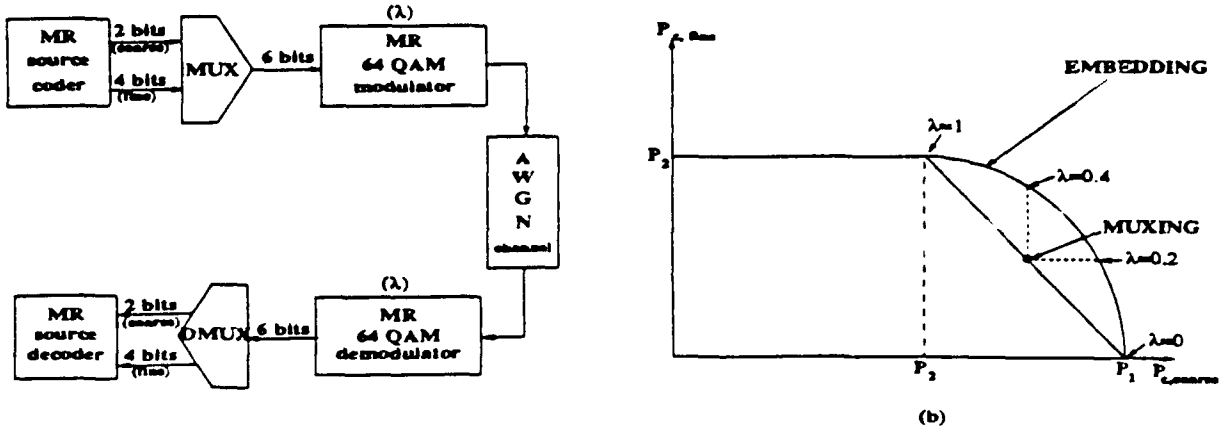


Figure 8: Typical broadcast environment. (a) MR QAM system block diagram (Example A). Note that the modulator and demodulator are operated at transmission parameter λ (see Figure 3). (b) Achievable performance (no packet loss probabilities) for practical system of Fig. 6(a). See analogy with theoretical curve of Fig. 4(b).

selected cloud. By "matching" the relative distances between intra-cloud constellation points (D_1) and inter-cloud points (D_2), whose ratio is a design parameter λ, to the relative "information contents" of the two bitstreams, one obtains an efficiently designed joint MR source/ MR transmission system. One could determine an optimal "broadcast λ" if a meaningful cost function over the broadcast area (which would probably include factors like population density) is available. On the other hand, due to the difficulty of this model, as pointed out in Section 2, one may instead choose, as an operating point, the maximum value of λ that meets the full-resolution coverage range requirement, as will be discussed in Section 5.

The appendix contains the mathematical analysis of the coarse and detail channel performance of the MR 64 QAM of Figure 7, on which the curves shown in Figure 9 are based. Note that those curves reflect packet error rates for the two channels, where a composite packet of length 1080 bits (with 1/3 coarse and 2/3 detail information embedded in it) is used to prevent error propagation.

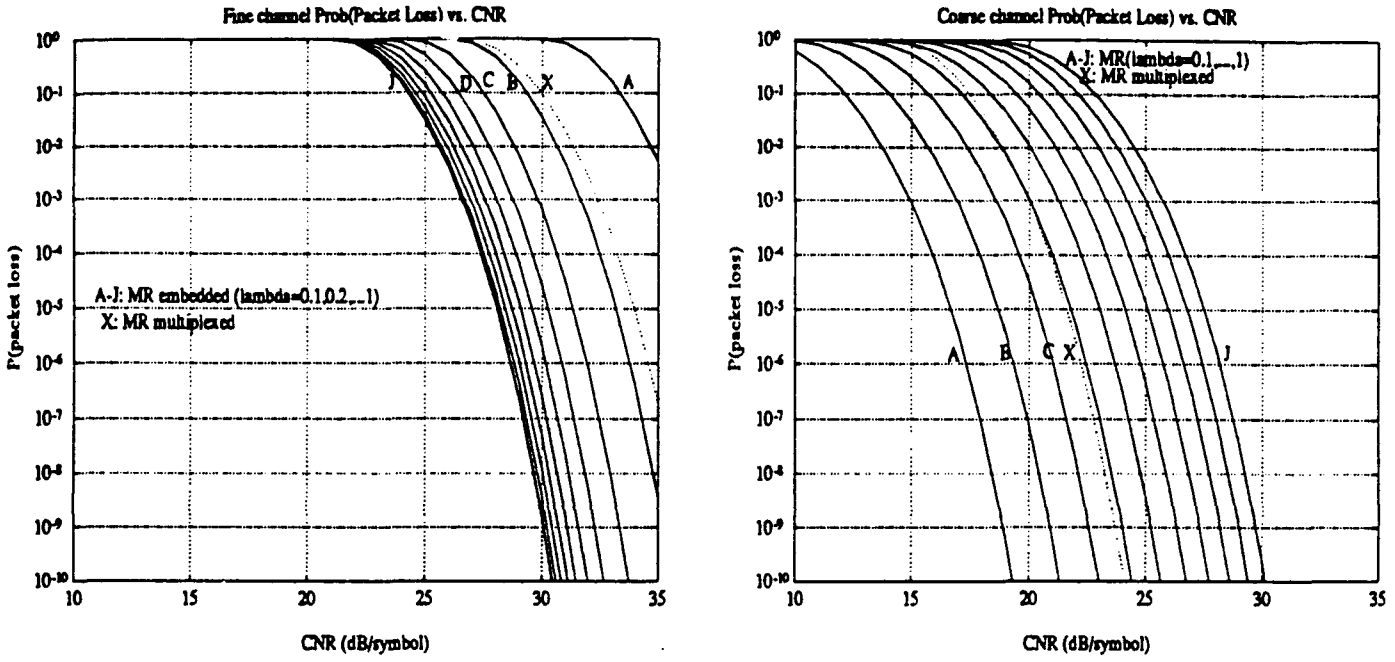
While the details are provided in the appendix, it is important to mention a few salient features. Note that $d^k_{intra}(i)$ and $d^k_{inter}(i)$ represent half the Euclidean distances between signal point i and its nearest coarse and detail neighbors, respectively, in the k-direction. Also, it must be emphasized that the topology of the equivalent constellation

at the broadcast receiver is a function of the CNR and λ. Qualitatively, the CNR affects the "radius" of the constellation as seen at the receiver for a fixed noise variance, while λ affects the relative distances between intercloud and intracloud points. As λ goes from 0 to 1, the intracloud and intercloud thresholds decrease and increase, respectively, for a fixed power budget, indicating the quantitative tradeoffs involved in coarse and detail channel robustness as shown in Figure 9. Also, note that as we can always form a Gray-code fine channel digital bit-mapping of the constellation points exactly as in Karnaugh maps used in digital logic design [20] (see Figure 7), we can guarantee that every point in the constellation is at Hamming distance one away from each of its intracloud nearest neighbors. Thus, assuming that single bit errors dominate when symbol errors occur, we can equate symbol errors with bit errors. This leads to an efficient mapping, besides aiding in the mathematical analysis.

Due to favored protection of the coarse stream via the parameter λ, it is possible for the fine packet component to be corrupted, while the coarse packet component is received reliably *for the same composite packet.* The dotted curves in Figure 9 refer to a "naive" multiplexing of the broadcast channel between the coarse and detail information streams, under conditions of equal power, bandwidth, and average spectral efficiency, as will be explained in Section 5. The curves clearly show the superiority of embedding over multiplexing. For example, for values of λ from 0.2 to about 0.4, *both coarse and detail channel performances are better* than those of the multiplexed case. The particular multiplexing point shown in the figure is obtained when the power in the coarse and detail constellations are made equal, though similar performance improvement can be obtained by embedding over any other multiplexing point also, corresponding to different values of λ. This is a verification of the information theoretical result that embedding outperforms multiplexing. See Figure 8 (b).

4.3 Embedded TCM constellation

In order to increase reliability of reception over the demanding broadcast channel and to increase coverage area, it may be necessary to add more redundancy to protect the broadcast information. As is well known, convolutional codes deploying a Euclidean distance metric can achieve better performance for the same complexity than the more commonly used block ECC's, which use a "hard limiting" Hamming distance metric. Convolutional (trellis) codes achieve coding gain by using soft decoding with the Viterbi algorithm[21]. Conventional convolutional coding, like block coding, would require an increase in bitrate to accommodate the redundant bits, which must come at the expense of lowered source coding quality, for a fixed total throughput. However, it is possible to achieve almost all the coding gain theoretically possible, i.e. to approach the Shannon



(a) (b)

Figure 9: Example A: Probability of packet error vs. Receiver CNR over the entire range of transmission parameter λ for the embedded MR 64 QAM case and a composite packet length of 1080. (a) Fine channel packet loss. (b) Coarse channel packet loss.

limit, by expanding the 2-D modulation constellation by a factor of 2, and deploying a redundant constellation via Trellis Coded Modulation, as established by Ungerboeck [22]. While multi-dimensional TCM [23, 24] can provide the same gain for a smaller expansion factor than the 2-D Ungerboeck constellations, we restrict ourselves to the latter, in the interest of simplicity of design and analysis. *The novelty here is that we combine the concept of multiresolution with the power of TCM to propose an embedded TCM modulation* for efficient broadcast of a MR source (see Figure 11 (a)).

An Ungerboeck TCM scheme requires an expansion factor of 2 in the constellation size. Thus, our original MR 64 QAM constellation would be expanded to 128 QAM, using the same power as the former. Of course, this large constellation size is for our specific example (Example B: see Figure 10): a more practical example for HDTV broadcast might be expansion of a MR 16 QAM scheme (with a 1:1 coarse to detail bitrate ratio, as in [25], using 2 bits/symbol for each resolution) into an embedded TCM 32 QAM, which is certainly practical in size. The principle of operation is what is important. The idea for the TCM 128 QAM scheme (see Figure 11 (a)) is that the coarse information retains preferential protection through λ, while the detail information gets expanded from 16 points to 32 points per cloud via a TCM coding scheme. Figure 11 (a) shows the first level set-partitioning for each 32 point cloud into the subset marked "a" and its complement (unmarked), each subset enjoying a 3 dB gain in squared Euclidean minimum distance over that of its parent, as needed for an Ungerboeck code.

Figure 11 (b) shows the coding gain for the fine channel (the coarse channel remains unchanged) for $\lambda = 0.3$ for trellises with 4, 8, and 16 states. The coding gain over the unexpanded MR-64 QAM constellation is seen to be consistent with that tabulated in [22]. Thus, the simple 4 state trellis is seen to provide a coding gain of 3 dB/symbol in CNR. Identical gains in detail channel protection will occur for any desired value of λ. Thus, for an efficient end-to-end MR design, one may ensure the coarse channel robustness through λ, while using a TCM code of acceptable complexity to achieve the desired full-resolution coverage area. An important feature of our MR system is that due to inclusion of error concealment techniques at the decoder (see Section 6.1), it is possible to obtain indistinguishable full-resolution quality even at a fine-channel packet loss rate exceeding 10^{-1}. As seen from Figure 11(b), at this high loss rate, one gets marginal return from using trellises over 4 states, thus making our TCM design nearly optimal with only a simple 4-state trellis! It is important to note that this scheme permits operation with *no decrease in source coding bit rate* over that of an uncoded system.

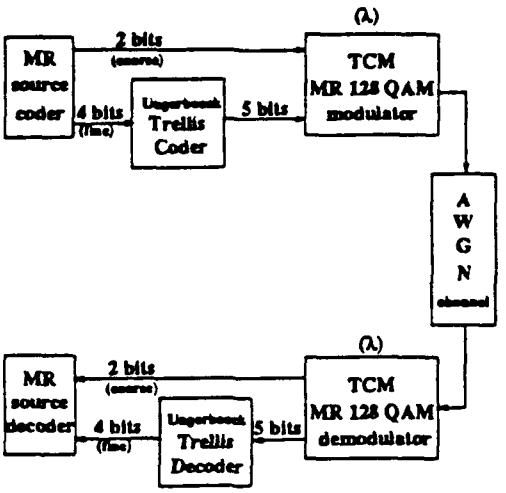
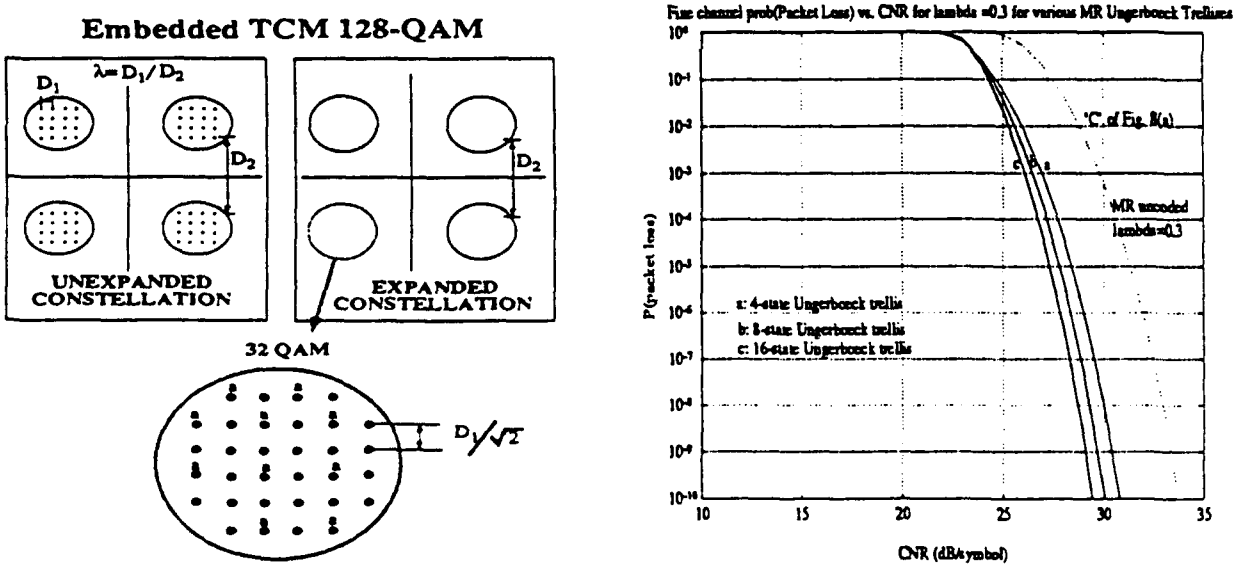


Figure 10: Example B: Block diagram of an embedded MR TCM system using a 128 QAM constellation. Note that it consists of 4 clouds of trellis-code-modulated 32 QAM constellations.



(a) (b)

Figure 11: Example B: (a) Expansion of MR 64 QAM into MR-TCM 128 QAM with an expansion in constellation points of each cloud from 16 to 32. Note that the coarse channel is unaffected. (b) Coding gain over MR 64 QAM for the detail channel using MR-TCM 128 QAM for $\lambda = 0.3$.

4.4 Embedding in the ECC domain: UEP codes

While the effect of providing unequal degrees of robustness for the coarse and refinement channels via the analog parameter λ was discussed above, the algebraic coding expert may argue correctly that one can achieve similar results by using digital techniques like error correcting codes with unequal error protection (UEP) [26, 27, 28, 29, 13]. While the TCM constellation mentioned earlier is indeed efficient, there may be practical limitations to expanding the constellation size. Moreover, one may need ECC's to "bridge" any mismatches in rate between the source coder and the channel modulator (see Figure 2).

It can be seen that embedding in the modulation and ECC domains are essentially equivalent. In the ECC domain, codewords of length n in $(GF(2))^n$ are clustered into "clouds" whose members ("satellites") are closer in Hamming distance with respect to one another, than to members of other clouds. Codes having this behavior would then be called *Unequal Error Protection* (UEP) codes.

A UEP code can be described as an (n, k_1, k_2, t_1, t_2) code (where t_i represents the number of channel errors the code can withstand for the information k_i). It has to be noted that using a UEP code is by no means the only way to provide unequal error protection. As a first approach, one could use two different codes for each category of information, but it is essential to note that embedding the codes can yield better (in terms of the rate: k/n) codes than using two separate codes. In other words, combining two (n_1, k_1, t_1) and (n_2, k_2, t_2) codes to obtain a $(n_1+n_2, k_1, k_2, t_1, t_2)$ code can potentially be outperformed by a (n, k_1, k_2, t_1, t_2) embedded code. As an example, consider a (63,12,24,5,3) binary cyclic UEP code, listed in [30]. Alternatively, one can consider two smaller BCH codes, with characteristics (31,11,5) and (31,12,3). The BCH codes can provide the same protection but clearly their rates are worse than those obtained with the embedded code. To further the analogy with the modulation domain, the use of different codes for the different classes of information (as in [11]) can be likened to the "naive" multiplexing for transmission for the two user broadcast channel.

The advantages, in terms of rate, of using embedded UEP codes are clear, but they come at a high price. Indeed, UEP codes require, in their design, a comparatively much higher effort than the usual single resolution codes. Substantial work has been done in designing the UEP codes and, in particular, on providing bounds for the attainable rates for codes, specially linear UEP codes (LUEP), having these properties. However, no structured method, that does not require brute force computer search, has been described to design these codes. See Lin et. al. [30] for a tabulation of all possible embedded ECC's of odd lengths up to 65. The codes listed in [30] are not appropriate for the application considered in that, of those codes with ratio of coarse to detail information (k_2/k_1) close to 2, few are efficient (i.e. with rates, $(k_1 + k_2)/n$, close enough to 1). Figure 12 (Example

C) presents the block diagram of a scheme that uses embedded UEP codes, and Figure 13 shows the results, in terms of packet loss, for different CNR's, when several of the codes tabulated in [30] are used.

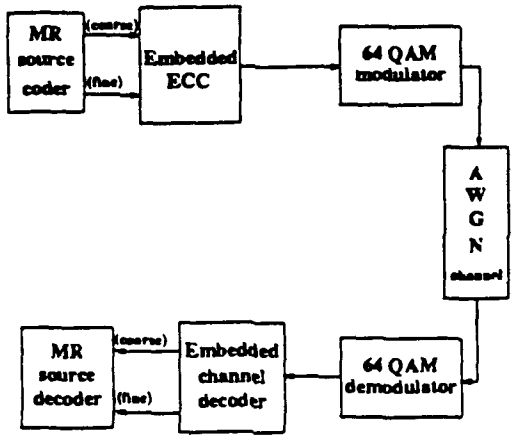
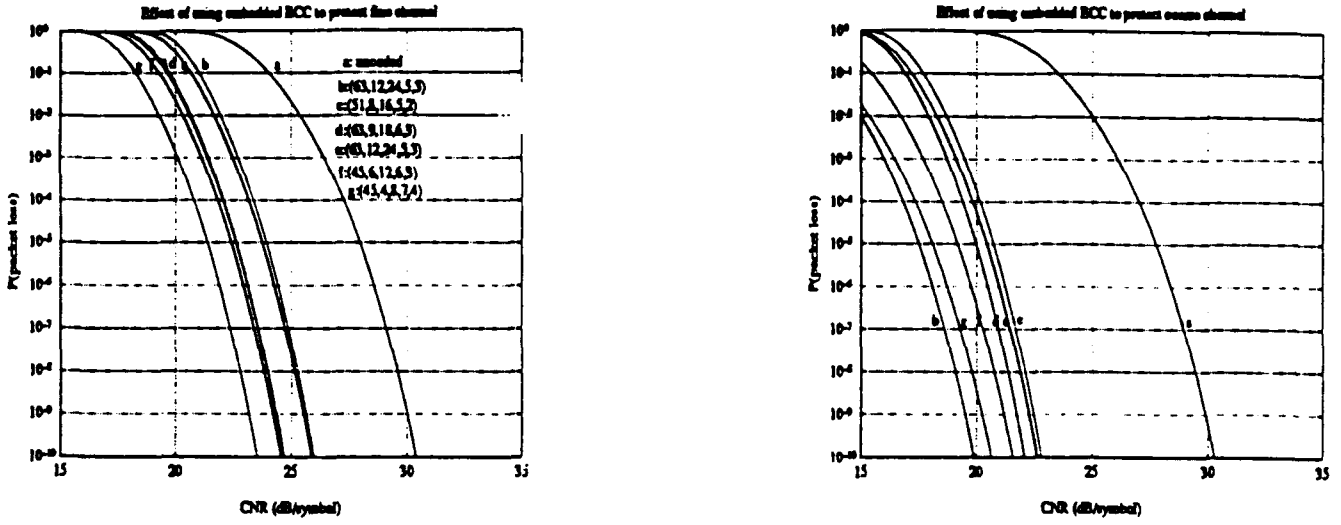


Figure 12: Example C: Block diagram of a MR system with embedded ECC's for the coarse and detail channels.

Thus, while UEP codes can be used instead of MR modulation schemes to perform the MR transmission, the issue of designing good UEP codes is largely open and involves a high degree of complexity. Following the above considerations, for our application, we consider unequally error protected ECC's designed *independently* for the coarse and detail information channels. Using the same coarse packetsize of 360 user bits (k) and various levels of redundancy ($n-k$), we simulated the performance of various (n,k,t) ECC's. (Example D). This example consists of protection of only the coarse channel to varying degrees of robustness, while leaving the detail channel uncoded. See Figure 14 (a). Figure 14 (b) shows how using ECC's lowers the probability of coarse packet loss over the range of CNR's of interest.

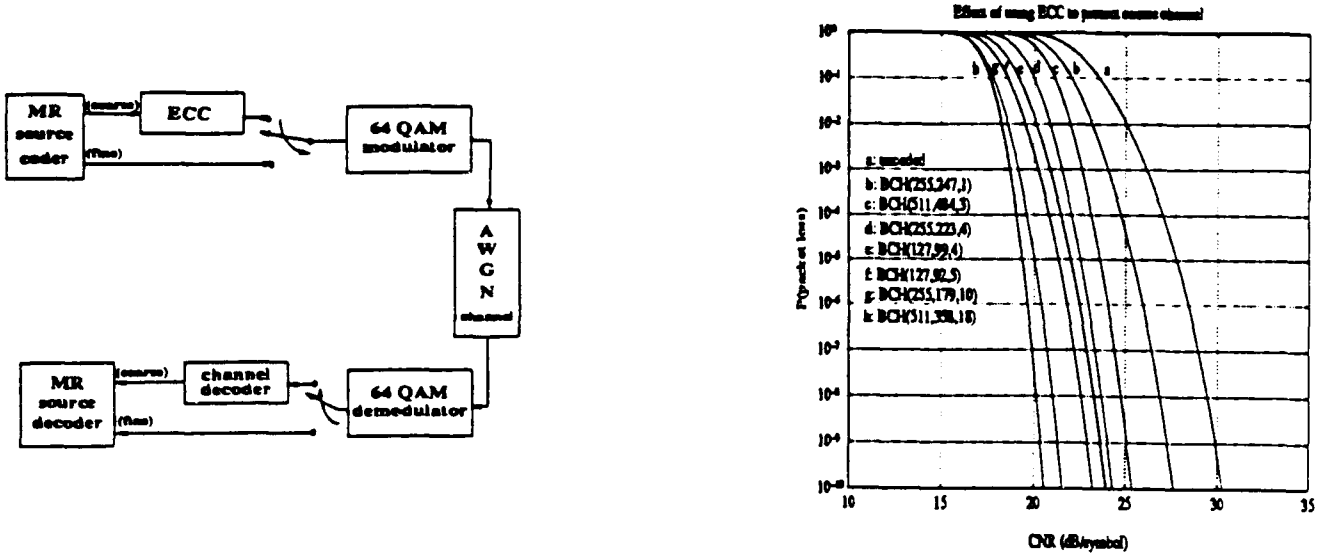
4.5 Hybrid embedded modulation/ECC scheme

It must be noted that an efficient end-to-end system might need to deploy both ECC and MR embedded modulation in tandem to jointly "match" the source coder. For example, one could use a non-uniform QAM scheme rather than the uniform QAM constellation in Figure 14 (a) (Example D). Thus, the ECC scheme could be used as a "bridge" to achieve



(a) (b)

Figure 13: Example C: Probability of packet error vs. Receiver CNR for some known embedded ECC codes. Note that the 5-tuple (n, k_1, k_2, t_1, t_2) listed refers to the embedded code length, the coarse bits per block, the detail bits per block, the error-correction capability for the coarse bits per block, and the error-correction capability for the detail bits per block, resp. The packetsizes of the coarse and detail channels are 360 and 720 bits resp. (a) Fine channel packet loss. (b) Coarse channel packet loss.



(a) (b)

Figure 14: Example D: Multiplexed (non-embedded) ECC using a family of BCH codes. Note that error correction is applied to the coarse channel only, with the detail channel sent unprotected. Note that the 3-tuple (n, k, t) listed refers to the code length, the coarse bits per block, and the error-correction capability for the coarse bits per block. The packet loss rate refers to a coarse packet length of 360 bits. (a) Block diagram. (b) Simulation of coarse channel performance.

a match between the bitrates (coarse and fine) required by the MR constellation and the bitrates (source bits plus ECC bits) sent through each of the channels (see Figure 2). Also, embedding in both the ECC domain and the MR modulation domain would lead to an efficiently designed MR joint source-channel system with more than two resolutions, without resorting to a complex "fractal" modulation constellation. This could be accomplished, for example, for a three resolution design, by having the two coarsest resolutions being embedded in the ECC domain, and the resultant composite coarse bitstream being embedded in the third (detail) channel bitstream in the modulation domain as a 2 layer embedding (see Figure 15).

5 An efficient end-to-end system design

In the previous section, we have illustrated, by way of examples, the different tools one can employ to design an efficient broadcast system. Here, we undertake a comparison of the tradeoffs involved in the various schemes, and then provide a general recipe to help solve the problem, as stated in Section 2.

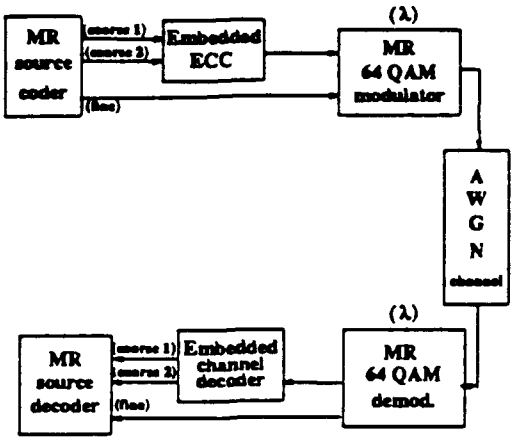


Figure 15: Example E: Block diagram of a MR system with 3 levels of resolution using both embedded ECC's and embedded modulation to make overall design efficient and practical.

5.1 Comparison of λ-Modulation, TCM, and ECC schemes

As pointed out in the previous section through Examples A-E, a number of UEP schemes can be invoked to ensure efficient MR transmission. Table 1 gives the coordinates of Examples A-D of our paper.

The λ-modulation scheme of Example A might be used to provide a desired coverage range for the coarse-resolution signal, and a "basic" coverage for the fine channel, with the MR TCM scheme of Example B used to increase the full-resolution coverage area using an embedded TCM for the fine channel. While the ECC scheme of Example D can be used instead to make the coarse resolution channel more robust, it comes at the cost of reduced quality, for a fixed total bit rate budget for source and channel coding. The TCM scheme of Example B, on the other hand, does not sacrifice source coding quality compared to that of the uncoded system, but it requires an expanded modulation constellation. However, as seen from Figure 11(b), for a probability of fine channel packet loss of 10^{-1} (reasonable to get good full resolution quality if the coarse channel is near-perfect and error concealment is invoked: see Section 6.1 and Figure 21), one needs only a simple 4-state Ungerboeck trellis to get most of the coding gain. Thus, Example B seems like an attractive solution if it can meet the coverage and quality demands.

The hybrid scheme of Example E may be necessary to address a particular broadcast

Example	A	B	C	D
Description	MR QAM	MR TCM	Embedded ECC	Multiplexed ECC
Section	4.2	4.3	4.4	4.4
Block Diagram	Fig. 8	Fig. 10	Fig. 12	Fig. 14
Simulation	Fig. 9	Fig. 11	Fig. 13	Fig. 14

Table 1: Summary of presented alternatives.

problem, especially if more than two grades of service are required. The scheme of Example C (embedded ECC's), while efficient in an information theoretical sense, is unlikely to meet the bit rate ratios of the different resolutions required of most practical HDTV schemes, and is hence omitted from our discussion.

Table 2 gives a comparison of Examples A, B, and D for a typical problem. We fix the coarse channel quality and coverage requirement for receiver CNR's above 20 dB/symbol at a delivered packet error rate (PER) of less than 10^{-4}, and compare the full resolution quality and coverage range for the different schemes. As seen, all schemes perform well with respect to an uncoded system. Note the benefit of error concealment used by the MR source decoder to increase robustness. Example B, operating at $\lambda = 0.3$ with an embedded 4-state trellis is the best choice if constellation expansion is tolerable, while Example A is a good low-complexity solution. Example D gives the same coverage as Example B, but it requires a complicated ECC which also results in 15% reduced fine channel bit rate, and therefore a degraded full-resolution quality.

We now compare the λ-modulation scheme of Example A with the ECC scheme of Example D. Figure 14 (b) of Example D shows how using ECCs lowers the probability of coarse packet loss over the range of CNR's of interest. A comparison with Figure 9 shows how we can achieve similar performances via either the channel modulation parameter λ or an appropriate ECC.

Figure 16 shows the different performances obtained with a modulation domain UEP achieved for λ=0.5 and an ECC domain UEP obtained by protecting the coarse channel

Example		Uncoded	A	B	D
Coarse	PER	$\leq 10^{-4}$	$\leq 10^{-4}$	$\leq 10^{-4}$	$\leq 10^{-4}$
	CNR (dB/symbol) Range	≥ 27	≥ 20	≥ 20	≥ 20
	Low Resolution Quality	-	Same as uncoded	Same as uncoded	Same as uncoded
Fine	PER	$\leq 10^{-4}$ (*)	$\leq 10^{-1}$	$\leq 10^{-1}$	$\leq 10^{-1}$
	CNR (dB/symbol) Range	≥ 27	≥ 27	≥ 24	≥ 24
	High Resolution Quality	-	Same as uncoded	Same as uncoded	Fine channel bitrate 15% less than uncoded (**)
Design parameter		-	$\lambda = 0.3$	$\lambda = 0.3$ 4-state trellis	BCH(255,179,10)
Complexity		-	Same order as uncoded	Higher than uncoded	Higher than uncoded
Increase in coverage over uncoded		-	Coarse: +7dB Full: +0dB	Coarse: +7dB Full: +3dB	Coarse: +7dB Full: +3dB

Table 2: Comparison of schemes A, B and D. We require Packet Error Rate (PER) less than 10^{-4} for the coarse channel at 20 dB/symbol CNR. We compare performance for fine channel PER less than 10^{-1}.

(*) For the uncoded system, the fine channel error rate cannot be 10^{-1} as error concealment requires "perfect" coarse channel performance at that fine channel error rate. See Section 6.1.

(**) The reduction in bitrate available for source coding is due to the use of an ECC.

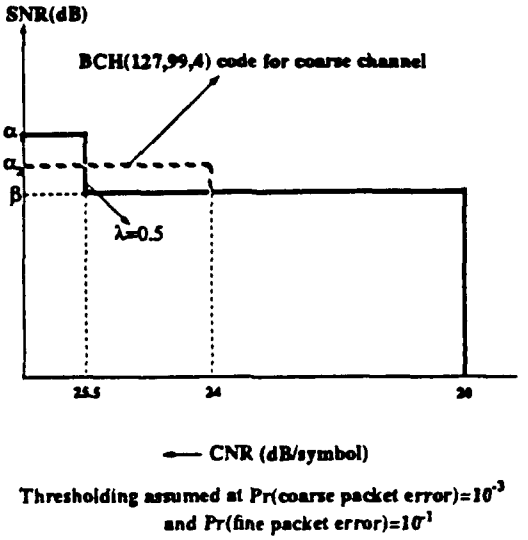


Figure 16: Tradeoff between using modulation domain protection via transmission parameter $\lambda = 0.5$, and ECC domain protection using a BCH(127,99,4) code applied to the coarse channel.

with a BCH(127,99,4) code. The two UEP schemes provide identical coarse channel performance, with CNR's below 20 dB/symbol receiving no signal, and the crossover from coarse resolution quality (SNR= β) to full resolution occurring at 24 dB/symbol and 26 dB/symbol respectively for the ECC (Example D) and λ-modulation (Example A) schemes. Note however, that the parity bits needed by the ECC-protected coarse channel must necessarily come at the expense of a lower fine channel bitrate, resulting in degraded full resolution quality, as noted in Table 2.

Thus, if a comparison is to be made on the basis of equal bandwidth, the ECC scheme would necessarily have lower full resolution quality (SNR=α_1) than the MR modulation scheme (SNR=α_2) for all receiver CNR's better than 25.5 dB/symbol, but the full resolution gain in CNR is 1.5 dB for the ECC scheme (24 dB vs. 25.5 dB). The assessment of the tradeoff depends on the values of α_1, α_2, and β, which in turn depend on the source coding used.

The following points of comparison between the two schemes of Example A (MR modulation) and D (ECC scheme) are worthy of note:

- The coverage tradeoff is between the modulation scheme's degradation of quality by $(\alpha_2 - \beta)$ dB for receivers between 24 dB/symbol and 25.5 dB/symbol CNR, versus the ECC scheme's degradation of full-resolution quality by $\alpha_1 - \alpha_2$ dB for all

receivers with CNR better than 25.5 dB/symbol.

- Note the complexity disparity in the two schemes, with the ECC scheme resorting to a complicated BCH code, while the MR embedded modulation QAM scheme comes at relatively little excess cost over that of a uniform QAM scheme which must be used for transmission anyway.

- The MR modulation parameter λ, being a continuous variable, also affords any desired operating point over the range of CNR's of interest, while the ECC scheme, being discrete in nature, may not afford a solution at any desired operating point.

- In an information-theoretic sense, an embedded MR coding scheme outperforms a non-embedded one, and embedding is accomplished much more easily in the modulation domain. As the ECC scheme uses a Hamming distance metric compared to a softer Euclidean distance criterion for the modulation scheme, the latter is more efficient.

5.2 An efficient choice of system parameters

Assume that the chosen modulation constellation constrains the coarse and fine channels to operate at bit rates of R_c and R_f (for our MR 64 QAM example, we must have $R_c : R_f$=2:4). Note that R_c and R_f represent the combined bit budget to be allocated between source coding and channel coding (i.e. error correction) for the coarse and fine channels respectively. Refer to Section 2 and Figure 3. As the coarse channel is the "anchor" channel for the MR system, it should meet the desired coverage requirement d_c of Figure 3 at the maximum low-resolution quality that the bitrate constraint will permit. To this end, a sensible strategy would be to allocate the coarse channel bit budget *completely for source coding*, while using the embedded λ-modulation scheme of Example A to provide the needed robustness for the coarse channel. Thus, it would be efficient to operate at the maximum value of λ for which the coarse-distance coverage range requirement d_c is satisfied with the desired reliability. This approach is reasonable because the coarse channel represents the fallback mode and not only provides a minimum quality in those areas where the fine channel cannot be reliably decoded, but also allows for better error concealment in the transition area between full resolution and lower resolution coverage (See Section 6.1.). Moreover, the higher the low resolution quality, the higher the full resolution quality will be, for a given modulation-fixed fine channel bit budget.

If the above approach results in an impractically low value of λ, one could resort to a hybrid scheme using ECC's and a practical λ-constellation. This could be used to "boost" the coarse channel coverage, albeit at the expense of a lower coarse-resolution

quality, since part of the total budget must now be diverted from source coding to channel coding.

Now, the R_f bits of the fine channel have to be allocated between error protection ($R_{f,c}$) and source coding ($R_{f,s}$), where $R_f = R_{f,s} + R_{f,c}$. As was discussed in Section 4, an efficient strategy is to protect the fine channel with the MR TCM scheme of Example B. If the constellation expansion is reasonable, this is efficient, as it costs no error protection bits (although, the subtlety lies in the fact that we have an expanded constellation): i.e. $R_{f,c} = 0$. However, if the constellation is not practical in size, or if an increase in full-resolution coverage is desired, we may need an ECC to help satisfy coverage range.

The essential point is that once the coverage distances (i.e. d_c and d_f) have been fixed, the joint source channel coding problem has been converted into a simpler one, thus enabling us to determine the remaining free variables of the system, given d_c, d_f, R_c and R_f.

(Step 1) Use the budget allocated to the coarse channel, R_c, to maximize the quality of the source coder for that channel. (One could for example adopt the guidelines of the optimal source bit allocation strategy described in [31] to achieve optimality for the pyramidal coder considered in this paper.)

(Step 2) Use a MR modulation scheme (Example A) and set λ to the maximum value for which, at distance d_c from the emitter, the error rate for the coarse channel is below the desired threshold.

(Step 3) Use a MR TCM scheme (Example B) to protect the fine channel if the constellation size is reasonable. An embedded multi-dimensional TCM scheme may be deployed [23] to reduce the expansion factor, if complexity permits. Else, and if an additional increase in coverage (coding gain) is desired beyond that affordable by TCM, find an efficient error correction code (convolutional or block) that satisfies the desired fine channel error probability at distance d_f. This code will use $R_{f,c}$ bits and therefore the remaining $R_{f,s} = R_f - R_{f,c}$ will be used for source coding. However, due to the effect of error concealment techniques feasible with an MR design, it is unlikely that the fine channel would require additional protection (see Section 6.1). Note that if the MR TCM scheme will suffice to meet the requirements, no extra channel bits would be needed and $R_{f,s} = R_f$.

(Step 4) Finally, use the remaining $R_{f,s}$ bits for the fine channel source allocation, in an efficient manner, as in Step 1.

6 Comparison of MR embedded, MR independent, and SR constellations

Simulations were carried out for an Additive White Gaussian Noise (AWGN) channel for the multiresolution embedded constellation, the multiresolution non-embedded constellation (i.e. independently transmitted constellations for the two resolutions), and the single resolution constellation, as shown in Figure 17. The independent case refers to separate transmission of the coarse and fine channels using "naive multiplexing" of the frequency spectrum. To ensure fairness of comparison, all three cases were tailored to operate under conditions of equal average power (it can be shown that the comparison under equal peak power constraint would be similar) and equal spectral efficiency (i.e. throughput/bandwidth).

To compare the MR vs. independent constellations, a MR 64-QAM (of free parameter λ), and a 16/256 QAM (coarse/fine) independent constellation pair were picked. The independent channels have a spectral efficiency of 4 bits/symbol and 8 bits/symbol, or an average spectral efficiency (6 bits/symbol) identical to that of the MR 64-QAM. Recall the curves of Figure 9. Also shown on these curves is the performance of the non-embedded MR scheme for the independent constellations of 16 QAM (coarse) and 256 QAM (fine). As was mentioned in Section 4.2, for the range of values of λ from about 0.2 to 0.4, *the embedded MR scheme outperforms the multiplexed MR scheme for both coarse and fine channels.* In order to get a comprehensive picture of the situation, a plot of received quality (SNR) vs. receiver CNR is shown in Figure 18(a), using perceptually consistent thresholding of the curves of Figure 9 at coarse and fine packet loss probabilities of 10^{-3} and 10^{-1} respectively, as justified earlier. As can be seen from Figure 18 (a) (and Figure 8(b)), the MR constellation outperforms the independent one over all ranges of CNR's for some λ values (e.g. λ=0.2).

In our comparison, we assume that the SR source coder is 16% more efficient than the MR coder. This is a "worst case" analysis from the MR point of view, as an empirical comparison using the popular "Lenna" image even in a non-MR-friendly framework like the still-image coding standard JPEG [32] revealed only a 16% increase in source compression for the SR JPEG scheme. Under these conditions, the SR channel could afford a 32-QAM modulation scheme for the same transmitter power as the MR 64-QAM scheme due to a source compression advantage of roughly 5/6. For fairness of comparison, the SR scheme received the same thresholding (10^{-3}) as the coarse resolution packet stream, as they both achieve transitions from the region of no signal (oblivion) to the region of discernible signal.

Note that for the MR 64-QAM scheme, the coarse and fine packetized channels could

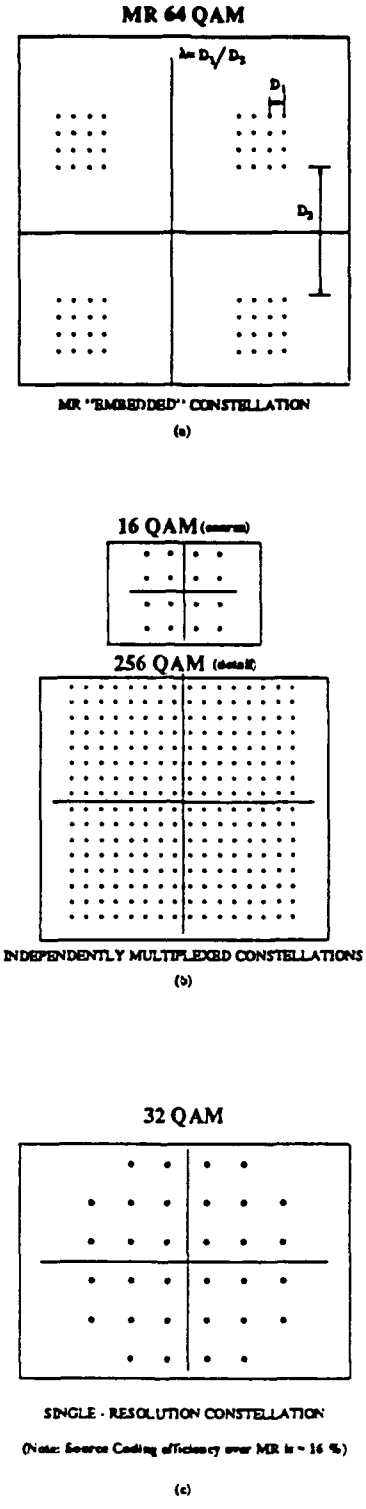


Figure 17: (a) MR 64-QAM constellation of parameter λ. (b) Independent modulation constellations (16/256 QAM) for coarse and fine channels. (c) Single resolution 32-QAM constellation. All constellations use equal power.

be interpreted as entering virtual independent buffers with throughputs in the ratio of 1:2, with instantaneous temporal mismatches in the input channel rates being absorbed by the buffers and if necessary, to prevent overflow or underflow, resolved by exchange of data between the buffers, resulting in minimal degradation for slight mismatches.

The results shown in Figure 18 indicate the tradeoffs involved. As can be seen by comparing the SR scheme with, say, the MR embedded scheme with $\lambda = 0.5$, the broadcast coverage area is much greater for the MR scheme, at the price of some mid-region suboptimality.

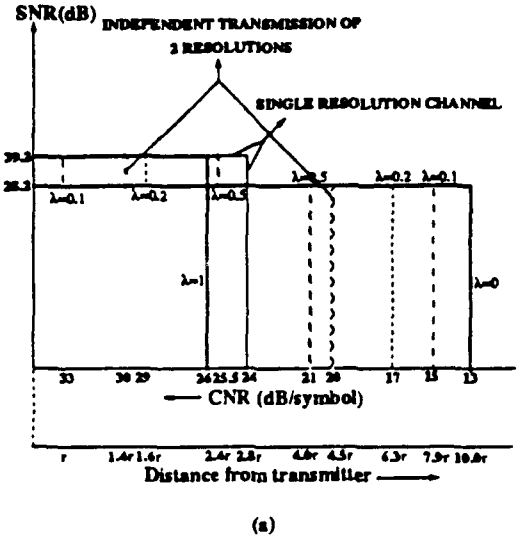
A point to note in favor of the MR scheme is the increase in full-resolution quality coverage area made possible by performing error concealment techniques to be described next. The SR scheme loses this advantage as it has no coarse resolution channel to fall back upon.

6.1 Error concealment

Due to the nature of the broadcast communication, it is impossible (or perhaps impractical) to achieve error–free transmission. Therefore, any real system has to be able to function in the presence of transmission errors, which may range from occasional bit errors in a satellite system to packet losses in digital networks. Communication systems also vary in their resilience to error and speed of recovery. Bitstreams are often packetized to speed up resynchronization in case of a channel error, but a single bit error still renders the whole packet unusable. Recursive systems (motion–compensated hybrid DCT being the typical example) take much longer to recover, specifically until the next restart of the prediction loop. An error concealment scheme is often required to mask those errors and provide a gracefully degrading picture.

The source coder we have used is based on a finite memory structure, and errors would not accumulate but die out within a few time samples. The structure used in conjunction with the MR modulation also allows very successful error concealment. As can be seen in Figure 9, for typical values of λ, at the same CNR's for which the fine channel packet error rate is greater than 10^{-1}, *the coarse channel is almost perfect* (packet error rate less than 10^{-9}). Concealment strategies are typically based on a much lower packet loss rate (on the order of 10^{-5}, as in [33]), since transmission systems based on prediction loops are extremely fragile.

Therefore, most of the errors will occur in the fine detail and a coarse version and motion vectors will be available for concealment. The concealment strategy differs slightly for the frames that are interpolated spatially or temporally, and assumes that the information transmitted in the coarse channel is intact. The spatially interpolated frames of the finest layer are called anchor frames, as they have no temporal dependence on any



(a)

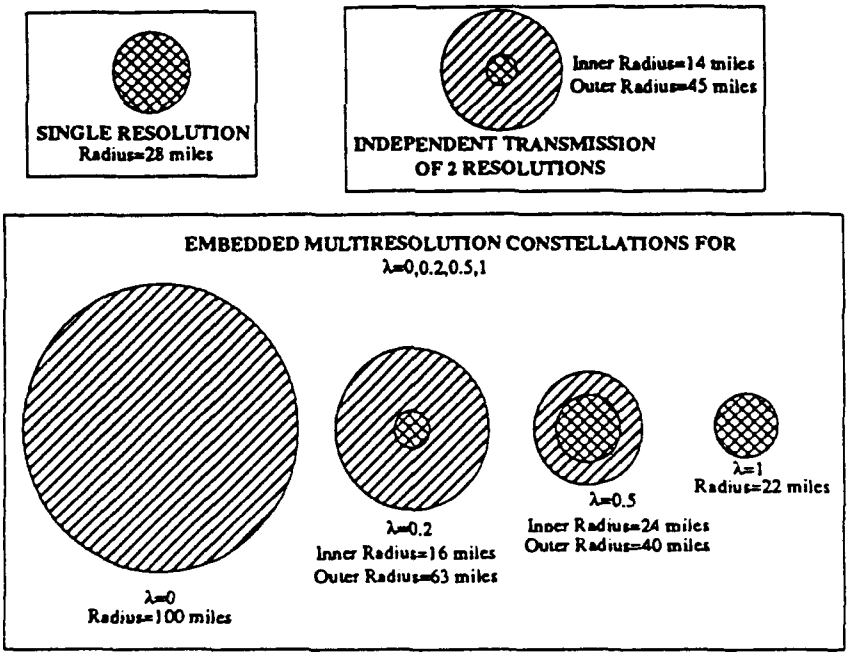
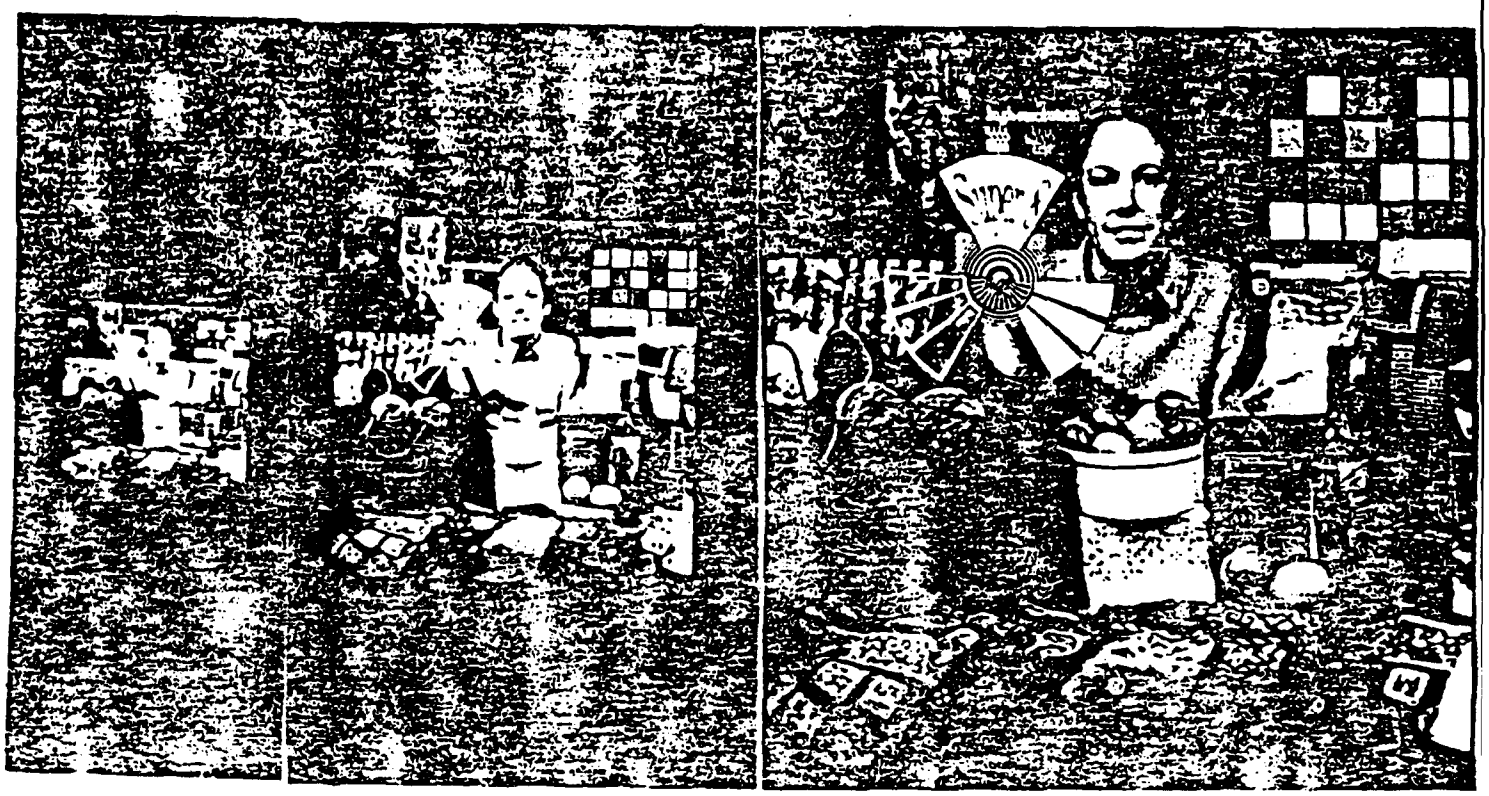
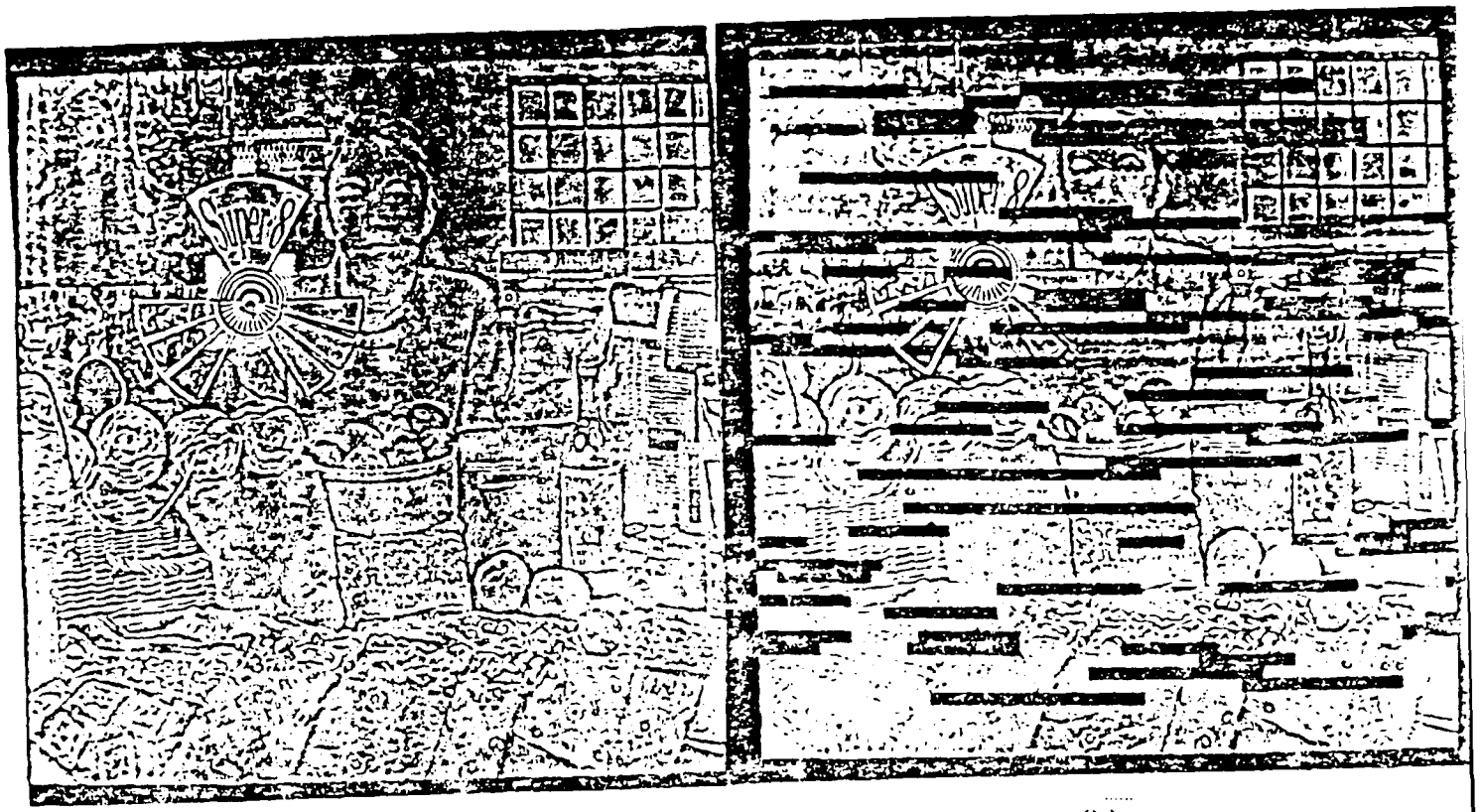


Figure 18: Typical broadcast environment (a) SNR vs. Receiver CNR. (b) Broadcast ranges for the different constellations.



(a) (b) (c)

Figure 19: Resolutions of the pyramid (a) Coarsest layer. (b) Intermediate layer. (c) Full-resolution layer. Images are of size 128x128, 256x256 and 512x512, respectively.

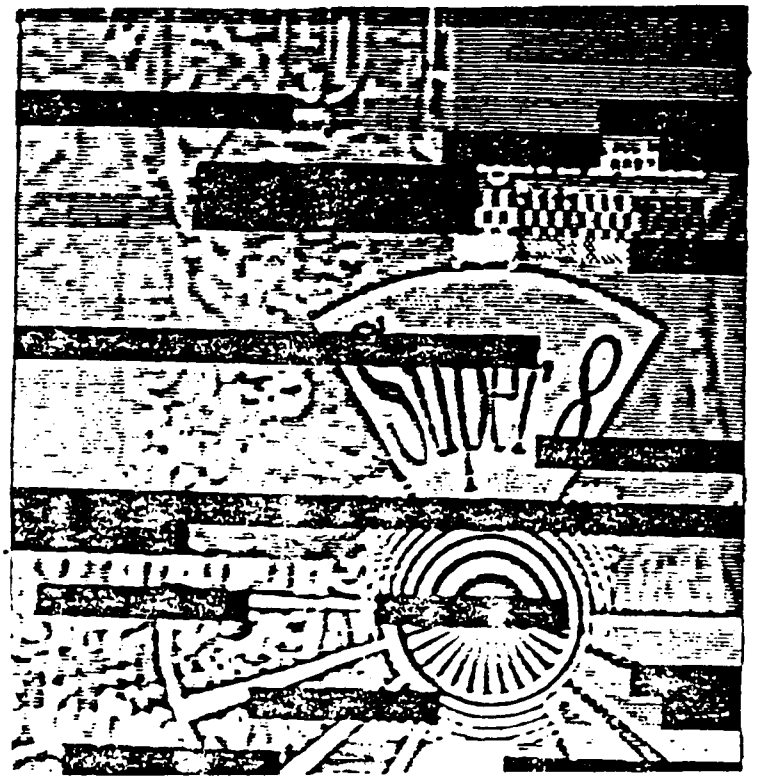


(a) (b)

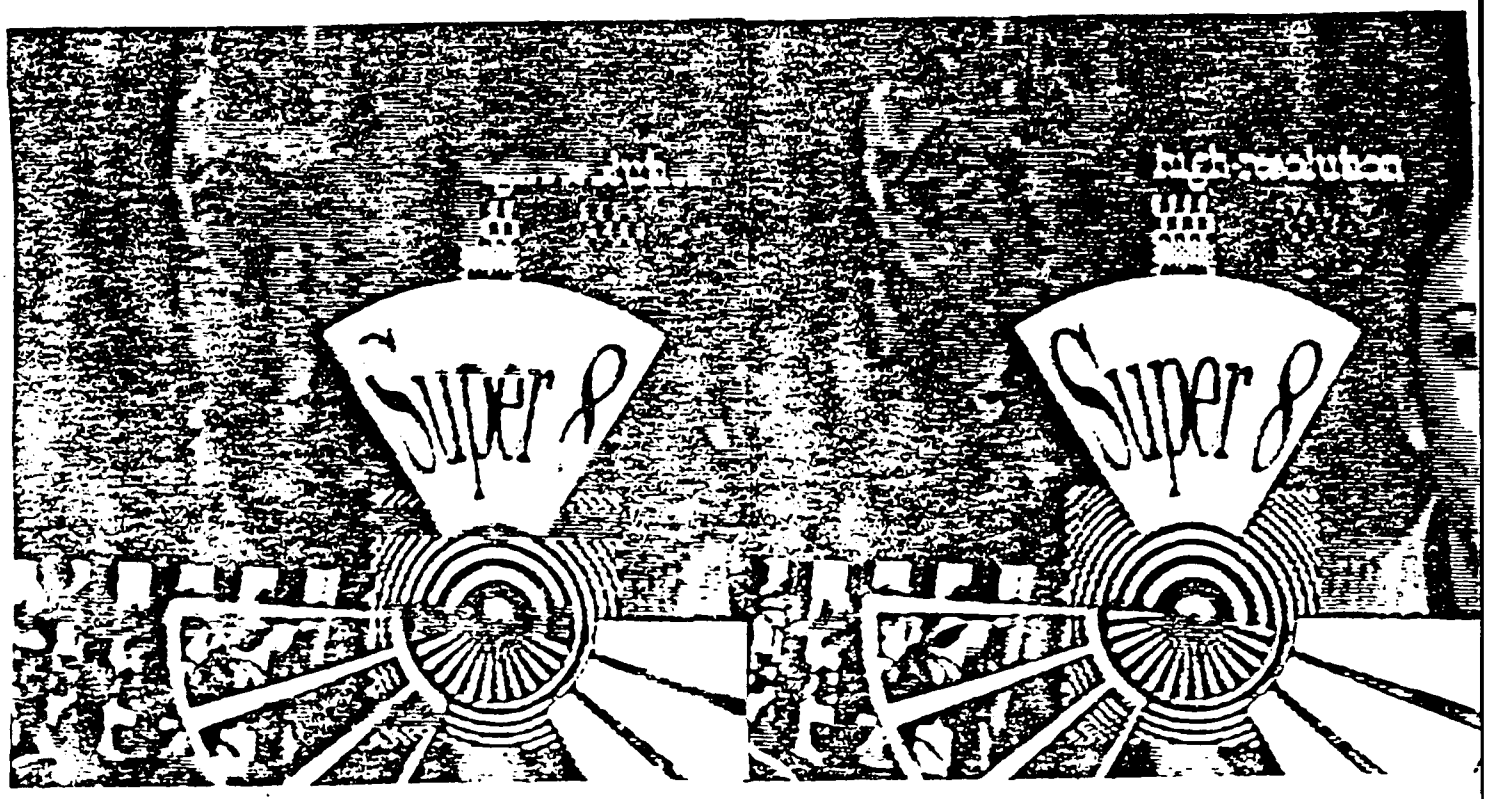


(c)

Figure 20: Effect of channel noise for λ=0.5, CNR=25.5 dB/symbol (15% fine-channel packet loss): (a) Spatial residual frame. (b) Packets corrupted by channel. (c) Full resolution reconstruction after error concealment. Images are of size 512x512.



(a)



(b) (c)

Figure 21: Effect of error concealment for 15% fine-channel packet loss (blow up of Figure 20): (a) Corrupted spatial residual frame. (b) Reconstruction without error concealment. (c) Reconstruction with error concealment.

other frame.

For the temporally interpolated frames, motion vectors and the selected interpolation mode for each block are available, but the actual interpolation error (or the residual) is lost. In the packetized transmission we have implemented, the typical region affected by a packet loss is a narrow strip 8 pixels in height, and 1000–2000 pixels long. Because the encoder uses a smooth motion vector field (enforced by the hierarchical motion estimation algorithm), and both previous and next frames are available for interpolation, errors tend to be very small. Most artifacts show up as "blockiness" and are almost invisible, even in a still frame. Since these frames are not used to predict any other frames, the errors do not need to be processed further.

The errors are more visible, and potentially last longer for the spatially interpolated (anchor) frames. The artifacts appear as blurred blocks or decreased spatial resolution, and are clearly visible in still-frame. (See Figure 21(b).) Furthermore, since the previous and next frames (which are temporally interpolated) are based on this frame, errors can be annoying in real-time. The concealment is based on replacing the region affected by the lost packet from the previous anchor frame. Since motion vectors are not available for this frame, an approximation is computed based on the motion vectors of the previous frame. Then, this is used to interpolate blocks from the previous anchor frame. This works very well in practice, as motion vectors resemble the true motion.

This concealment strategy gives excellent results even in extreme cases of packet loss. Complete loss of a frame can be tolerated, and sustained 15% packet loss rate causes no visible loss in quality. Figure 20 shows the effect of 15% fine-packet loss (obtained for λ=0.5, CNR=25.5 dB/symbol) on the spatial residual of the sequence, with Figure 20(c) showing the reconstructed quality, while Figure 21 illustrates the power of error concealment in a MR environment.

7 Conclusion

We have demonstrated a multiresolution (MR) joint source channel coding system, where using a source coder matched to an embedded trellis-coded modulation constellation (with/without error correction coding) has been shown to provide an efficient end-to-end MR system. The threshold effect plaguing single resolution (SR) systems is softened by a stepwise graceful degradation reminiscent of analog systems, without sacrificing the source coding advantage of digital schemes. We show the superiority of an embedded MR transmission scheme over independent transmissions of the MR source resolutions, and point out the tradeoffs in robustness and broadcast area coverage of low and high resolutions between embedded MR and SR digital systems for QAM constellations, highlighting

the benefits of deploying joint MR source and channel coding.

Acknowledgements: The authors would like to thank Prof. W. Schreiber of MIT for pointing out the importance of spectrum efficiency for television broadcast. We also thank Dr. R. Calderbank of AT&T Bell Labs for fruitful discussions on multiplexed versus embedded modulation.

References

[1] W. F. Schreiber, "Considerations in the design of HDTV systems for terrestrial broadcasting," *Electronic Imaging*, Oct. 1990.

[2] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[3] K. M. Uz, K. Ramchandran, and M. Vetterli, "Multiresolution source and channel coding for digital broadcast of HDTV," in *Proceedings Fourth International Workshop on HDTV and beyond, Torino*, Sept. 1991.

[4] K. M. Uz, M. Vetterli, and D. LeGall, "Interpolative multiresolution coding of advanced television with compatible subchannels," *IEEE Transactions on CAS for Video Technology, Special Issue on Signal Processing for Advanced Television*, vol. 1, pp. 86–99, Mar. 1991.

[5] D. Anastassiou and M. Vetterli, "All digital multiresolution coding of HDTV," in *Proceedings of the National Association of Broadcasting (NAB), Las Vegas, Nevada.*, pp. 210–216, Apr. 1991.

[6] "Digital Spectrum Compatible HDTV System." AT&T/Zenith Technical Report, Sept. 1991.

[7] A. Netravali, E. Petajan, S. Knauer, K. Matthews, R. Safranek, and P. Westerink, "A high performance digital HDTV codec," in *Proceedings of the National Assocation of Broadcasters (NAB), Las Vegas, Nevada.*, Apr. 1991.

[8] "Advanced Digital Television : System Description." Sarnoff/ NBC/ Philips/ Thomson Technical Report, Feb. 1991.

[9] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *AT&T Technical Memorandum*, 1992.

[10] W. F. Schreiber, "All-digital HDTV terrestrial broadcasting in the U.S. : Some problems and possible solutions," *Workshop on Advanced Television, ENST, Paris*, May 1991.

[11] J. Modestino, D.G.Daut, and A. Vickers, "Combined source-channel coding of images using the block cosine transform," *IEEE Transactions on Communications*, vol. COM-29, pp. 1261–1274, Sept. 1981.

[12] G. Karlsson and M. Vetterli, "Sub-band coding of video for packet networks," *Optical Engineering*, vol. 27, pp. 574–586, July 1988.

[13] K. Fazel and J. L'Huillier, "Application of unequal error protection codes on combined source-channel coding," in *IEEE International Conference on Communications, ICC'90, Atlanta*, pp. 320.5.1–6, Apr. 1990.

[14] M. Vetterli and K. M. Uz, "Multiresolution coding techniques for digital video: a review," *Special Issue on Multidimensional Processing of Video Signals, Multidimensional Systems and Signal Processing*, Mar. 1992. Invited paper, to appear.

[15] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, 1948.

[16] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell System Tech. Journal*, Nov. 1974.

[17] W. H. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, vol. 37, pp. 269–275, Mar. 1991.

[18] J. M. Shapiro, "An embedded wavelet hierarchical image coder," *Proceedings of ICASSP*, Mar. 1992. To appear.

[19] A. E. Gamal and T. Cover, "Multiple user information theory," *IEEE Transactions on Information Theory*, vol. 68, pp. 1466–1483, Dec. 1980.

[20] M. M. Mano, *Digital logic and computer design.* Prentice-Hall, 1979.

[21] J. G. Proakis, *Digital Communications.* McGraw-Hill, 1989.

[22] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[23] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Transactions on Information Theory*, vol. IT-33, pp. 483–501, July 1987.

[24] A. R. Calderbank and N. J. A. Sloane, "Four-dimensional modulation with an eight-state trellis code," *AT&T Tech. Journal*, vol. 64, pp. 1005–1018, May 1985.

[25] T. Chiang and D. Anastassiou, "HDTV coding with a compatible Standard/TV subchannel," *Journal on Selected Areas in Communications*, 1992. To be submitted.

[26] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 600–607, Oct. 1967.

[27] L. Dunning and W. Robbins, "Optimal encodings of linear block codes for unequal error protection," *Information Contributions*, vol. 37, pp. 150–177, 1978.

[28] W. J. VanGils, "Two topics on linear unequal error protection codes: bounds on their length and cyclic code classes," *IEEE Transactions on Information Theory*, vol. IT-29, pp. 866–876, Nov. 1983.

[29] T.Kasami, S.Lin, V. Wei, and S. Yamamura, "Coding for the binary symmetric broadcast channel with two receivers," *IEEE Transactions on Information Theory*, vol. IT-31, pp. 616–625, Sept. 1985.

[30] M.-C. Lin, C.-C. Lin, and S. Lin, "Computer search for binary cyclic UEP codes of odd length up to 65," *IEEE Transactions on Information Theory*, vol. 36, pp. 924–935, July 1990.

[31] K. Ramchandran and M. Vetterli, "Efficient bit allocation for a pyramid coder using arbitrary quantizers," *IEEE Transactions on Image Processing*, 1992. To be submitted.

[32] "JPEG technical specification: Revision (DRAFT), joint photographic experts group, ISO/IEC JTC1/SC2/WG8, CCITT SGVIII," Aug. 1990.

[33] F.-C. Jeng and S. H. Lee, "Concealment of bit error and cell loss in inter-frame coded video transmission," in *IEEE International Conference on Communications, ICC*, 1991.

Appendix

Analysis of MR QAM of section 4.2.1

Let us introduce some definitions associated with Figure 7.

- S= {set of all constellation points in the modulation scheme}.
- $N = |S|$ is the number of signals in the constellation.

- D= {set of all "directions" (N,S,E,W) representing the one-sided independent degrees of freedom for the additive Gaussian noise, with unit directional vectors $(\mathbf{u_N}, \mathbf{u_S}, \mathbf{u_E}, \mathbf{u_W})$ respectively.}
- C_i= $\{j | j \in S$ and i, j are in the same cloud$\}$
  i.e. the set of all points which are in the same cloud as signal i.
- $d^k_{intra}(i)(d^k_{inter}(i))$= half the Euclidean distance between i and its nearest "fine" ("coarse") neighbor in the (positive) $\mathbf{u_k}$ direction. Thus, $\{d^k_{intra}(i)\}$ $(\{d^k_{inter}(i)\})\forall k \in D, \forall i \in S$ is the minimum instantaneous noise amplitude component in the $\mathbf{u_k}$ direction that will cause the receiver to incorrectly decode the intracloud (intercloud) information in that direction. Note also that if a signal point should have no neighbor in the positive $\mathbf{u_k}$ direction, then its corresponding nearest-neighbor distance will be ∞.

From these definitions, using the simple Gaussian error function, we can obtain closed form solutions to the probabilities of fine and coarse channel bit errors ($P^f_{e,b}$ and $P^c_{e,b}$, respectively). It is easy to show that the probability of fine bit error for a given λ and CNR are given by:

$$P^f_{e,b}(\lambda, CNR) = \sum_{i \in S(\lambda, CNR)} q(i) \sum_{k \in D} [0.5 * \mathrm{erfc}(d^k_{intra}(i)/\sqrt{2})] \tag{1}$$

where erfc(x) above is the standard complementary Gaussian error function defined as

$$\mathrm{erfc}(x) = (2/\sqrt{\pi}) \int_{-\infty}^{x} e^{-t^2} dt \tag{2}$$

and $q(i)$ in Equation 1 refers to the symbol probabilities, which, if assumed to be equal, would simplify it to:

$$P^f_{e,b}(\lambda, CNR) = \frac{1}{2N} \sum_{i \in S(\lambda, CNR)} \sum_{k \in D} \mathrm{erfc}(d^k_{intra}(i)/\sqrt{2}) \tag{3}$$

Similarly, for the coarse bitstream, we have, assuming equiprobable symbols:

$$P^c_{e,b}(\lambda, CNR) = \frac{1}{2N} \sum_{i \in S(\lambda, CNR)} \sum_{k \in D} \mathrm{erfc}(d^k_{inter}(i)/\sqrt{2}) \tag{4}$$

In order to prevent error propagation, we packetize the streams into a composite length of L bits/packet, comprising $L/3$ bits of coarse data and $2L/3$ bits of fine information (as demanded by the 1:2 ratio in coarse to fine bit rate). In the absence of ECC, we assume that a single bit error anywhere in an entire packet corrupts that packet and causes it to get lost. As was shown, due to the Karnaugh mapping, single bit errors will dominate.

Defining the packet error probabilities as $P^c_{e,p}$ and $P^f_{e,p}$ respectively for the coarse and fine channels, we have:

$$P^f_{e,p}(\lambda, CNR) = 1 - (1 - P^f_{e,b}(\lambda, CNR))^{L/6} \tag{5}$$

and,

$$P^c_{e,p}(\lambda, CNR) = 1 - (1 - P^c_{e,b}(\lambda, CNR))^{L/6} \tag{6}$$

See Figure 9 for a plot of the curves for L=1080 using the MR 64 QAM constellation for both coarse and fine packet probability of loss performance as a function of the broadcast area CNR for a multitude of λ values encompassing its region of definition from 0 to 1.

©1992 IEEE. Reprinted, with permission, from *IEEE Transcations on Information Theory*, Vol. 38, No. 2, pp. 776-784, March 1992.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commerical advantage, the IEEE copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Institute of Electrical and Electronics Engineers. To copy otherwise, or to republish, requires a fee and specific permission.

MODELING AND ESTIMATION OF MULTIRESOLUTION STOCHASTIC PROCESSES

Michele Basseville[1], Albert Benveniste[1], Kenneth C. Chou[2], Stuart A. Golden[2], Ramine Nikoukhah[3] and Alan S. Willsky[2]

Abstract

In this paper, we provide an overview of the several components of a research effort aimed at the development of a theory of multiresolution stochastic modeling and associated techniques for optimal multiscale statistical signal and image processing. As we describe, a natural framework for developing such a theory is the study of stochastic processes indexed by nodes on lattices or trees in which different depths in the tree or lattice correspond to different spatial scales in representing a signal or image. In particular we will see how the wavelet transform directly suggests such a modeling paradigm. This perspective then leads directly to the investigation of several classes of dynamic models and related notions of "multiscale stationarity" in which scale plays the role of a time-like variable. In this paper we focus primarily on the investigation of models on homogeneous trees. In particular we describe the elements of a dynamic system theory on trees and introduce two notions of stationarity. One of these leads naturally to the development of a theory of multiscale autoregressive modeling including a generalization of the celebrated Schur and Levinson algorithms for order-recursive model building. The second, weaker notion of stationarity leads directly to a class of state space models on homogeneous trees. We describe several of the elements of the system theory for such models and also describe the natural, extremely efficient algorithmic structures for optimal estimation that these models suggest: one class of algorithms has a multigrid relaxation structure; a second uses the scale-to-scale whitening property of wavelet transforms for our models; and a third leads to a new class of Riccati equations involving the usual predict and update steps and a new "fusion" step as information is propagated from fine to coarse scales. As we will see, this framework allows us to consider in a very natural way the fusion of data from sensors with differing resolutions. Also, thanks to the fact that wavelet transforms do an excellent job of "compressing" large classes of covariance kernels, we will see that these modeling paradigms appear to have promise in a far broader context than one might expect.

[1]Institut de Recherche en Informatique et Systemes Aleatoires (IRISA), Campus de Beaulieu, 35042 Rennes, CEDEX, FRANCE. M.B. is also with the Centre National de la Recherche Scientifique (CNRS) and A.B. is also with the Institut National de Recherche en Informatique et en Automatique (INRIA). The research of these authors was also supported in part by Grant CNRS G0134.

[2]Laboratory for Information and Decision Systems and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge,MA 02139, USA. The work of these authors was also supported in part by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, by the National Science Foundation under Grants MIP-9015281 and INT-9002393, and by the Office of Naval Research under Grant No. N00014-91-J-1004. In addition some of this research was performed while KCC and ASW were visitors at IRISA and while ASW received support from INRIA.

[3]INRIA, Domaine de Voluceau, Rocquencourt, BP105, 78153 Le Chesnay, CEDEX, FRANCE.

1 Introduction

In recent years there has been considerable interest and activity in the signal and image processing community in developing multi-resolution processing algorithms. Among the reasons for this are the apparent or claimed computational advantages of such methods and the fact that representing signals or images at multiple scales is an evocative notion– it seems like a "natural" thing to do. One of the more recent areas of investigation in multiscale analysis has been the emerging theory of multiscale representations of signals and wavelet transforms [10, 21, 22, 23, 24, 28, 33, 34, 38, 49]. This theory has sparked an impressive flurry of activity in a wide variety of technical areas, at least in part because it offers a common unifying language and perspective and perhaps the promise of a framework in which a rational methodology can be developed for multiscale signal processing, complete with a theoretical structure that pinpoints when multiresolution methods might be useful and why.

It is important to realize, however, that the wavelet transform by itself is not the only element needed to develop a methodology for signal analysis. To understand this one need only look to another orthonormal transform, namely the Fourier transform which decomposes signals into its frequency components rather than its components at different resolutions. The reason that such a transform is useful is that its use simplifies the description of physically meaningful classes of signals and important classes of transformations of those signals. In particular stationary stochastic processes are *whitened* by the Fourier transform so that individual frequency components of such a process are statistically uncorrelated. Not only does this greatly simplify their analysis, but, it also allows us to deduce that frequency-domain operations such as Wiener or matched filtering–or their time domain realizations as linear shift-invariant systems–aren't just convenient things to do. *They are in fact the right–* i.e., the statistically optimal– *things to do.* In analogy, what is needed to complement wavelet transforms for the construction of a rational framework for multi-resolution signal analysis is the identification of a rich class of signals and phenomena whose description is simplified by wavelet transforms. Having this, we then have the basis for developing a methodology for *scale domain* filtering and signal processing, for

deducing that such operations are indeed the right ones to use, and for developing a new and potentially powerful set of insights and perspectives on signal and image analysis that are complementary to those that are the heritage of Fourier.

In this paper we describe the several components of our research into the development of a theory for multiresolution stochastic processes and models aimed at achieving the objectives of describing a rich class of phenomena and of providing the foundation for a theory of optimal multiresolution statistical signal processing. In developing this theoretical framework we have tried to keep in mind the three distinct ways in which multi-resolution features can enter into a signal or image analysis problem. First, the phenomenon under investigation may possess features and physically significant effects at multiple scales. For example, fractal models have often been suggested for the description of natural scenes, topography, ocean wave height, textures, etc. [5, 35, 36, 41]. Also, anomalous broadband transient events or spatially-localized features can naturally be thought of as the superposition of finer resolution features on a more coarsely varying background. As we will see, the modeling framework we describe is rich enough to capture such phenomena. For example, we will see that 1/f -like stochastic processes as in [50, 51] are captured in our framework as are surprisingly useful models of many other processes. Secondly, whether the underlying phenomenon has multi-resolution features or not, it may be the case that the data that has been collected is at several different resolutions. For example the resolutions of remote sensing devices operating in different bands– such as IR, microwave, and various band radars– may differ. Furthermore, even if only one sensor type is involved, measurement geometry may lead to resolution differences (for example, if zoomed and un-zoomed data are to be fused or if data is collected at different sensor-to-scene distances). As we will see, the framework we describe provides a natural way in which to design algorithms for such multisensor fusion problems.

Finally, whether the phenomenon or data have multi-resolution features or not, the signal analysis *algorithm* may have such features motivated by the two principal manifestations of the at least superficially daunting complexity of many image processing problems. The first and more well-known of these is the use of multi-resolution algorithms to combat the computational demands of such problems by solving coarse

(and therefore computationally simpler) versions and using these to guide (and hopefully speed up) their higher resolution counterparts. Multigrid relaxation algorithms for solving partial differential equations are of this type as are a variety of computer vision algorithms. As we will see, the stochastic models we describe lead to several extremely efficient computational structures for signal processing.

The second and equally important issue of complexity stems from the fact that a multi-resolution formalism allows one to exercise very direct control over "greed" in signal and image reconstruction. In particular, many imaging problems are, in principle, ill-posed in that they require reconstructing more degrees of freedom then one has elements of data. In such cases one must "regularize" the problem in some manner, thereby guaranteeing accuracy of the reconstruction at the cost of some resolution. Since the usual intuition is precisely that one should have higher confidence in the reconstruction of lower resolution features, we are led directly to the idea of reconstruction at multiple scales, allowing the resolution-accuracy tradeoff to be confronted directly. As we will see the algorithms arising in our framework allow such multi-scale reconstruction and provide the analytical tools both for assessing resolution versus accuracy and for correctly accounting for fine scale fluctuations as a source of "noise" in coarser scale reconstructions.

While there are several ways in which to introduce and motivate our modeling framework, one that provides a fair amount of insight begins with the wavelet tranforms. However, the key for modeling is not to view the transform as a method for analyzing signals but rather as a mechanism for *synthesizing* or *generating* such signals beginning with coarse representations and adding fine detail one scale at a time. Specifically let us briefly recall the structure of multiscale representations associated with orthonormal wavelet transforms [22, 33]. For simplicity we do this in the context of $1-D$ signals (i.e. signals with one independent variable), but the extension to multidimensional signals and images introduces only notational rather than mathematical complexity.

The multiscale representation of a continuous signal $f(x)$ consists of a sequence of approximations of that signal at finer and finer scales where the approximations of $f(x)$ at the mth scale consists of a weighted sum of shifted and compressed (or

dilated) versions of a basic scaling function $\phi(x)$:

$$f_m(x) = \sum_{n=-\infty}^{+\infty} f(m,n)\phi(2^m x - n) \tag{1.1}$$

In order for the $(m+1)$st approximation to be a refinement of the mth, we require $\phi(x)$ to be representable at the next scale:

$$\phi(x) = \sum_n h(n)\phi(2x - n) \tag{1.2}$$

As shown in [22], $h(n)$ must satisfy several conditions for (1.1) to be an orthonormal series and for several other properties of the representation to hold. In particular $h(n)$ must be the impulse response of a quadrature mirror filter (QMF) [22, 44]. The simplest example of such a ϕ, h pair is the Haar approximation with

$$\phi(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & otherwise \end{cases} \tag{1.3}$$

and

$$h(n) = \begin{cases} 1 & n = 0, 1 \\ 0 & otherwise \end{cases} \tag{1.4}$$

By considering the incremental detail added in obtaining the $(m+1)$st scale approximation from the mth, we arrive at the wavelet transform. Such a transform is based on a single function $\psi(x)$ that has the property that the full set of its scaled translates $\left\{2^{m/2}\psi(2^m x - n)\right\}$ form a complete orthonormal basis for L^2. In [22] it is shown that ϕ and ψ are related via an equation of the form

$$\psi(x) = \sum_n g(n)\phi(2x - n) \tag{1.5}$$

where $g(n)$ and $h(n)$ form a *conjugate mirror filter* pair [44], and that

$$f_{m+1}(x) = f_m(x) + \sum_n d(m,n)\psi(2^m x - n) \tag{1.6}$$

Thus, $f_m(x)$ is simply the partial orthonormal expansion of $f(x)$, up to scale m, with respect to the basis defined by ψ. For example if ϕ and h are as in (1.3), (1.4), then

$$\psi(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & otherwise \end{cases} \tag{1.7}$$

$$g(n) = \begin{cases} 1 & n = 0 \\ -1 & n = 1 \\ 0 & otherwise \end{cases} \tag{1.8}$$

and $\left\{2^{m/2}\psi(2^m x - n)\right\}$ is the *Haar basis.*

One of the appealing features of the wavelet transforms for the analysis of signals is that they can be computed recursively in scale, from fine to coarse. Specifically, if we have the coefficients $\{f(m+1,\cdot)\}$ of the $(m+1)$st-scale representation we can "peel off" the wavelet coefficients at this scale and at the same time carry the recursion one complete step by calculating the coefficients $\{f(m,\cdot)\}$ at the next somewhat coarser scale:

$$f(m,n) = \sum_k h(2n-k) f(m+1,k) \tag{1.9}$$

$$d(m,n) = \sum_k g(2n-k) f(m+1,k) \tag{1.10}$$

Reversing this process we obtain the synthesis form of the wavelet transform in which we build up finer and finer representations via a coarse-to-fine scale recursion:

$$f(m+1,n) = \sum_k h(2k-n) f(m,k) + \sum_k g(2k-n) d(m,k) \tag{1.11}$$

Thus we see that the synthesis form of the wavelet transform defines a *dynamical* relationship between the coefficients $f(m,n)$ at one scale and those at the next. Indeed this relationship defines a lattice on the points (m,n), where $(m+1,k)$ is connected to (m,n) if $f(m,n)$ influences $f(m+1,k)$. The simplest example of such a lattice is the dyadic tree illustrated in Figure 1, where each node t corresponds to a particular scale/shift pair (m,n). As with all these lattices, the scale index is indeed time-like, with each level of the tree corresponding to a representation of signals or phenomena at a particular scale. In this paper we focus for the most part on this tree structure and on dynamic models and stochastic processes defined on it[1]. Note that this setting has a natural association with the Haar transform in which the value at a particular

[1] In Sections 4 and 5 we briefly describe some aspects of the more general case.

node $t = (m, n)$ is obtained from the average of the values at the two descendant nodes $(m + 1, 2n)$ and $(m + 1, 2n + 1)$. However, while the Haar transform indeed plays an important role in our analysis, the dyadic tree and the pyramidal structure it captures should be viewed in a broader sense as providing a natural setting for capturing representations of signals at multiple resolutions where the relationships between the representations at different resolutions need not be constrained to the rigid equalities in (1.9) - (1.11). Rather, if we view these multiscale representations more abstractly, much as in the notion of state, as capturing the features of signals up to a particular scale that are relevant for the "prediction" of finer-scale approximations, we can define rich classes of stochastic processes and models that contain the multiscale wavelet representations of (1.9) - (1.11) as special (and in a sense degenerate) cases.

Carrying this a bit farther, let us return to the point made earlier that for wavelet transforms to be useful it should be the case that their application simplifies the description or properties of signals. For example, this clearly would be the case for a stochastic process that is whitened by (1.9), (1.10), i.e. for which the wavelet coefficients $\{d(m, \cdot)\}$ at a particular scale are white and uncorrelated with the lower resolution version $\{f(m, \cdot)\}$ of the signal. In this case (1.11) represents a first-order recursion in scale that is driven by white noise. However, as we know from time series analysis, white-noise-driven first-order systems yield a comparatively small class of processes which can be broadened considerably if we allow higher-order dynamics. Also, in sensor fusion problems one wishes to consider collectively an entire set of signals or images from a suite of sensors. In this case one is immediately confronted with the need to use higher-order models in which the actually observed signals may represent samples from such a model at *several* scales, corresponding to the differing resolutions of individual sensors.

In this paper we describe two stochastic modeling paradigms for multiresolution processes that have as their motivation the preceding observations as well as the desire to investigate and develop multiscale counterparts to the notions of stationarity and rationality that have proven to be of such value in time series analysis. The first step in doing this is the introduction of dynamics and concepts of shift-invariance on dyadic trees, and in the next section we outline the elements of this formalism and

in particular introduce two notions of (second-order) shift-invariance for stochastic processes on dyadic trees. In Section 3 we then use the stronger of these two notions to develop a theory of multiscale autoregressive modeling and in particular we describe a generalization of the celebrated Schur and Levinson algorithms for the efficient construction of such models. Figure 2 illustrates the output of a third-order model of this type displaying some of the fractal-like, multi-scale characteristics that can be captured by this class of models. An alternate modeling paradigm—coinciding with that of Section 3 only for first-order models—is described in Section 4. This formalism, which generalizes finite-dimensional state models to dyadic trees, also can be used to capture fractal-like behavior and indeed includes the 1/f-like models developed in [50, 51] as a special case. Moreover these models provide surprisingly accurate descriptions of a broad variety of stochastic processes and also lead to extremely efficient and highly parallelizable algorithms for optimal estimation and for the fusion of multiresolution measurements using multiscale, scale-recursive generalizations of Kalman filtering and smoothing. For example, Figure 3(a) illustrates the sample path of a process with a 1/f-like spectrum and its optimal estimation based on noisy measurements of the process collected only at the two ends of the data interval. Figure 3(b) illustrates the use of our methodology for the estimation of the process based on these noisy data augmented with coarser resolution measurements– i.e. the formalism we describe allows us, with relative ease, to use coarse scale data to optimally guide the interpolation of fine-scale but sparsely-collected data. Figures 3(c) and 3(d) display analogous results for the case of a standard Gauss-Markov process in which an approximate multiscale model for this process is used to design the coarse/fine data fusion and interpolation algorithm.

Due to the limitations of space our presentation of the various topics we have mentioned is of a summary nature. References to complete treatments are given, and, in addition, in Section 5 we briefly discuss several important issues, current lines of investigation, and open questions.

2 Stochastic Processes and Dynamic Models on Dyadic Trees

In this section we introduce the machinery needed for specifying linear models of random processes on the dyadic tree, that is for stochastic processes y_t where t is an element of the set of nodes, $\mathcal{T}$, of the tree of Figure 1. As indicated in the introduction, we have several objectives in developing such models. Our first objective is to introduce models that can be specified by finitely many parameters in order to provide associated effective algorithms. That is, we would like to develop models analogous to those specified by finite-order difference equations or finite-dimensional state models– i.e. those corresponding to rational system functions– which have provided the setting for a vast array of powerful methods of signal and system analysis. Also, recursive models of this type are naturally associated with a notion of causality. In our context we will also seek recursive structures where the associated notion of causality will be in scale, from coarse to fine as in the wavelet transform synthesis equation (1.11).

Finally, another notion from time series that we will want to adapt to our context is that of shift-invariance or stationarity. To understand what is involved in this, let us recall the usual notion of stationarity[2] for a discrete-time, zero-mean stochastic processes y_t, where in this case $t \epsilon Z$, the integers. Such a process, with covariance function

$$r_{t,s} = E[y_t y_s] \tag{2.1}$$

is stationary if $r_{t+n,s+n} = r_{t,s}$ for all integers n. That is, shifting the time index of the process by n leaves the statistics invariant. Since it is also obviously time that $r_{s,t} = r_{t,s}$, we can immediately deduce that

$$r_{s,t} = r_{d(s,t)} \tag{2.2}$$

where $d(s,t) = |t - s|$.

[2] In this paper we focus completely on linear models and second-order properties, which, of course, yield complete descriptions if the processes considered are Gaussian.

In order to understand how we might generalize these ideas to the dyadic tree, we need to make several observations. The first is that the integers Z and our dyadic tree are both examples of homogeneous trees. Specifically a homogeneous tree of multiplicity q is an infinite acyclic graph such that each node has exactly q+1 branches to other nodes representing its neighbors. In the case of Z, $q = 1$, and the neighbors of an integer t are simply $t - 1$ and $t + 1$. For the case of $\mathcal{T}$, $q = 2$. However, Figure 1 isn't the easiest way in which to see this or to understand notions of stationarity. Specifically, in considering the usual notion of stationarity we are compelled to consider processes defined on all of Z, and the same is true in our context. Thus, we must be able to extend our tree in all directions capturing in particular the fact that there is neither a finest nor a coarsest scale of description. A much more convenient representation of $\mathcal{T}$ that allows such extensions is depicted in Figure 4. As we will see, both Figures 1 and 4 will prove of use to us.

An important fact about trees is that there is a natural notion of distance $d(s,t)$ between two nodes, s and t, namely the number of branches on the path from s to t, which reduces to $|t - s|$ for Z. This allows us to define the notion of an isometry, that is a one-to-one and onto map of the tree onto itself that preserves distances. For Z the only isometries are shifts, $t \longmapsto t + n$ i.e. and reversals, i.e. $t \longmapsto -t$ (and concatenations of these), so that a useful way (for us!) in which to define the usual notion of stationarity is that the statistics of the process are invariant under any isometry on the index set, i.e. $r_{t,s} = r_{\tau(t),\tau(s)}$ for any isometry.

It is this type of notion that we seek to generalize to the dyadic tree. However, the tree $\mathcal{T}$ has many isometries. For example consider an isometry pivoting on the node denoted "$s \wedge t$" in Figure 4, where all nodes below and to the right of this point are left unchanged but the upper left-hand portion of the tree is "flipped" in that the two branches extending from $s \wedge t$ are interchanged (so that, for example, u is mapped into s). Obviously we can do the same thing pivoting at any node. We refer the reader to [14] for complete treatments of the nature and structure of isometries.

The preceding discussion suggests a first notion of shift-invariance for a stochastic process y_t which we refer to as isotropy. Specifically y_t is an isotropic process if its statistics remain invariant under any isometry on the index set. As shown in [3, 6, 7, 8]

y_t is isotropic if any only if its covariance $r_{t,s}$, as defined in (2.1) (with $t, s \epsilon \mathcal{T}$), satisfies (2.2). Thus, as with a standard temporally-stationary process, an isotropic process on $\mathcal{T}$ is characterized by a covariance sequence $r_0, r_1, r_2, \ldots$ and, as in the standard case we have two natural questions: (1) when does such a sequence of numbers correspond to a valid covariance sequence for a process on $\mathcal{T}$; and (2) how can we construct dynamic models for the construction of an isotropic process corresponding to such a valid sequence. A first form of the answer to the first question can actually be stated a bit more generally. Specifically, if S is any index set, and if $\{y_t, t \epsilon S\}$ is a zero-mean process defined on S then its covariance $r_{s,t}$ must satisfy the following : select an arbitrary finite family $\{t_i\}_{i=1,\ldots,I}$ in S; then the $I \times I$ matrix whose (i,j)-element is r_{t_i,t_j} must be non-negative definite since

$$\text{cov} \begin{bmatrix} y_{t_1} \\ \ldots \\ y_{t_I} \end{bmatrix} = \left[r_{t_i,t_j} \right]_{i,j=1,\ldots,I} \tag{2.3}$$

This property of r, which is necessary and sufficient for it to be the covariance of such a process, will be referred to as *positive definiteness* in the sequel. For general index sets it is not possible to find more useful criteria or characterizations of positive definiteness. However for stationary time series, i.e. for $S = Z$ and $r_{t,s}$ satisfying (2.2) much more can be said. In particular the celebrated Bochner spectral representation theorem states that a sequence $r_n, n = 0, 1, \ldots$ is the covariance function of a stationary time series if and only if there exists a nonnegative, symmetric spectral measure $S(d\omega)$ so that

$$r_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega n} S(d\omega)$$

As shown in [2, 3] there is a corresponding generalized Bochner theorem for a sequence r_n to be the covariance of an isotropic process on $\mathcal{T}$. Note that we can obviously find a subset of $\mathcal{T}$ isomorphic to Z — i.e. a sequence of nodes extending infinitely in both directions, and y_t restricted to such a set is essentially a temporally-stationary process. Thus for r_n to be a valid covariance of an isotropic process on $\mathcal{T}$ it must certainly be a valid covariance for a temporally-stationary process. However

there are additional constraints for isotropic processes– for example in $\mathcal{T}$ we can find three nodes which are all a distance two from one another (e.g. u, v, and $s \wedge t$ in Figure 4), and this implies an additional constraint on r_n. The impact of these additional constraints can be seen in the Bochner theorem in [2, 3] and also in the results described in the next section.

While the Bochner theorem is a powerful characterization result for time series and for processes on trees, it does not provide a computational procedure for testing positive definiteness or for constructing models for such processes. However for time series we do have such a method, namely the Wold representation of stationary processes via causal, autoregressive (AR) models. This representation and the well-known Levinson algorithm for its construction not only provide a procedure for testing positive-definiteness but also for constructing rational, finite-order models for stationary processes. The subject of Section 3 of this paper is the extension of this methodology to isotropic processes on trees. An important point in doing this is to realize that such a construction for time series produces a model that treats time asymmetrically (by imposing causality) in order to represent a process whose statistics do not have inherent temporal asymmetry. This is not a point that is typically highlighted since the geometry of Z is so simple. However the situation for $\mathcal{T}$ is decidedly more complex, and to carry out our program we need the following development which in essence relates the pictorial representations of Figures 1 and 4 and provides the basis for defining causal systems in scale.

An important concept associated with any homogeneous tree is the notion of a *boundary point* [2, 3, 6, 14, 15] of a tree. Consider the set of infinite sequences of nodes on such a tree, where any such sequence consists of a set of distinct nodes $t_1, t_2, \ldots$ where $d(t_i, t_{i+1}) = 1$. A boundary point is an equivalence class of such sequences where two sequences are equivalent if they differ by a finite number of nodes. For the case of Z there are two boundary points corresponding to paths toward $\pm\infty$. For $\mathcal{T}$ there are many. Let us choose one boundary point in $\mathcal{T}$ which we denote by $-\infty$. Note that from any node t there is a unique path in the equivalence class defined by $-\infty$ (i.e. a unique path from t "towards" $-\infty$ - see Figure 4). Then if we take any two nodes s and t, their paths to $-\infty$ must differ only by a finite number of points

and thus must meet at some node which we denote by $s \wedge t$ (see Figure 4). Thus, we can define a notion of *relative distance* of two nodes to $-\infty$:

$$\delta(s,t) = d(s, s \wedge t) - d(t, s \wedge t) \tag{2.4}$$

so that

$$s \preceq t \text{ ("s is at least as close to } -\infty \text{ as t") if } \delta(s,t) \leq 0 \tag{2.5}$$

$$s \prec t \text{ ("s is closer to } -\infty \text{ than t") if } \delta(s,t) < 0 \tag{2.6}$$

This also yields an equivalence relation on nodes of $\mathcal{T}$:

$$s \asymp t \leftrightarrow \delta(s,t) = 0 \tag{2.7}$$

For example, the points s, v, and u in Figure 4 are all equivalent. The equivalence classes of such nodes are referred to as *horocycles*. These equivalence classes are best visualized as in Figure 1 by redrawing the tree, in essence by picking the tree up at $-\infty$ and letting the tree "hang" from this boundary point. In this case the horocycles appear as points on the same horizontal level and $s \preceq t$ means that s lies on a horizontal level above or at the level of t. Note that in this way we make explicit the dyadic structure of the tree as depicted in Figure 1 and provide the basis for defining multiscale dynamic models.

In order to define dynamics on trees, let us again step back to take a more careful look at the usual formalism that is used for time series. Specifically, in specifying a temporal system in terms of a difference equation we make essential use of the notion of shifts or moves - e.g. in an AR model we relate y_t to y_{t-1}, y_{t-2}, *etc.* where the backward shift $z^{-1} : t \longmapsto t-1$ obviously plays an essential role in expressing the "local" dynamics, i.e. the relationship of a signal at a particular point to its values at nearby points. Moreover, thanks to the simple structure of Z, we have the luxury of using the symbol z^{-1} for two additional purposes. In particular, the backward shift z^{-1} is an isometry and in fact it and its inverse, the forward shift, generate all translations. Furthermore,we also use the symbol z^{-1} and its positive and negative powers to code signals— i.e. we represent the signal y_t by its z-transform - and all of these properties provides us with the powerful transform domain formalism for analyzing stationary, i.e. translation-invariant systems.

The situation is decidedly more complex on $\mathcal{T}$. To see this let us begin by defining moves on $\mathcal{T}$ that will be needed to provide a "calculus" for stochastic processes, i.e. for specifying local dynamics. Such moves are illustrated in Figure 1 and are introduced next :

- 0 the identity operator (no move)
- $\overline{\gamma}$ the backward shift (move one step toward $-\infty$)
- α the left forward shift (move one step away from $-\infty$ toward the left)
- β the right forward shift (move one step away from $-\infty$ toward the right)
- δ the interchange operator (move to the nearest point in the same horocycle)

Note that the richer structure of $\mathcal{T}$ requires a richer collection of moves. Also, unlike its counterpart z^{-1}, the backward shift $\overline{\gamma}$ is not an isometry (it is onto but not one-to-one), and it has two forward shift counterparts, α and β, which are one-to-one but not onto. Also, while these shifts allow us to move up and down in scale, (i.e. from one horocycle to the next), it is necessary to introduce another operator, δ, in order to define purely translational shifts at a given level. Note also that 0 and δ are isometries and that these operators satisfy the following relations (where the convention is that the left-most operator is applied first)[3]:

$$\alpha\overline{\gamma} = \beta\overline{\gamma} = 0 \tag{2.8}$$

$$\delta\overline{\gamma} = \overline{\gamma} \tag{2.9}$$

$$\delta^2 = 0 \tag{2.10}$$

$$\beta\delta = \alpha \tag{2.11}$$

Arbitrary moves on the tree can then be encoded via finite strings or *words* using these symbols as the alphabet and the formulas (2.8)–(2.11). Specifically define the language

$$\mathcal{L} = (\overline{\gamma})^* \cup (\overline{\gamma})^*\delta\{\alpha,\beta\}^* \cup \{\alpha,\beta\}^* \tag{2.12}$$

[3]Our convention will be to write operators on the right, e.g. $t\alpha, t\delta\beta$

where K^* denotes arbitrary sequences of symbols in K including the empty sequence which we identify with the operator 0. Then any move on $\mathcal{T}$ is uniquely represented by a word of this language. It is straightforward to define a *length* $|w|$ for each word in $\mathcal{L}$, corresponding to the number of shifts required in the move specified by w. Note that

$$|\overline{\gamma}| = |\alpha| = |\beta| = 1$$
$$|0| = 0 \quad , \quad |\delta| = 2 \tag{2.13}$$

Thus $|\overline{\gamma}^n| = n$, $|w_{\alpha\beta}|$ = the number of α's and β's in $w_{\alpha\beta} \in \{\alpha, \beta\}^*$, and $|\overline{\gamma}^n \delta w_{\alpha\beta}| = n + 2 + |w_{\alpha\beta}|$. This notion of length will be useful in defining the *order* of dynamic models on $\mathcal{T}$. We will also be interested exclusively in *causal* models, i.e. in models in which the output at some scale (horocycle) does not depend on finer scales. For this reason we are most interested in moves that either involve pure ascents on the tree, i.e. all elements of $\{\overline{\gamma}\}^*$, or elements $\overline{\gamma}^n \delta w_{\alpha\beta}$ of $\{\overline{\gamma}\}^* \delta \{\alpha, \beta\}^*$ in which the descent is no longer than the ascent, i.e. $|w_{\alpha\beta}| \leq n$. We use the notation $w \preceq 0$ to indicate that w is such a causal move. Note that we include moves in this causal set that are not strictly causal in that they shift a node to another on the *same* horocycle. We use the notation $w \asymp 0$ for such a move. The reasons for this will become clear when we examine autoregressive models.

Also, on occasion we will find it useful to use a simplified notation for particular moves. Specifically, we define $\delta^{(n)}$ recursively, starting with $\delta^{(1)} = \delta$ and

$$\text{If } t = t\overline{\gamma}\alpha, \text{ then } t\delta^{(n)} = t\overline{\gamma}\delta^{(n-1)}\alpha$$
$$\text{If } t = t\overline{\gamma}\beta, \text{ then } t\delta^{(n)} = t\overline{\gamma}\delta^{(n-1)}\beta \tag{2.14}$$

What $\delta^{(n)}$ does is to map t to another point on the same horocycle in the following manner: we move up the tree n steps and then descend n steps; the first step in the descent is the opposite of the one taken on the ascent, while the remaining steps are the same. That is if $t = t\overline{\gamma}^{n-1} w_{\alpha,\beta}$ then $t\delta^{(n)} = t\overline{\gamma}^{n-1}\delta w_{\alpha\beta}$. For example, referring to Figure 1, $s = u\delta^{(2)}$.

The preceding development provides us with the move structure required for the specification of local dynamics on trees. Let us turn next to the specification of "shift-invariant" systems and processes. The most general linear input/output relationship

for signals defined on the tree is simply

$$y_t = \sum_{s \in \mathcal{T}} h_{t,s} u_s \triangleq (Hu)_t \tag{2.15}$$

As with temporal systems, one would expect the requirements of various notions of shift-invariance to impose constraints on the weighting coefficients $h_{t,s}$. To see this let us first adopt an abuse of notation commonly used for time series. Specifically, if τ is an isometry of $\mathcal{T}$, we use the same notation to denote an operation on signals over $\mathcal{T}$, i.e.

$$\tau(y)_t = y_{\tau(t)} \tag{2.16}$$

(analogous to $z^{-1}y_t = y_{t-1}$). A first, rather strong notion of shift-invariance might be that if $\tau(u)$ is applied to the system for any isometry τ, then the output is $\tau(y)$, where y is the response to u. It is not difficult to check that for this to be the case we must have that

$$h_{t,s} = h(d(s,t)) \tag{2.17}$$

Note, however, that this is an exceedingly strong condition and indeed generalizes the notion of zero-phase LTI systems, i.e. systems with impulse responses such that $h(t,s) = h(|t-s|)$. Such systems obviously are not causal, and in fact are far too constrained in that they require invariance to too many isometries. In particular such an LTI system has the property that it is not only translation-invariant but also reversal invariant (i.e. $u(-t)$ yields $y(-t)$). In the case of time series we overcome this by using the smaller group of isometries generated by the shift z^{-1}. On $\mathcal{T}$, however, the shifts $\overline{\gamma}, \alpha$, and β are not isometries. For this reason it is necessary to introduce a subgroup of isometries of $\mathcal{T}$ corresponding to the other role played by z^{-1}, that of defining backward, causal, translations.

Specifically, let $(t_n)_{n \in Z}$ denote an infinite path extending in $\mathcal{T}$ back toward $-\infty$ (as $n \to -\infty$). A (one step) translation with skeleton (t_n) is an isometry of $\mathcal{T}$ that has the property that

$$\tau(t_n) = t_{n+1} \tag{2.18}$$

Since there are many such paths (t_n) there obviously are many translations, and indeed for any particular (t_n) there are numerous translations (see Figure 5). Never-

theless the class of translations represents a proper subset of all isometries, and does allow us to define a very useful notion of shift invariance:

Definition 1 (stationary systems) *A linear system H as in (2.15), acting on signals on $\mathcal{T}$, is said to be a stationary system if*[4]

$$H \circ \tau = \tau \circ H \tag{2.19}$$

for any translation τ.

A fundamental result proven in [9] is that H is stationary if and only if its weighting pattern satisfies.

$$h_{t,s} = h\left[d(t, s \wedge t), d(s, s \wedge t)\right] \tag{2.20}$$

Thus a stationary system is specified by a 2-D sequence $h(n,m), n, m \geq 0$ and, referring to Figure 1, we see that (2.20) has an intuitively appealing interpretation. Specifically $s \wedge t$ denotes the "parent" node of s and t, i.e. the finest scale node that has both s and t as descendants, and (2.20) states that $h_{t,s}$ depends only on the distances in scale from this parent node to s and to t. Roughly speaking the influence of the input at node s on the output at node t in a stationary system depends on the differences in scale and in temporal offset of the scale/shift pairs represented by t and s.

Obviously, a system satisfying (2.17) (and thus corresponding to a system that commutes with all isometries) also satisfies (2.20) (this is easily seen since $d(s,t) = d(s, s \wedge t) + d(t, s \wedge t)$). The reverse is certainly not true indicating that we have a far larger class of stationary systems as defined in Definition 1. Similarly, we can define a larger class of shift-invariant processes:

Definition 2 (stationary stochastic processes) *A zero mean (scalar) stochastic process y is said to be stationary if its covariance function is translation-invariant, i.e.*

$$r_{s,t} = r_{\tau(s),\tau(t)} \tag{2.21}$$

for any translation τ.

[4] $\circ$ denotes the composition of maps.

As shown in [9] a process is stationary if and only if

$$r_{s,t} = r[d(s, s \wedge t), d(t, s \wedge t)] \tag{2.22}$$

Thus a stationary process is specifies by a 2-D sequence $r(n,m), n,m \geq 0$. Also isotropic processes— i.e. processes for which (2.21) is satisfied for all isometries and for which (2.2) holds– are obviously stationary, but the reverse implication is not true, so that stationary processes represent a richer class of processes. Furthermore the covariance structure (2.22) in essence says that the statistical relationship between the values of a stationary process at two nodes depends on the differences in scale and in temporal offset of the two nodes. In particular from (2.22) it follows that the statistical behavior of the restriction of a stationary process to any scale (i.e. horocycle) does not depend on the scale, indicating that the concept of stationarity on the tree appears to be a natural and convenient one for capturing a notion of statistical self-similarity. Moreover, as we will see, the Haar transform yields the eigenstructure of the process at any scale, providing another tie back to wavelet transforms. In Section 4, we expand on these and related points in the context of the investigation of a class of finite-dimensional state models on dyadic trees that, in the constant-coefficient case, provides us with the class of rational linear systems satisfying the notion of stationarity we have introduced.

Let us close this discussion with a few comments. First, as shown in [9], the notions of systems and stochastic stationarity introduced in Definitions 1 and 2 are compatible in the sense that the output of a stationary system driven by a stationary input is itself stationary. In general, however, an isotropic process driving an arbitrary stationary system does not yield an isotropic output, and thus we might expect that we will have to work harder to pinpoint the class of systems that does generate isotropic processes. Furthermore, as we have indicated we are interested in constructing causal models, i.e. systems as in (2.15) with

$$h_{t,s} = 0 \text{ for } t \prec s \tag{2.23}$$

For stationary systems this corresponds to requiring

$$h(d(t, s \wedge t), d(s, s \wedge t)) = 0 \text{ for } d(t, s \wedge t) < d(s, s \wedge t) \tag{2.24}$$

Finally, let us make a brief comment about the generalization of the third use of z^{-1}, namely to define transforms. Specifically, as discussed in [6, 7, 8, 9], natural objects to consider in this context are noncommutative formal power series of the form:

$$S = \sum_{w \in \mathcal{L}} s_w \cdot w \tag{2.25}$$

We will use such tranforms in the next section in order to encode correlation functions in our generalization of the Schur recursions. In addition transforms of this type can be used to encode convolutional systems. Specifically, we can think of (2.25) as defining the system function of a system in the following manner: if the input to this system is $u_t, t \in \mathcal{T}$, then the output is given by the generalized convolution:

$$(Su)_t = \sum_{w \in \mathcal{L}} s_w u_{tw} \tag{2.26}$$

Note that in this context causality corresponds to $s_w = 0$ for all $0 \prec w$. Also it is important to realize that while (2.25), (2.26) would seem to correspond to a general class of shift-invariant systems, both classes of systems we have described– stationary and isotropic– require further restrictions. In particular for S in (2.25), (2.26) to be stationary we must have that if $\omega = \overline{\gamma}^n \delta\omega_{\alpha\beta}$, then s_ω depends only on n and $|\omega_{\alpha\beta}|$. Similarly, S is isotropic if s_ω depends only on $|w|$. Finally, for future reference we use the notation $S(0)$ to denote the coefficient of the empty word in S. Also it will be necessary for us to consider particular shifted versions of S:

$$\overline{\gamma}[S] = \sum_{w \in \mathcal{L}} s_{w\overline{\gamma}} \cdot w \tag{2.27}$$

$$\delta^{(k)}[S] = \sum_{w \in \mathcal{L}} s_{w\delta^{(k)}} \cdot w \tag{2.28}$$

where we use (2.8)–(2.11) and (2.14) to write $w\overline{\gamma}$ and $w\delta^{(k)}$ as elements of $\mathcal{L}$. Notice that, because of the relations (2.8)–(2.11), the operators $S \longrightarrow \overline{\gamma}[S]$ and $S \longrightarrow \delta[S]$ can *not* be thought of as multiplication operators on formal power series.

3 Isotropic Processes and Multiscale Autoregressive Models

In this section we investigate how multiscale isotropic processes may be finitely parametrized and how properties of processes may be checked on their associated parametrizations. In particular, as for time series it is of considerable interest to develop white-noise-driven models for processes on trees and, more specifically, models that are in some sense of finite-order. Also, as we have discussed in the preceding section, we are interested in developing a framework for constructing models that possess a causal structure in scale. Motivated by the theory of AR representations for temporally-stationary stochastic processes, we focus attention here on the class of multiscale AR models, where the pth-order version of such a model has the form

$$y_t = \sum_{\substack{w \prec 0 \\ |w| \leq p}} a_w y_{tw} + \sigma W_t \tag{3.1}$$

where W_t is white noise (i.e. it is uncorrelated from node to node) with unit variance. The form of (3.1) deserves some comment. A first question that arises is: why not look instead at models in which y_t depends only on its "strict" past, i.e. on point of the form $t\overline{\gamma}^n$. As shown in [6, 7, 8], the only model of this type that yields an isotropic output is the first-order version of (3.1), i.e.

$$y_t = a y_{t\overline{\gamma}} + \sigma W_t \tag{3.2}$$

Indeed higher-order versions of such a model yield stationary processes in the sense of Definition 2.2 and as considered in the next section. Secondly, note that the constraints placed on ω in the summation of (3.1) state that y_t is a linear combination of the white noise W_t and the values, y_{tw}, at nodes that are both at distances at most p from t (i.e. $|\omega| \leq p$) and also on the same or previous horocycles ($w \preceq 0$). Thus the model (3.1) is not strictly "causal" and is indeed an implicit specification since values of y on the same horocycle depend on each other through (3.1). For example, consider the AR(2) process, which specializing (3.1), has the form

$$y_t = a_1 y_{t\overline{\gamma}} + a_2 y_{t\overline{\gamma}^2} + a_3 y_{t\delta} + \sigma W_t \tag{3.3}$$

Note first that this is indeed an implicit specification, since if we evaluate (3.3) at $t\delta$ rather than t we see that

$$y_{t\delta} = a_1 y_{t\bar{\gamma}} + a_2 y_{t\bar{\gamma}^2} + a_3 y_t + \sigma W_{t\delta} \tag{3.4}$$

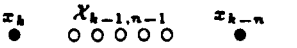
The structure of (3.3), (3.4) reveals that for a second-order model we need to consider simultaneously the coupled propagation of pairs of values $y_t, y_{t\delta}$. It also suggests that perhaps the implicit representation of (3.1) is not the most ideal one. To add further credence to this, note that the second-order AR(2) model has *four* coefficients—three a's and σ, while for second-order time series there would only be two a's. Indeed this disparity grows with increasing order as the number of coefficients a_w in (3.1) grows geometrically with p. On the other hand, as shown in [6] the constraints of isotropy place nonlinear and rather unwieldy constraints on these coefficients. For these reasons there is strong motivation to consider an alternate representation for isotropic AR models. Again it is useful to contrast the situation on $\mathcal{T}$ with that on Z. In particular, there are two equally useful parametrizations for pth order AR models for stationary time series: in terms of the p lagged coefficients $a_n, 1 \leq n \leq p$ or in terms of the p reflection or partial correlation (PARCOR) coefficients $k_n, 1 \leq n \leq p$ used in lattice filter representation of AR models. For time series, increasing the order by one increases the number of a's and k's by one. For multiscale AR models, increasing the order by one doubles the number of a's, although these are subject to a (growing!) number of nonlinear constraints. However, as we will see, if we switch to the alternate PARCOR representation, we will again only need to add only one new coefficient and will avoid completely the need for nonlinear constraints.

To begin, recall that the basic idea behind the Levinson algorithm for the construction of AR models of increasing order for stationary time series involves the consideration of both forward and backward predictions of the series based on increasing intervals of data. Specifically, consider an ordinary time series x_k and introduce the spaces $\mathcal{X}_{k,n} = \mathcal{H}\{x_k, ..., x_{k-n}\}$ where $\mathcal{H}\{\cdots\}$ denotes the linear span of the random variables indicated between the braces. Forward and backward prediction errors or "residuals" are defined as $e_{k,n} = x_k - E\{x_k|\mathcal{X}_{k-1,n-1}\}$ and $f_{k,n} = x_{k-n} - E\{x_{k-n}|\mathcal{X}_{k,n-1}\}$

respectively. The formulae

$$\begin{aligned} e_{k,n+1} &= x_k - E\{x_k|\mathcal{X}_{k-1,n}\} \\ &= x_k - E\{x_k|\mathcal{X}_{k-1,n-1}\} \\ &\quad + E\{x_k|\mathcal{X}_{k-1,n-1}\} - E\{x_k|\mathcal{X}_{k-1,n}\} \\ &= e_{k,n} - E\{x_k|\mathcal{X}_{k-1,n} \ominus \mathcal{X}_{k-1,n-1}\}\} \\ &= e_{k,n} - E\{e_{k,n}|f_{k-1,n}\}] \\ &= e_{k,n} - k_n f_{k-1,n} \end{aligned} \tag{3.5}$$

where $\mathcal{U} \ominus \mathcal{V}$ denotes the orthogonal complement of $\mathcal{V}$ in $\mathcal{U}$, show that the key to the calculation of the $(n+1)$st-order prediction error $e_{k,n+1}$ is the computation of the prediction of the forward residual $e_{k,n}$ given the backward one $f_{k-1,n}$. Similarly, the prediction of the backward residual given the forward one is needed for the calculation of backward residuals of increasing order. It is a remarkable property of stationary time series that *both prediction operators are identical*, i.e. that the same coefficient k_n in (3.5) also appears in the corresponding equation for the backward residual. This fact, which then leads to the celebrated Levinson recursions, stems from the fact that the statistics of a stationary time series are invariant under the isometry $k \longmapsto -k$. The correlation coefficient k_n of the two involved residuals is also known as the *PARCOR* coefficient of x_k and x_{k-n} given $\mathcal{X}_{k-1,n-1}$. This is illustrated in the following diagram :



Since $e_{k,0} = f_{k,0} = x_k$, we find that (3.5) and the associated Levinson recursion provide us with a method for constructing models for x_n of increasing order. In particular,if $e_{k,n}$ and $f_{k,n}$ are white, (so that all higher-order PARCOR coefficients are 0), we obtain an nth order AR model for x_n constructed in lattice form, i.e. one first-order section (specified by one PARCOR coefficient) at a time.

Let us now consider the extension of these ideas to the dyadic tree. As one might expect from the preceding discussion of AR(2) and as developed in detail in [6, 7, 8], construction of models of increasing order requires the consideration of <u>vectors</u> of forward and backward residuals of dimension that increases with model order. To

begin, let y_t be an isotropic process on a tree, and define the (nth-order) past of the node t on $\mathcal{T}$:

$$\mathcal{Y}_{t,n} \triangleq \mathcal{H}\{y_{tw} : w \preceq 0, |w| \leq n\} \tag{3.6}$$

In analogy with the time series case, the backward innovations or prediction error space, which we denote by $\mathcal{F}_{t,n}$, are defined as the variables spanning the new information in $\mathcal{Y}_{t,n}$ which are orthogonal to $\mathcal{Y}_{t,n-1}$:

$$\mathcal{Y}_{t,n} = \mathcal{Y}_{t,n-1} \oplus \mathcal{F}_{t,n} \tag{3.7}$$

so that $\mathcal{F}_{t,n}$ is the orthogonal complement of $\mathcal{Y}_{t,n-1}$ in $\mathcal{Y}_{t,n}$ (i.e. $\mathcal{F}_{t,n} = \mathcal{Y}_{t,n} \ominus \mathcal{Y}_{t,n-1}$ for $n > 0$, while $\mathcal{F}_{t,0} = \mathcal{Y}_{t,0}$). A basis for $\mathcal{F}_{t,n}$ can be obtained by defining the backward prediction errors for the "new" elements of the "past" introduced at the nth step, i.e. for $w \preceq 0$ and $|w| = n$, define

$$F_{t,n}(w) \triangleq y_{tw} - E(y_{tw}|\mathcal{Y}_{t,n-1}) \tag{3.8}$$

Then

$$\mathcal{F}_{t,n} = \mathcal{H}\{F_{t,n}(w) : |w| = n, w \preceq 0\} \tag{3.9}$$

Similarly we introduce the forward innovations or prediction error space, which we denote by $\mathcal{E}_{t,n}$. For $n = 0$, $\mathcal{E}_{t,0} = \mathcal{H}\{y_t\}$, while for $n > 0$

$$\mathcal{E}_{t,n} \triangleq (\mathcal{Y}_{t,n-1} + \mathcal{Y}_{t\overline{\gamma},n-1}) \ominus \mathcal{Y}_{t\overline{\gamma},n-1} \tag{3.10}$$

Note that $\mathcal{Y}_{t,n-1} + \mathcal{Y}_{t\overline{\gamma},n-1}$ is used here instead of $\mathcal{Y}_{t,n}$; while both spaces are equal in the case of ordinary time series (in which $\overline{\gamma}$ is replaced by z^{-1}), they differ here[5]. To obtain a basis for $\mathcal{E}_{t,n}$, we define the forward innovations

$$E_{t,n}(w) \triangleq y_{tw} - E(y_{tw}|\mathcal{Y}_{t\overline{\gamma},n-1}) \tag{3.11}$$

where w ranges over a set of words such that tw is on the same horocycle as t and at a distance at most $n-1$ from t (so that $\mathcal{Y}_{t\overline{\gamma},n-1}$ is the past of that point as well), i.e. $|w| < n$ and $w \succ 0$. Then

$$\mathcal{E}_{t,n} = \mathcal{H}\{E_{t,n}(w) : |w| < n \text{ and } w \succ 0\} \tag{3.12}$$

[5]For example $\mathcal{Y}_{t,2}$ consists of y_t, $y_{t\overline{\gamma}}$, $y_{t\overline{\gamma}^2}$, and $y_{t\delta}$. However, $\mathcal{Y}_{t,1}$ consists of y_t and $y_{t\overline{\gamma}}$, while $\mathcal{Y}_{t\overline{\gamma},1}$ consists of $y_{t\overline{\gamma}}$ and $y_{t\overline{\gamma}^2}$.

Let $E_{t,n}$ and $F_{t,n}$ denote column vectors of the elements $E_{t,n}(w)$ and $F_{t,n}(w)$, respectively. As n increases the dimensions of these residual vectors grow geometrically. Levinson recursions for isotropic processes involve the recursive computation of $F_{t,n}$ and $E_{t,n}$ as n increases. Since $F_{t,0}$ and $E_{t,0}$ both equal y_t, these recursions yield lattice structures for AR models of increasing order. As developed in [6] and as the reader may guess from the results for time series, the key to these recursions are all PARCOR coefficients involving an arbitrary pair $\{\square, \diamondsuit\}$ given the space spanned by the $\bigcirc$ in Figure 6. Furthermore, it can be verified that suitable combinations of the elementary isometries shown in this figure provide isometries

- leaving the space $\mathcal{Y}_{t\overline{\gamma},3}$ (circles) globally invariant
- exchanging two arbitrary $\square$'s or the two $\diamondsuit$.

From this it follows that *all pairs* $\{\square, \diamondsuit\}$ *possess the same PARCOR coefficients given the space spanned by the circles.* Hence, as for time series, we can show in general that *a single PARCOR or reflection coefficient is involved in each stage of the Levinson recursions.* Similar uses of the symmetries of the tree and the correlation structure of isotropic processes allows us to show that only the barycenters of the forward and backward prediction error vectors are needed to compute these reflection coefficients. These barycenters are defined as follows :

$$e_{t,n} = 2^{-[\frac{n-1}{2}]} \sum_{|w|<n, w \asymp 0} E_{t,n}(w)$$
$$f_{t,n} = 2^{-[\frac{n}{2}]} \sum_{|w|=n, w \preceq 0} F_{t,n}(w)$$

In particular in [6] the following results are proven providing a generalization of the Levinson recursions to the barycentric prediction errors for isotropic processes on $\mathcal{T}$:

Theorem 1 (barycentric Levinson recursions) *For n even:*

$$e_{t,n} = e_{t,n-1} - k_n f_{t\overline{\gamma},n-1} \tag{3.13}$$
$$f_{t,n} = \frac{1}{2}\left(f_{t\overline{\gamma},n-1} + e_{t\delta(\frac{n}{2}),n-1}\right) - k_n e_{t,n-1} \tag{3.14}$$

where

$$\begin{aligned} k_n &= cor(e_{t,n-1}, f_{t\overline{\gamma},n-1}) \\ &= cor\left(e_{t\delta(\frac{n}{2}),n-1}, e_{t,n-1}\right) \\ &= cor\left(e_{t\delta(\frac{n}{2}),n-1}, f_{t\overline{\gamma},n-1}\right) \end{aligned} \tag{3.15}$$

and $cor\,(x,y) = E(xy)/\left[E(x^2)E(y^2)\right]^{1/2}$.

For n odd:

$$e_{t,n} = \frac{1}{2}\left(e_{t,n-1} + e_{t\delta(\frac{n-1}{2}),n-1}\right) - k_n f_{t\overline{\gamma},n-1} \tag{3.16}$$

$$f_{t,n} = f_{t\overline{\gamma},n-1} - \frac{1}{2}k_n\left(e_{t,n-1} + e_{t\delta(\frac{n-1}{2}),n-1}\right) \tag{3.17}$$

where

$$k_n = cor\left(\frac{1}{2}\left(e_{t,n-1} + e_{t\delta(\frac{n-1}{2}),n-1}\right), f_{t\overline{\gamma},n-1}\right) \tag{3.18}$$

Corollary: *The variances of the barycenters satisfy the following recursions.*
For n even

$$\sigma^2_{e,n} = E\left(e^2_{t,n}\right) = \left(1 - k_n^2\right)\sigma^2_{n-1} \tag{3.19}$$

$$\sigma^2_{f,n} = E\left(f^2_{t,n}\right) = \left(\frac{1+k_n}{2} - k_n^2\right)\sigma^2_{n-1} \tag{3.20}$$

where k_n *must satisfy*

$$-\frac{1}{2} \le k_n \le 1 \tag{3.21}$$

For n odd

$$\sigma^2_{e,n} = \sigma^2_{f,n} = \sigma^2_n = \left(1 - k_n^2\right)\sigma^2_{f,n-1} \tag{3.22}$$

where

$$-1 \le k_n \le 1 \tag{3.23}$$

As we had indicated previously, the constraint of isotropy represents a significantly more severe constraint on the covariance sequence $r(n)$ of an isotropic process than on that for a stationary time series. It is interesting to note that these additional

constraints appear in the preceding development only in the form of the simple modification (3.21) of the constraint on k_n for n even over the form (3.23) that one also finds in the corresponding theory for time series. Also, as with the usual Levinson recursions for time series we can use the formulae in Theorem 1 and its corollary to obtain explicit recursions for the computation of the k_n sequence directly from the given covariance data, $r(n)$. These recursions also contain some differences from the usual results reflecting the constraints of isotropy on the tree. Rather than displaying these we describe here an alternative computational procedure generalizing the so-called Schur recursions [30, 43] for the cross-spectral densities between a given time series and its forward and backward prediction errors. In considering the generalization of these recursions to isotropic processes on trees, we must replace the z-transform power series for cross-spectral densities by corresponding formal power series of the type introduced in Section 2. Specifically for $n \geq 0$ define P_n and Q_n as:

$$P_n \triangleq \operatorname{cov}(y_t, e_{t,n}) \triangleq \sum_{w \preceq 0} E(y_t e_{tw,n}) \cdot w \tag{3.24}$$

$$Q_n \triangleq \operatorname{cov}(y_t, f_{t,n}) \triangleq \sum_{w \preceq 0} E(y_t f_{tw,n}) \cdot w \tag{3.25}$$

where we begin with P_0 and Q_0 specified in terms of the correlation function r_n of y_t:

$$P_0 = Q_0 = \sum_{w \preceq 0} r(|w|) \cdot w \tag{3.26}$$

Recalling the definitions (2.27), (2.28) of $\overline{\gamma}[S]$ and $\delta^{(k)}[S]$ for S a formal power series and letting $S(0)$ denote the coefficient of $w = 0$, we have the following generalization of the Schur recursions, proven in [6]:

Theorem 2 (Schur recursions) *The following Schur recursions on formal power series yield the sequence of reflection coefficients.*
For n even

$$P_n = P_{n-1} - k_n \overline{\gamma}[Q_{n-1}] \tag{3.27}$$

$$Q_n = \frac{1}{2}\left(\overline{\gamma}[Q_{n-1}] + \delta^{(\frac{n}{2})}[P_{n-1}]\right) - k_n P_{n-1} \tag{3.28}$$

where

$$k_n = \frac{\overline{\gamma}[Q_{n-1}](0) + \delta^{(\frac{n}{2})}[P_{n-1}](0)}{2P_{n-1}(0)} \tag{3.29}$$

For n odd

$$P_n = \frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})}[P_{n-1}]\right) - k_n\overline{\gamma}[Q_{n-1}] \tag{3.30}$$

$$Q_n = \overline{\gamma}[Q_{n-1}] - k_n\frac{1}{2}\left(P_{n-1} + \delta^{(\frac{n-1}{2})}[P_{n-1}]\right) \tag{3.31}$$

where

$$k_n = \frac{2\overline{\gamma}[Q_{n-1}](0)}{P_{n-1}(0) + \delta^{(\frac{n-1}{2})}[P_{n-1}](0)} \tag{3.32}$$

Theorems 1 and 2 provide us with the right way in which to parametrize isotropic processes. Furthermore, as developed in [6, 7, 8], we can build on these results to provide a complete generalization of the Wold decomposition of an isotropic process. In particular, lattice structures can be constructed for whitening filters, i.e. for the computation of the prediction error vectors $E_{t,n}$ and $F_{t,n}$ as outputs when y_t is taken as input. Similarly lattice forms are derived in [7] for modeling filters, i.e. systems whose output is the isotropic process when the input is the corresponding-order prediction error. Figure 2 illustrates the output, along one horocycle of a third-order modeling filter (i.e. an AR(3)-model) driven by a white $E_{t,3}$ process. We note that a major difference between these lattice structures and the usual ones for time series is that they involve lattice blocks of growing dimension, capturing the coupling along a horocycle for AR processes of higher order. Also, as with time series, statistical properties of isotropic processes may be checked using the parametrization via reflection coefficients. The main results are now listed and we again refer the reader to [7, 8] for more precise formulations of these results and their proofs.

Theorem 3 (checking properties via reflection coefficients)

1. **Characterization of AR processes** : *an isotropic process is AR(n) if and only if its reflection coefficients of order $> n$ are all zero.*

2. **Schur criterion** : *if the sequence* (r_n) *is the covariance function of an isotropic process, then the Schur recursions must yield reflection coefficients satisfying the inequalities*

$$-1 \leq k_{2n+1} \leq +1 \ , \ -\frac{1}{2} \leq k_{2n} \leq +1 \tag{3.33}$$

3. **Parametrizing AR processes** : *conversely, a finite family of coefficients satisfying the above* **strict** *inequalities (3.33) defines a unique isotropic AR process.*

4. **Regular and singular processes** : *If the sequence* (r_n) *satisfies the strict inequalities (3.33) and furthermore the condition*

$$\sum_{n=1}^{\infty} k_{2n+1}^2 + |k_{2n}| \ < \infty$$

holds true, then it is the reflection coefficient sequence of a regular (i.e. purely nondeterministic) isotropic process.

The first three of these results represent easily understood generalizations of results for time series. For example they imply that the nth and higher-order prediction error vectors of an AR(n) process are white noise processes. The fourth statement concerns itself with the issue of whether or not the value of y_t can be perfectly prediction based on data in its (infinite) past. Specifically, an isotropic process y_t is *regular* or *purely nondeterministic* if

$$\sigma^2 > 0 \tag{3.34}$$

holds, where

$$\sigma^2 \triangleq \inf \| \left(\sum_{w \succeq 0} \mu_w y_{tw} \right) - E \left(\left(\sum_{w \succeq 0} \mu_w y_{tw} \right) | \mathcal{Y}_{t\gamma^{-1},\infty} \right) \|^2 \tag{3.35}$$

and the infimum ranges over all collections of scalars $(\mu_w)_{w \succeq 0}$ where only finitely many of the μ_w are nonzero and the condition $\sum \mu_w^2 = 1$ is satisfied. In other words, no nonzero linear combination of the values of y_t on any given horocycle can be predicted exactly with the aid of knowledge of Y in the strict past, $\mathcal{Y}_{t\gamma^{-1},\infty}$ and the associated

prediction error is uniformly bounded from below. It is interesting to note that the condition for regularity for isotropic processes involves the absolute sum rather than sum of squares of the even reflection coefficients and thus is a stronger condition. This implies that there is apparently a far richer class of singular processes on $\mathcal{T}$ than on Z. This appears to be related to the complications arising in the Bochner theorem for isotropic processes on $\mathcal{T}$ and to the large size of its boundary. We refer the reader to [6, 7, 8] for further discussions of these and other points related to isotropic processes and their AR representations.

4 System Theory and Estimation for Stationary Processes and State Models

In this section we describe some of the basic concepts associated with the analysis of stationary systems and processes on the dyadic tree. To begin, let us introduce the following basic systems on $\mathcal{T}$:

$$(\gamma.u)_t = \frac{1}{2}(u_{t\alpha} + u_{t\beta}) \tag{4.1}$$

$$(\overline{\gamma}.u)_t = u_{t\overline{\gamma}} \tag{4.2}$$

It is not difficult to check that each of these systems is stationary. The system $\overline{\gamma}$ can be naturally thought of as a "backward" shift towards $-\infty$, corresponding to the coarse-to-fine interpolation operation in the fine-to-coarse Haar transform, whereas γ is a "forward–and–average" shift corresponding to the "Haar smoother". Using these operators, it is not difficult to show that a stationary system can be represented as

$$H = \sum_{i,j \geq 0} s_{i,j}\, \overline{\gamma}^i \gamma^j \tag{4.3}$$

Such a system is causal if and only if $s_{i,j}$ is nonzero only over the set $\{(i,j) : i \geq j\}$, i.e. only past inputs can influence the considered output.

The representation in (4.3) is one of two extremely useful transform-like representations of stationary systems. This one is, in particular, of use in providing a generalization of time series results on the effect of linear systems on power spectra and cross-spectra. Specifically, consider two jointly stationary processes x and y, with covariance function

$$E(x_s y_t) = r^{xy}[d(s, s \wedge t), d(t, s \wedge t)] \tag{4.4}$$

Let us define the *cross-spectrum* of x and y as the following power series:

$$R^{xy} \stackrel{\Sigma}{=} \sum_{i,j \geq 0} r^{xy}[i,j]\, \overline{\gamma}^i \gamma^j$$

Also, given a stationary transfer function as in (4.3), we introduce the following notion of an "adjoint" :

$$H^* = \sum s_{j,i} \overline{\gamma}^i \gamma^j \tag{4.5}$$

Then as shown in [9], if H and K are stationary transfer functions, the processes Hx and Ky are also jointly stationary[6], and we have the following generalization of a well-known result :

$$R^{(H_x)(K_y)} = H^* R^{xy} K^* \tag{4.6}$$

Let us now turn to the question of internal, "state" realizations of stationary systems. In this case an alternate representation to (4.3) is also of value. To define this we introduce the following family of operators which perform a smoothing of data on the same horocycle:

$$\sigma^{[i]} = \overline{\gamma}^i \gamma^i \tag{4.7}$$

This operator provides an average of the values of a signal at the 2^i nearest points on the same horocycle. For example, $(\sigma.u)_t = 1/0(u_t + u_{t\delta})$ where $\sigma = \sigma^{[1]}$ and $(\sigma^{[2]}.u)_t = \frac{1}{4}(u_t + u_{t\delta} + u_{t\delta^{(2)}} + u_{t\delta^{(2)}\delta})$. Note also that each $\sigma^{[i]}$ is an *idempotent* operator. As shown in [9] operators may be used to encode any stationary causal system via a representation of the form :

$$H = \sum_{i,j \geq 0} h_{i,j}\, \overline{\gamma}^i \sigma^{[j]} \tag{4.8}$$

In order to develop a realization theory for stationary systems, let us note that both formulae (4.3) and (4.8) are strikingly similar to the forms of system functions studied in standard 2-D system theory. While there are obvious differences – e.g. we have the relation $\gamma\overline{\gamma} = 1$ between the two variables in (4.3) and the symbol $\sigma^{[2]}$ is not simply interpretable as the square of σ– it is indeed possible to build on standard 2-D realization theories. Note in particular that even though (4.3) includes noncausal multiscale systems, it has the appearance of a 2-D quadrant-causal system, as does (4.8) since the summations are restricted to $i, j \geq 0$. Let us begin with, (4.3). Building on the 2-D analogy, if we interpret γ as the row operator and $\overline{\gamma}$ as the column generator, then it is natural to consider row–by–row scanning to define a total ordering on the $2D$ index space. This corresponds to decomposing the transfer function H according to the following two steps:

[6]This of course, is true only if Hx and Ky are well-defined, i.e. if they are finite-variance processes. As one might expect, this requires some notion of stability for the systems. We return to this point later in this section in the context of state models.

1. a bottom–up (i.e. fine–to–coarse) smoothing, followed by

2. a top-down (i.e. coarse–to–fine) propagation.

$2D$–system theory for systems having separable denominator [4, 32] may be applied here. Rational transfer functions in this latter case are of the following form:

$$H = C\,(I - \overline{\gamma} A_{\overline{\gamma}})^{-1}\, P\,(I - \gamma A_{\gamma})^{-1}\, B \tag{4.9}$$

which yields the following state space form

$$\left\{ \begin{array}{rcl} v_t & = & A_\gamma \left(\frac{v_{t\alpha} + v_{t\beta}}{2} \right) + B u_t \\ z_t & = & P_2 v_t \\ x_{t\alpha} & = & A_{\overline{\gamma}} x_t + P_1 z_{t\alpha} \\ x_{t\beta} & = & A_{\overline{\gamma}} x_t + P_1 z_{t\beta} \\ y_t & = & C x_t \end{array} \right. \tag{4.10}$$

where $P = P_1 P_2$. The first two equations define a purely "anticausal" process, whereas the last three equations define a causal zero depth process. Later in this section we describe an optimal multiscale estimation algorithm that has precisely this structure.

Now let us turn to the representation of multiscale causal systems in (4.8). Here we interpret the sequence $\sigma^{[i]}$ as the powers of the row operator and $\overline{\gamma}$ as the column operator. Then again we consider row–by–row scanning to define a total ordering of the $2D$ index space. This corresponds to decomposing the transfer function H according to the following two steps:

1. a smoothing along the considered horocycle (i.e. constant scale smoothing), followed by

2. a top-down (i.e. coarse–to–fine) propagation.

$2D$–system theory for systems having separable denominator [4, 32] may again be applied here. Rational transfer functions in this latter case are of the following form:

$$H = C\,(I - \overline{\gamma} A_{\overline{\gamma}})^{-1}\, P\,(I - \sigma A_{\sigma})^{-1}\, B \tag{4.11}$$

where it is understood that, in expanding such a formula into a power series, σ^i should be replaced by $\sigma^{[i]}$. This latter unusual feature has as a consequence the fact that no simple "time domain" translation of the "frequency domain" formula (4.11) is available. However, if A_σ is nilpotent so that $(I - \sigma A_\sigma)^{-1}$ is a finite series, we do obtain the following explicit representation for what we refer to as the *finite depth* case :

$$\begin{cases} x_{t\alpha} &= A_{\overline{\gamma}} x_t + D\left(1, \sigma, ..., \sigma^{[i]}\right) u_{t\alpha} \\ x_{t\beta} &= A_{\overline{\gamma}} x_t + D\left(1, \sigma, ..., \sigma^{[i]}\right) u_{t\beta} \\ y_t &= C x_t \end{cases} \tag{4.12}$$

where $D\left(1, \sigma, ..., \sigma^{[i]}\right)$ is a linear combination of the listed operators.

The dynamics (4.12) represent a finite-extent smoothing along each horocycle and a generalized coarse-to-fine interpolation. For example, as discussed in Section 1, the synthesis form of the Haar transform can be placed exactly in this form. It can also be shown that stationary finite depth scalar transfer functions may be equivalently expressed in the following *ARMA* form

$$H = A^{-1} D \tag{4.13}$$

where A is a causal function of *finite support* and $D = D\left(1, \sigma, ..., \sigma^{[k]}\right)$ is as in (4.12). This *ARMA* form includes as a special case the AR modeling filters for "isotropic" processes introduced in Section 3.

The preceding development, as well as the interpretation of the synthesis form of the wavelet transform provides ample motivation for the studies in [16, 17, 18, 19, 20, 48, 52] of properties and estimation algorithms for multiscale state models of the form:

$$x(t) = A(t)x(t\overline{\gamma}) + B(t)w(t) \tag{4.14}$$

$$y(t) = C(t)x(t) + v(t) \tag{4.15}$$

where $w(t)$ and $v(t)$ are independent vector white noise processes with covariances I and $R(t)$, respectively. The model class described in (4.14),(4.15) represents a noise-driven generalization of the zero-depth, causal, stationary model (4.12). Specifically we obtain such a stationary model if all of the parameters, A, B, C, and R are

constant. There are, however, important reasons to consider the more general case (and, in addition, its consideration does not complicate our analysis). First of all, one important intermediate case is that in which the system parameters are constant at each scale but may vary from scale to scale. If we let $m(t)$ denote the scale, i.e. the horocycle, on which the node t lies, we abuse notation in this case by writing $A(t) = A(m(t))$, etc. Such a model is useful for capturing the fact that data may be available at only particular scales (i.e. $C(m) \neq 0$ only for particular values of m); for example in the original context of wavelet analysis, we actually have only one measurement set, corresponding to $C(m)$ being nonzero only at the finest scale in our representation. [7] Also, by varying $A(m), B(m)$, and $R(m)$ with m we can capture a variety of scale-dependent effects. For example, dominant scales might correspond to scales with larger values of $B(m)$. Also, by building a geometric decay in scale into $B(m)$ it is possible to capture 1/f-like, fractal behavior as shown and studied in [16, 47, 50]. Finally, the general case of t-varying parameters has a number of potential uses. For example such form for $C(t)$ is clearly required to capture the situation depicted in Figure 3 in which fine scale measurements are not available at all locations. Also, it is our belief that such models will prove useful in modeling transient events localized in scale and time or space and to capture changing signal or image characteristics.

As with standard temporal state models, the second-order statistics of $x(t)$ are easily computed. In particular the covariance $P_x(t) = E[x(t)x^T(t)]$ evolves according to a Lyapunov equation on the tree:

$$P_x(t) = A(t)P_x(t\bar{\gamma})A^T(t) + B(t)B^T(t) \tag{4.16}$$

Specializing to the case in which $A(t) = A(m(t))$ and $B(t) = B(m(t))$, we can obtain a covariance that allows dependence only on scale, i.e. $P_x(t) = P_x(m(t))$, and indeed in this case we have a standard Lyapunov equation in scale :

$$P_x(m+1) = A(m)P_x(m)A^T(m) + B(m)B^T(m) \tag{4.17}$$

[7] It is important to emphasize here that the wavelet transform of this fine scale measurement— which we use as well as in the sequel– does not correspond to measurements as in (4.15) at several scales. Rather (4.15) corresponds to independent measurements at various nodes.

Also, as shown in [16, 19] the full covariance function in this case is given by

$$K_{xx}(t,s) = \Phi(m(t), m(s \wedge t)) P_x(m(s \wedge t)) \Phi^T(m(s), m(s \wedge t)) \tag{4.18}$$

where $\Phi(m,n)$ is the state transition matrix associated with $A(m)$. Specializing further to the constant coefficient case we have the following [16, 19]: if A is stable and if P_x is the unique solution to the algebraic Lyapunov equation

$$P_x = A P_x A^T + B B^T \tag{4.19}$$

then our state model generates the stationary covariance

$$K_{xx}(t,s) = A^{d(t,s\wedge t)} P_x (A^T)^{d(s,s\wedge t)} \tag{4.20}$$

Note that in the scalar case our constant coefficient model is exactly the AR(1) model introduced in the preceding section and indeed (4.19)-(4.20) reduce to

$$K_{xx}(t,s) = \left\{ \frac{B^2}{1-A^2} \right\} A^{d(s,t)} \tag{4.21}$$

In the vector case (4.20) is stationary but not, in general, isotropic. However, it is interesting to note that we do obtain an isotropic model if $AP_x = P_x A^T$, precisely the condition arising in the study of temporally-reversible vector stochastic models [1]. Let us turn now to the problems of estimating the state of (4.14) based on the measurements (4.15). Note that this framework allows us to consider not only the fusion of measurements at multiple resolutions but also the reconstruction of processes at multiple scales. Indeed in this way we can consider the resolution-accuracy tradeoff directly and can also assess the impact of fine-scale fluctuations on the accuracy of coarser scale reconstructions, a problem of some importance in applications such as the fusion of satellite IR measurement of ocean temperature variations with point measurements from ships in order to produce temperature maps at an intermediate scale. To be specific in the following development we consider the problem of optimal estimation on a finite portion of $\mathcal{T}$. This corresponds to estimation of a temporal process on a compact interval so that there is a coarsest scale (and hence a top to our subtree) denoted by $m = 0$, and a finest scale, denoted by $m = M$, at which

measurements may be available and/or reconstructions desired. As developed in [16, 17, 18, 52], the model structure (4.14), (4.15) leads to three efficient, highly parallelizable algorithmic structures for optimal multiscale estimation. A first of these is an iterative algorithm taking advantage of the fact that (4.14) defines a Markov random field structure on $\mathcal{T}$. Specifically, let Y denote the full set of measurements at all scales. Then, thanks to Markovianity we have that

$$\begin{aligned} E[x(t)|Y] &= E\{E[x(t)|x(t\overline{\gamma}), x(t\alpha), x(t\beta), Y]|Y\} \\ &= E\{E[x(t)|x(t\overline{\gamma}), x(t\alpha), x(t\beta), y(t)]|Y\} \end{aligned} \tag{4.22}$$

where the second equality in (4.22) states that given $x(t\overline{\gamma})$, $x(t\alpha)$, $x(t\beta)$, only the measurement at node t provides additional useful information about $x(t)$. From (4.22) we can then obtain an explicit representation for the optimal estimate of $x(t)$ in terms of the optimal estimates at its parent node, $t\overline{\gamma}$, at its immediate descendant nodes, $t\alpha$ and $t\beta$, and the single measurement at node t. This implicit specification is then perfectly set up for solution via Gauss-Seidel or Jacobi iteration which can be organized to have exactly the same structure as multigrid relaxation algorithms, with coarse-to-fine and fine-to-coarse sweeps that in multigrid terminology [11, 12, 26, 29, 37, 39] lead to so-called V- and W-cycle iterations. Furthermore, in such iterations all of the calculations at any particular scale can be carried out in parallel. In addition this methodology carries over completely not only to the case of nonzero depth models as in (4.12), with the additional inter-node connectivity implied by the coupling introduced by the horocycle-smoothing operator D, but also to state models on more general lattices corresponding to the interpretation of (1.11) as defining a scale-to-scale dynamic relationship for any finitely-supported QMF pair $h(n)$, $g(n)$ and thus for any compactly-supported wavelet transform. We refer the reader to [16, 19] for details and further development of this multigrid estimation methodology.

A second estimation structure applies to the case in which all system parameters depend only on scale (i.e. $A(t) = A(m(t))$, etc.). In this case, as shown in [16, 17, 19], the Haar transform, applied to each scale of the state process $x(t)$ and the measurement data $y(t)$ yields a decoupled set of estimation problems for each of the scale components. Specifically, let $x(m)$ denote the vector of all 2^m values of $x(t)$

at the mth scale, and let $y(m)$, $w(m)$, and $v(m)$ similarly. Then in this case (4.14), (4.15) can be rewritten in scale-to-scale form:

$$x(m+1) = \mathcal{A}_{m+1}x(m) + \mathcal{B}_{m+1}w(m+1) \tag{4.23}$$

$$y(m) = \mathcal{C}_m x(m) + v(m) \tag{4.24}$$

where

$$\mathcal{A}_{m+1} = \begin{bmatrix} A(m+1) & 0 & 0 & \cdots & 0 \\ A(m+1) & 0 & 0 & \cdots & 0 \\ 0 & A(m+1) & 0 & \cdots & 0 \\ 0 & A(m+1) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & A(m+1) \\ 0 & 0 & 0 & \cdots & A(m+1) \end{bmatrix} \tag{4.25}$$

$$\mathcal{B}_{m+1} = diag(B_{m+1}, ..., B_{m+1}) \tag{4.26}$$

$$\mathcal{C}_m = diag(C(m), ..., C(m)) \tag{4.27}$$

Note that $x(m)$ has half as many elements as $x(m+1)$, reflecting the fine-to-coarse decimation that occurs in multiscale representations. As shown in [16, 19], the covariances of $x(m)$ and $y(m)$ as well as the cross-covariance between x at different scales have (block-) eigenstructures specified by the Haar transform. For example if $x(t)$ is a scalar process and we look at $x(3)$, which is 8-dimensional, we find that the covariance of this vector has as its eigenvectors the columns of the following orthonormal matrix, corresponding to the (8-dimensional) discrete Haar basis consisting of vectors

representing "dilated, translated, and scaled" versions of the vector $[1,-1]^T$:

$$V_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} & -\frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \end{bmatrix} \tag{4.28}$$

Analogous bases can be defined for any dimension that is a power of two, and when $x(t)$ is a vector each of the elements of matrices as in (4.28) is replaced by a correspondingly-scaled version of the identity matrix of dimension equal to that of $x(t)$ (e.g. the $(1,1)$ block of such a matrix would be $(1/\sqrt{2})I$).

As a consequence of these observations, one would expect considerable simplification if we consider the Haar-transformed version of our estimation problem. Specifically, define the transformed variables

$$s(m) = V_x^T(m)x(m), \quad z(m) = V_y^T(m)y(m) \tag{4.29}$$

where $V_x(m)$ ($V_y(m)$) is the block-Haar transform matrix of block-size equal to the dimension of $x(t)$ ($y(t)$). In this transformed representation the system and measurement equations block-decouple completely. Specifically, the vector $s(m)$ can be decomposed into 2^m subvectors each of the same dimension as $x(t)$, and we index these as $s_{00}(m)$, $s_{01}(m)$, and $s_{ij}(m)$ for $1 \leq i \leq m-1$, $1 \leq j \leq 2^i$. Here $s_{00}(m)$ is the component corresponding to the right-most (block) basis component in $V_x(m)$ (refer to (4.28))- i.e. it is the average of the $x(t)$ at the mth horocycle (scaled by $2^{-m/2}$); $s_{01}(m)$ is then the coarsest resolution first difference coefficient (see the next-to-last column in (4.28)), while for $i \geq 1$, the s_{ij} correspond to the ith resolution first difference coefficients (note in (4.28) that there are four such coefficients at the finest resolution and two at the next, coarser scale). In a similar fashion we define the components of $z(m)$. With these definitions we find that we have a set of *completely*

decoupled standard dynamic systems in the time-like variable m:

$$s_{ij}(m+1) = A(m+1)s_{ij}(m) + B(m+1)w_{ij}(m+1),\ 0 \le i \le m-1 \tag{4.30}$$

$$s_{mj}(m+1) = B(m+1)w_{mj}(m+1) \tag{4.31}$$

$$z_{ij}(m) = C(m)s_{ij}(m) + v_{ij}(m) \tag{4.32}$$

Here $w_{ij}(m)$ and $v_{ij}(m)$ are white in all three indices, with covariances I and $R(m)$, respectively.

Recall that the dimension of $x(m)$ increases with m, indicative of the increasing detail available at finer scales. In the transformed basis this is made absolutely explicit in that we see that the dynamics (4.30), (4.31) consists of two parts: the interpolation of coarse features to finer scales (4.30) and the initiation, at each scale, of new components (4.31) representing levels of detail that can be captured at this (but not at any coarser) scale. Thus for any pair of indices i, j we have a dynamic system in m, initiated at scale $m = i$, and thus we can use standard state space smoothing techniques independently for each such system, leading to a highly parallel algorithm in which (a) we transform the available measurement data $y(m)$ to obtain $z(m)$ as in (4.29); (b) we then use standard smoothing techniques on the individual components; and (c) we inverse transform the resulting estimates of $s(m)$ to obtain the optimal estimates of $x(t)$ at all nodes. Note that the fact that each s_{ij} is initiated only at the ith scale implies that the corresponding smoother works on data only from this and finer scales, leading to a set of smoothing algorithms of different (scale) length. This is consistent with the intuition that data at any particular scale provides useful information at that scale and at coarser scales (by averaging) but not at finer scales.

We refer the reader to [16, 17] for details of this procedure and for its generalization to the case of nonzero-depth models and to arbitrary lattices associated with other wavelet transforms–i.e. to dynamic system as in (1.11) (and a significant extension of these) with other choices for the QMF's $h(n)$ and $g(n)$ than the Haar pair. Again one finds that the wavelet transformed – modified appropriately to deal with the windowing effect of smoothing multiscale measurements over a compact interval – yields a set of decoupled smoothing problems in scale. Since the wavelet transform can be computed quite quickly, this leads to an extremely efficient overall procedure. We note

also that by specializing our model to the case in which process noise variances decrease exponentially in scale we obtain a generalization of the procedure developed in [51] for the estimation of 1/f-like processes. In particular, what we have just described provides a procedure for fusing multiresolution measurements of such processes. Finally, we note that the interpretation of our models as scale-to-scale Markov processes and the dual viewpoint that the wavelet transform for such a model whitens the data in scale suggest the problems of (a) optimizing wavelet transforms in order to achieve maximal scale-to-scale decorrelation; and (b) approximating stochastic processes by such scale-to-scale Markov models. The former of these problems is discussed in [27] and the latter is touched upon in [16, 17, 27]. In particular in [17, 27] we construct approximate models of this type for a standard first-order Gauss-Markov process (i.e. with temporal correlation function of the form $\sigma^2 e^{-\alpha|t|}$) and demonstrate their fidelity in several ways including their use as the basis for the fusion and smoothing of multiresolution measurements of Gauss-Markov processes. In Figure 7 we depict the correlation function of such a unit-variance first-order Gauss-Markov process – i.e. viewing a set of 2^m samples of this process as the values of $x(m)$, Figure 7 displays the matrix of correlation coefficients of the elements of this vector. In contrast in Figure 8 we display the correlation coefficients of the elements of $s(m)$ obtained as in (4.29), but using an 8-tap QMF $h(n)$ rather than the 2-tap $h(n)$ – i.e. first the corresponding orthogonal matrix for this $h(n)$ is applied to $x(m)$, and then the resulting covariance of $s(m)$ is modified by dividing its (i,j) element by the square-root of the product of the (i,i) and (j,j) elements, yielding the matrix of correlation coefficients. As one would expect from the work on transforming kernels of integral operators in [10], the result is an almost-diagonal matrix, implying nearly perfect scale-to-scale whitening. This is further substantiated in [16, 17] (see also Figure 3) by demonstration of the high quality estimates produced if such remaining inter-scale correlation is neglected.

While the preceding algorithm provides a very efficient procedure for multiscale fusion, its use does require that all model parameters vary only with scale and thus are constant on each horocycle. For example this implies that if any measurement is available at any particular scale, than a full set of measurements is available at that scale. In contrast, the result shown in Figure 3 (a),(b) corresponds to a situation in which we

have only sparse, fine scale measurements from a 1/f-like model of the type described in [50, 51], together with full-coverage, but coarser-resolution measurements, while Figure 3 (c) and (d) correspond to the analogous situation for a first-order Gauss-Markov process. In particular in each case 16 fine scale measurements are taken at each end of the 64-point signal, together with coarse measurements of 4-point averages of this signal. While the wavelet-transform-based smoothing algorithm does not apply to this case, the multigrid method described previously does (using in the case of (c) and (d) an approximate model of the form of (4.14), (4.15) for the Gauss-Markov process), as does the following approach which not only provides an extremely efficient algorithm for multiscale fusion but also illuminates several system-theoretic issues on dyadic trees. Specifically, as developed in detail in [16, 18, 19], there is a nontrivial generalization of the so-called Rauch-Tung-Striebel (RTS) smoothing algorithm for causal state models [42]. Recall that the standard RTS algorithm involves a forward Kalman filtering sweep followed by a backward sweep to compute the smoothed estimates. The generalization to our models on trees has the same structure, with several important differences. First for the standard RTS algorithm the procedure is completely symmetric with respect to time – i.e. we can start with a reverse-time Kalman filtering sweep followed by a forward smoothing sweep. For processes on trees, the Kalman filtering sweep *must* proceed from fine-to-coarse followed by a coarse-to-fine smoothing sweep[8].

Furthermore the Kalman filtering sweep, is somewhat more complex for processes on trees. In particular one full step of the Kalman filter recursion involves a measurement update, *two* parallel backward predictions (corresponding to backward prediction along both of the paths descending from a node), and the *fusion* of these predicted estimates. Specifically, as depicted in Figure 9, the fine-to-coarse Kalman filter step has as its goal the recursive computation of $\hat{x}(t|t)$, the best estimate of $x(t)$ based on data in the descendant subtree with root node t. As in usual Kalman filtering if $\hat{x}(t|t+)$ denotes the best estimate based on all of the same data *except the*

[8]The reason for this is not very complex. To allow the measurement on the tree at one point to contribute to the estimate at another point on the same level of the tree, one must use a recursion that first moves up and then down the tree.

measurement at node t, we obtain a straightforward update step to produce $\hat{x}(t|t)$:

$$\hat{x}(t|t) = \hat{x}(t|t+) + K(t)[y(t) - C(t)\hat{x}(t|t+)] \tag{4.33}$$

$$K(t) = P(t|t+)C^T(t)V^{-1}(t) \tag{4.34}$$

$$V(t) = C(t)P(t|t+)C^T(t) + R(t) \tag{4.35}$$

and

$$P(t|t) = [I - K(t)C(t)]P(t|t+) \tag{4.36}$$

Here $P(t|t)$ and $P(t|t+)$ are the error covariances associated with $\hat{x}(t|t)$ and $\hat{x}(t|t+)$, respectively. Working back one-step, we see that $\hat{x}(t|t+)$ represents the *fusion* of information in the subtree under $t\alpha$ and under $t\beta$. Thus we might expect that $\hat{x}(t|t+)$ could be computed from the one-step-backward-predicted estimates $\hat{x}(t|t\alpha)$ and $\hat{x}(t|t\beta)$ of $x(t)$ based separately on the information in the subtrees with root $t\alpha$ and root $t\beta$, respectively. Indeed as shown in [16, 19]

$$\hat{x}(t|t) = P(t|t+)[P^{-1}(t|t\alpha)\hat{x}(t|t\alpha) + P^{-1}(t|t\beta)\hat{x}(t|t\beta)] \tag{4.37}$$

$$P(t|t+) = [P^{-1}(t|t\alpha) + P^{-1}(t|t\beta) - P_x^{-1}(t)]^{-1} \tag{4.38}$$

Finally to complete the recursion, $\hat{x}(t|t\alpha)$ and $\hat{x}(t|t\beta)$ are computed from $\hat{x}(t\alpha|t\alpha)$ and $\hat{x}(t\beta|t\beta)$, respectively, in identical fashions. Specifically, each of these calculations represents a one-step-backward prediction. It is not surprising, then that a backward version of the model (4.14) plays a role here. Indeed, as shown in [16]

$$\hat{x}(t|t\alpha) = F(t\alpha)\hat{x}(t\alpha|t\alpha) \tag{4.39}$$

$$P(t|t\alpha) = F(t\alpha)P(t\alpha|t\alpha)F^T(t\alpha) + Q(t\alpha) \tag{4.40}$$

where

$$F(t) = A^{-1}(t)[I - B(t)B^T(t)P_x^{-1}(t)] \tag{4.41}$$

$$Q(t) = A^{-1}(t)B(t)Q(t)B^T(t)A^{-T}(t) \tag{4.42}$$

$$Q(t) = I - B^T(t)P_x^{-1}(t)B(t) \tag{4.43}$$

The prediction (4.39-4.43) and update (4.33-4.36) steps correspond to the analogous steps in the usual Kalman filter (although here we *must* use the backward model in

the prediction step), while the fusion step (4.37)-(4.38) has no counterpart in usual Kalman filtering. The interpretation of (4.37)-(4.38) is that we are fusing together two estimates each of which incorporates one set of information that is independent of that used in the other—i.e. the measurements in the $t\alpha$ and $t\beta$ subtrees– and one common information source, namely the prior statistics of $x(t)$. Eq. (4.38) ensures that this common information is accounted for only once in the fused estimate. Once the top of the overall tree is reached we, of course, have the optimal smoothed estimate at that node. As shown in [16, 18, 19], it is then possible to compute the optimal smoothed estimated in a recursive fashion moving down the tree, from coarse to fine. This recursion combines the smoothed estimate $\hat{x}_s(t\overline{\gamma})$ with the filtered estimates from the upward sweep to produce $\hat{x}_s(t)$:

$$\hat{x}_s(t) = \hat{x}(t|t) + P(t|t)F^T(t)P^{-1}(t\overline{\gamma}|t)[\hat{x}_s(t\overline{\gamma}) - \hat{x}(t\overline{\gamma}|t)] \tag{4.44}$$

Note that this algorithm also has a highly parallel, and in this case pyramidal, structure, since all calculations, on either the fine-to-coarse or coarse-to-fine sweep can be computed in parallel.

Equations (4.34-4.36), (4.38), and (4.40-4.43) define, in essence a Riccati equation on the dyadic tree. As for standard Riccati equations, it is possible to relate properties of the solution of this equation to system-theoretic properties. For example, one can show that suitably defined notions of uniform complete reachability and uniform complete observability imply upper and lower positive-definite bounds on the error covariance. Here since the Riccati equation propagates up the tree, the analysis of reachability and observability relate to systems defined recursively from fine-to-coarse scale—i.e. noncausal systems as in the first two equations of (4.10). One might also expect that one could obtain results on the stability of the error dynamics and asymptotic behavior in the constant parameter case. This is indeed the case, but there are several issues that complicate the analysis. Specifically, in standard Kalman filtering analysis the Riccati equation for the error covariance can be viewed simply as the covariance of the error equation, which can be analyzed directly without explicitly examining the state dynamics, since the error evolves as a state process itself. This is not the case here in general. First, while the process $x(t)$ is defined recursively moving

down the tree, the filtered estimate $\hat{x}(t|t)$ is defined by a recursion in the opposite direction. This difficulty cannot be overcome in general simply by reversing one of these processes, as the reversal process does not, in general, produce a system driven by white noise.[9] Also, unlike the standard situation, our Riccati equation explicitly involves the prior state covariance $P_x(t)$, arising as we've seen to prevent the double counting of prior information.

There is, however, a way in which these difficulties can be avoided, essentially by setting P_x^{-1} to zero. In particular, as discussed in [16, 18] if we do this in (4.33)-(4.43), the estimates produced have the interpretation as maximum likelihood (ML) estimates. A variation of the RTS algorithm we have described here uses this ML procedure to propagate to the top of the tree, at which point prior information is then incorporated, followed by the coarse-to-fine sweep (4.44). To see what happens to the Riccati equation and error dynamics in this case, let us focus on the scale-varying case, i.e. the case in which all parameters depend only on $m(t)$. In this case the same is true of the error covariances, yielding the following Riccati equation in scale:

$$\begin{aligned} P_{ML}(m|m+1) &= A^{-1}(m+1)P_{ML}(m+1|m+1)A^{-T}(m+1) \\ &+ G(m+1)Q(m+1)G^T(m+1) \qquad (4.45) \\ P_{ML}^{-1}(m|m) &= 2P_{ML}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \qquad (4.46) \end{aligned}$$

where

$$G(m) = -A^{-1}(m)B(m) \qquad (4.47)$$

This Riccati equation differs from the usual equation only in the presence of the factor of 2 in (4.46), representing the doubling of information arising in the fusion step. In this case we can also write a direct fine-to-coarse state form for the ML estimation error $\tilde{x}_{ML}(t|t) = x(t) - \hat{x}_{ML}(t|t)$:

$$\tilde{x}_{ML}(t|t) = \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))A^{-1}(m(t)+1)(\tilde{x}_{ML}(\alpha t|\alpha t) + \tilde{x}_{ML}(\beta t|\beta t))$$

[9] In particular the backward models used in [16, 18, 19] to write $x(t)$ in terms of $x(t\alpha)$ and in terms of $x(t\beta)$ yield driving noises which are martingale differences *with respect to the partial order defined on the tree.*

$$- \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))G(m(t)+1)(w(\alpha t) + w(\beta t)) - K_{ML}(m(t))v(t) \tag{4.48}$$

$$K_{ML}(m) = P(m|m)C^T(m)R^{-1}(m) \tag{4.49}$$

In [16, 18] we provide a detailed analysis of (4.45)-(4.49). In particular the stability of the error dynamics (4.48) under reachability and observability conditions is established. The notion of stability, however, deserves further comment. Intuitively what we would like stability to mean is that the state of the recursion up the tree decays to 0 as we propagate farther and farther away from the initial level of the tree. Note, however, that as we move up the tree the state at any node is influenced by a geometrically increasing number of nodes at the initial level. Thus in order to study asymptotic stability it is *necessary* to consider an infinite dyadic tree, with an infinite set of initial conditions corresponding to all nodes at the initial level. The implications of this are most easily seen in the constant parameter case. In this case we have that if (A, B) is a reachable pair and (C, A) observable, then

$$\begin{aligned} \overline{P}_\infty &= \frac{1}{2}A^{-1}\overline{P}_\infty A^{-T} + \frac{1}{2}GQG^T \\ &- K_\infty(\frac{1}{2}CA^{-1}\overline{P}_\infty A^{-T}C^T + \frac{1}{2}CGQG^TC^T + R)K_\infty^T \end{aligned} \tag{4.50}$$

where

$$K_\infty = \overline{P}_\infty C^T R^{-1} \tag{4.51}$$

Moreover, the autonomous dynamics of the steady-state ML filter, i.e.

$$e(t) = \frac{1}{2}(I - K_\infty C)A^{-1}(e(\alpha t) + e(\beta t)) \tag{4.52}$$

is exponentially l_2 stable, i.e. the l_2 norm of all values of $e(t)$ along an entire horocycle converges exponentially to zero as $m(t) \to 0$. As shown in [16, 18] this is equivalent to all eigenvalues of the Kalman filter error dynamics matrix

$$\frac{1}{2}(I - K_\infty C)A^{-1} \tag{4.53}$$

having magnitude less than $\frac{\sqrt{2}}{2}$.

5 Conclusions

In this paper we have outlined a mathematical framework for the multiresolution modeling and analysis of stochastic processes. As we have discussed, the theory of multiscale signal analysis and wavelet transforms leads naturally to the investigation of multiscale statistical representations and dynamic models on dyadic trees and lattices. The rich structure of the dyadic tree requires that we take some care in the specification of such models and in the generalization of standard time series notions. In particular, we have seen that in this context there are two natural concepts of shift invariance which provide new ways in which to capture notions of scale-invariant statistical descriptions. In addition, the observation that the scale variable is time-like in nature leads to a natural notion of "causal" dynamics in scale: from fine to coarse; however the tree provides only a partial ordering of points, requiring that we take some care in defining the "past".

In part of our work we have described the multiscale autoregressive modeling of isotropic processes, i.e. processes satisfying our stronger notion of statistical shift-invariance. As we have seen, the usual AR representation of time series is not a particularly convenient one thanks both to the geometric explosion of points in the "past" as we increase system order and to the nonlinear constraints isotropy imposes on the AR coefficients. In contrast, we have seen that it is possible to construct a generalization of the reflection-coefficient-based lattice representation for such models, including generalized Levinson and Schur recursions. As we have illustrated such models can be used to generate fractal-like signals.

The other part of our work was motivated by our weaker notion of stationarity which in essence says that the correlation between two values in our multiscale representation depends on the difference in scale and location of the two points. As we have seen, this framework leads to state models evolving from coarse-to-fine scales on dyadic trees. We have described some of our work on a basic system theory for such models and have also discussed an estimation framework that allows us to capture the fusion of measurements at differing resolutions. In addition the structure of these models leads to several extremely efficient and highly parallel estimation structures:

a multiscale iterative algorithm that can be arranged so as to have the same form as well-known multigrid algorithms for solving partial differential equations; an algorithm using wavelet transforms to decouple the estimation procedure into a large set of far simpler parallel estimation algorithms; and a pyramidal algorithm that introduces a generalization of the Kalman filter and the associated Riccati equation.

As we have discussed and illustrated, these models appear to be useful for a rich variety of processes including the 1/f-like models as introduced in [50, 51] and standard first-order Gauss-Markov processes. Much, of course, remains to be done in developing this theory, in investigating the processes that can be conveniently and accurately represented within this framework, and in applying these results to problems of practical importance such as sensor fusion, noise rejection, multisensor or multiframe data registration and mapping, and segmentation. Among the theoretical topics under investigation are the development of model fitting and likelihood function-based methods for parameter estimation and segmentation and the development of a detailed theory of approximation of stochastic processes including a specification of those processes that can be "well"-approximated by models of the type we have introduced. Of particular interest is the dynamic interpretation of so-called wave packet transforms [21] in which the wavelet coefficients are subjected to further decomposition through the same filter pair used in the wavelet transform. Viewing this from our dynamic synthesis perspective, this would appear to correspond to a class of higher-order models. Identifying and analyzing this model class, however, remains for the future.

References

[1] B. Anderson and T. Kailath, "Forwards, backwards, and dynamically reversible Markovian models of second-order processes," *IEEE Trans. Circuits and Systems*, CAS-26, no. 11, 1978, pp. 956-965.

[2] J. Arnaud, "Fonctions spheriques et fonctions definies-positives sur l'arbre homogene", C.R. Acad. Sc., Serie A, 1980, pp. 99-101.

[3] J. Arnaud and B. Letac, "La formule de representation spectrale d'un processus gaussien stationnaire sur un arbre homogene," Laboratoire de Stat. et. Prob.-U.A.-CNRS 745, Toulouse.

[4] S. Attasi, "Modeling and Recursive Estimation for Double Indexed Sequences," in *System Identification: Advances and Case Studies*, R.K. Mehra and D.G. Lainiotis, eds., Academic Press, NY 1976.

[5] M. Barnsley, *Fractals Everywhere*, Academic Press, San Diego, 1988.

[6] M. Basseville, A. Benveniste, and A.S. Willsky "Multiscale Autoregressive Processes, Part I: Schur-Levinson Parametrizations", submitted to *IEEE Transactions on ASSP.*

[7] M. Basseville, A. Benveniste, and A.S. Willsky "Multiscale Autoregressive Processes, Part II: Lattice Structures for Whitening and Modeling", submitted to *IEEE Transactions on ASSP.*

[8] M. Basseville, A. Benveniste, A.S. Willsky, and K.C. Chou, "Multiscale Statistical Processing: Stochastic Processes Indexed by Trees," in *Proc. of Int'l Symp. on Math. Theory of Networks and Systems*, Amsterdam, June 1989.

[9] A. Benveniste, R. Nikoukhah, and A.S. Willsky, "Multiscale System Theory", Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu, HI, December 1990.

[10] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast Wavelet Transforms and Numerical Algorithms I", to appear in *Comm. Pure and Appl. Math.*

[11] A. Brandt, "Multi-level adaptive solutions to boundary value problems," Math. Comp. Vol. 13,1977, pp.333-390.

[12] W. Briggs, "*A Multigrid Tutorial*, SIAM, Philadelphia, PA, 1987.

[13] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, vol. 31, pp. 482–540, 1983.

[14] P. Cartier, "Harmonic analysis on trees", Proc. Symos. Pure Math., Vol 26, Amer. Math. Soc. Providence, R.I., 1974, pp.419-424.

[15] P. Cartier, "Geometrie et analyse sur les arbres",Seminaire Bourbaki, 24eme annee, Expose no. 407, 1971/72.

[16] K.C. Chou, *A Stochastic Modeling Approach to Multiscale Signal Processing*, MIT, Department of Electrical Engineering and Computer Science, Ph.D. Thesis, (in preparation).

[17] K.C. Chou, S. Golden and A.S. Willsky, "Modeling and Estimation of Multiscale Stochastic Processes", *Int'l Conference on Acoustics, Speech, and Signal Processing"*, Toronto, April 1991.

[18] K.C. Chou and A.S. Willsky, "Multiscale Riccati Equations and a Two-Sweep Algorithm for the Optimal Fusion of Multiresolution Data", *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, December 1990.

[19] K.C. Chou, A.S. Willsky, A. Benveniste, and M. Basseville,"Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," Proc. 28th IEEE Conf. on Dec. and Cont., Tampa, Dec. 1989.

[20] S.C. Clippingdale and R.G. Wilson, "Least Squares Image Estimations on a Multiresolution Pyramid", Proc. of the 1989 Int'l Conf. on Acoustics, Speech, and Signal Proceeding.

[21] R.R. Coifman, Y. Meyer, S. Quake and M.V. Wickehauser, "Signal Processing and Compression with Wave Packets", preprint, April 1990.

[22] I. Daubechies,"Orthonormal bases of compactly supported wavelets", Comm. on Pure and Applied Math. 91, 1988, pp. 909-996.

[23] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Information Theory*, 36, 1990, pp. 961-1005.

[24] I. Daubechies, A. Grossman, and Y. Meyer, "Painless non-orthogonal expansions," J. Math. Phys. 27, 1986, pp. 1271-1283.

[25] P. Flandrin, "On the Spectrum of Fractional Brownian Motions", *IEEE Transactions on Information Theory*, Vol. 35, 1989, pp. 197-199.

[26] J. Goodman and A. Sokal, "Multi-grid Monte Carlo I. conceptual foundations," Preprint, Dept. Physics New York University, New York, Nov. 1988; to be published.

[27] S. Golden, *Identifying Multiscale Statistical Models Using the Wavelet Transform*, S.M. Thesis, M.I.T. Dept. of EECS, May 1991.

[28] A. Grossman and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", SIAM J. Math. Anal. 15, 1984, pp. 723-736.

[29] W. Hackbusch and U. Trottenberg, Eds., *Multigrid Methods and Applications*, Springer-Verlag, N.Y., N.Y., 1982.

[30] T. Kailath, "A theorem of I. Schur and its impact on modern signal processing", in *Schur Methods in Operator Theory and Signal Processing,* I. Gohberg Ed., Operator theory: advances and Applications, Vol. 18, Birkhäuser (Basel, Boston, Stuttgart), 1986.

[31] M. Kim and A.H. Tewfik, "Fast Multiscale Detection in the Presence of Fractional Brownian Motions", *Proceedings of SPIE Conference on Advanced Algorithms and Architecture for Signal Processing V*, San Diego, CA, July 1990.

[32] T. Lin, M. Kawamata and T. Higuchi, "New necessary and sufficient conditions for local controllability and observability of 2-D separable denominator systems," *IEEE Trans. Automat. Control*, AC-32, pp. 254-256, 1987.

[33] S.G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Transactions on Pattern Anal. and Mach. Intel.*, Vol. PAMI-11, July 1989, pp. 674-693.

[34] S.G. Mallat, "Multifrequency Channel Decompositions of Images and Wavelet Models", *IEEE Transactions on ASSP*, Vol. 37, December 1989, pp. 2091-2110.

[35] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, New York, 1982.

[36] B. B. Mandelbrot and H.W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications", *SIAM Review*, Vol. 10, October 1968, pp. 422-436

[37] S. McCormick, *Multigrid Methods*, Vol. 3 of the SIAM Frontiers Series, SIAM, Philadelphia, 1987.

[38] Y. Meyer, "L'analyse par ondelettes", Pour la Science, Sept. 1987.

[39] D. Paddon and H. Holstein, Eds. *Multigrid Methods for Integral and Differential Equations*, Clarendon Press, Oxford, England, 1985.

[40] A.J. Pentland, "Fast Surface Estimation Using Wavelet Bases", MIT Media Lab Vision and Modeling Group TR-142, June 1990.

[41] A.P. Pentland, "Fractal-Based Description of NaturalScenes", *IEEE Transactions on Patt. Anal. and Mach. Intel.*, Vol. PAMI-6, November 1989, 661-674.

[42] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems," AIAA Journal, Vol. 3, No. 8, Aug. 1965, pp. 1445-1450.

[43] E.A. Robinson, S. Treitel, "Maximum Entropy and the Relationship of the Partial Autocorrelation to the Reflection Coefficients of a Layered System," *IEEE Trans. on ASSP*, vol. 28 Nr 2, 224–235, 1980.

[44] M.J. Smith and T.P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders", IEEE Trans. on ASSP 34, 1986, pp. 434-441.

[45] R. Szeliski, "Fast Surface Interpolation Using Hierarchical Basis Function", *IEEE Transactions on PAMI*, Vol. 12, No. 6, June 1990, pp. 513-528.

[46] D. Terzopoulos, "Image Analysis Using Multigrid Relaxation Methods", *IEEE Transaction on PAMI*, Vol. PAMI-8, No. 2, March 1986, pp. 129-139.

[47] A.H. Tewfik and M. Kim, "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motions", submitted to *IEEE Transactions on Information Theory.*

[48] M. Todd and R. Wilson, "An Anisotropic Multi-Resolution Image Data Compression Algorithm", Proc. of the 1989 Int'l Conf. on Acoustics, Speech, and Signal Processing.

[49] M. Vetterli, and C. Herley, "Wavelet and Filter Banks: Relationships and New Results", *Proceedings of the ICASSP*, Albuquerque, NM, 1990.

[50] G.W. Wornell, "A Karhunen-Loeve-Like Expansion for 1/f Processes via Wavelets", *IEEE Transactions on Information Theory*, Vol. 36, No. 9, July 1990, pp. 859-861.

[51] G.W. Wornell and A.V. Oppenheim, "Estimation of Fractal Signs from Noisy Measurements Using Wavelets", submitted to *IEEE Transactions on ASSP.*

[52] A.S. Willsky, K.C. Chou, A. Benveniste, and M. Basseville,"Wavelet Transforms, Multiresolution Dynamical Models, and Multigrid Estimation Algorithms", *1990 IFAC World Congress*, Tallinn, USSR, August 1990.

[53] A. Witkin, D. Terzopoulos and M. Kass, "Signal Matching Through Scale Space", *Int. J. Comp. Vision*, Vol 1, 1987, pp. 133-144.

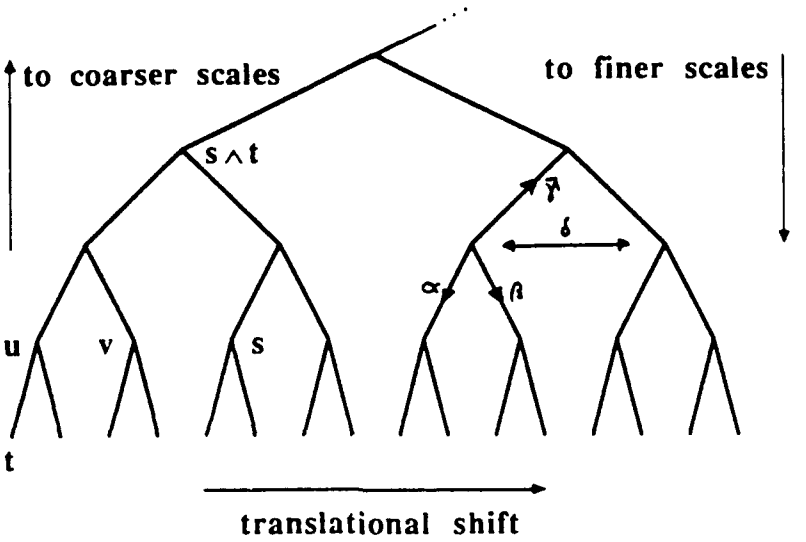


Figure 1: The dyadic tree, in which each level of the tree corresponds to a single scale in a multiscale representation. The nodes here correspond to scale/shift pairs (m, n).

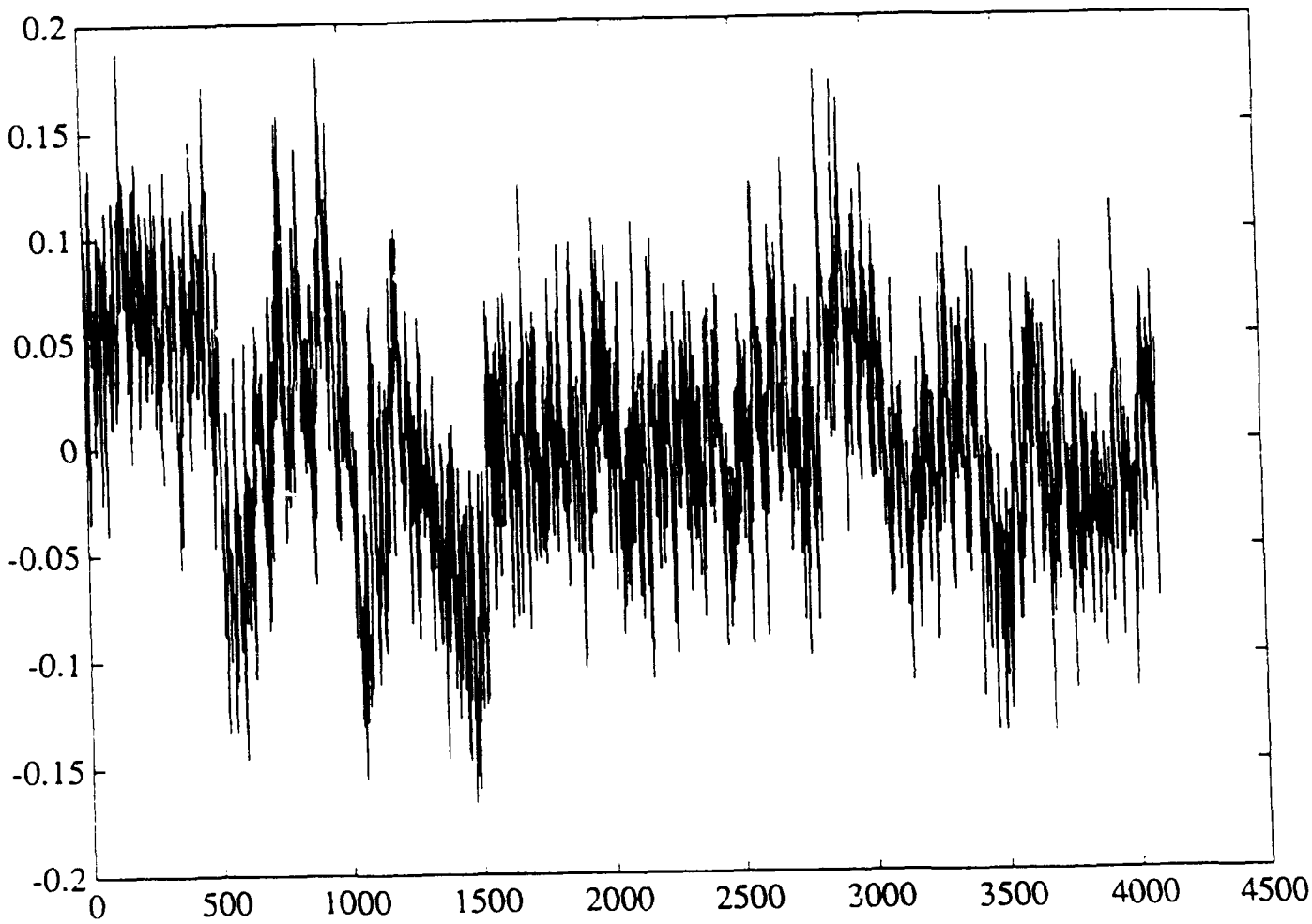


Figure 2: A signal generated by a third-order multiscale autoregressive model, as described in Section 3.

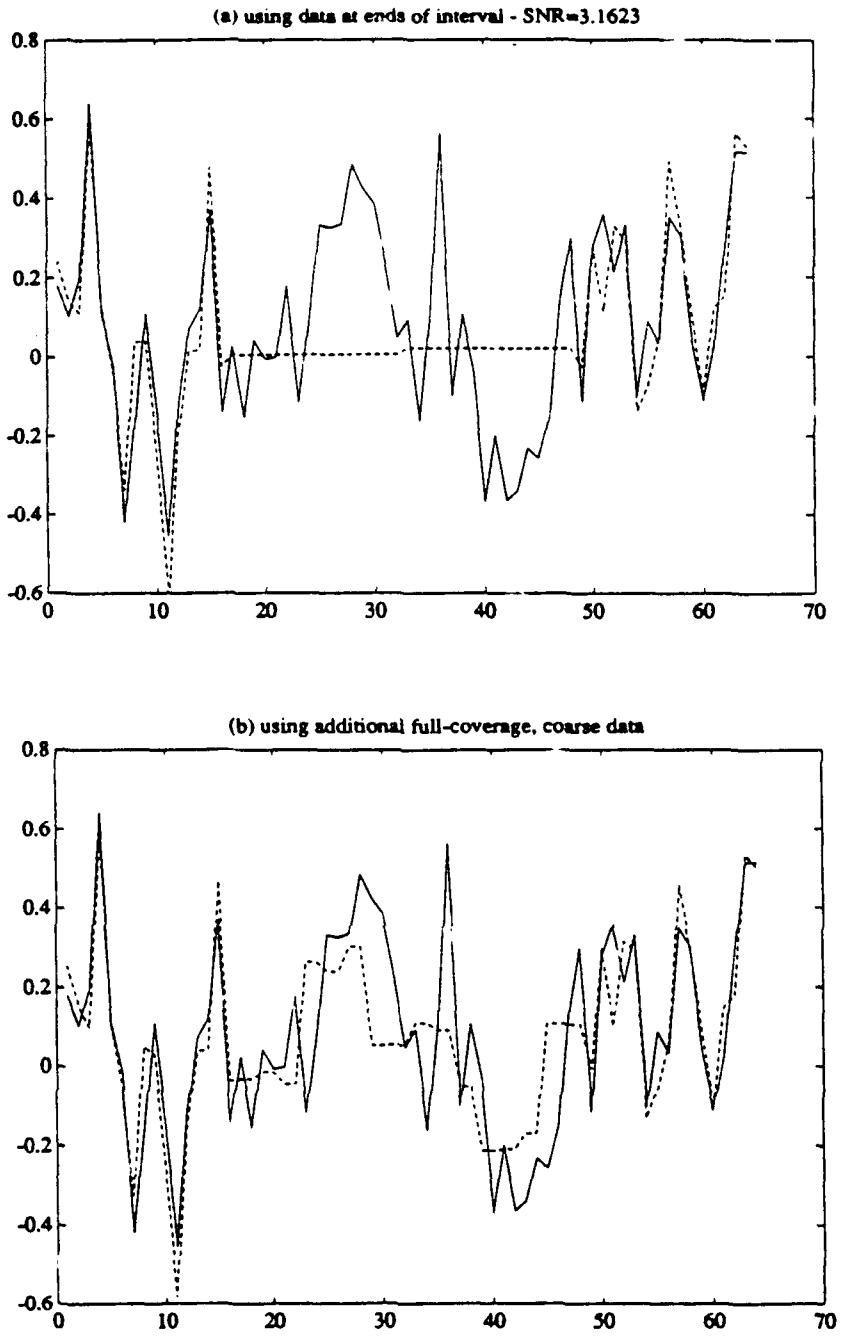
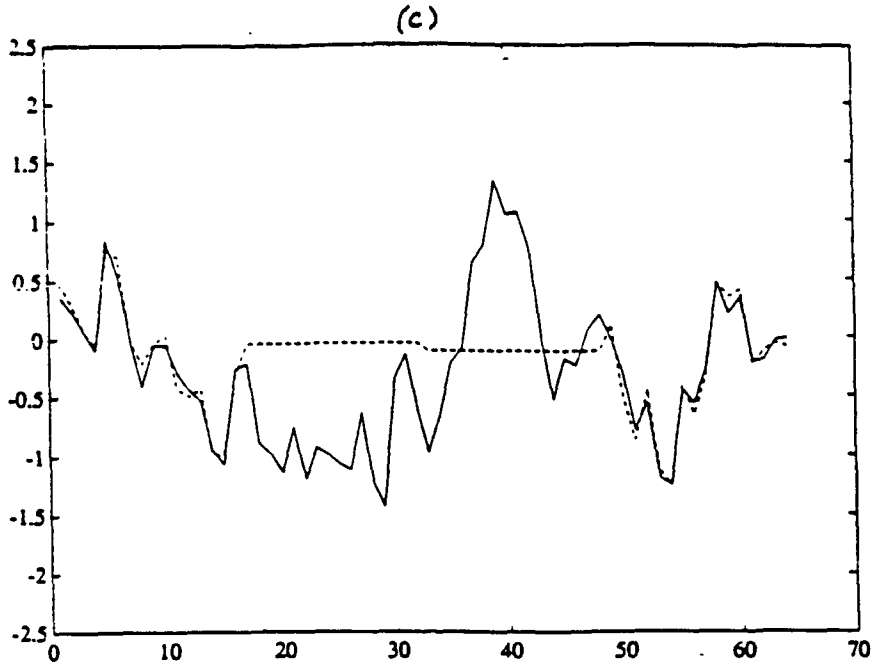


Figure 3: Illustrating multiscale data fusion using the techniques described in Section 4. In (a) and (b) a signal with a 1/f-like spectrum (as described in [50]), shown as a solid line in both plots, is reconstructed based on measurements. In (a) data is available only at the two ends of the interval, while in (b) coarse scale (i.e. locally averaged) measurements are fused to improve signal interpolation. In (c) and (d) analogous results are shown for the multiscale data fusion and interpolation of a Gauss-Markov process.



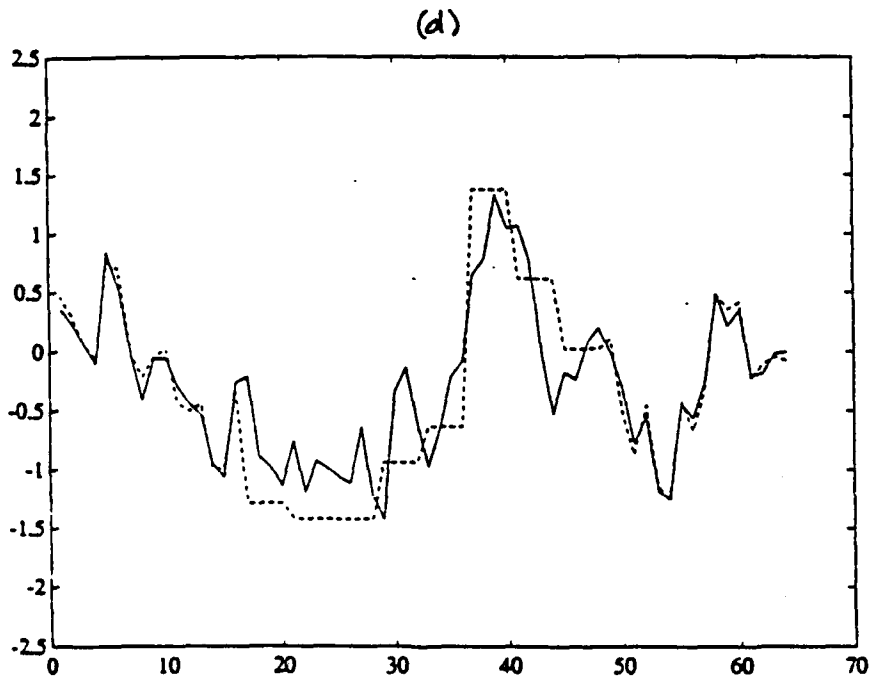


Figure 3: (continued)

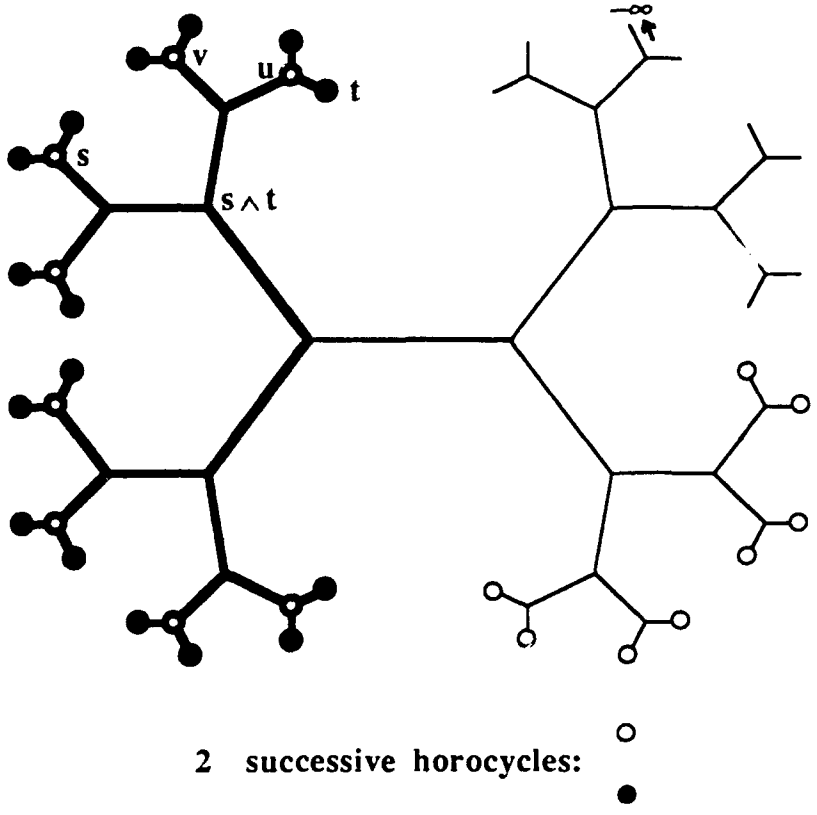


Figure 4: A more symmetric depiction of the dyadic tree, illustrating the notion of a boundary point $-\infty$, horocycles, and the "parent" $s \wedge t$ of nodes s and t (see the text for explanations).

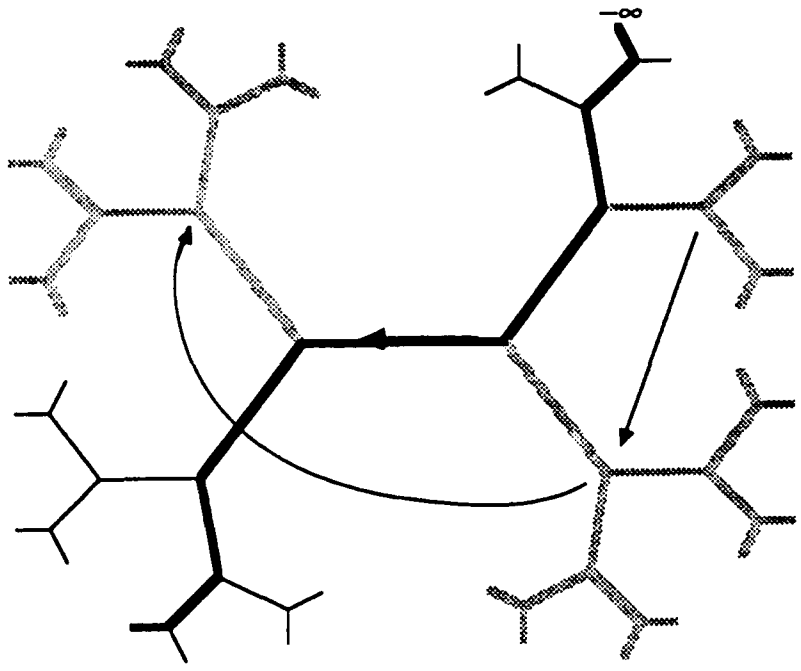


Figure 5: Illustrating (in bold) the skeleton of a translation. As indicated in the figure, any translation with this skeleton must map the subtree extending away from any node on the skeleton onto the corresponding subtree of the next node. There are, however, many ways in which this can be done (e.g. by "pivoting" isometries within any of these subtrees).

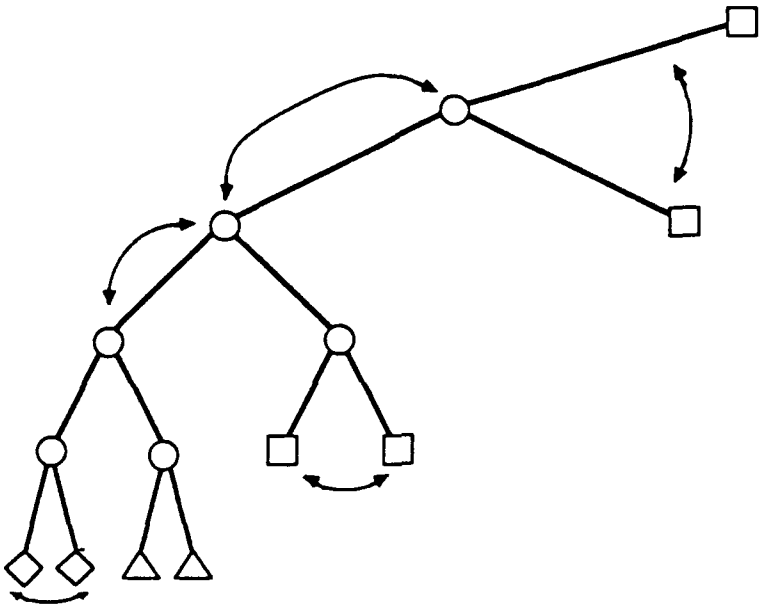


Figure 6: Illustrating the nature of the construction required in developing recursions for $E_{t,n}$ and $F_{t,n}$. Here if t is the node in the lower left-hand corner, then the elements of $E_{t,4}$ are the prediction errors at the two points indicated by diamonds given the data $\mathcal{Y}_{t\overline{\gamma},3}$ spanned by the circles. The elements of $F_{t\overline{\gamma},4}$ are the prediction errors at the four points indicated by squares given again the data in $\mathcal{Y}_{t\overline{\gamma},3}$. The elementary "pivoting" isometries indicated in the figure allow us to obtain the result on PARCOR coefficients described in the text.

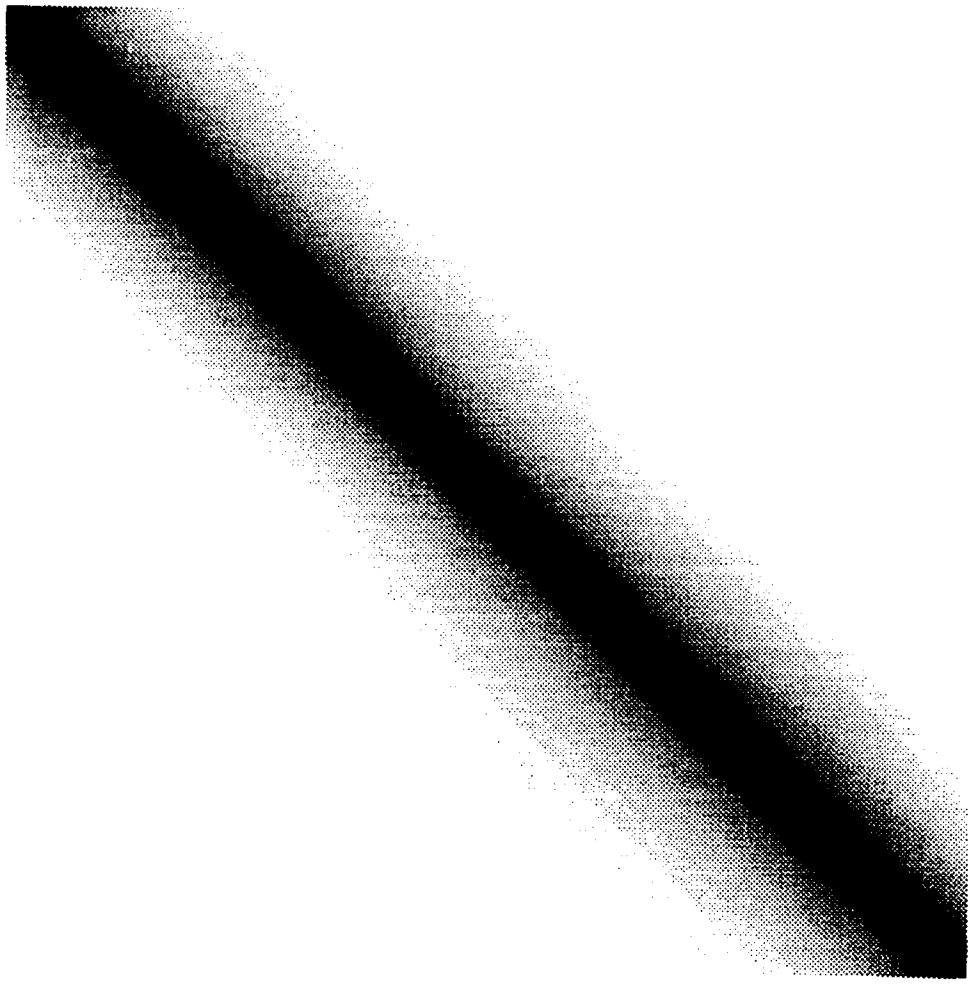


Figure 7: Illustrating the covariance matrix of a set of samples of a first-order Gauss-Markov process with covariance of the form $\exp^{-\alpha|t|}$. Black corresponds to a value of 1 with lighter shades representing smaller values. The covariance of this process decays exponentially as we move away from the main diagonal.

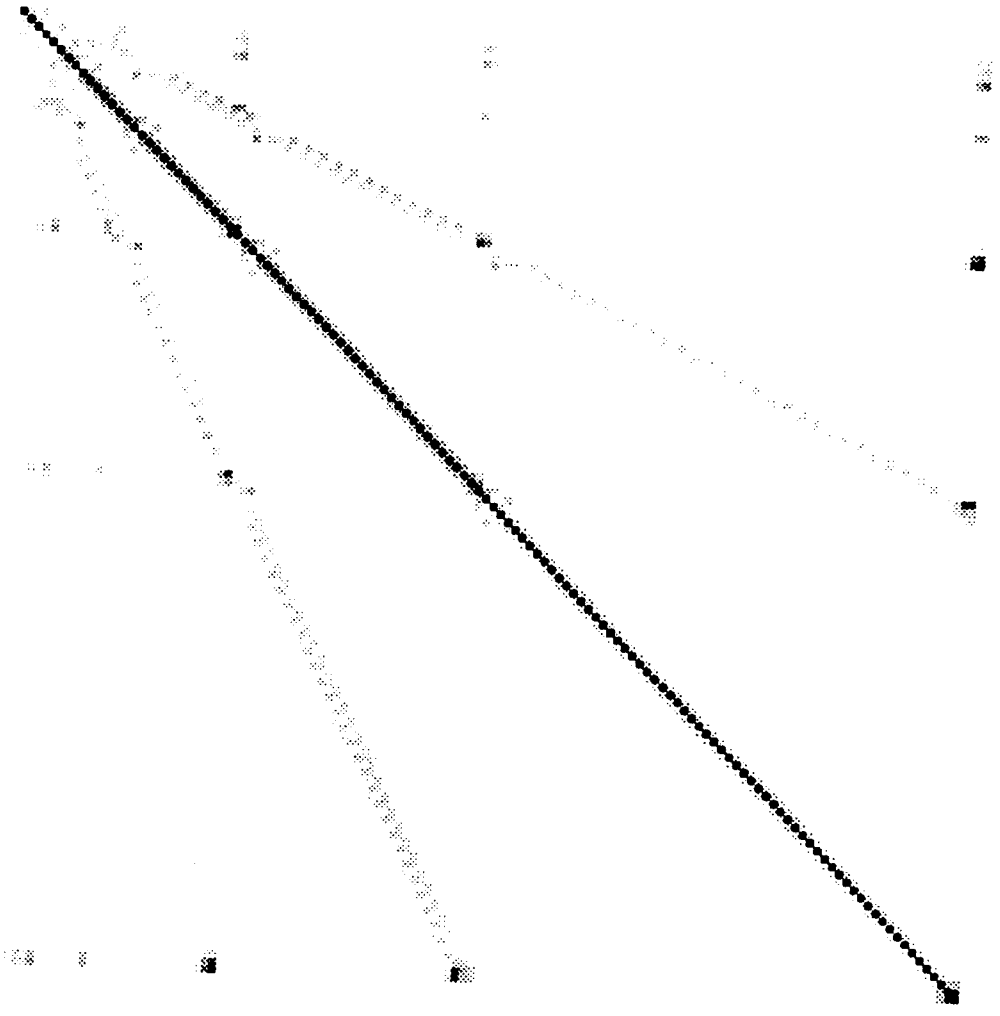


Figure 8: The matrix of correlation coefficients (i.e. covariance divided by the square root of the product of variances) for the wavelet-transform of the Gauss-Markov process of Figure 7 using an 8-tap QMF.

©1992 IEEE. Reprinted, with permission, from *IEEE Transactions on Information Theory*, Vol. 38, No. 2, pp. 785-800, March 1992.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commerical advantage, the IEEE copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Institute of Electrical and Electronics Engineers. To copy otherwise, or to republish, requires a fee and specific permission.

Wavelet-Based Representations for a Class of Self-Similar Signals with Application to Fractal Modulation

Gregory W. Wornell and Alan V. Oppenheim

Abstract

A potentially important family of self-similar signals is introduced based upon a deterministic scale-invariance characterization. These signals, which we refer to as "dy-homogeneous" signals because they generalize the well-known homogeneous functions, have highly convenient representations in terms of orthonormal wavelet bases. In particular, wavelet representations can be exploited to construct orthonormal "self-similar" bases for these signals. The spectral and fractal characteristics of dy-homogeneous signals make them appealing candidates for use in a number of applications. As one potential example, we consider their use in a communications-based context. Specifically, we develop a strategy for embedding information into a dy-homogeneous waveform on multiple time-scales. This multirate modulation strategy, which we term "fractal modulation," is potentially well-suited for use with noisy channels of simultaneously unknown duration and bandwidth. Computationally efficient modulators and demodulators are suggested for the scheme, and the results of a preliminary performance evaluation are presented. Although not yet a fully developed protocol, fractal modulation represents a potentially viable paradigm for communication.

Index Terms— fractals, wavelets, modulation theory, spread spectrum

1 Introduction

Signals with self-similar properties, *i.e.*, signals which retain many of their essential characteristics under time scaling arise frequently in physical processes and also are potentially important in signal generation for communications, remote sensing, and many other applications. The most extensively studied class of such signals are those random processes which exhibit *statistical* self-similarity, *e.g.*, processes whose autocorrelation functions remain invariant to within an amplitude factor under arbitrary scalings of the time axis. An

This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract No. N00014-89-J-1489, and the Air Force Office of Scientific Research under Grant No. AFOSR-91-0034.

The authors are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.

important family of such random processes are typically referred to as $1/f$ processes. These processes are often used in modeling natural landscapes, the distribution of earthquakes, ocean waves, turbulent flow, the pattern of errors on communication channels, and many other natural phenomena.

In this paper we consider signals that exhibit *deterministic* self-similarity, whereby the signal itself remains invariant to within an amplitude factor under arbitrary scaling of the time axis. This class of signals, referred to as homogeneous signals [1], is fairly restricted. However, by generalizing the class of homogeneous signals to require self-similarity only under time scaling by integer powers of two, a family of signals results with potential use as waveforms in a range of engineering applications. As an example of one promising direction for applications, we consider the use of homogeneous signal sets in a communications-based context. Specifically, we develop an approach for embedding information into homogeneous waveforms which we term "fractal modulation." Because the resulting waveforms have the property that the information is contained within multiple time scales and frequency bands, we are able to show that such signals are well-suited for transmission over noisy channels of simultaneously unknown duration and bandwidth. This a reasonable model not only for many physical channels, but also for the receiver constraints inherent in many point-to-point and broadcast communication scenarios. While this proposed modulation scheme is very preliminary and there are many unresolved issues to be explored, it is suggestive of potential ways in which homogeneous signals can perhaps be exploited.

Our approach to the analysis and representation of homogeneous signals is based on the use of orthonormal wavelet bases. These bases, which have the property that all basis functions are dilations and translations of some prototype function, are in many respects ideally suited for use with self-similar signals [2]. Furthermore, because wavelet transformations can be implemented in a computationally efficient manner, the wavelet transform is not only a theoretically important tool, but a practical one as well.

In Section 2, we briefly summarize the notation and properties of wavelet bases to be used in the remainder of the paper. Section 3 introduces and develops the generalized family of homogeneous signals defined in terms of a dyadic scale-invariance property. We distinguish between two classes: energy-dominated and power-dominated, and develop their

spectral properties. We show that orthonormal self-similar bases can be constructed for homogeneous signals using wavelets. Using these representations, we then derive efficient discrete-time algorithms for synthesizing and analyzing homogeneous signals. Section 4 develops the concept of fractal modulation. In particular, we use the orthonormal self-similar basis expansions derived in Section 3 to develop an approach for modulating information sequences onto homogeneous signals. After developing the corresponding optimal receiver, we evaluate the performance of the resulting scheme in the context of a particular channel model and make comparisons to more traditional forms of modulation. Finally, Section 5 summarizes the principal contributions of the paper and suggests some interesting and potentially important directions for future research.

2 Wavelet Notation

In this section, we establish the notational conventions and terminology for the aspects of wavelet theory we shall exploit in this paper. For a more general review of the theory of orthonormal wavelet bases, see, *e.g.*, the classic references [3] [4].

An orthonormal wavelet transformation of a signal $x(t)$ is described in terms of the synthesis/analysis equations[1]

$$x(t) = \sum_m \sum_n x_n^m \, \psi_n^m(t) \tag{1a}$$

$$x_n^m = \int_{-\infty}^{\infty} x(t) \, \psi_n^m(t) \, dt \tag{1b}$$

and has the special property that the orthogonal basis functions are all dilations and translations of a single function referred to as the *basic wavelet* $\psi(t)$. In particular,

$$\psi_n^m(t) = 2^{m/2} \, \psi(2^m t - n) \tag{2}$$

where m and n are the dilation and translation indices, respectively.

The Fourier transform of the basic wavelet, denoted $\Psi(\omega)$, often has a bandpass charac-

[1] We shall assume throughout that all summations over m and n extend from $-\infty$ to ∞ unless otherwise noted.

ter, at least roughly. As a consequence, wavelet decompositions may be interpreted rather naturally in terms of a critically sampled generalized constant-Q or octave-band filter bank. In fact, an example of a wavelet basis, and one which will play an important role in this paper, is the ideal bandpass wavelet basis. In this specific case, the Fourier transform of the wavelet, which we denote by $\tilde{\Psi}(\omega)$, is

$$\tilde{\Psi}(\omega) = \begin{cases} 1 & \pi < |\omega| \leq 2\pi \\ 0 & \text{otherwise} \end{cases} . \tag{3}$$

In many applications, it is useful to impose some degree of regularity on the wavelet basis. As is well-known [4], a sufficient condition for a wavelet basis to possess Rth-order regularity

$$\Psi(\omega) \sim \mathcal{O}\left(|\omega|^{-R}\right), \qquad |\omega| \to \infty$$

where R is some positive integer, is that the wavelet have R vanishing moments, *i.e.*,

$$\int_{-\infty}^{\infty} t^r \, \psi(t)\, dt = (j)^r \Psi^{(r)}(0) = 0, \qquad r = 0, 1, \ldots, R-1.$$

Many examples of wavelets with such regularity have been developed in the literature; see, *e.g.*, [4].

A broad class of orthonormal wavelet bases may also be conveniently interpreted in terms of multiresolution signal analysis. Associated with each such wavelet basis is a corresponding scaling function $\phi(t)$ having a Fourier transform $\Phi(\omega)$ that is at least roughly lowpass. The scaling function associated with the ideal bandpass wavelet basis, in fact, has an ideal lowpass Fourier transform

$$\tilde{\Phi}(\omega) = \begin{cases} 1 & |\omega| \leq \pi \\ 0 & \text{otherwise} \end{cases} .$$

A resolution-limited approximation $A_m x(t)$ to a signal $x(t)$ in which details on scales 2^m and finer are discarded is obtained via the orthonormal expansion

$$A_m x(t) = \sum_n a_n^m \, \phi_n^m(t) \tag{4}$$

where the $\phi_n^m(t)$ are also all dilations and translations of one another, *viz.*,

$$\phi_n^m(t) = 2^{m/2}\,\phi(2^m t - n),$$

and where the coefficients a_n^m are obtained by projection:

$$a_n^m = \int_{-\infty}^{\infty} x(t)\,\phi_n^m(t)\,dt. \tag{5}$$

For these signal approximations, the detail signal $D_m x(t)$ capturing the information in $x(t)$ between scales 2^m and 2^{m+1} has the orthonormal expansion

$$D_m x(t) = A_{m+1}x(t) - A_m x(t) = \sum_n x_n^m\,\psi_n^m(t).$$

The multiresolution signal analysis interpretation of wavelet bases also leads to efficient discrete-time algorithms for implementing wavelet transformations. In particular, associated with every wavelet-based multiresolution analysis is a quadrature mirror filter (QMF) pair whose unit-sample responses $h[n]$ and $g[n]$ have at least roughly lowpass and highpass discrete-time Fourier transforms $H(\omega)$ and $G(\omega)$, respectively. These filters are exploited in the following filter-downsample analysis algorithm

$$a_n^m = \sum_l h[l-2n]\,a_l^{m+1} \tag{6a}$$

$$x_n^m = \sum_l g[l-2n]\,a_l^{m+1} \tag{6b}$$

which may be applied recursively to extract the wavelet coefficients x_n^m at successively coarser scales. In a complementary manner, the following upsample-filter-merge synthesis algorithm

$$a_n^{m+1} = \sum_l \{h[n-2l]\,a_l^m + g[n-2l]\,x_l^m\} \tag{6c}$$

may be applied recursively to reconstruct the coefficients a_n^m of an increasingly fine-scale approximation to a signal $x(t)$. Collectively eqs. (6) constitute what has become known as the Discrete Wavelet Transform (DWT).

3 Deterministically Self-Similar Signals

Signals $x(t)$ satisfying the deterministic scale-invariance property

$$x(t) = a^{-H}x(at) \tag{7}$$

for all $a > 0$, are generally referred to in mathematics as *homogeneous* functions of degree H. As shown by Gel'fand [1], homogeneous functions can be parameterized with only a few constants. As such, they constitute a rather limited class of signal models in many contexts.

A comparatively richer class of signal models is obtained by considering waveforms which are required to satisfy (7) only for values of a that are integer powers of two, *i.e.*, signals that satisfy the dyadic self-similarity property

$$x(t) = 2^{-kH}x(2^k t) \tag{8}$$

for all integers k. While we shall use the generic term "homogeneous signal" to refer to signals satisfying (8), when there is risk of confusion in our subsequent development we will specifically refer to signals satisfying (8) as *dy-homogeneous*.

Homogeneous signals have spectral characteristics very much like those of $1/f$ processes and, in fact, have fractal properties as well. Specifically, although all non-trivial homogeneous signals have infinite energy and many have infinite power, there are nevertheless some such signals with which one can associate a generalized $1/f$-like Fourier transform, and others with which one can associate a generalized $1/f$-like power spectrum. We distinguish between these two classes of homogeneous signals in our subsequent treatment, denoting them *energy-dominated* and *power-dominated* homogeneous signals, respectively. As we develop in Sections 3.1 and 3.2, orthonormal wavelet basis expansions constitute particularly convenient and efficient representations for these two classes of signals.

3.1 Energy-Dominated Homogeneous Signals

Definition 1 *A dy-homogeneous signal $x(t)$ is said to be energy-dominated if when $x(t)$ is*

filtered by an ideal bandpass filter with frequency response

$$B_0(\omega) = \begin{cases} 1 & \pi < |\omega| \leq 2\pi \\ 0 & \textit{otherwise} \end{cases} \tag{9}$$

the resulting signal $\tilde{x}_0(t)$ *has finite-energy, i.e.,*

$$\int_{-\infty}^{\infty} \tilde{x}_0^2(t)\, dt < \infty.$$

The choice of passband edges at π and 2π in our definition is, in fact, somewhat arbitrary. In particular, substituting in the definition any passband that includes one entire frequency octave but does not include $\omega = 0$ or $\omega = \infty$ leads to precisely the same class of signals. However, our particular choice is sufficient and is made in anticipation of the representation of this class of signals in terms of a wavelet basis.

The class of energy-dominated homogeneous signals includes both reasonably regular functions, such as the constant $x(t) = 1$, the ramp $x(t) = t$, the time-warped sinusoid $x(t) = \cos[2\pi \log_2 t]$, and the unit step function $x(t) = u(t)$, as well as singular functions, such as $x(t) = \delta(t)$ and its derivatives. We denote by E^H the collection of all energy-dominated homogeneous signals of degree H. The following theorem allows us to interpret the notion of spectra for such signals. A straightforward but detailed proof is provided in Appendix A.

Theorem 2 *When an energy-dominated homogeneous signal* $x(t)$ *of degree* H *is filtered by an ideal bandpass filter with frequency response*

$$B(\omega) = \begin{cases} 1 & \omega_L < |\omega| \leq \omega_U \\ 0 & \textit{otherwise} \end{cases} \tag{10}$$

for arbitrary $0 < \omega_L < \omega_U < \infty$, *the resulting signal* $y(t)$ *has finite energy and a Fourier transform of the form*

$$Y(\omega) = \begin{cases} X(\omega) & \omega_L < |\omega| \leq \omega_U \\ 0 & \textit{otherwise} \end{cases} \tag{11}$$

where $X(\omega)$ *is some function that is independent of* ω_L *and* ω_U *and has octave-spaced ripple,*

i.e., for all integers k,

$$|\omega|^{H+1}X(\omega) = |2^k\omega|^{H+1}X(2^k\omega). \tag{12}$$

Since in this theorem $X(\omega)$ does not depend on ω_L or ω_U, this function may be interpreted as the generalized Fourier transform of $x(t)$. Furthermore, (12) implies that the generalized Fourier transform of signals in E^H obeys a $1/f$-like (power-law) relationship, *viz.*,

$$|X(\omega)| \sim \frac{1}{|\omega|^{H+1}}.$$

We note that because (11) excludes $\omega = 0$ and $\omega = \infty$, the mapping

$$x(t) \longleftrightarrow X(\omega)$$

is not one to one. As an example, $x(t) = 1$ and $x(t) = 2$ are both in E^H for $H = 0$, yet both have $X(\omega) = 0$ for $\omega > 0$. In order to accommodate this behavior in our subsequent theoretical development, all signals having a common $X(\omega)$ will be combined into an equivalence class. For example, two homogeneous functions $f(t)$ and $g(t)$ are equivalent if they differ by a homogeneous function whose frequency content is concentrated at the origin, for example t^H in the case that H is an integer.

Because the dyadic self-similarity property (8) of dy-homogeneous signals is very similar to the dyadic scaling relationship between basis functions in an orthonormal wavelet basis, wavelets provide a particularly nice representation for this family of signals. Specifically, with $x(t)$ denoting an energy-dominated homogeneous signal, the expansion in an orthonormal wavelet basis is

$$x(t) = \sum_m \sum_n x_n^m \psi_n^m(t) \tag{13a}$$

$$x_n^m = \int_{-\infty}^{\infty} x(t)\psi_n^m(t). \tag{13b}$$

Since $x(t)$ satisfies (8) and since $\psi_n^m(t)$ satisfies (2), it easily follows from (13b) that for homogeneous signals

$$x_n^m = \beta^{-m/2}x_n^0 \tag{14}$$

where

$$\beta = 2^{2H+1} = 2^{\gamma}. \tag{15}$$

Denoting x_n^0 by $q[n]$, (13a) then becomes

$$x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t), \tag{16}$$

from which we see that $x(t)$ is completely specified in terms of $q[n]$. We term $q[n]$ a *generating sequence* for $x(t)$ since, as we shall see, this representation leads to techniques for synthesizing useful approximations to homogeneous signals in practice.

Let us now specifically choose the ideal bandpass wavelet basis, whose basis functions we denote by

$$\tilde{\psi}_n^m(t) = 2^{m/2}\, \tilde{\psi}(2^m t - n)$$

where $\tilde{\psi}(t)$ is the ideal bandpass wavelet whose Fourier transform is given by (3). If we sample the output $\tilde{x}_0(t)$ of the filter in Definition 1 at unit rate, we obtain the sequence $\tilde{q}[n] = \tilde{x}_n^0$, where $\tilde{x}_n^m$ denotes the coefficients of expansion of $x(t)$ in terms of the ideal bandpass wavelet basis. Since $\tilde{x}_0(t)$ in Definition 1 has the orthonormal expansion

$$\tilde{x}_0(t) = \sum_n \tilde{q}[n] \tilde{\psi}_n^0(t) \tag{17}$$

we have

$$\int_{-\infty}^{\infty} \tilde{x}_0^2(t)\, dt = \sum_n \tilde{q}^2[n]. \tag{18}$$

Consequently, a homogeneous function is energy-dominated if and only if its generating sequence in terms of the ideal bandpass wavelet basis has finite energy, *i.e.*,

$$\sum_n \tilde{q}^2[n] < \infty.$$

A convenient inner product between two energy-dominated homogeneous signals $f(t)$ and $g(t)$ can be defined as

$$\langle f, g \rangle_{\tilde{\psi}} = \int_{-\infty}^{\infty} f_0(t)\, g_0(t)\, dt$$

where the signals $f_0(t)$ and $g_0(t)$ are the responses of the bandpass filter (9) to $f(t)$ and $g(t)$, respectively. Exploiting (17) we may more conveniently express this inner product in terms of $\tilde{a}[n]$ and $\tilde{b}[n]$, the respective generating sequences of $f(t)$ and $g(t)$ under the bandpass wavelet basis, as

$$\langle f, g \rangle_{\tilde{\psi}} = \sum_n \tilde{a}[n]\,\tilde{b}[n]. \tag{19}$$

With this inner product, $\boldsymbol{E}^H$ constitutes a Hilbert space and the induced norm on $\boldsymbol{E}^H$ is

$$\|x\|_{\tilde{\psi}} = \int_{-\infty}^{\infty} \tilde{x}_0^2(t)\,dt = \sum_n \tilde{q}^2[n]. \tag{20}$$

One can readily construct "self-similar" bases for $\boldsymbol{E}^H$. Indeed, the ideal bandpass wavelet (16) immediately provides an orthonormal basis for $\boldsymbol{E}^H$. In particular, for any $x(t) \in \boldsymbol{E}^H$, we have the synthesis/analysis pair

$$x(t) = \sum_n \tilde{q}[n]\,\tilde{\theta}_n^H(t) \tag{21a}$$

$$\tilde{q}[n] = \langle x, \tilde{\theta}_n^H \rangle_{\tilde{\psi}} \tag{21b}$$

where one can easily verify that the basis functions

$$\tilde{\theta}_n^H(t) = \sum_m \beta^{-m/2} \tilde{\psi}_n^m(t) \tag{22}$$

are self-similar, orthogonal, and have unit norm.

The fact that the ideal bandpass basis is unrealizable means that (21a) is not a practical mechanism for synthesizing or analyzing homogeneous signals. However, more practical wavelet bases are equally suitable for defining an inner product for the Hilbert space $\boldsymbol{E}^H$. In fact, we shall show that a broad class of wavelet bases can be used to construct such inner products, and that, as a consequence, some highly efficient algorithms arise for processing homogeneous signals.

Not every orthonormal wavelet basis can be used to define inner products for $\boldsymbol{E}^H$. In order to determine which orthonormal wavelet bases can be used to define inner products

for E^H, we must determine for which wavelets $\psi(t)$

$$q[n] = \int_{-\infty}^{\infty} x(t)\,\psi_n^0(t)\,dt \in \mathrm{l}^2(\mathbf{Z}) \Leftrightarrow x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t) \in E^H.$$

That is, we seek conditions on a wavelet basis such that the sequence

$$q[n] = \int_{-\infty}^{\infty} x(t)\,\psi_n^0(t)\,dt$$

has finite energy whenever the homogeneous signal $x(t)$ is energy-dominated, and simultaneously such that the homogeneous signal

$$x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t)$$

is energy-dominated whenever the sequence $q[n]$ has finite energy. Our main result is presented in terms of the following theorem. A proof of this theorem is provided in Appendix B.

Theorem 3 *Consider an orthonormal wavelet basis such that $\psi(t)$ has R vanishing moments for some integer $R \geq 1$, i.e.,*

$$\Psi^{(r)}(0) = 0, \qquad r = 0, 1, \ldots, R-1 \tag{23}$$

and let

$$x(t) = \sum_m \sum_n \beta^{-m/2} q[n] \psi_n^m(t)$$

be a dy-homogeneous signal whose degree H is such that $\gamma = \log_2 \beta = 2H + 1$ satisfies $0 < \gamma < 2R - 1$. Then $x(t)$ is energy-dominated if and only if $q[n]$ has finite energy.

This theorem implies that we may choose for our Hilbert space E^H from among a large number of inner products whose induced norms are all equivalent. In particular, for any wavelet $\psi(t)$ with sufficiently many vanishing moments, we may define the inner product between two functions $f(t)$ and $g(t)$ in E^H whose generating sequences are $a[n]$ and $b[n]$, respectively, as

$$\langle f, g \rangle_\psi = \sum_n a[n] b[n]. \tag{24}$$

Of course, this collection of inner products is almost surely not exhaustive. Even for wavelet-based inner products, Theorem 3 asserts only that the vanishing moment condition is sufficient to ensure that the inner product generates an equivalent norm. It seems unlikely that the vanishing moment condition is a necessary condition.

The wavelet-based norms for E^H constitute a highly convenient and practical collection from which to choose in applications involving the use of homogeneous signals. Indeed, each associated wavelet-based inner product leads immediately to an orthonormal self-similar basis for E^H: if $x(t) \in E^H$, then

$$x(t) = \sum_n q[n]\,\theta_n^H(t) \tag{25a}$$

$$q[n] = \langle x, \theta_n^H \rangle_\psi \tag{25b}$$

where, again, the basis functions

$$\theta_n^H(t) = \sum_m \beta^{-m/2}\,\psi_n^m(t) \tag{26}$$

are all self-similar, mutually orthogonal, and have unit norm.

Finally, we remark that wavelet-based characterizations also give rise to a convenient expression for the generalized Fourier transform of an energy-dominated homogeneous signal, $x(t)$. In particular, if we take the Fourier transform of (16) we get, via some routine algebra,

$$X(\omega) = \sum_m 2^{-(H+1)m}\,\Psi(2^{-m}\omega)Q(2^{-m}\omega) \tag{27}$$

where $Q(\omega)$ is the discrete-time Fourier transform of $q[n]$. This spectrum is to be interpreted in the sense of Theorem 2, *i.e.*, $X(\omega)$ defines the spectral content of the output of a bandpass filter at every frequency ω within the passband.

3.2 Power-Dominated Homogeneous Signals

Energy-dominated homogeneous signals have infinite energy. In fact, most have infinite power as well. However, there are other infinite power homogeneous signals that are not energy-dominated. In this section, we consider a more general class of infinite-power homo-

geneous signals referred to as power-dominated homogeneous signals which will find application in Section 4. The definition and properties closely parallel those for energy-dominated homogeneous signals.

Definition 4 *A dy-homogeneous signal $x(t)$ is said to be power-dominated if when $x(t)$ is filtered by an ideal bandpass filter with frequency response (9) the resulting signal $\tilde{x}_0(t)$ has finite power, i.e.,*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \tilde{x}_0^2(t)\, dt < \infty.$$

The notation $\boldsymbol{P}^H$ will be used to designate the class of power-dominated homogeneous signals of degree H. Moreover, while our definition necessarily includes the energy-dominated signals, which have zero power, insofar as our discussion is concerned they constitute a degenerate case.

Analogous to Theorem 2 for the energy-dominated case, we can establish the following theorem describing the spectral properties of power-dominated homogeneous signals.

Theorem 5 *When a power-dominated homogeneous signal $x(t)$ is filtered by an ideal bandpass filter with frequency response (10), the resulting signal $y(t)$ has finite power and a power spectrum of the form*

$$S_y(\omega) = \lim_{T\to\infty} \frac{1}{2T} \left| \int_{-T}^{T} y(t)\, e^{-j\omega t}\, dt \right|^2 = \begin{cases} S_x(\omega) & \omega_L < |\omega| \le \omega_U \\ 0 & \textit{otherwise} \end{cases} \tag{28}$$

where $S_x(\omega)$ is some function that is independent of ω_L and ω_U and has octave-spaced ripple, i.e., for all integers k,

$$|\omega|^{2H+1} S_x(\omega) = |2^k\omega|^{2H+1} S_x(2^k\omega). \tag{29}$$

The details of the proof of this theorem are contained in Appendix C, although it is identical in style to the proof of its counterpart, Theorem 2. Note that since $S_x(\omega)$ in the theorem does not depend on ω_L or ω_U, this function may be interpreted as the generalized power spectrum of $x(t)$. Furthermore, the relation (29) implies that signals in $\boldsymbol{P}^H$ have a generalized time-averaged power spectrum that is $1/f$-like, *i.e.*,

$$S_x(\omega) \sim \frac{1}{|\omega|^\gamma} \tag{30}$$

where, via (15), $\gamma = 2H + 1$.

Theorem 5 directly implies that a homogeneous signal $x(t)$ is power-dominated if and only if its generating sequence $\tilde{q}[n]$ in the ideal bandpass wavelet basis has finite power, *i.e.*,

$$\lim_{L\to\infty} \frac{1}{2L+1} \sum_{n=-L}^{L} \tilde{q}^2[n] < \infty.$$

Similarly we can readily deduce from the results of Section 3.1 that, in fact, for any orthonormal wavelet basis with sufficiently many vanishing moments R so that $0 < \gamma < 2R - 1$, the *generating sequence* for a homogeneous signal of degree H in that basis has finite power if and only if the signal is power-dominated. This implies that when we use (25a) with such wavelets to synthesize a homogeneous signal $x(t)$ using an arbitrary finite power sequence $q[n]$, we are assured that $x(t) \in P^H$. Likewise, when we use (25b) to analyze any signal $x(t) \in P^H$, we are assured that $q[n]$ has finite power.

Remarks

Energy-dominated homogeneous signals of arbitrary degree H can be highly regular, at least away from $t = 0$. In contrast, power-dominated homogeneous signals typically have a fractal structure similar to the statistically self-similar $1/f$ processes of corresponding degree H, whose power spectra are also of the form (30) with $\gamma = 2H + 1$. One might reasonably conjecture that power-dominated homogeneous signals and $1/f$ processes of the same degree also have identical Hausdorff-Besicovitch dimensions [5], when defined. Indeed, despite their obvious structural differences, power-dominated homogeneous signals and $1/f$ processes "look" remarkably similar in a qualitative sense. This is apparent in Fig. 1, where we depict the sample path of a $1/f$ process along side a power-dominated homogeneous signal of the same degree whose generating sequence has been taken from a white random process. We stress that in Fig. 1(a), the self-similarity of the $1/f$ process is statistical, *i.e.*, it does not satisfy (8) but its autocorrelation function does. In Fig. 1(b), the self-similarity of the homogeneous signal is deterministic. In fact, while the wavelet coefficients

of homogeneous signals are identical from scale to scale to within an amplitude factor, *i.e.*,

$$x_n^m = \beta^{-m/2} q[n]$$

the wavelet coefficients of $1/f$ processes have only the same second-order statistics from scale to scale to within an amplitude factor, *i.e.*,

$$E\left[x_n^m x_l^m\right] = \beta^{-m} \rho[n-l]$$

for some function $\rho[n]$ that is independent of m [2] [6].

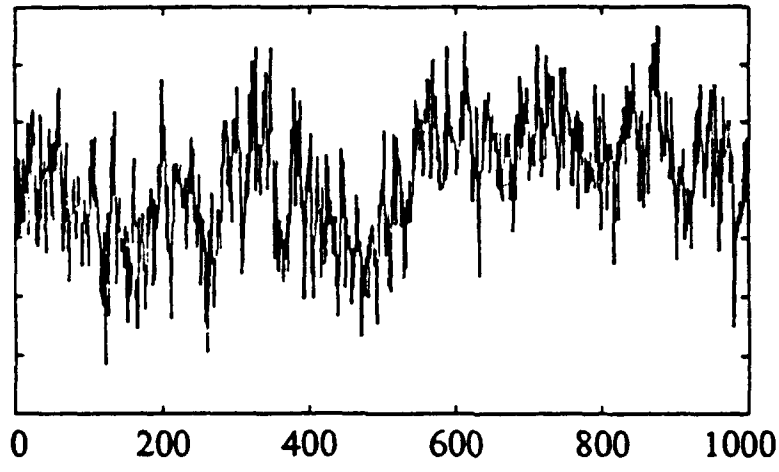
Finally, we remark that not all power-dominated homogeneous signals have spectra that are bounded on $\pi \leq \omega \leq 2\pi$. An interesting subclass of power-dominated homogeneous signals with such unbounded spectra will, in fact, arise in our development of fractal modulation. For these signals, $\tilde{x}(t)$ as defined in Definition 4 is *periodic*, so we refer to this class of power-dominated homogeneous signals as *periodicity-dominated*. It is straightforward to establish that these homogeneous signals have the property that when passed through an arbitrary bandpass filter of the form (10) the output is periodic as well. Furthermore, their power spectra consist of impulses whose *areas* decay according to a $1/|\omega|^\gamma$ relationship. An important class of periodicity-dominated homogeneous signals can be generated through a wavelet-based synthesis of the form (16) in which the generating sequence $q[n]$ is periodic.

3.3 Discrete-Time Algorithms for Processing Homogeneous Signals

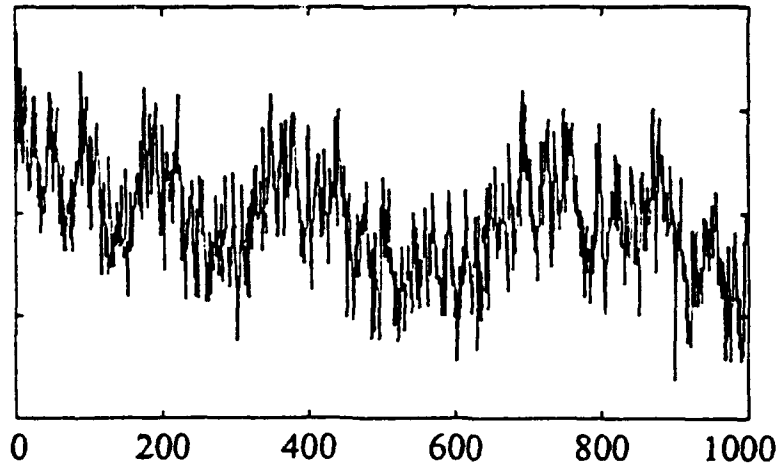
Orthonormal wavelet representations provide some useful insights into homogeneous signals. For instance, because the sequence $q[n]$ is replicated at each scale in the representation (16) of a homogeneous signal $x(t)$, the detail signals

$$D_m x(t) = \beta^{-m/2} \sum_n q[n] \psi_n^m(t)$$

representing $q[n]$ modulated into a particular octave band are simply time-dilated versions of one another, to within an amplitude factor. The corresponding time-frequency portrait of a homogeneous signal is depicted in Fig. 2, from which the scaling properties are apparent.



(a) A sample function of a $1/f$ process.



(b) A power-dominated homogeneous signal.

Figure 1: *Comparison between the sample path of a* $1/f$ *process and a power-dominated homogeneous signal. Both correspond to* $\gamma = 1$ *(i.e.,* $H = 0$*).*

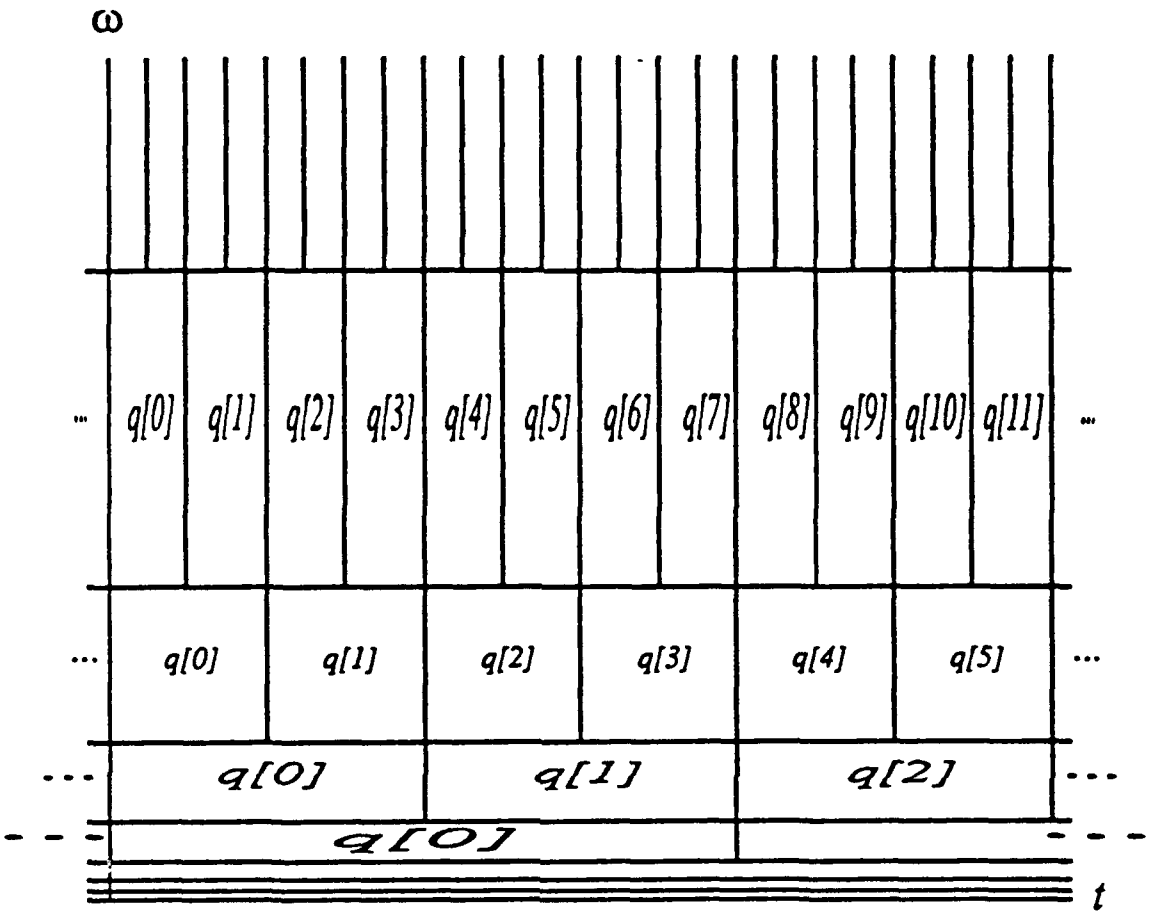


Figure 2: *The time-frequency portrait of a homogeneous signal of degree* $H = -1/2$.

For purposes of illustration, the signal in this figure has degree $H = -1/2$ (*i.e.*, $\beta = 1$), which corresponds to the case in which $q[n]$ is scaled by the same amplitude factor in each octave band. Clearly, the partitioning in such time-frequency portraits is idealized: in general, there is both spectral and temporal overlap between cells.

Wavelet representations also lead to some highly efficient algorithms for synthesizing, analyzing, and processing homogeneous signals just as they do for $1/f$ processes as discussed in [7]. The signal processing structures we develop in this section are a consequence of applying the DWT algorithm to the highly structured form of the wavelet coefficients of homogeneous signals.

We have already encountered one discrete-time representation for a homogeneous signal $x(t)$, namely that in terms of a generating sequence $q[n]$ which corresponds to the coefficients

of the expansion of $x(t)$ in an orthonormal basis $\{\theta_n^H(t)\}$ for $\mathcal{E}^H$. When the $\theta_n^H(t)$ are derived from a wavelet basis according to (26), another useful discrete-time representation for $x(t)$ is available, which we now discuss.

Consider the coefficients a_n^m characterizing the resolution-limited approximation $A_m x(t)$ of a homogeneous signal $x(t)$ with respect to a particular wavelet-based multiresolution signal analysis. Since these coefficients are the projections of $x(t)$ onto dilations and translations of the scaling function $\phi(t)$ according to (5), it is straightforward to verify that they, too, are identical at all scales to within an amplitude factor, i.e.,

$$a_n^m = \beta^{-m/2} a_n^0. \tag{31}$$

Consequently, the sequence a_n^0 is an alternative discrete-time characterization of $x(t)$, since knowledge of it is sufficient to reconstruct $x(t)$ to arbitrary accuracy. For convenience, we refer to a_n^0 as the *characteristic sequence* and denote it as $p[n]$. As is true for the generating sequence, the characteristic sequence associated with $x(t)$ depends upon the particular multiresolution analysis used; distinct multiresolution signal analyses generally yield different characteristic sequences for any given homogenous signal. We shall require that the wavelet associated with any multiresolution analysis we consider have sufficiently many vanishing moments that it meets the conditions of Theorem 3.

The characteristic sequence $p[n]$ is associated with a resolution-limited approximation to the corresponding homogeneous signal $x(t)$. Specifically, $p[n]$ represents unit-rate samples of the output of the filter, driven by $x(t)$, whose frequency response is the complex conjugate of $\Phi(\omega)$. Because frequencies in the neighborhood of the spectral origin, where the spectrum of $x(t)$ diverges, are passed by such a filter, $p[n]$ will often have infinite energy or, worse, infinite power, even when the generating sequence $q[n]$ has finite energy.

The characteristic sequence can, in fact, be viewed as a *discrete-time* homogeneous signal, and a theory can be developed following an approach directly analogous to that used in Sections 3.1 and 3.2 for the case of continuous-time homogeneous signals. The

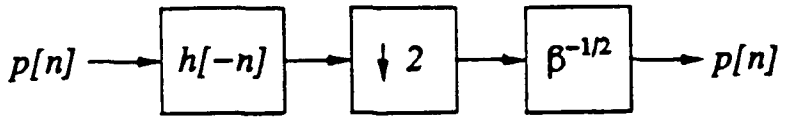


Figure 3: *The discrete-time self-similarity identity for a characteristic sequence* $p[n]$.

characteristic sequence satisfies the discrete-time self-similarity relation[2]

$$\beta^{1/2} p[n] = \sum_k h[k - 2n] p[k] \tag{32}$$

which is readily obtained by substituting for a_n^m in the DWT analysis equation (6a) using (31). Indeed, as depicted in Fig. 3, (32) is a statement that when $p[n]$ is lowpass filtered with the conjugate filter whose unit-sample response is $h[-n]$ and then downsampled, we recover an amplitude-scaled version of $p[n]$. Although characteristic sequences are, in an appropriate sense, "generalized sequences," when highpass filtered with the corresponding conjugate highpass filter whose unit-sample response is $g[-n]$, the output is a finite energy or finite power sequence, depending on whether $p[n]$ corresponds to a homogeneous signal $x(t)$ that is energy-dominated or power-dominated, respectively. Consequently, we can analogously *classify the sequence* $p[n]$ as energy-dominated in the former case, and power-dominated in the latter case. In fact, when the output of such a highpass filter is downsampled at rate two, we recover the characteristic sequence $q[n]$ associated with the expansion of $x(t)$ in the corresponding wavelet basis, *i.e.*,

$$\beta^{1/2} q[n] = \sum_k g[k - 2n] p[k]. \tag{33}$$

This can be readily verified by substituting for a_n^m and x_n^m in the DWT analysis equation (6b) using (31) and (14), and by recognizing that $a_0^m = p[n]$ and $x_0^m = q[n]$.

From a different perspective, (33) provides a convenient mechanism for obtaining the representation for a homogeneous signal $x(t)$ in terms of its generating sequence $q[n]$ from

[2]Relations of this type may be considered discrete-time counterparts of the *dilation equations* considered by Strang in [8].

one in terms of its corresponding characteristic sequence $p[n]$, i.e.,

$$p[n] \longrightarrow q[n].$$

To obtain the reverse mapping

$$q[n] \longrightarrow p[n]$$

is less straightforward. For an arbitrary sequence $q[n]$, the associated characteristic sequence $p[n]$ is the solution to the linear equation

$$\beta^{-1/2}p[n] - \sum_k h[n-2k]p[k] = \sum_k g[n-2k]q[k], \tag{34}$$

as can be verified by specializing the DWT synthesis equation (6c) to the case of homogeneous signals. There appears to be no direct method for solving this equation. However, the DWT synthesis algorithm suggests a convenient and efficient iterative algorithm for *constructing* $p[n]$ *from* $q[n]$. In particular, denoting the estimate of $p[n]$ on the ith iteration by $p^{[i]}[n]$, the algorithm is

$$p^{[0]}[n] = 0 \tag{35a}$$

$$p^{[i+1]}[n] = \beta^{1/2} \sum_k \left\{ h[n-2k]p^{[i]}[k] + g[n-2k]q[k] \right\}. \tag{35b}$$

This recursive upsample-filter-merge algorithm, depicted in Fig. 4, can be interpreted as repeatedly modulating $q[n]$ with the appropriate gain into successively lower octave bands of the frequency interval $0 \leq |\omega| \leq \pi$. Note that the precomputable quantity

$$q_*[n] = \sum_k g[n-2k]q[k]$$

represents the sequence $q[n]$ modulated into essentially the upper half band of frequencies.

Any real application of homogeneous signals can ultimately exploit scaling properties over only a *finite* range of scales, so that it suffices in practice to modulate $q[n]$ into a finite range of contiguous octave bands. Consequently, only a finite number of iterations of the algorithm (35) are be required. More generally, this also means that many of the theoretical

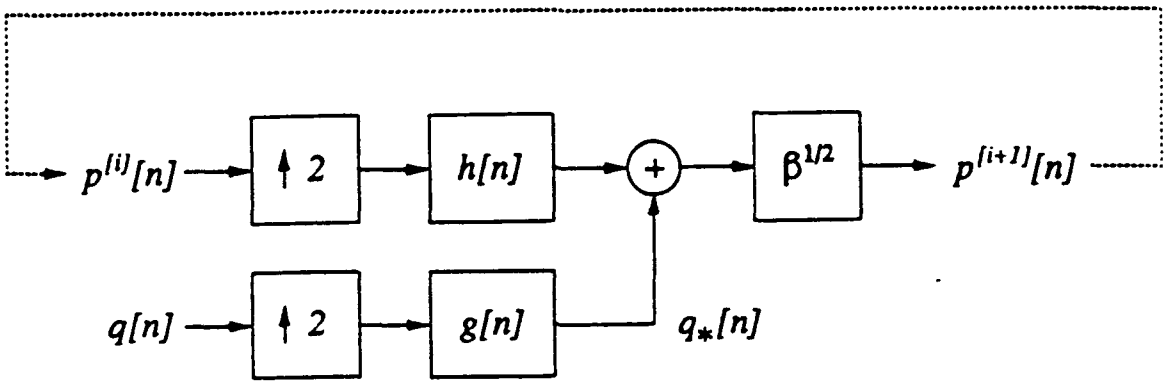


Figure 4: *Iterative algorithm for the synthesis of the characteristic sequence $p[n]$ of a homogeneous signal $x(t)$ from its generating sequence $q[n]$. The notation $p^{[i]}[n]$ denotes the value of $p[n]$ at the ith iteration.*

issues associated with homogeneous signals concerning singularities and convergence do not present practical difficulties in the application of these signals, as will be apparent in our developments of Section 4.

Before turning to a potential application of homogeneous signal sets, we mention that there would appear to be important connections to be explored between the theory of self-similar signals described here and the work of Barnsley, *et al.*, [9] on deterministically self-affine signals. Interestingly, the recent work of Malassenet and Mersereau [10] has shown that these signals, which are conveniently generated using so-called "iterated function systems" have efficient representations in terms of wavelet bases as well.

4 Fractal Modulation

In this section, we consider the use of homogeneous signals as modulating waveforms in a communications-based context as an example of the direction that some applications may take. Beginning with an idealized but fairly general channel model, we demonstrate that the use of homogeneous waveforms in such channels is at least natural, if not optimal, and leads to a multirate modulation strategy in which data is transmitted simultaneously at multiple rates. While it is a preliminary proposal, the modulation has a number of properties that seem appealing.

Our problem involves the design of a communication system for transmitting a contin-

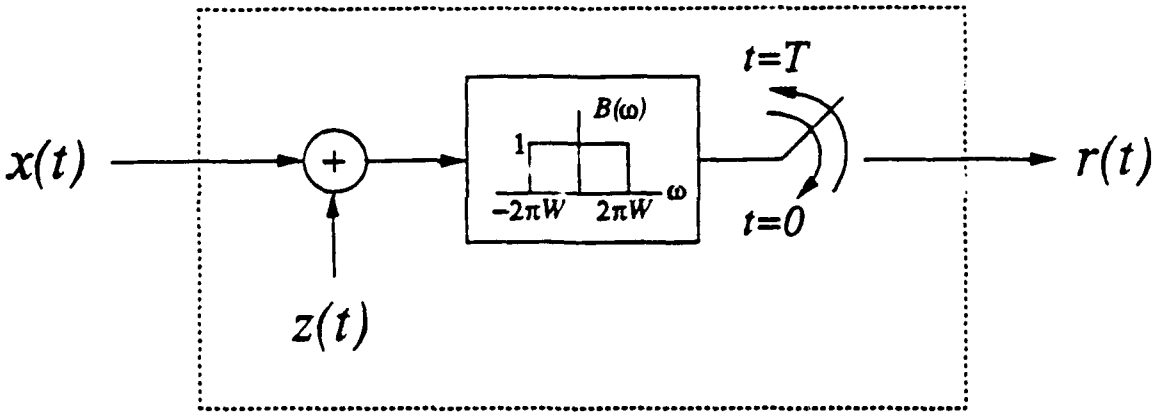


Figure 5: *The channel model for a typical communications scenario.*

nous- or discrete-valued data sequence over a noisy and unreliable continuous-amplitude, continuous-time channel. We must therefore design a modulator at the transmitter that embeds the data sequence $q[n]$ into a signal $x(t)$ to be sent over the channel. At the receiver, a demodulator must be designed for processing the distorted signal $r(t)$ from the channel to extract an optimal estimate of the data sequence $\hat{q}[n]$.

In a typical communication scenario, the channel would be "open" for some time interval T, during which it has a particular bandwidth W and signal-to-noise ratio (SNR). Such a channel model can be used to capture both characteristics of the transmission medium and constraints inherent in one or more receivers. When the noise characteristics are additive, the overall channel model is as depicted in Fig. 5, where $z(t)$ represents the noise process.

When either the bandwidth or duration parameters of the channel are known *a priori*, there are many well-established methodologies for designing an efficient and reliable communication system. However, we shall restrict our attention to the case in which *both* the bandwidth and duration parameters are either unknown or not available to the transmitter. This case, by contrast, has received comparatively less attention in the communications literature, although it encompasses a range of both point-to-point and broadcast communication scenarios involving, for example, jammed and fading channels, multiple access channels, covert and low probability of intercept (LPI) communication, and broadcast communication to disparate receivers.

We shall require the communication system we design for such channels to satisfy the

following performance characteristics:

1. Given a duration-bandwidth product $T \times W$ that exceeds some threshold, we must be able to transmit $q[n]$ without error in the absence of noise, *i.e.*, $z(t) = 0$.

2. Given increasing duration-bandwidth product in excess of this threshold, we must be able to transmit $q[n]$ with increasing fidelity in the presence of noise. Furthermore, in the limit of infinite duration-bandwidth product, perfect transmission should be achievable at any finite SNR.

The first of these requirements implies that, at least in principle, we ought to be able to recover $q[n]$ using arbitrarily narrow receiver bandwidth given sufficient duration, or, alternatively, from an arbitrarily short duration segment given sufficient bandwidth. The second requirement implies that we ought to be able to obtain better estimates of $q[n]$ the longer a receiver is able to listen, or the greater the bandwidth it has available. Consequently, the modulation must contain redundancy of a type that can be exploited for the purposes of error correction. As we shall demonstrate, the use of homogeneous signals for transmission appears to be rather naturally suited to fulfilling both these system requirements.

The minimum achievable duration-bandwidth threshold in such a system is a measure of the efficiency of the modulation. Actually, because the duration-bandwidth threshold $T \times W$ is a function of the length L of the data sequence, it is more convenient to transform the duration constraint T into a symbol rate constraint $R = L/T$ and phrase the discussion in terms of a rate-bandwidth threshold R/W that is independent of sequence length. Then, the maximum achievable rate-bandwidth threshold constitutes the *spectral efficiency* of the modulation, which we shall denote by η. The spectral efficiency of a transmission scheme using bandwidth W is, in fact, defined as

$$\eta = R_{\max}/W$$

where $R_{\max}$ is the maximum rate at which perfect communication is possible in the absence of noise. Hence, the higher the spectral efficiency of a scheme, the higher the rate that can be achieved for a given bandwidth, or, equivalently, the smaller the bandwidth that is required to support a given rate.

When the available channel bandwidth is known *a priori*, a reasonably spectrally efficient, if impractical, modulation of a data sequence $q[n]$ involves expanding the sequence in terms of an ideally bandlimited orthonormal basis. Specifically, with W_0 denoting the channel bandwidth, a transmitter produces

$$x(t) = \sum_n q[n] \sqrt{W_0}\, \mathrm{sinc}\,(W_0 t - n)$$

where

$$\mathrm{sinc}\,(t) = \begin{cases} 1 & t = 0 \\ \dfrac{\sin \pi t}{\pi t} & \text{otherwise} \end{cases}.$$

In the absence of noise, a receiver may recover $q[n]$ from the projections

$$q[n] = \int_{-\infty}^{\infty} x(t) \sqrt{W_0}\, \mathrm{sinc}\,(W_0 t - n)\, dt$$

which can be implemented as a sequence of filter-and-sample operations. Since this scheme achieves a rate of $R = W_0$ symbols/sec using the double-sided bandwidth of $W = W_0$ Hz, it is characterized by a spectral efficiency of

$$\eta_0 = 1 \text{ symbol/sec/Hz}. \tag{36}$$

However, because the transmitter is assumed to have perfect knowledge of the rate-bandwidth characteristics of the channel, this modulation does not constitute a viable solution to our communications problem. Indeed, in order to accommodate a decrease in available channel bandwidth, the transmitter would have to be accordingly reconfigured by decreasing the parameter W_0. Similarly, for the system to maintain a spectral efficiency of $\eta_0 = 1$ when the available channel bandwidth increases, the transmitter must be reconfigured by correspondingly increasing the parameter W_0. Nevertheless, while not a solution to our communications problem, this benchmark modulation provides a useful performance baseline in evaluating the fractal modulation strategy we develop.

We now turn our attention to the problem of designing a modulation strategy that maintains its spectral efficiency over a broad range of rate-bandwidth combinations using

a fixed transmitter configuration. A rather natural solution to this problem arises out of the concept of embedding the data to be transmitted into a homogeneous signal. Due to the fractal properties of the transmitted signals, we refer to the resulting scheme as "fractal modulation."

4.1 Transmitter Design: Modulation

To embed a finite-power sequence $q[n]$ into a dy-homogeneous waveform $x(t)$ of degree H, it suffices to consider using $q[n]$ as the coefficients of an expansion in terms of a wavelet-based orthonormal self-similar basis of degree H, *i.e.*,

$$x(t) = \sum_n q[n]\,\theta_n^H(t)$$

where the basis functions $\theta_n^H(t)$ are constructed according to (26). When the basis is derived from the ideal bandpass wavelet, as we shall generally assume in our analysis, the resulting waveform $x(t)$ is a power-dominated homogeneous signal whose idealized time-frequency portrait has the form depicted in Fig. 2. Consequently, we may view this as a *multirate modulation* of $q[n]$ where in the mth frequency band $q[n]$ is modulated at rate 2^m using a double-sided bandwidth of 2^m Hz. Furthermore, the energy per symbol used in successively higher bands scales by $\beta = 2^{2H+1}$. Using a suitably designed receiver, $q[n]$ can, in principle, be recovered from $x(t)$ at an arbitrary rate 2^m using a baseband bandwidth of 2^{m+1} Hz. Consequently, this modulation has a spectral efficiency of

$$\eta_F = (1/2) \text{ symbol/sec/Hz.}$$

We emphasize that in accordance with our channel model of Fig. 5, it is the baseband bandwidth that is important in defining the spectral efficiency since it defines the highest frequency available at the receiver.

While the spectral efficiency of this modulation is half that of the benchmark scheme (36), this loss in efficiency is, in effect, the price paid to enable a receiver to use any of a range of rate-bandwidth combinations in demodulating the data. Fig. 6 illustrates the rate-bandwidth tradeoffs available to the receiver. In the absence of noise the receiver can,

in principle, perfectly recover $q[n]$ using rate-bandwidth combinations lying on or below the solid curve. The stepped character of this curve reflects the fact that only rates of the form 2^m can be accommodated, and that full octave increases in bandwidth are required to enable $q[n]$ to be demodulated at successively higher rates. For reference, the performance of our benchmark modulation is superimposed on this plot using a dashed line. We emphasize that in contrast to fractal modulation, the transmitter in the benchmark scheme requires perfect knowledge of the rate-bandwidth characteristics of the channel.

Although it considerably simplifies our analysis, the use of the ideal bandpass wavelet to synthesize the orthonormal self-similar basis in our modulation strategy is impractical due to the poor temporal localization in this wavelet. However, we may, in practice, replace the ideal bandpass wavelet with one having not only comparable frequency domain characteristics and better temporal localization, but sufficiently many vanishing moments to ensure that the transmitted waveform is power-dominated as well. Fortunately, there are many suitable wavelets from which to choose, among which are those due to Daubechies [4]. When such wavelets are used, the exact spectral efficiency of the modulation depends on the particular definition of bandwidth employed. Nevertheless, using any reasonable definition of bandwidth, we would expect to be able to achieve, in practice, a spectral efficiency close to (1/2) symbols/sec/Hz with this modulation, and, as a result, we shall assume $\eta_F \approx 1/2$ in subsequent analysis.

Another apparent problem with fractal modulation as initially proposed is that it requires infinite transmitter power. Indeed, as Fig. 2 illustrates, $q[n]$ is modulated into an infinite number of octave-width frequency bands. However, in a practical implementation, only a finite collection of contiguous bands $\mathcal{M}$ would, in fact, be used by the transmitter. As a result, the transmitted waveform

$$x(t) = \sum_n q[n] \sum_{m \in \mathcal{M}} \beta^{-m/2} \psi_n^m(t) \tag{37}$$

would exhibit self-similarity only over a range of scales, and demodulation of the data would be possible at one of only a finite number of rates. In terms of Fig. 6, the rate-bandwidth characteristic of the modulation would extend over a finite range of bandwidths chosen to

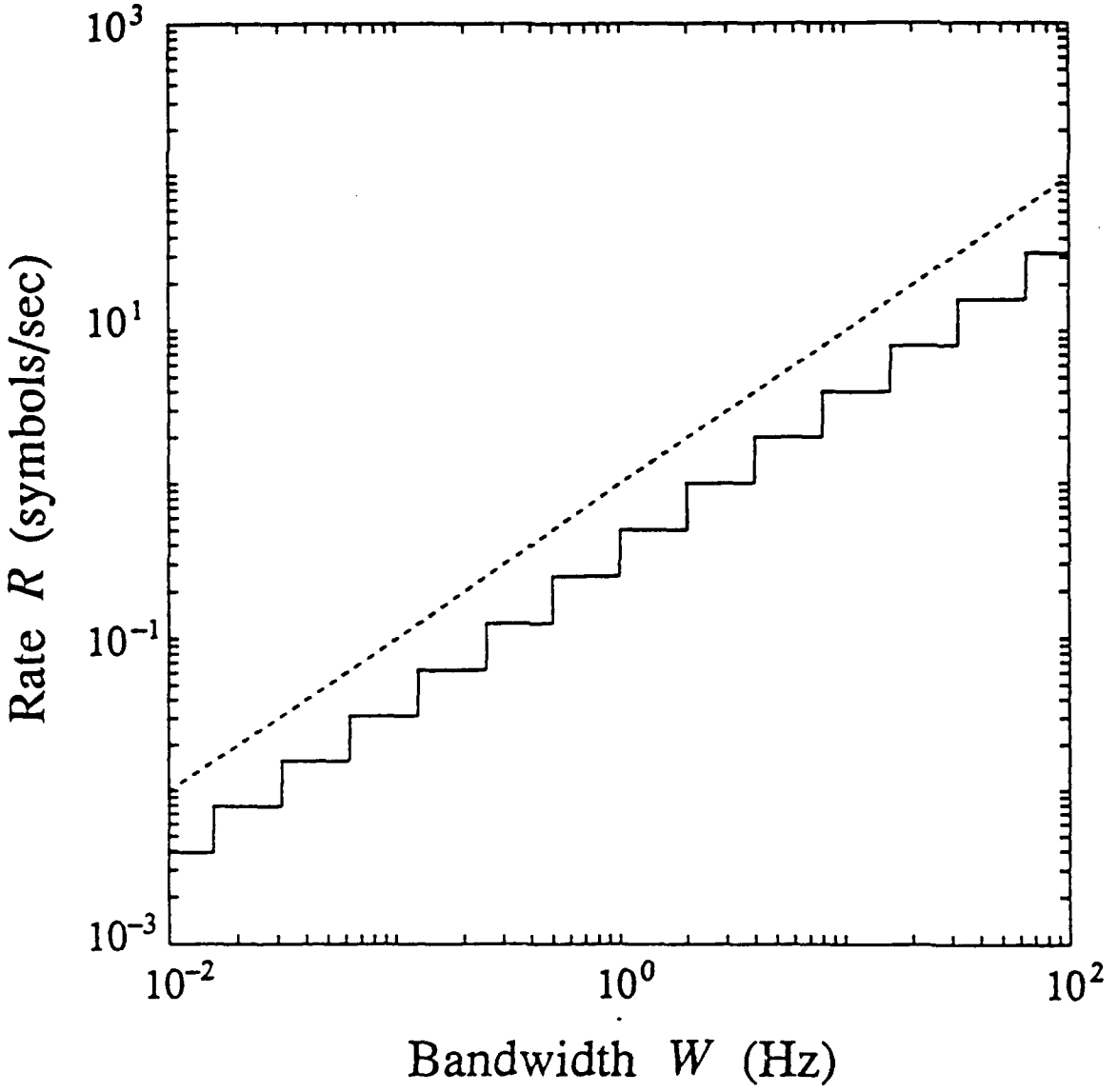


Figure 6: *Spectral efficiency of fractal modulation. At each bandwidth B, the solid curve indicates the maximum rate at which transmitted data can be perfectly recovered in the absence of noise. The dashed curve indicates the corresponding performance of the benchmark scheme.*

cover extremes anticipated for the system.

The fractal modulation transmitter can be implemented in a computationally highly efficient manner, since much of the processing can be performed using the discrete-time algorithms of Section 3.3. For example, synthesizing the waveform $x(t)$ given by (37) for $\mathcal{M} = \{0, 1, \ldots, M-1\}$ involves two stages. In the first stage, which involves only discrete-time processing, $q[n]$ is mapped into M consecutive octave-width frequency bands to obtain the sequence $p^{[M]}[n]$. This sequence is obtained using M iterations of the synthesis algorithm (35) with the QMF filter pair $h[n], g[n]$ appropriate to the wavelet basis. The second stage then consists of a discrete- to continouous-time transformation in which $p^{[M]}[n]$ is modulated into the continuous-time frequency spectrum via the appropriate scaling function according to

$$x(t) = \sum_n p^{[M]}[n]\, \phi_n^M(t) = \sum_n p^{[M]}[n]\, 2^M \phi(2^M t - n).$$

It is important to point out that because a batch-iterative algorithm is employed, potentially large amounts of data buffering may be required. Hence, while the algorithm may be computationally efficient, it may be considerably less so in terms of storage requirements. However, in the event that $q[n]$ is *finite length*, it is conceivable that the algorithm may be modified so as to be memory-efficient as well. Such potential remains to be explored.

The transmission of finite length sequences using fractal modulation more generally raises a variety of issues and, therefore, requires some special consideration. In fact, as initially proposed, fractal modulation is rather inefficient in this case, in essence because successively higher frequency bands are increasingly underutilized. In particular, we note from the time-frequency portrait in Fig. 2 that if $q[n]$ has finite length, *e.g.*,

$$q[n] = 0, \qquad n < 0,\ n > L-1,$$

then the mth band will complete its transmission of $q[n]$ and go idle in half the time it takes the $(m-1)$st band, and so forth. However, finite length messages may be accommodated rather naturally and efficiently by modulating their periodic extensions $q[n \bmod L]$ thereby

generating a transmitted waveform

$$x(t) = \sum_n q[n \bmod L]\, \theta_n^H(t)$$

which constitutes a periodicity-dominated homogeneous signal of the type discussed in Section 3.2. If we let

$$\mathbf{q} = \{q[0]\ q[1]\ \cdots\ q[L-1]\}$$

denote the data vector, then the time-frequency portrait associated with this signal is shown in Fig. 7. Using this enhancement of fractal modulation, we not only maintain our ability to make various rate-bandwidth tradeoffs at the receiver, but we acquire a certain flexibility in our choice of time origin as well. Specifically, as is apparent from Fig. 7, the receiver need not begin demodulating the data at $t = 0$, but may *more generally choose* a *time-origin* that is some multiple of LR when operating at rate R. Additionally, this strategy can, in principle, be extended to accommodate data transmission on a block-by-block basis.

The final aspect of fractal modulation that remains to be considered in this section concerns the specification of the parameter H. While H has no effect on the spectral *efficiency of* fractal modulation, it does affect the power efficiency of the scheme. Indeed, it controls the relative power distribution between frequency bands and, hence, the overall transmitted power spectrum, which takes the form (30) where $\gamma = 2H + 1$. Consequently, the selection of H is important when we consider the presence of additive noise in the channel.

For traditional additive stationary Gaussian noise channels of known bandwidth, the appropriate spectral shaping of the transmitted signal is governed by a "water-filling" procedure [11] [12] which is also the method by which the capacity of such channels is computed [13]. Using this procedure, the available signal power is distributed in such a way that proportionally more power is located at frequencies where the noise power is smaller.

When there is uncertainty in the available bandwidth, the water-filling approach leads to poor worst-case performance. As an example, for a channel in which the noise power is very small only in some fixed frequency band $0 < \omega_L < \omega_U < \infty$, the water-filling recipe will locate the signal power predominantly within this band. As a result, the overall SNR

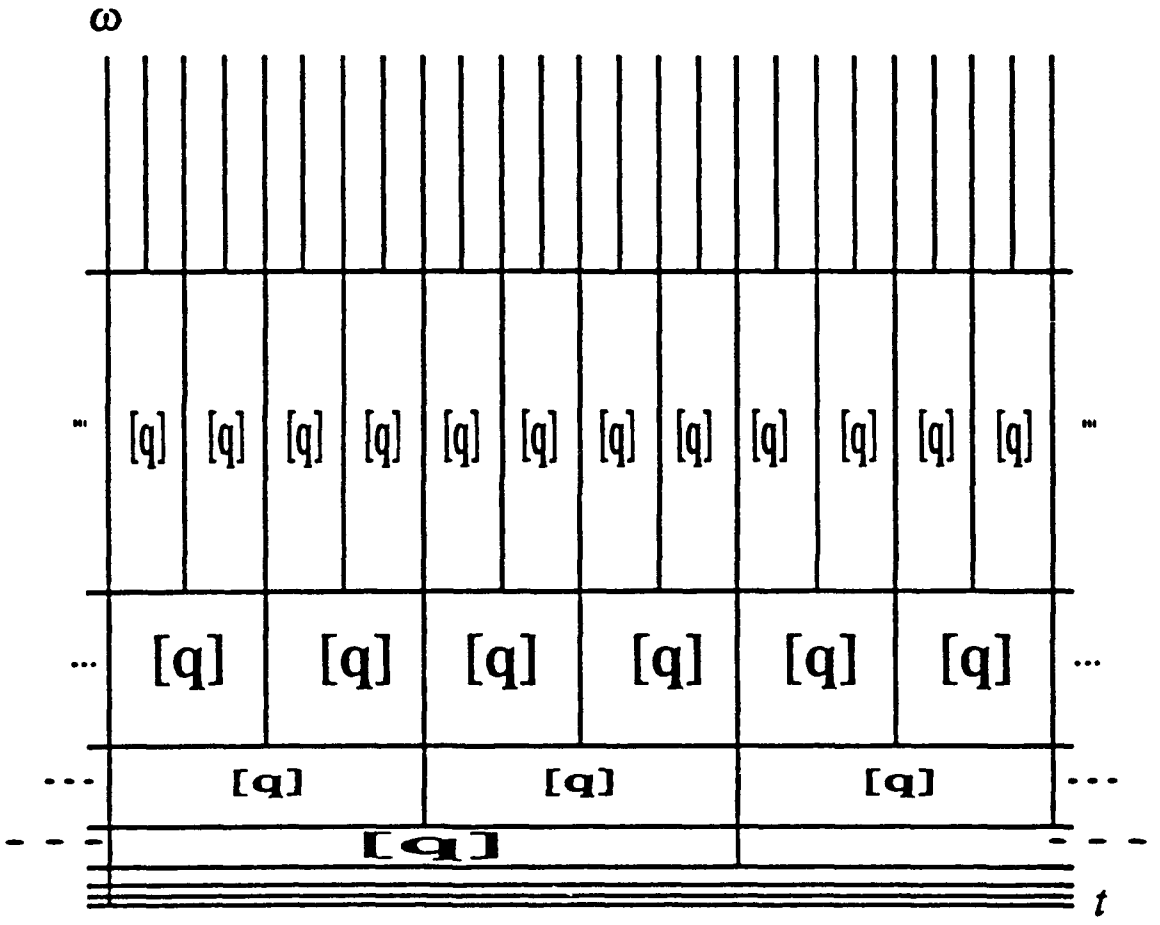


Figure 7: *A portion of the time-frequency portrait of the transmitted signal for fractal modulation of a finite-length data vector* $\mathbf{q}$. *The case* $H = -1/2$ *is shown for convenience.*

in the channel will strongly depend on whether the channel bandwidth is such that these frequencies are passed. By contrast, the distribution of power according to a spectral-matching rule that maintains an SNR that is independent of frequency leads to a system whose performance is uniform with variations in bandwidth and, in addition, is potentially well-suited for LPI communication. Since power-dominated homogeneous signals have a power spectrum of the form of (30), the spectral-matching rule suggests that fractal modulation may be naturally suited to channels with additive $1/f$ noise whose degree H is the same as that of the transmitted signal. This rather broad class of statistically self-similar processes includes not only classical white Gaussian noise ($H = -1/2$) and Brownian motion ($H = 1/2$), but, more generally, a range of rather prevalent nonstationary noises which exhibit strong long-term statistical dependence [14].

In this section, we have developed a modulation strategy that satisfies the first of the two system requirements described at the outset of Section 4. In the next section, where we turn our attention to the problem of designing optimal receivers for fractal modulation, we shall see that fractal modulation also satisfies the second of our system requirements.

4.2 Receiver Design: Demodulation

Consider the problem of recovering a finite length message $q[n]$ from band-limited, time-limited, and noisy observations $r(t)$ of the transmitted waveform $x(t)$ consistent with our channel model of Fig. 5. We shall assume that the noise $z(t)$ is a Gaussian $1/f$ process of degree $H_z = H$, and that the degree H_x of the homogeneous signal $x(t)$ has been chosen according to our spectral-matching rule, i.e.,

$$H_x = H_z = H. \tag{38}$$

We remark that if it is necessary that the transmitter measure H_z in order to perform this spectral matching, the robust and efficient parameter estimation algorithms for $1/f$ processes developed in [7] may be exploited.

Depending on the nature of the message being transmitted, there are a variety of different optimization criteria from which to choose in designing a suitable receiver. As a

representative example, we consider the case in which the transmitted message is a random bit stream of length L represented by a binary-valued sequence

$$q[n] \in \{+\sqrt{E_0}, -\sqrt{E_0}\}$$

where E_0 is the energy per bit. For this data, we develop a receiver that demodulates $q[n]$ so as to minimize the probability of a bit-error. Demodulation of non-binary discrete-valued sequences is achieved using a straightforward extension of our results, and demodulation of continuous-valued sequences under a minimum mean-square error criterion is described in [2].

An efficient implementation of the optimum receiver processes the observations $r(t)$ in the wavelet domain by first extracting the wavelet coefficients r_n^m using the DWT (6). These coefficients take the form

$$r_n^m = \beta^{-m/2} q[n \bmod L] + z_n^m \tag{39}$$

where the z_n^m are the wavelet coefficients of the noise process, and where we have assumed that in accordance with our discussion in Section 4.1 the periodic replication of the finite length sequence $q[n]$ has been modulated. To simplify our analysis, we shall further assume that the ideal bandpass wavelet is used in the transmitter and receiver, although we reiterate that comparable performance can be achieved when more practical wavelets are used.

The duration-bandwidth characteristics of the channel will in general affect which observation coefficients r_n^m may be accessed. In particular, if the channel is bandlimited to 2^{M_U} Hz for some integer M_U, this precludes access to the coefficients at scales corresponding to $m > M_U$. Simultaneously, the duration-constraint in the channel results in a minimum allowable decoding rate of 2^{M_L} symbols/sec for some integer M_L, which precludes access to the coefficients at scales corresponding to $m < M_L$. As a result, the collection of coefficients available at the receiver is

$$\mathbf{r} = \{r_n^m, m \in \mathcal{M}, n \in \mathcal{N}(m)\}$$

where

$$\mathcal{M} = \{M_L, M_L + 1, \ldots, M_U\}$$

$$\mathcal{N}(m) \;=\; \{0,1,\ldots,L\,2^{m-M_L}-1\}.$$

This means that we have available

$$K = \sum_{m=M_L}^{M_U} 2^{m-M_L} = 2^{M_U-M_L+1}-1 \tag{40}$$

noisy measurements of each of the L non-zero samples of the sequence $q[n]$. The specific relationship between decoding rate R, bandwidth W, and redundancy K can, therefore, be expressed in terms of the spectral efficiency of the modulation η_F as

$$\frac{R}{W} = \frac{2\eta_F}{K+1}, \tag{41}$$

where, as discussed earlier, $\eta_F \approx 1/2$. Note that $M_U = M_L$ when $K = 1$, and (41) attains its maximum value, η_F.

The optimal decoding of each bit can be described in terms of a binary hypothesis test on the set of available observation coefficients $\mathbf{r}$. Denoting by H_1 the hypothesis in which $q[n] = +\sqrt{E_0}$, and by H_0 the hypothesis in which $q[n] = -\sqrt{E_0}$, we may construct the likelihood ratio test for the optimal decoding of each symbol $q[n]$. The derivation is particularly straightforward because of the fact that, in accordance with the wavelet-based models for $1/f$ processes developed in [15] [7] [2], the z_n^m in (39) may be modeled as independent zero-mean Gaussian random variables with variances

$$\mathrm{Var}\, z_n^m = \sigma_z^2 \beta^{-m} \tag{42}$$

for some variance parameter $\sigma_z^2 > 0$. Consequently, the likelihood ratio test reduces to the test

$$\ell = \sum_{m=M_L}^{M_U} \sum_{l=0}^{2^{m-M_L}-1} \frac{r_{n+lK}^m \cdot \sqrt{E_0}\beta^{-m/2}}{\sigma_z^2 \beta^{-m}} \underset{H_0}{\overset{H_1}{\gtrless}} 0.$$

under the assumption of equally likely hypotheses, *i.e.*, a random bit stream. The bit-error

probability associated with this optimal receiver is readily derived, and can be expressed as

$$\Pr(\varepsilon) = \Pr(\ell > 0 | H_0) = Q\left(\frac{1}{2}\sqrt{K\sigma_c^2}\right) \tag{43}$$

where $Q(\cdot)$ is defined by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-v^2/2}\, dv.$$

and where σ_c^2 is the SNR in the channel, *i.e.*

$$\sigma_c^2 = \frac{E_0}{\sigma_z^2}.$$

Substituting for K in (43) via (41) we can rewrite this error probability in terms of the channel rate-bandwidth ratio as

$$\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\sigma_c^2\left[\frac{2\eta_F}{R/W} - 1\right]}\right), \tag{44}$$

where, again, $\eta_F \approx 1/2$. Note that the performance of fractal modulation is independent of the spectral exponent of the noise process when we use spectral matching.

To establish a performance baseline, we shall also evaluate a modified version of our benchmark modulation in which we incorporate repetition-coding, *i.e.*, in which we add redundancy by transmitting each sample of the message sequence K times in succession. This comparison scheme is not particularly power efficient both because signal power is distributed uniformly over the available bandwidth irrespective of the noise spectrum, and because much more effective redundancy schemes can be used with channels of known bandwidth (see, *e.g.*, [16]). Nevertheless, with these caveats in mind, such comparisons do lend some insight into the *relative* power efficiency of fractal modulation.

In our modified benchmark modulation, incorporating redundancy reduces the effective decoding rate per unit bandwidth by a factor of K, *i.e.*,

$$\frac{R}{W} = \frac{\eta_0}{K}, \tag{45}$$

where η_0 is the efficiency of the modulation without coding, *i.e.*, unity. When the channel

adds stationary white Gaussian noise, for which $H = -1/2$, the optimum receiver for this scheme demodulates the received data and averages together the K symbols associated with the transmitted bit, thereby generating a sufficient statistic. When this statistic is positive, the receiver decodes a 1-bit, and a 0-bit otherwise. The corresponding performance is, therefore, given by

$$\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\sigma_\varepsilon^2 K}\right) = Q\left(\frac{1}{2}\sqrt{\sigma_\varepsilon^2\left[\frac{\eta_0}{R/W}\right]}\right), \tag{46}$$

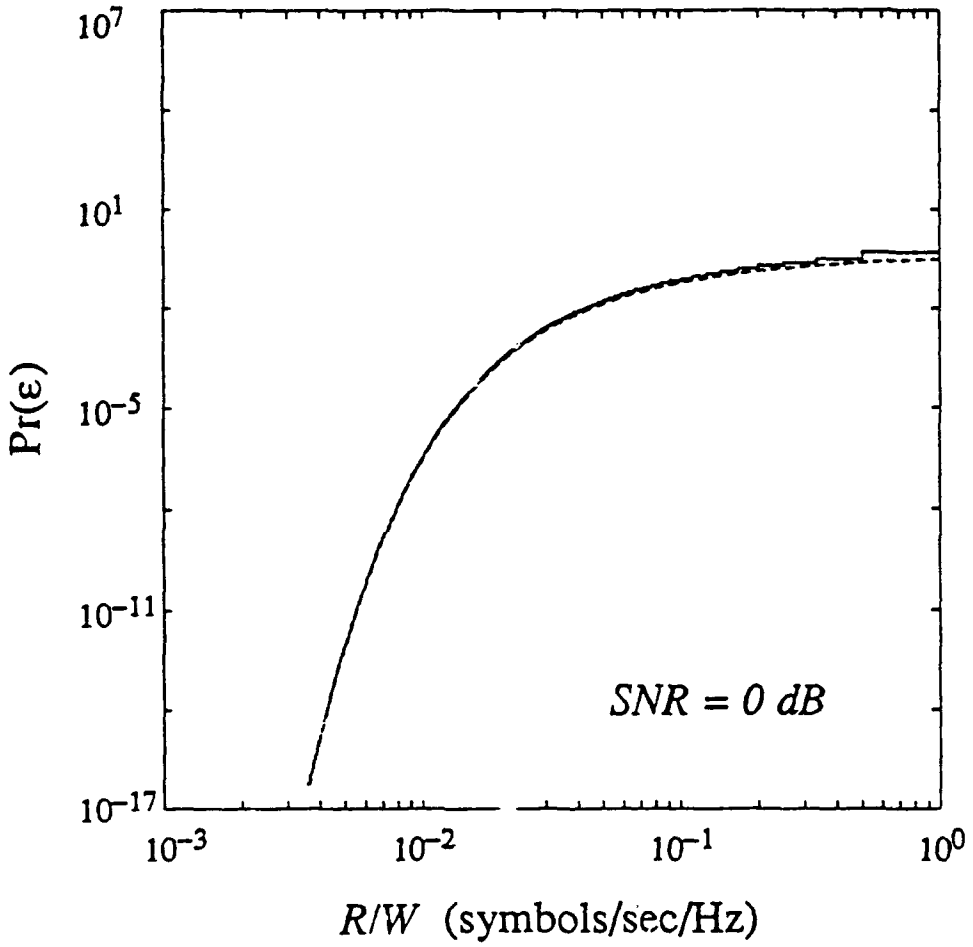
where the last equality results from substituting for K via (45).

Comparing (46) with (44), we note that since $\eta_0 \approx 2\eta_F$, the asymptotic bit-error performances of fractal modulation and the benchmark scheme are effectively equivalent for $R/W \ll \eta_F$, as is illustrated in Fig. 8. In Fig. 8(a), $\Pr(\varepsilon)$ is shown as a function of R/W at a fixed SNR of 0 dB ($\sigma_\varepsilon^2 = 1$), while in Fig. 8(b), $\Pr(\varepsilon)$ is shown as a function of SNR at a fixed $R/W = 0.1$ bits/sec/Hz. Both these plots reveal strong thresholding behavior whereby the error probability falls off dramatically at high SNR and low R/W. We emphasize that comparisons between the two schemes are appropriate only for the case in which the noise has parameter $H = -1/2$, corresponding to the case of stationary white Gaussian noise. For other values of H, the performance of the benchmark modulation is not only difficult to evaluate, but necessarily poor as well because of inefficient distribution of power among frequencies.

5 Concluding Comments

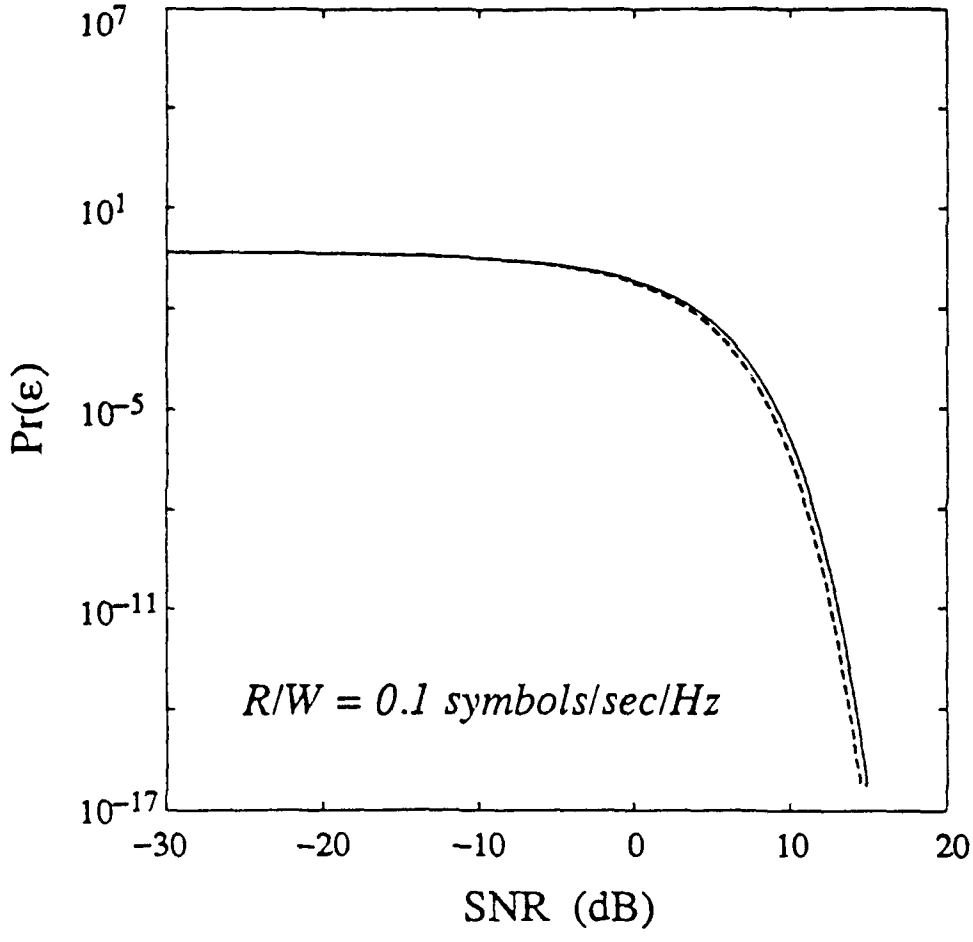
We have developed convenient, efficient, and robust wavelet-based representations for a generalized class of homogeneous signals, and explored their properties. Furthermore, we have explored their potential for use as modulating waveforms in a communications-based application, and demonstrated that fractal modulation would appear to be well-suited for use with noisy channels of simultaneously uncertain duration and bandwidth.

While our development of fractal modulation considered many issues, many others, such as synchronization and buffering, remain to be investigated. Furthermore, there are many potential refinements to be explored. One might involve the incorporation of block or trellis



(a) Bit-error probability Pr(ϵ) as a function of Rate/Bandwidth ratio R/W at 0 dB SNR.

Figure 8: *Bit-error rate performance of fractal modulation. Solid lines indicate the performance of fractal modulation, while dashed lines indicate the performance of the benchmark modulation with repetition coding.*



(b) Bit-error probability $\Pr(\epsilon)$ as a function of SNR at $R/W = 0.1$ symbols/sec/Hz.

Figure 8: *Continued.*

coding techniques to improve the power efficiency of the modulation. It would seem that coding of this type cannot be incorporated without sacrificing properties of the transmission scheme. In particular, the simple redundancy scheme apparent in Fig. 7 enables the recovery of a message $\mathbf{q}$ from observations corresponding to *any* single cell of the time-frequency plane. Nevertheless, it would be important to identify the tradeoffs involved.

The potential of fractal modulation in LPI applications also remains to be explored. While we have argued that the second-order statistics of homogeneous signals are effectively indistinguishable from those of $1/f$ noises, a more comprehensive study of the detectability of homogeneous signals is warranted. In the process, some potentially useful extensions to fractal modulation may arise. As an example, drawing from the notions underlying direct-sequence spread spectrum, one technique for more effectively concealing the modulation from unintended receivers might involve premultiplying the entire wavelet coefficient field x_n^m of the signal $x(t)$ prior to transmission by a pseudorandom bit field known to both transmitter and receiver.

Finally, we remark that there would appear to be many additional applications for the self-similar signals we have introduced in this paper. In many respects, identifying and exploring other potentially promising applications represents perhaps the most exciting direction for future research.

Acknowledgement

The authors wish to thank Prof. Alan S. Willsky for many valuable discussions, comments and suggestions regarding this work. We also wish to thank the anonymous reviewers for their careful reading and thoughtful criticisms of the original manuscript, which led to significant improvements in the presentation of this work.

A Proof of Theorem 2

To show that $y(t)$ has finite energy, we exploit an equivalent synthesis for $y(t)$ as the output of a cascade of filters driven by $x(t)$, the first of which is an ideal bandpass filter whose passband includes $\omega_L < |\omega| < \omega_U$, and the second of which is the filter given by (10).

Let $b_m(t)$ b: the impulse response of a filter whose frequency response is given by

$$B_m(\omega) = \begin{cases} 1 & 2^m\pi < |\omega| \leq 2^{m+1}\pi \\ 0 & \text{otherwise} \end{cases}, \tag{47}$$

and let $b(t)$ be the impulse response corresponding to (10). Furthermore, choose finite integers M_L and M_U such that $2^{M_L}\pi < \omega_L$ and $\omega_U < 2^{M_U+1}\pi$. Then, using $*$ to denote convolution,

$$\begin{aligned} y(t) &= b(t) * \left[\sum_{m=M_L}^{M_U} b_m(t) \right] * x(t) \\ &= b(t) * \sum_{m=M_L}^{M_U} \tilde{x}_m(t) \end{aligned} \tag{48}$$

where

$$\tilde{x}_m(t) = x(t) * b_m(t) = 2^{-mH}\tilde{x}_0(2^m t), \tag{49}$$

and where the last equality in (49) results from an application of the self-similarity relation (8) and the identity

$$b_m(t) = 2^m b_0(2^m t).$$

Because $x(t)$ is energy-dominated, $\tilde{x}_0(t)$ has finite energy. Hence, (49) implies that every $\tilde{x}_m(t)$ has finite energy. Exploiting this fact in (48) allows us to conclude that $y(t)$ must have finite energy as well.

To verify the spectrum relation (11), we express (48) in the Fourier domain. Exploiting the fact that we may arbitrarily extend the limits in the summation in (48), we get

$$Y(\omega) = B(\omega) \sum_{m=-\infty}^{\infty} \tilde{X}_m(\omega) = \begin{cases} X(\omega) & \omega_L < |\omega| < \omega_U \\ 0 & \text{otherwise} \end{cases}$$

where $\tilde{X}_m(\omega)$ denotes the Fourier transform of $\tilde{x}_m(t)$, and where

$$X(\omega) \triangleq \sum_{m=-\infty}^{\infty} \tilde{X}_m(\omega). \tag{50}$$

The right-hand side of (50) is, of course, pointwise convergent because for each ω at most one term in the sum is non-zero. Finally, exploiting (49) in (50) gives

$$X(\omega) = \sum_m 2^{-m(H+1)} \tilde{X}_0(2^{-m}\omega),$$

which, as one can readily verify, satisfies (12) ∎

B Proof of Theorem 3

To prove the "only if" statement, we suppose $x(t) \in E^H$ and begin by expressing $x(t)$ in terms of the ideal bandpass wavelet basis. In particular, we let

$$x(t) = \sum_m \tilde{x}_m(t)$$

where

$$\tilde{x}_m(t) = \beta^{-m/2} \sum_n \tilde{q}[n] \tilde{\psi}_n^m(t)$$

and where $\tilde{q}[n]$, the generating sequence in this basis, has energy $\tilde{E} < \infty$. The new generating sequence $q[n]$ can then be expressed as

$$q[n] = \sum_m q_m[n] \tag{51}$$

where

$$q_m[n] = y_m(t)|_{t=n}$$

and

$$y_m(t) = \tilde{x}_m(t) * \psi(-t).$$

For each m, since $\tilde{x}_m(t)$ is bandlimited, $y_m(t)$ and $q_m[n]$ each have finite energy and Fourier transforms $Y_m(\omega)$ and $Q_m(\omega)$ respectively. Hence,

$$Q_m(\omega) = \sum_k Y_m(\omega - 2\pi k) \tag{52}$$

where

$$Y_m(\omega) = \begin{cases} (2\beta)^{-m/2}\Psi^*(\omega)\tilde{Q}(2^{-m}\omega) & 2^m\pi < |\omega| \leq 2^{m+1}\pi \\ 0 & \text{otherwise} \end{cases}$$

with $\tilde{Q}(\omega)$ denoting the Fourier transform of $\tilde{q}[n]$, and $\Psi^*(\omega)$ the complex conjugate of $\Psi(\omega)$.

In deriving bounds on the energy E_m in each sequence $q_m[n]$ for a fixed m, it is convenient to consider the cases $m \leq -1$ and $m \geq 0$ separately. When $m \leq -1$, the sampling by which $q_m[n]$ is obtained involves no aliasing. Since on $|\omega| \leq \pi$ we then have

$$Q_m(\omega) = Y_m(\omega)$$

we may deduce that $q_m[n]$ has energy

$$E_m = \sum_n |q_m[n]|^2 = \frac{(2\beta)^{-m}}{\pi} \int_{2^m\pi}^{2^{m+1}\pi} |\Psi(\omega)|^2 \, |\tilde{Q}(2^{-m}\omega)|^2 \, d\omega. \tag{53}$$

Because $\psi(t)$ has R vanishing moments, there exists a $0 < \epsilon_0 < \infty$ such that

$$|\Psi(\omega)| \leq \epsilon_0 |\omega|^R \tag{54}$$

for all ω. Exploiting this in (53) we obtain

$$E_m \leq C_0 2^{(2R-\gamma)m} \tilde{E} \tag{55}$$

for some $0 \leq C_0 < \infty$.

Consider, next, the case corresponding to $m \geq 0$. Since $\psi(t)$ has R vanishing moments, there also exists a $0 < \epsilon_1 < \infty$ such that

$$|\Psi(\omega)| \leq \epsilon_1 |\omega|^{-R} \tag{56}$$

for all ω. Hence, on $2^m\pi < |\omega| \leq 2^{m+1}\pi$,

$$|Y_m(\omega)| \leq \epsilon_1 \, \pi^{-R} \, 2^{-(\gamma+1+2R)m/2} \, |\tilde{Q}(2^{-m}\omega)|. \tag{57}$$

From (52), we obtain

$$|Q_m(\omega)| \leq \epsilon_1 \pi^{-R} 2^{-(\gamma+1+2R)m/2} \sum_{k=0}^{2^m-1} |\tilde{Q}(2^{-m}\omega + 2\pi k 2^{-m})| \tag{58}$$

by exploiting, in order, the triangle inequality, the bound (57), the fact that only 2^m terms in the summation in (52) are non-zero since $y_m(t)$ is bandlimited, and the fact that $\tilde{Q}(\omega)$ is 2π-periodic. In turn, we may use, in order, (58), the Schwarz inequality, and again the periodicity of $\tilde{Q}(\omega)$ to conclude that

$$\begin{aligned} E_m &\leq \epsilon_1^2 \pi^{-2R} 2^{-(\gamma+1+2R)m} \left[\sum_{k=0}^{2^m-1} \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{Q}(2^{-m}\omega + 2\pi k 2^{-m})|^2 \, d\omega} \right]^2 \\ &\leq C_1 2^{-(\gamma-2+\omega R)m} \tilde{E} \end{aligned} \tag{59}$$

for some $0 \leq C_1 < \infty$.

Using (51), the triangle inequality, and the Schwarz inequality, we obtain the following bound on the energy in $q[n]$

$$E = \sum_n |q[n]|^2 \leq \left[\sum_m \sqrt{E_m} \right]^2$$

which, from (59) and (55) is finite provided $0 < \gamma < 2R$ and $R \geq 1$.

Let us now show the converse. Suppose $q[n]$ has energy $E < \infty$, and express $x(t)$ as

$$x(t) = \sum_m x_m(t)$$

where

$$x_m(t) = \beta^{-m/2} \sum_n q[n] \, \psi_n^m(t).$$

If we let

$$\tilde{y}_m(t) = b_0(t) * x_m(t)$$

where $b_0(t)$ is the impulse response of the ideal bandpass filter in Definition 1, it suffices to

show that

$$\tilde{y}(t) = \sum_m \tilde{y}_m(t) \tag{60}$$

has finite energy.

For each m, we begin by bounding the energy in $\tilde{y}_m(t)$, which is finite because $x_m(t)$ has finite energy. Since $\tilde{y}_m(t)$ has Fourier transform

$$\tilde{Y}_m(\omega) = \begin{cases} (2\beta)^{-m/2} Q(2^{-m}\omega)\, \Psi(2^{-m}\omega) & \pi < \leq < 2\pi \\ 0 & \text{otherwise} \end{cases}$$

where $Q(\omega)$ is the discrete-time Fourier transform of $q[n]$, we get that

$$\tilde{E}_m = \frac{2^{-\gamma m}}{\pi} \int_{2-m\pi}^{2^{-m+1}\pi} |Q(\omega)|^2 \, |\Psi(2^{-m}\omega)|^2 \, d\omega.$$

Again, it is convenient to consider the cases corresponding to $m \leq -1$ and $m \geq 0$ separately. For $m \leq -1$, most of the energy in $x_m(t)$ is at frequencies below the passband of the bandpass filter. Hence, using the bound (56) and exploiting the periodicity of $Q(\omega)$ we obtain

$$\tilde{E}_m \leq \tilde{C}_0 2^{(2R-1-\gamma)m} E. \tag{61}$$

for some $0 \leq \tilde{C}_0 < \infty$. For $m \geq 0$, most of the energy in $x_m(t)$ is at frequencies higher than the passband of the bandpass filter. Hence, using the bound (54) we obtain

$$\tilde{E}_m \leq \tilde{C}_1 2^{-(\gamma+2R+1)m} E. \tag{62}$$

for some $0 \leq \tilde{C}_1 < \infty$.

Finally, using (60), the triangle inequality, and the Schwarz inequality, we obtain the following bound on the energy in $\tilde{y}(t)$

$$\tilde{E} = \int_{-\infty}^{\infty} |\tilde{y}(t)|^2 \, dt \leq \left[\sum_m \sqrt{\tilde{E}_m} \right]^2$$

which, from (62) and (61) is finite provided $0 < \gamma < 2R - 1$ since $R \geq 1$ ■

C Proof of Theorem 5

Following an approach analogous to the proof of Theorem 2, let $b_m(t)$ be the impulse response of a filter whose frequency response is given by (47), and let $b(t)$ be the impulse response corresponding to (10). By choosing finite integers M_L and M_U such that $2^{M_L}\pi < \omega_L$ and $\omega_U < 2^{M_U+1}\pi$, we can again express $y(t)$ in the form of eq. (48). Because $x(t)$ is power-dominated, $\tilde{x}_0(t)$ has finite power. Hence, (49) implies that every $\tilde{x}_m(t)$ has finite power. Exploiting this fact in (48) allows us to conclude that $y(t)$ must have finite power as well.

To verify the spectrum relation (28), we use (48) together with the fact that the $\tilde{x}_m(t)$ are uncorrelated for different m to obtain

$$S_y(\omega) = |B(\omega)|^2 \sum_{m=-\infty}^{\infty} S_{\tilde{x}_m}(\omega) = \begin{cases} S_x(\omega) & \omega_L < |\omega| < \omega_U \\ 0 & \text{otherwise} \end{cases}$$

where $S_{\tilde{x}_m}(\omega)$ denotes the power spectrum of $\tilde{x}_m(t)$, and where

$$S_x(\omega) \triangleq \sum_{m=-\infty}^{\infty} S_{\tilde{x}_m}(\omega). \tag{63}$$

Again we have exploited the fact that the upper and lower limits on the summation in (48) may be extended to ∞ and $-\infty$, respectively. The right-hand side of (63) is, again, pointwise convergent because for each ω at most one term in the sum is non-zero. Finally, exploiting (49) in (63) gives

$$S_x(\omega) = \sum_m 2^{-\gamma m} S_{\tilde{x}_0}(2^{-m}\omega)$$

which, as one can readily verify, satisfies (29) ∎

References

[1] I. M. Gel'fand, G. E. Shilov, N. Y. Vilenkin, and M. I. Graev, *Generalized Functions*. New York, NY: Academic Press, 1964.

[2] G. W. Wornell, "Synthesis, analysis, and processing of fractal signals," RLE Tech. Rep. No. 566, M. I. T., Cambridge, MA, Oct. 1991.

[3] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 674–693, July 1989.

[4] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, Nov. 1988.

[5] B. B. Mandelbrot, *The Fractal Geometry of Nature.* San Francisco, CA: Freeman, 1982.

[6] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 197–199, Jan. 1989.

[7] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, 1992. To appear.

[8] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Rev.*, vol. 31, pp. 614–627, Dec. 1989.

[9] M. F. Barnsley, *Fractals Everywhere.* New York, NY: Academic Press, 1988.

[10] F. J. Malassenet and R. M. Mersereau, "Wavelet representations and coding of self-affine signals," in *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1991.

[11] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Comm. Mag.*, pp. 5–14, May 1990.

[12] I. Kalet, "The multitone channel," *IEEE Trans. Commun.*, vol. COM-37, pp. 119–124, Feb. 1989.

[13] R. G. Gallager, *Information Theory and Reliable Communication.* New York, NY: John Wiley and Sons, 1968.

[14] M. S. Keshner, "$1/f$ noise," *Proc. IEEE*, vol. 70, pp. 212–218, Mar. 1982.

[15] G. W. Wornell, "A Karhunen-Loève-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 859–861, July 1990.

[16] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. SAC-2, pp. 632–647, Sept. 1984.