JONG-HWA KIM

Lossless Wideband Audio Compression: Prediction and Transform

Verlustfreie Kompression für Breitband-Audiosignale: Prädiktion und Transformation

Von der Fakultät I - Geisteswissenschaften
der Technischen Universität Berlin
zur Verleihung des akademischen Grades

Doktor der Philosophie
- Dr. phil. -

genehmigte Dissertation

Technische Universität Berlin

Lossless Wideband Audio Compression: Prediction and Transform

Verlustfreie Kompression für Breitband-Audiosignale: Prädiktion und Transformation

vorgelegt von
Master of Electronic Engineering
Jong-Hwa Kim
aus Seoul / Korea

Von der Fakultät I - Geisteswissenschaften
der Technischen Universität Berlin
zur Verleihung des akademischen Grades

Doktor der Philosophie
- Dr. phil. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Eckart Mensching
Berichter: Prof. Dr. Klaus Hobohm
Berichterin: Prof. Dr. Helga de la Motte-Haber

Tag der wissenschaftlichen Aussprache: 19. Dezember 2003

Berlin 2004
D 83

ABSTRACT

Lossless Wideband Audio Compression: Prediction and Transform

This thesis studies lossless audio compression. In the domain of lossless compression, research takes place on two broad development sections, signal modeling and coding algorithm. The former is concerned with the *understanding* of the source signal, while coding is the more tightly specified task of efficiently representing a single symbol as a code. The focus of this thesis is the evaluation and the development of signal modeling techniques for lossless compression. Related with the modeling method used to decorrelate a signal, the data compression schemes are generally divided in two categories, predictive modeling and transform-based modeling. In the thesis, all two categories are investigated in depth and handled from the lossless viewpoint.

The first contribution of the thesis is an exploration of the general audio compression systems including the lossy compression system. In predictive modeling, the structures of various linear prediction filters are introduced by presenting the fundamental autoregressive modeling. The prediction filters including the approaches to the nonstationary signal modeling and to the adaptive linear prediction filters are explored and evaluated by testing within a prototypical lossless audio compression system.

For transform modeling, two well-known subband transform coding methods, Laplacian pyramid and subband coding scheme, are first described, and then the design methods of perfect reconstruction multirate filter banks are studied. Concerning with the modulated lapped orthogonal transform, the efficiency of linear prediction from subband and from fullband is formally compared and empirically examined. Wavelet transform is in depth studied from the various viewpoints in order to find the theoretical relationship between the wavelet and the multirate filter banks. Theoretical and practical aspects of reversible transforms are discussed by introducing the S-transform, S+P transform, and RTS transform. The lifting method is examined as a means to realize the biorthogonal wavelets. Integer lifting scheme with rounding-off method is investigated to construct reversible version of wavelet transforms and its performance is validated by applying to lossless audio compression. Finally, some of the more important results presented in this thesis are summarized with the suggesting directions for future research.

ZUSAMMENFASSUNG

Verlustfreie Kompression für Breitband-Audiosignale: Prädiktion und Transformation

Das Ziel dieser Dissertation ist die Untersuchung von verlustfreier Audiokompression. Im Bereich der verlustfreien Kompression teilt sich die Forschung in zwei große Entwicklungsabschnitte, nämlich Signalmodellierung und Codierungsverfahren. Die Signalmodellierung befasst sich mit dem *Verstehen* des Quellsignals, während sich die Codierung mit dem speziellen Problem der effizienten Repräsentation von Einzelsymbolen befaßt. Der Fokus dieser Dissertation liegt auf der Evaluation und Entwicklung von Signalmodellierungstechniken für die verlustfreie Kompression. Mit Bezug auf die Modellierungsmethode zur Dekorrelation des Signals fallen die Datenkompressionssysteme in zwei Kategorien, nämlich die prädiktive Modellierung und die auf Transformation basierende Modellierung. In der Dissertation wurden beide Kategorien aus der Sicht verlustfreier Kompression ausführlich untersucht.

Der erste Beitrag der Dissertation ist der Erforschung von Audiokompressionssystemen einschließlich der verlustbehafteten Kompressionssysteme gewidmet. Bei der prädiktiven Modellierung werden die verschiedenen Strukturen der Linear-Prädiktionsfilter durch Vorlage der fundamentalen autoregressiven Modellierung vorgestellt. Die Prädiktionsfilter einschließlich der Ansätze zur Modellierung des nicht-stationären Signals und adaptiven Linear-Prädiktionsfiltern werden vorgestellt, und deren Effizienz im Experiment mit einem prototypischen verlustfreien Audiokompressionssystem evaluiert.

Bezüglich der Transformationsbasierten Modellierung werden zwei prominente Codierungsmethoden beschrieben, die auf Subbandtransformation basieren, nämlich Laplacian Pyramide und Subbandcodierung, beschrieben, und die Entwurfsverfahren für perfekt rekonstruierbare Multirate-Filterbänke werden dargestellt. Die Effizienz linearer Prädiktion von Vollbandsignalen und Subbandsignalen wird theoretisch untersucht und empirisch validiert. Die Wavelet-Transformation wird nach verschiedenen Aspekten gründlich untersucht, um die theoretische Verknüpfung zwischen der Wavelet-Transformation und der Multirate-Filterbank zu diskutieren. Theoretische und praktische Aspekte von reversiblen Transformationen werden in Zusammenhang mit S-Transformation, S+P Transformation, und RTS Transformation untersucht. Die Liftingmethode, die die biorthogonalen Wavelets realisiert, wird untersucht. Um eine reversible Wavelet-Transformation zu konstruieren, wird das mit Ganzzahlen operierende Liftingsystem durch Abrundungsverfahren entwickelt. Dessen Funktion wird durch Anwendung auf die verlustfreie Audiokompression validiert. Zum Schluss sind einige wichtige Untersuchungsergebnisse der Dissertation mit Vorschlägen für zukünftige Forschungsarbeiten zusammengefasst.

ACKNOWLEDGEMENTS

This thesis is dedicated to the memory of my supervisor Prof. Dr.-Ing. Manfred Krause. As a teacher and a partner, he introduced me to this intriguing field of research and continuously taught me precise scientific thinking and writing. I feel deeply grateful for his academic support and moral endorsement through the years. The scientific community of audio researchers suffered from a great loss of a person who contributed outstanding ardour and affection to the area, when he passed away a few months ago.

I would like to thank Prof. Dr. Klaus Hobohm and Prof. Dr. de la Motte-Haber for reviewing this thesis and numerous valuable comments and suggestions. I also thank Dr.-Ing. Axel Röbel who works currently at IRCAM, France, for fruitful discussions in this work. Doris Grasse deserves special thanks for all arrangements of my postgraduate life in the Institute.

My current colleagues at Multimedia Lab of University Augsburg have also been very supportive for successful conclusion of this work. I would like to thank Prof. Dr. Elisabeth André who gave me the opportunity to continue and finish my work I started in Berlin. Thanks to Dr. Martin Müller for being my best friend.

Finally, my deepest thanks for beaming love from my parents and brothers. Thanks to Jongkwan for bringing me to the sunny places. Thanks, Jongihn, gathering our family together always in a good atmosphere. Thanks, Jongho, for reminding how science and art go together, and being my best friend. Thanks to my mother who always understood the value of continuing my long schooling. Thanks to my father for encouraging me to follow my own path to happiness, letting me know you are always there for me, and I deeply respect your honest life of 40 years as a schoolteacher.

Eternal gratitude goes to my wife Heejin whose unfathomable love and understanding during this work made it possible for me to keep going. Thanks to our son Nova who does not pose any questions related to the contents of this thesis yet.

Contents

List of Figures

List of Tables

List of Abbreviations

ADPCM	Adaptive Pulse Code Modulation
AR	Autoregressive
ARMA	Autoregressive Moving Average
BPF	Band-pass Filter
CD	Compact Disc
CDF	Cohen-Daubechies-Feauveau wavelet
CMF	Conjugate Mirror Filter
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DD	Deslauriers-Duduc wavelet
DST	Discrete Sine Transform
DVD	Digital Versatile Disc
DWT	discrete Wavelet Transform
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FIR-CM	FIR Filter with Context-based Error Modeling
HPF	High-pass Filter
IIR	Infinite Impulse Response
IRWT	Integer Reversible Wavelet Transform
ISO	International Organization for Standardization
KLT	Karhunen-Loève Transform
LMS	Least Mean Square
LOT	Lapped Orthogonal Transform
LP	Linear Prediction
LPC	Linear Predictive Coding
LPF	Low-pass Filter
LSI	Linear Shift Invariant
MA	Moving Average Model
MDCT	Modified Discrete Transform
MLT	Modulated Lapped Transform
MMSE	Minimum Mean Square Error

MPEG	Motion Picture Experts Group
MRA	Multiresolution Analysis
NLMS	Normalized LMS
PAP	Polynomial Approximation
PCM	Pulse Code Modulation
PMF	Probability Mass Function
PRF	Perfect Reconstruction Filter
PSD	Power Spectral Density
QMF	Quadrature Mirror Filter
RLS	Recursive Least Square
RTS	Reversible Two-Six Transform
SFM	Spectral Flatness Measure
SMR	Signal-to-Mask Ratio
SPL	Sound Pressure Level
SQAM-CD	Sound Quality Assessment Material CD
S-Transform	Sequential Transform
WLP	Warped Linear Prediction
WT	Wavelet Transform

Chapter 1

Introduction

Recently a number of new technologies related to the storage capacity and the transmission bandwidth are emerging. From transmission part, for example, ADSL (Asymmetric Digital Subscriber Line) provides several Mbps transmission bandwidth for normal telephone line on down stream side. On storage part, hard disk capacity has being increased dramatically. A new high-density disc such as DVD (Digital Versatile Disc) also provides huge storage capability for audio and video sources. However, despite such tremendous growth of the storage capacity and the transmission bandwidth, the demand for higher quality of multimedia associated with audio, image, and video continues to outpace it. For instance, the required data rate satisfying the high-quality audio (*more word size, more sample rate, and more channel*) will be continuously increased, unless we give up to enjoy the digital audio world. Hence the importance of data compression is not likely to diminish, as a key technology to allow efficient storage and transmission.

The general idea behind data compression is to remove the redundancy present in the data to find more compact representations. Two families of algorithms exist in compression. When the information can be exactly recovered from the bits, the source coding or compression is called *lossless*; otherwise, it is called *lossy*. To achieve higher compression ratios, lossy algorithms remove information from the original in a way that comes *close* to the original or that is not perceptible. In this case, therefore, we allow approximate representations of the original, instead of trying to represent the original exactly, and have only a modified version of the original after transmission. In contrast, lossless algorithms respect the integrity of the original signal. After transmission and reconstruction, an exact copy of the original signal is available.

Two main drawbacks are related with lossless audio compression. The first one is the time varying compression ratio, which makes difficult the allocation of a fixed bandwidth to the transmitted audio data and also complicates the editing of compressed material. The second one is the poor compression rate. The lossless audio compression usually achieves compression ratio 2 to 3 without loss any quality, whereas lossy compression can

achieve compression ratio 8 to 40 or higher. While achieving higher compression ratios, the lossy audio compression is highly objectionable in high fidelity audio compression applications, due to unexpected artifacts introduced even by the most heavily engineered schemes which use perceptual auditory models. It is more problematical whenever the audio signal undergoes multiple encoding-decoding operations.

Pure-Audio DVD standards and new audio format for high-density discs are receiving serious consideration from audio community. Since the sample rate is going up to 96kHz and quantization level up to 24 bits, lossless audio compression as a *reformatting* of the standard PCM audio offers an opportunity to make a much better product in that more precision and more channels can be provided. Another application where lossless compression is currently used is archiving large corpus of speech data for distributing it to the research community. Certain scenarios make use of both lossy and lossless audio compression: an audio on demand application will use lossy compression for browsing and ordering the desired tune, while lossless compression should be used for releasing the full length audio material. All of these enlighten the lossless audio compression. Moreover, as we can see today, it is fact that the growth of the storaging capacity is substantially rapider than that of the network bandwidth. In this sense, it should not be impetuous to say that we will have no reason to accept the lossy compressed audio data and that, therefore, the lossy compression will be no more dominating technology, at least in the storage branch today.

This thesis studies lossless audio compression. In the domain of lossless compression, research takes place on two broad development sections, signal modeling techniques and coding algorithms. The former is concerned with the *understanding* of the source signal, while the coding is the more tightly specified task of, given a set of estimated symbol probabilities, efficiently representing a single symbol as a code, usually in binary. The focus of this thesis is the development of signal modeling techniques. Related with the modeling method used to decorrelate a signal, the data compression schemes are generally divided in two categories, predictive modeling and transform-based modeling. In the thesis, all two categories, predictive coding and transform coding, are investigated in depth and handled from the lossless viewpoint.

The first contribution of the thesis is an exploration of the general audio compression systems including the lossy compression systems. Parallelly we state the problems that we will discuss through the thesis from various angles. In Chapter 3, the structures of various linear prediction filters are introduced by presenting the fundamental autoregressive modeling. We review the approaches to the nonstationary signal modeling and describe the adaptive linear prediction filters. There are infinitely many alternative ways to form a linear combination of signal history. As a good example of modified linear prediction filter, frequency-warped linear prediction is introduced by the generalization of the allpole linear prediction filter. To remove the redundancy still remaining in the prediction errors, we consider the context-based error modeling method. An experimental comparison

of all these prediction filters is presented by testing within a prototypical lossless audio compression system.

In Chapter 4, we begin to treat the transform-based coding methods. We first review two well-known subband transform coding methods, Laplacian pyramid and subband coding scheme, and then describe the design methods of perfect reconstruction multirate filter banks. In principle, the block transforms e.g., discrete cosine transform and wavelet transform, qualify also as a subband transform. To solve the blocking effect, which is the discontinuity between block boundaries in block transforms, we discuss the lapped orthogonal transform. The efficiency of prediction from subbands and from fullband is formally compared and empirically investigated by designing a hybrid audio compression system.

We explore the wavelet transform in Chapter 5. From the signal processing viewpoint, it is clear that all block transforms and all subband transforms including wavelet transform can be defined and implemented by the filter design and windowing theory. It is also the fact that the terms in the literature, e.g., subband coding, filter bank method, and multiresolution transform, that are used to eventually mean the wavelet transform, are defined ambiguously and used heedlessly, strictly speaking. However, the wavelet transform, that is one of the most important new phenomena in signal processing area since the Fourier transform has long ruled over the area, has obviously different history and characteristics. Hence, we study the wavelet transform in detail by beginning to review the principles of time-frequency representation and describe the theoretical relationship between the wavelet and the multirate filter banks.

In Chapter 6, both theoretical and practical aspects of reversible transforms are discussed. First, the reversibility property is defined and the reason why transforms with this property are desirable is explained. Some good reversible transforms such as the S+P transform are introduced. Next, lifting is examined as a means to realize the biorthogonal wavelets. Integer lifting scheme with fixed-point arithmetic is proposed to construct reversible version of wavelet transform. The chapter also evaluates the performance of several reversible transforms for lossless audio compression.

Finally, Chapter 7 summarizes some of the more important results presented in this thesis along with the contributions it makes. The chapter concludes by suggesting directions for future research.

Chapter 2

High-Quality Digital Audio Compression

In this chapter, we review the fundamental issues concerning the high-quality digital audio and describe general framework for audio compression schemes. Parallelly we state the problems that we will discuss in the thesis from various angles.

2.1 Digital Audio Representation

2.1.1 Wideband audio signal

Sound starts as a continuous vibration in air and we perceive it through an aural hearing mechanism. As shown in Figure 2.1.1, sounds are characterized by sound pressure level (SPL) and frequency limits. The SPL limit is bounded by the threshold of hearing and feeling, while the perceptible frequency is limited by bandwidth of about 10Hz to 20KHz. Wideband audio, which we will treat in this thesis, has a bandwidth of about 20-20KHz. Note that the speech signal can also be considered a wideband audio signal.

The conversion from the analog to the digital domain begins by sampling the audio input in regular, discrete intervals of time and quantizing the sampled values into a discrete number of evenly spaced levels. As a result, digital audio signal consists of a sequence of binary values representing the number of quantizer levels for each audio sample. The conventional format that represents each sample with an independent code word is linear pulse code modulation (PCM).
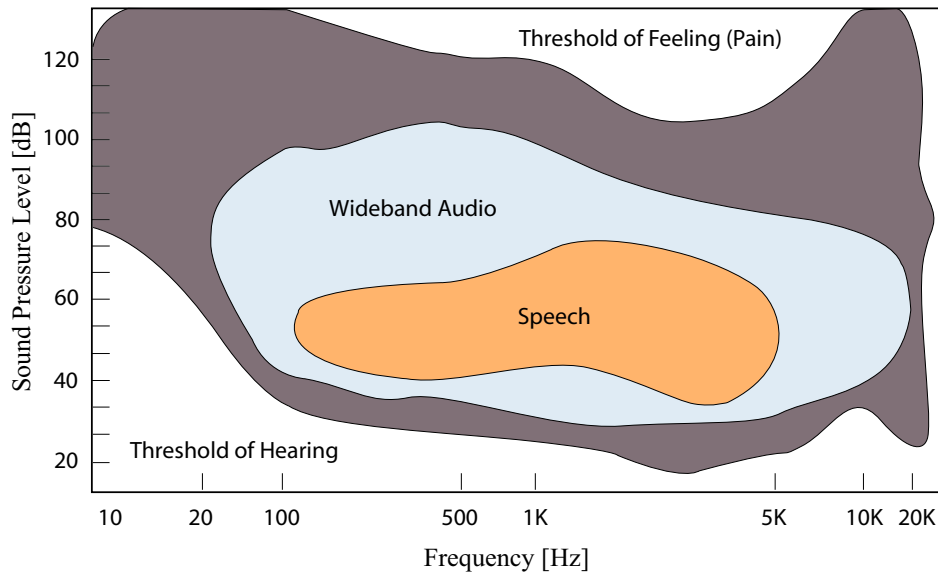
Figure 2.1.1: Limits of aural perception.

Compared with the analogue form, the digital representation of audio signal offers many advantages, e.g., high noise immunity, stability, and reproducibility. Audio information in digital form can be transmitted without loss or the introduction of interference and manipulated simply without distortions which are problematic in the case of analogue processing.

2.1.2   Bandwidth, capacity, and high-quality audio

Compact disc (CD), which has become a de facto standard for digital audio representation, carries the PCM audio sampled at 44.1kHz and quantized into 16-bit amplitude level. The bit rate of a CD stereo signal is then roughly 1.4Mbit/s. It is clear that the bit rate is too large for a real-time transmission over a restricted network bandwidth these days.

For storing the PCM audio, the problem seemed to be nearly solved by introducing new high-density CD formats, such as the Digital Versatile Disc (DVD) which has capacity of 4.7GB (single layer), 8.5GB (dual layer), and 17GB (double sided dual layer), while capacity of the conventional CD amounts 650 MB. Moreover, this capacity of DVDs will vastly be multiplied in the near future. Practically the replacing the red laser or infrared light, used to read the current CDs and DVDs, with *blue laser* having a shorter wavelength than the red variety enables to read and write more bits of information (nearly six times

the capacity of the current discs) on a given area of disc space. It is fact that the growth of the storaging capacity is substantially rapider than that of the network bandwidth. In this sense, it should not be impetuous to say that we will have no reason to accept the lossy compressed audio data and that, therefore, the lossy compression will be no more dominating technology, at least in the storage branch today.

Such high-density discs are a great evolution, indeed. On the other hand, the new aspirations for high-quality audio to achieve finer audio processing and more professional reproduction forces researchers to discuss issue of higher sample rate, larger word size, and more channels (practically, the term of High-Quality Audio means these three crucial factors) for such a high-density audio carrier. Consequently the first DVD-Audio standard (Table 2.1) approved even the sample rate up to 192kHz with word size up to 24-bit. Moreover the number of channel for the DVD-Audio extended to eight channels. The maximum data rate in this case amounts to 9.1Mbit/s. It is obvious that the required data rate satisfying the high-quality audio will be increased continuously, unless we give up enjoying the digital audio world.

Specification		DVD-Audio	CD
Audio Format		PCM	PCM
Disk Capacity		4.7Gb - Single layer 8.5Gb - Dual Layer 17Gb - Double Sided Dual Layer	650Mb
Channels		Up to 6	2 (stereo)
Frequency Response		0 - 96kHz (max)	5 - 20kHz
Dynamic Range		144dB	96dB
Sampling Rate	2 Channel	44.1, 88.2, 176.4kHz or 48, 96, 192kHz	44.1kHz
	Multichannel	44.1, 88.2kHz or 48, 96kHz	n/a
Sample Size		12, 16, 20, or 24bits	16bits
Maximum Data Rate		9.6Mbps	1.4Mbps

Table 2.1: DVD-Audio specification (version 1.0)

Lossless audio compression as a *reformatting* of the standard PCM audio offers an opportunity to make a much better product in that more precision and more channels can be provided. Hence it will continue to be a key technology for efficient transmission and storaging, in terms of the high-resolution audio on the high-density discs.

2.2   General Framework of Audio Compression System

Compression (source coding) is to represent information in smaller number of bits than of original, without any distortion or loss of information. Audio compression takes advantage of the continuity of digital audio, e.g., the statistical redundancy inherent in the signal and the tendency of samples to be similar to their neighbors. Many techniques for lossless compression exist. The most effective of these belong to a class of coders commonly called "entropy coders". These methods have proven effective for text compression, but perform poorly on most kinds of audio data, as they fail to exploit the high correlation that typically exists among the data samples. Therefore, preprocessing the data to achieve decorrelation is a desirable main step for the audio compression. Figure 2.2.1 shows a general framework of typical lossless audio compression system.
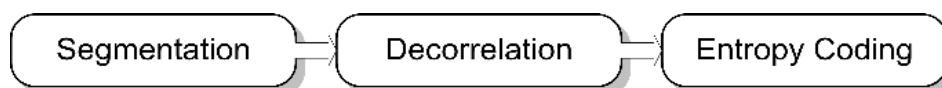
Figure 2.2.1: General framework of audio compression system

2.2.1   Segmentation of audio stream

At the first stage in Fig. 2.2.1, the PCM audio data stream is divided or grouped into blocks (frames) with appropriate block length. The modeling is then carried out locally and therefore each block has its own prefixed header that defines the parameters for decoding, which can change on block-to-block basis to follow the varying characteristics of the input signal over the time. The block length should not be too short, because the overheads of transmitting the header information may outweigh the other advantages. On the other hand, if the block length is too long, the temporal modification of the audio signal characteristics would not be taken into account, implying less efficiency of the compression system. In the state-of-the-art audio compression systems, typical block length is 576 to 1152 samples, which result in 13 to 26 ms frames for a signal sampled at 44.1 kHz.

In practice, the choice of the block length has considerable influence on the efficiency of the modeling scheme and on the resulting compression ratios consequently, since the coefficients for predictor or transform used for the decorrelation are computed within the each block separately, as mentioned above. Hence, the problem of finding an efficient segmentation is crucial to the entire compression system and depends strongly on the characteristics of the input audio signal. We discuss this problem more in detail in

Chapter 3.

2.2.2   Signal decorrelation

Generally, audio signals show continuous behaviour in the time domain. So they are represented by a continuous waveform which is acoustic pressure variation as a function of time. This continuity of the audio signals implies that a sample of sequence obtained from discretizing the signal is more likely to be close (correlated) to its adjacent one. Hence, applying universal coding methods, such as Lempel-Ziv related algorithms that assume that all samples are independent each other, directly to the audio signal gives poor compression results with a higher entropy than real entropy.

To take advantage of the long-term correlation related to the dependencies within the audio sample sequence, most audio compression systems have a preprocessing stage prior to the entropy coding, regardless of lossy or lossless. In the preprocessing stage, the segmented signal is decorrelated by using a linear prediction or a linear transform in order to compact the energy of the signal into a few uncorrelated coefficients. In fact, the decorrelation methods used in preprocessing stage characterize the current state-of-the-art audio compression schemes that are generally divided into two natures, *predictive coding* and *transform coding*. If the signal decorrelation performs well, the obtained coefficients (or residues resulted from predictive coding) will be uncorrelated and have the reduced signal variance with a flat frequency spectrum. The coefficients instead original audio signal are quantized into as few bits as possible by using either scalar or vector quantization method.

Predictive modeling

In the linear prediction, current sample value is predicted by an estimate formed as a linear combination of previous sample values. Then the difference between the original sample value and the estimate value is quantized and coded. For decoding the predictor coefficients must be coded as part of the lossless representation. Since signal redundancy is represented by the predictor coefficients, the residual has a smaller variance and a flatter signal spectrum than those of the original signal and therefore is amenable to quantization and the entropy coding.

In lossless compression system, no loss may be allowed in the quantization stage, whereas in the case of lossy compression, most of compression with an intentional loss occurs in this stage. As the PCM audio data consist of integer number, it is main task for lossless compression to retain the form of integer number through all the compression stages. In predictive coding, we can easily make the compression stages operating with

integer number by quantizing the prediction filter coefficients previously. We then obtain the residues with integer number and reform these into an efficient bit level without any quantization errors.

Arguably, an infinite number of alternative ways to form a linear combination of previous samples and to predict using it can be found without any mathematical limitation. In chapter 3, we investigate the various structures of the linear prediction filter that are used for lossless audio compression and develop a novel method, a combination of the linear prediction model with the context based error modeling.

Transform modeling

In transform coding, the input audio signal is decomposed by using a linear transform to pack the energy of the signal into the uncorrelated coefficients. The success of transform modeling depends on how well the basis functions of the transform represent the features of the signal. The most commonly used transform is the Fourier transform, or some variant of it.

Theoretically, the Karhunen-Loève transform (KLT) is best orthonormal energy-compaction transform for a Gaussian source model. However, the KLT is rarely used in practice for a variety of reasons. One crucial reason is that the KLT is signal dependent, that is, the transform used in the encoder and decoder must be adjusted to correspond to the covariance of the source in order to maintain optimality. In addition, since the KLT has no special structure, the KLT requires more operations to compute than a harmonic transform such as discrete cosine transform (DCT).

Most transform coders today employ the DCT, which yields nearly optimal energy concentration and provides a superiority of the computational efficiency. The main disadvantage of the DCT is that its basis functions are non-local. The correlation present in a signal, however, is mostly local. Since the signal is segmented into blocks, as mentioned in previous section, and then each block is individually decomposed by the DCT, the resulted coefficients can reveal their geometry that produces the well-known blocking effects. The lapped orthogonal transform (LOT) could be substituted for the DCT to avoid the blocking effects. On the other hand, the Fourier like series are best suited for representing periodic signals, or signals with specific parity property, while each block does not necessarily satisfy those properties. This slows down the convergence and hinders high compression ratios.

All transforms result in real-valued coefficients with infinite precision. While most of the compression occurs in the quantization of these coefficients, where an unrecoverable loss arises, traditional transform coding is then *lossy*. To achieve a transform-based lossless compression, one could maintain the coding error which is difference between the original signal and its approximation signal that is reconstructed locally in coder side

by local inverse transformation and dequantization. The coding residue is quantized and entropy coded and transmitted along with the coded coefficients. At the decoder side, the original signal can be exactly reconstructed by adding the residue to the approximation signal. This scheme is called *lossy-based lossless compression* (see Fig. 2.3.5 in section 2.3.2). In this case, a precise quantization of the transform coefficients is not desirable, while the problem is complex in lossy compression scheme.

Subband transform can be used for the transform coding scheme. In this case, the signal is successively subdivided by analysis filters, of which impulse response is at most as long as the subsampling factor used in the subbands. In Chapter 4, we describe the subband coding scheme concerned with multirate filter banks. Wavelet transform, one of the most successful tools in the data compression and signal processing, is explored in depth in Chapter 5. In chapter 6, various reversible transforms including the biorthogonal filtering structure using the lifting factorization are discussed. Lifting scheme is an efficient algorithm for discrete wavelet transform and that enables us to easily make the wavelet transform that maps integers to integers. Therefore the system can be used direct for the lossless compression scheme, not as lossy-based lossless compression.

2.2.3   Entropy coding

Before transmitting the residue sequence or coefficients, which are assumed to be uncorrelated and therefore may be coded independently, entropy coding is used to further save the bits in an invertible fashion. A signal carries a certain amount of *information*. This information will be, in the theoretical side, large if the sample coming out of a certain source (audio in our case) is seldom and small if it is frequent. In fact, a very frequent sample will give very small "novelty" and thus will be easily predictable. On the other hand, a rare sample will give new information and will be hardly predictable. After having observed this phenomenon, Shannon [2] defined in 1948 his source entropy theory. He showed that for the best possible compression code, the output length contains a contribution of $-\log_2 p$ bits from the encoding of each symbol whose probability of occurrence is p. If we can do that, we have an optimal code.

Entropy coding typically involves run-length coding combined with well-known Huffman or arithmetic codes. We briefly review these statistical codes and describe Rice code which is less well known and used for our test compression system in this thesis.

Huffman coding

The fundamental idea of Huffman coding is to use a look-up table for the input words to convert them to transmitted code words of varying length, so that commonly occurring

input data words should be represented by short output code words, whereas rarely occurring input words may be represented by fairly long code words, so that the average coded word length is small.

Given the residue sequence, probabilities of each symbol (residue value) occurring in the segmented stream are determined. The symbols are then ordered according to the probabilities in decreasing order. It leads to a node, called *leaf node*. Two nodes with the smallest probabilities are combined into a new node whose probability is the sum of the two previous node probabilities. Each of the two branches going into the new node will be assigned either a one or a zero. This process will be repeated until there is only one node (*root node*) left. The code for a particular symbol can then be determined by reading all the branch values sequentially starting from the root node and ending at the symbol leaf nodes. For a fixed set of probabilities, this procedure is straightforward. In a dynamic setting, however, changing the codes can be very time-consuming since small changes to the probabilities can cause global changes to the optimal tree.

There are two classes of Huffman coders, fixed- and adaptive Huffman coders. Fixed Huffman coding involves using a static symbol table based either on the entire data sequence or on global information. Adaptive Huffman coding involves forming a new code table for each data sequence and encoding the table in addition to the data sequence. Alternatively, the adaptive Huffman coder may switch at intervals between previously-selected code tables, indicating at each interval the selected table. Adaptive Huffman coders generally exhibit better performance in terms of the compression ratio achieved, yet suffer from increased overhead. In real-time applications, fixed Huffman coders work more quickly and have simpler hardware implementations.

Although Huffman codes are optimal among prefix codes, the expected code length of a Huffman code is the same as the entropy only if all the probabilities are powers of $1/2$.

Arithmetic coding

The fundamental idea behind arithmetic coding is to map a string of symbols into a subinterval of the interval $[0, 1)$, where $[x, y)$ denotes a half open interval, which includes x but excludes y. There are two basic concepts in arithmetic coding, the probability of a symbol and encoding interval range for a symbol [3] [4]. The occurrence probabilities of source symbols determine the compression efficiency as well as the interval ranges of source symbols for the encoding process. These interval ranges are contained within the interval from zero to one and determine the compression output.

For example, assume that the source symbols are a, b and the probabilities of these symbols are 0.4 and 0.6, respectively. Based on these probabilities, the interval $[0, 1)$ can be divided into the two subintervals, $[0, 0.4)$ and $[0.4, 1)$. The range $[0, 0.4)$ is used

to further subdivide an interval when the symbol a is encountered in the input data and the range $[0.4, 1)$ is used when the symbol b is encountered. To encode the string $baab$ as a subinterval of $[0, 1)$, we begin with $[0, 1)$ as the working subinterval and use the encoding range of the first symbol b from the input string to narrow the working subinterval to the last $6/10$ of the current working subinterval. Therefore, the interval $[0, 1)$ is narrowed to $[0.4, 1)$. Since the range of the second symbol a is $[0, 0.4)$, the working subinterval $[0.4, 1)$ is reduced to its first $4/10$ portion to obtain the new working subinterval $[0.4, 0.64)$. The third input symbol is a, and its encoding reduces the current interval $[0.4, 0.64)$ to its first $4/10$ portion, that is to $[0.4, 0.496)$. Encoding the fourth symbol b produces the final interval $[0.4576, 0.496)$. The input string $baab$ is encoded as the interval $[0.4576, 0.496)$. From the symbol probabilities and the encoded interval, we can reconstruct the input data. For example, the interval $[0.4576, 0.496)$ is a subinterval of b's interval $[0.4, 1)$. So the first symbol of the source string is b and the working subinterval becomes $[0.4, 1)$. Next we see that $[0.4576, 0.496)$ is in the left $4/10$ of the interval $[0.4, 1)$. So the next symbol in the source data is a and the working subinterval becomes $[0.4, 0.64)$. Since $[0.4576, 0.496)$ lies wholly in a's portion of the working subinterval, the next source symbol is also a and the working subinterval becomes $0.4, 0.496)$. Then $[0.4576, 0.496)$ lies in the right $6/10$ of the working subinterval. So the next source symbol is b and the working subinterval becomes $[0.4576, 0.496)$. At this time decoding terminates because the working subinterval and the coded subinterval are the same. If each source data is assumed to end with a unique symbol (such as end-of-file), the coded data can be represented by any number in the final subinterval rather than by the endpoints of this interval. In practice, the arithmetic can be done by storing the current interval in sufficiently long integers rather than in floating point of exact rational numbers [4].

Main disadvantage of such basic approach to arithmetic coding is that the entire sequence must be coded before transmission (incremental transmission is not possible) and the representation for the symbol mapping table is cumbersome and can produce significant overhead. For these problems, an adaptive arithmetic coding scheme is developed [4][5]. For incremental transmission, the top-most bits are transmitted during the coding process, when they are equal at each end of the interval, i.e. the interval has become sufficiently narrowed. The second problem is alleviated by maintaining a running, reversible symbol mapping table which starts with equal initial probabilities for each possible symbol.

In theory, arithmetic coding gives an exactly optimal compression if we have a model that provides correct probabilities of the symbols comprising the input data. However, the cost of that is slow speed because of the exact arithmetic required, and the arithmetic codes lack the direct correspondence between the symbols in the input data set and bits or groups of bits in the coded output data.

Golomb-Rice coding

Golomb coding [6] produces optimal prefix codes for exponentially decaying probability distributions. It is assumed that the source probabilities (p) follow a geometric distribution of the form $Q_n = (1 - \rho)p^n$, where $0 < p < 1$. Golomb codes are distinguished from each other by a single parameter m, so dynamic updating is accomplished by estimating the value of the parameter. The symbols are arranged in descending probability order, and non-negative integers are assigned to the symbols, beginning with 0 for the most probable event. To encode integer n with given unique parameter m, Golomb code divides the integer n into two parts: a *binary* representation of $n \bmod m$ and a *unary* representation of $\lfloor n/m \rfloor$. The optimal value of m is given [7] by

$$m = \lceil \log(1 + p)/\log(p^{-1}) \rceil. \tag{2.2.1}$$

Rice codes, developed independently by Rice [8] [9] [10], are a special case of Golomb codes. Rice coding uses a subset of the parameter values, i.e. $m = 2^k$, and other coding procedure is the same as Golomb coding. However, choosing m to be a power of 2 leads to simpler coding procedure than Golomb coding: the code for integer n consists of the k least significant bits of n, followed by the number formed by the remaining higher order bits of n in unary representation. The length of this code is then $k + 1 + \lfloor n/2^k \rfloor$. Table 2.2 gives examples of Rice codes for $k = 3$. The resulting codes give less compression efficiency than Golomb codes, but they are even easier to implement, especially in hardware [11], since we can compute $n/2^k$ by shifting and $n \bmod 2^k$ by masking out all but the k low order bits.

Number	Sign Bit	Lower Bits	Number of 0's	Full Code
0	0	000	0	00001
18	0	010	2	0010001
-12	1	100	1	110001

Table 2.2: Example of Rice codes for $k = 3$

Golomb-Rice coding give the fast, flexible modeling obtained with arithmetic coding without the time-consuming arithmetic. Particularly, Rice coding gives faster coding even than Golomb coding or Huffman coding because of the especially simple prefix codes involved, and adaptive modeling is possible without the complicated data structure manipulations required in dynamic Huffman coding. Hence, the coding is extremely straightforward to implement in software and hardware. The main drawback to Rice coding is the limited compression performance because it is assumed that the model has

a probability distribution of a particular, as we discussed above. However, this is not a serious problem[1] because the probabilities of the more probable events will be estimated fairly well.

2.3   Lossy and Lossless Audio Compression

Two families of algorithms exist in compression. When the information can be exactly recovered from the bits, the source coding or compression is called lossless; otherwise, it is called lossy. Lossy algorithms remove information from the original in a way that is not perceptible. However, after transmission with lossy compression, only a modified version of the original signal is available. Lossless algorithms respect the integrity of the original signal. After transmission and reconstruction, an exact copy of the original signal is available.

2.3.1   Lossy compression

To achieve higher compression ratios, we give up on trying to represent the original signal exactly. Instead, we consider lossy compression, i.e. we allow approximate representations of the signal which come *close* to the original signal. Most lossy compression schemes are then controlled by a parameter that trades fidelity against limited data rate. Therefore, lossy compression is most appropriate for the data that is originally analogue rather than digital.

The many forms of lossy audio compression techniques offer a range of encoder and decoder complexity, compressed audio quality, and compression ratio. The well-known μ-law transformation and adaptive differential pulse code modulation (ADPCM) are simple approaches with low-complexity, low-compression, and medium audio quality [12]. The MPEG/Audio standard is a high complexity, high compression, and high audio quality algorithm.

In lossy audio compression, limitation of human ear perception can be exploited to omit components of the signal that will not be noticed. It is also referred to as a *perceptual coding*. Using a psychoacoustic model based on the human receiver, the perceptual coding schemes attempt to evaluate the *irrelevant* components of the audio signal that fall outside the hearing threshold, or that will be masked by adjacent events.

---

[1]unless one event's probability is close to 1

In the result, an essence of the sound will be transmitted rather than the original audio waveform.

Auditory perception

As a sound enters the ear, it is processed by overlapping, nonuniform filters along the basilar membrane. These filters are tuned to specific frequencies and are called the critical bands. The audible spectrum can be partitioned into the critical bands that reflect the resolving power of the ear as a function of frequency. Table 2.3 gives a listing of critical bandwidths.

Band	Frequency(Hz)[1]	Band	Frequency(Hz)[1]
0	50	14	1,970
1	95	15	2,340
2	140	16	2,720
3	235	17	3,280
4	330	18	3,840
5	420	19	4,690
6	560	20	5,440
7	660	21	6,375
8	800	22	7,690
9	940	23	9,375
10	1,125	24	11,625
11	1,265	25	15, 375
12	1,500	26	20,250
13	1,735		

[1]Frequencies are at the upper end of the band

Table 2.3: Approximate critical band boundaries

Masking is a frequency domain phenomenon of the ear that occurs whenever the presence of a strong audio signal makes a spectral neighborhood of weaker audio signals imperceptible. This masking phenomenon has been observed and corroborated through a variety of psychoacoustic experiments [13]. In particular, masking is strongest in the immediate vicinity of the masker, masking increases with increased power level, low frequencies are stronger maskers than high frequencies, and noise is a better masker than a tone. Using psychoacoustic equations, masking thresholds can be calculated for each critical band. Masking thresholds can then be used to eliminate redundant data and to estimate quantization levels. Because of the ear's limited frequency resolving

power, the masking threshold at any given frequency is solely dependent on the signal activity within a critical band of that frequency. As an example, Figure 2.3.1 illustrates the masking threshold for 1 kHz narrowband masker with 60 dB of SPL.
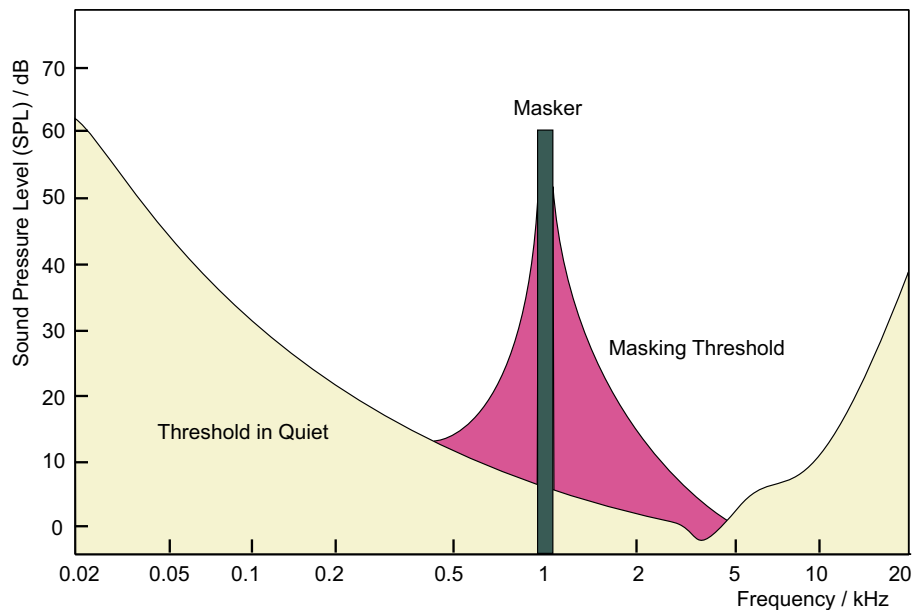
Figure 2.3.1: Threshold in quiet and example of masking threshold

Figure 2.3.1 shows that a signal is inaudible even without a masker if its SPL is below the threshold in quiet which depends on frequency and covers a dynamic range of more than 60 dB. The distance between the level of the masker and the masking threshold is called *signal-to-mask ratio* (SMR). Within a critical band, coding noise will not be audible as long as its signal-to-noise ratio (SNRm) resulted from m-bit quantization is higher than its SMR.

For audio compression, this property can be capitalized by transforming the audio signal into the frequency domain, then dividing the resulting spectrum into subbands that approximate critical bands, and finally quantizing each subband according to the audibility of quantization noise within that band. For optimal compression, each band should be quantized with no more levels than necessary to make the quantization noise inaudible.

MPEG/Audio coding

Most well-known lossy audio compression standard is the MPEG (motion picture experts group) compression algorithm of ISO (international organization for standardization) which provides high fidelity audio compression. This algorithm can achieve perceptually *transparent* compression[2]. The MPEG-1/Audio coding offers a subjective reproduction quality that is equivalent to CD quality at stereo rates at and above 128-256 kb/s for many types of music. The high performance of this algorithm is possible by the psychoacoustic modeling that determines where the quantization noise will be masked, as described above. Figure 2.3.2 shows the structure of the MPEG/audio encoder and decoder [14].

(a) MPEG/Audio Encoder

(b) MPEG/Audio Decoder
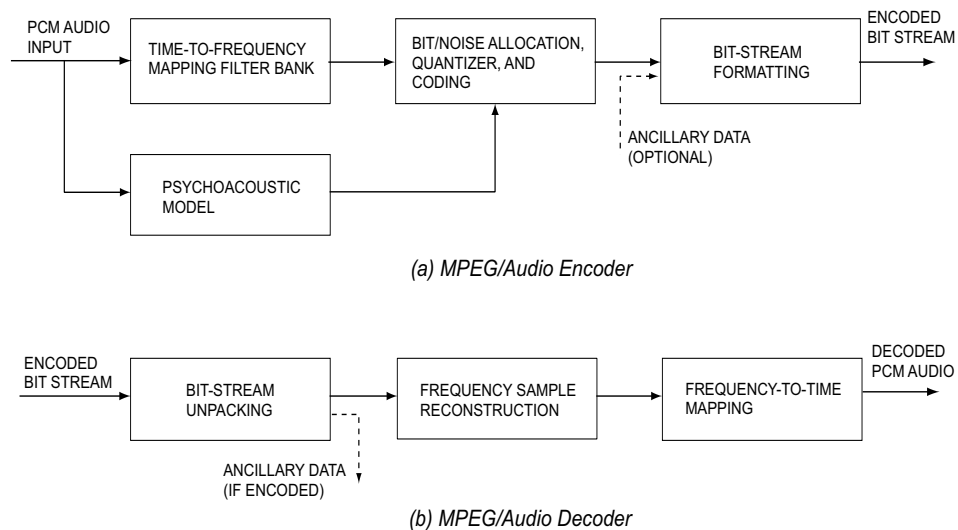
Figure 2.3.2: General structure of MPEG/Audio compression

The input audio stream passes through an analysis filter bank that divides the input into spectral multiple subbands. The input audio stream simultaneously passes through a psychoacoustic model that determines the SMR of each subband. The number of bits for each subband and a scale factor, a multiplier that sizes the samples to maximize the resolution of the quantizer, are determined on block-by-block basis. A dynamic bit allocation algorithm is used to decide the number of quantizer bits so that the quantization noise is kept below the masking threshold. Finally, the subband codewords, the scale factor, and

---

[2]The subject listening tests of MPEG/audio committee showed that even with a 6-to-1 compression ratio the listeners were unable to distinguish between coded and original audio clips under optimal listening condition. However, we are not prepared to accept the term "transparent", because even the human psychoacoustics used for such perceptual coding scheme is till in investigating.

the bit allocation information are multiplexed into one decodable bit stream including a header and optional ancillary data that is not necessarily related to the audio stream. The decoder simply reverses the formatting, then reconstructs the quantized subband values by using a synthesis filter bank, and finally transforms the set of subband values into a time-domain audio signal.

The MPEG/audio standard has three distinct layers for compression. Layer I forms the most basic algorithm, and Layers II and III are enhancements that use some elements found in Layer I. Each successive layer improves the compression performance but at the cost of greater encoder and decoder complexity. Layer I makes use of a subband filter bank that divides the audio signal into 32 constant-width frequency bands. The filter bank provides 32 frequency samples, one sample per band, for every 32 input audio samples. The Layer I algorithm groups together 12 samples from each of the 32 bands. Each group of 12 samples receives a bit allocation and, if the bit allocation is not zero, a scale factor. The Layer I encoder formats the 32 groups of 12 samples (i.e., 384 samples) into a frame.

The Layer II algorithm improves compression performance by coding data in larger groups. It forms frames of 3 by 12 by $32 = 1,152$ samples per audio channel, i.e., it codes the data in 3 groups of 12 samples for each subband, whereas Layer I codes in single groups of 12 samples for each subband. The Layer II algorithm also improves performance over Layer I by representing the bit allocation, the scale factor values, and the quantized samples with a more efficient code (see Fig. 2.3.3).
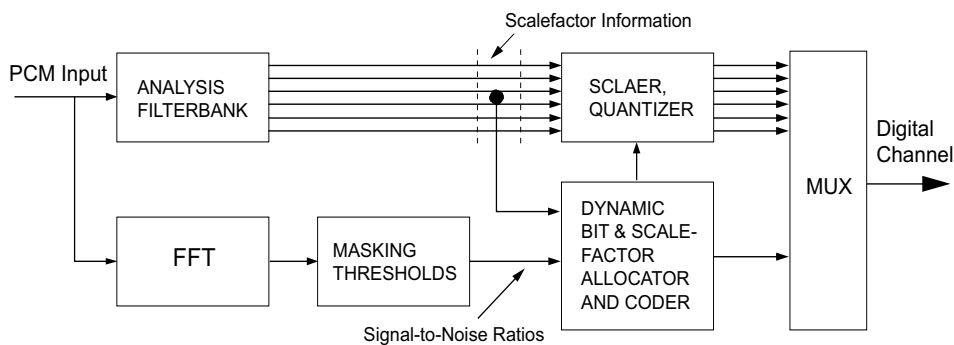
Figure 2.3.3: MPEG-1/Audio Layer II encoder

The Layer III algorithm compensates for some deficiencies of the filter bank found in Layer I and II by using a hybrid filter bank. This Layer III hybrid filter bank approach has become quite popular, in particular in Internet applications (MP3). Figure 2.3.4 shows a block diagram of the filter bank.
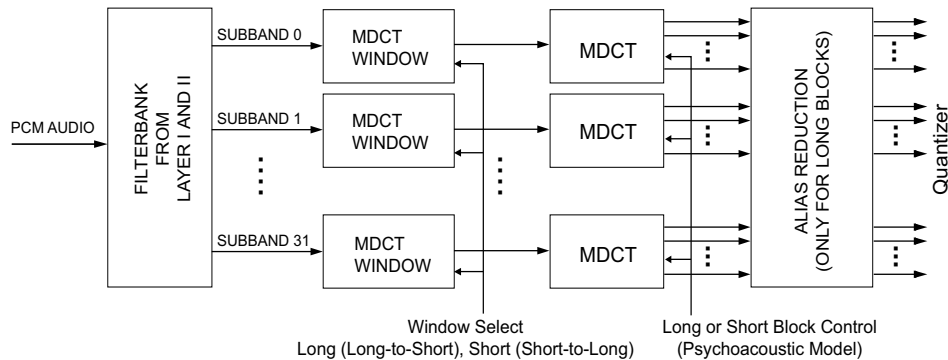
Figure 2.3.4: MPEG-1/Audio Layer III filter bank processing in encoder side

Applying modified discrete cosine transform (MDCT) to each of the 32 subband signals, the Layer III filter bank subdivides the filter bank outputs further in frequency content. This process provides a better spectral resolution closer to critical band partitions. Layer III specifies two different MDCT lengths, i.e. 6-point and 18-point block transform. Because of 50% overlapping transform, these lengths correspond to 12 and 36 subband samples in a window. Hence, when the short block length is used, then three short blocks replace a single long block. The short block length improves the time resolution to cope with transients. Note that the switch between long and short blocks is not instantaneous. A long block with a specialized long-to-short or short-to-long data window provides the transition mechanism from a long to a short block. In addition, Layer III employs an analysis-by-synthesis approach, an advanced pre-echo control, and nonuniform quantization with entropy coding. A buffer technique, called *bit reservoir*, leads to further savings in bit rate.

Problem of perceptual coding schemes

The structure of the MPEG/audio has several compromised concessions. Notably, although the idea of such perceptual coding is to exploit the phenomena that happen in our hearing system, the constant bandwidths of the filter bank do not accurately reflect the ear's critical bands. The bandwidth is too wide for the lower frequencies so the number of quantizer bits cannot be specifically tuned for the noise sensitivity within each critical band. Instead, the included critical band with the greatest noise sensitivity dictates the number of quantization bits required for the entire filter band. Second, because of a significant frequency overlap between the adjacent subbands, a signal at a single frequency can affect two adjacent filter bank outputs. Finally, the process of the filter bank and its

inverse are not lossless transformations, i.e. it is not perfect reconstruction filter (PRF). It implies that the inverse transformation would not perfectly recover the original input signal, even without quantization. In Chapter 4, we investigate the subband transformation and multirate filter bank that enable us a perfect reconstruction.

There is a growing concern around multiple coding issues when passing audio signal through several data reduction codecs. Particularly, the problems mentioned above, overlapping transformation and the non-reversible filter bank, can reveal the originally masked signal portions or can distort the signal uselessly by the multiple coding. This problem applies especially to the post-production industry.

2.3.2   Lossless compression

Unlike lossy compression, where the encoder throws away the components that are not psychoacoustically important, the encoder for lossless compression does not throw away any of the audio data. It merely packs[3] them more efficiently into the lower data rate, and the decoder can recover an exact bit-for-bit copy of the original. Therefore, it is very important to note that lossless compression can be regarded as a reformatting of the standard PCM audio representation. Binary-weighted PCM is just one of a number of possible formats for representing digital audio, and is not necessarily the most efficient in any particular case. Lossless coding minimizes the redundancy in the standard PCM representation and converts to a new format which needs only to be decoded. With this fact, the lossless compression provides the key not only to satisfying the demands for an increased data rate, but also to making it practical to design a system that will handle a wide variety of requirements very simply. In addition to the efficient transmission and storaging, the compressed format provides several supplementary advantages, for instance, it can lead to reduced levels of correlated jitter, which is a critical factor in high resolution digital audio systems.

Lossless compression has some obstacles to extend its practical application area. The main drawback of lossless audio compression is lower compression ratio, usually compression factors of 2 to 3, compared with the lossy compression that achieves compression factors ranging from 2.7 to 24 or higher, depending on the audio sampling rate. It should be also obvious that the data rate after lossless compression is not fixed, but depends on how much redundancy there is in the original audio signal. Some signals can be compressed more tightly into a smaller data rate than others, so lossless coding will by default produce a variable data rate which needs system that can handle a variable rate bit stream.

---

[3]Practically, the terminology "packing" is often used in literature in order to emphasize the difference between the lossy and the lossless compression,

More recently, a lossless compression scheme has been developed, that is optimized to produce a constant data rate in the compressed domain [15] [16]. Using this scheme, 16-bit 96kHz-sampled audio signals can almost always be losslessly compressed to 8 bits, and 16-bit 48kHz-sampled signals to 12 bits, with exact reconstruction of the original. In the result, the constant data rate of lossless compression costs lower compression ratio. On the other side, compressed 96kHz audio does not double the data rate of compressed 48kHz audio as expected. It proves that the lossless compression of PCM audio becomes more efficient as the sampling rate is increased. This is one of the important reasons why lossless compression will become a standard audio format and will be more importantly focused in the future, as the resolution of digital audio goes higher.

Apart from the importance of lossless compression, it is asserted in some literatures that current lossless audio coders have reached a limit in what can be achieved for lossless compression of audio. Indeed, by the nature of lossless coding, it could not be possible to achieve so high compression ratio as lossy compression. But several new decorrelation methods, for example, block-based prediction, parametric modeling or context-based modeling, show a capability to hurdle the present limit (factor of 2 to 3) of the lossless compression ratios.

Lossy-based lossless compression

In general, the spectrum resulting from a signal transform will be real-valued, even for an integer input signal. For an efficient transmission, the coefficients have to be quantized, which causes inevitable errors in the output signal. Therefore, except when the transform used for signal decorrelation is an integer operation entirely, lossless transform coding has to be considered as a combination of conventional lossy transform coding and additional transmission of the residual in respect to the original signal, as illustrated in Figure 2.3.5.

As showed in Figure 2.3.5, the lossy compressed signal is locally reconstructed in encoder side to obtain the residual. The reconstructed signal in this case is an approximation of the original and can be used as an estimate of the original. The lossy compressed signal can be transmitted in many different forms. It is obvious that the residual signal depends on the lossy coding algorithm. If the estimate is a good approximation of the original then the residual signal will contain small values. On the other hand, if lossy compression is not efficient, the residual will be large and correlated with redundancy. It means that the coding problem passes from the original signal to the residual signal. Thus, the most important task in such lossy-based lossless coding scheme is to find the best compromise between a high compression ratio of the lossy algorithm and a compressibility related to correlation state of the residual signal depending on lossy compression process.
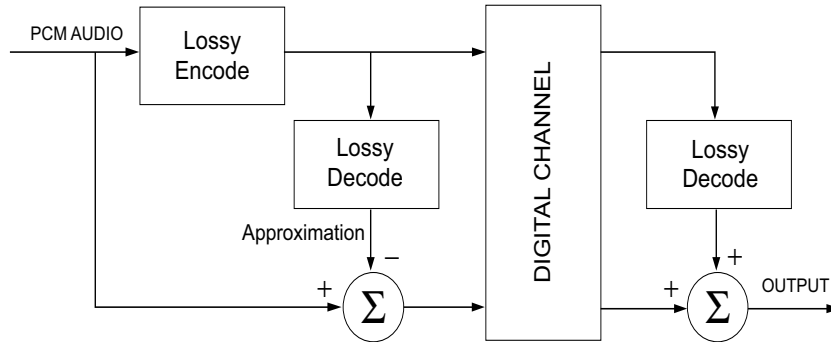
Figure 2.3.5: Lossy-based lossless compression using lossless transmission of error signal

A potential advantage of such lossy-based lossless coding schemes is that the lossy coded part of the input signal, which maintains also the signal wave shape in lower bit rate with no time shift, could be used for progressive transmission with varying data rate. The lossless JPEG compression standard is a simple version of lossy-based lossless approach.

2.4   Performance Criteria for Lossless Compression System

There are several criteria in quantifying the performance of a compression scheme. Compression efficiency is measured by the compression ratio or by the bit rate. Compression ratio is the ratio of the length of the input data sequence to the compressed output data sequence for a given compression method. This is the most important measure of performance for a compression technique. Compression bit rate is the number of bits per sample required to represent the compressed data sequence. The relation between both measurements is readily obtained as following

$$\text{compression ratio} = \frac{\text{bits / sample (original)}}{\text{bits / sample (compressed)}} \, . \tag{2.4.1}$$

The source entropy of original data sequence is used to compare the effectiveness of

a compression method. The entropy of a set of M samples is defined as

$$H_p = -\sum_{i=0}^{M-1} p_i \log_2 p_i \ \text{(bits/sample)}, \tag{2.4.2}$$

where p_i denotes individual probabilities of each sample. As the source entropy is a minimum number of bits required to encode a data stream, the effectiveness of a lossless compression method can be measured by determining how closely the bit rate of compressed data approximate the source entropy.

The variance of the zero-mean output of the decorrelation stage can be used to compare the effectiveness of signal decorrelation models used in a compression scheme. Given an N-point zero-mean data sequence $x[n]$, the variance σ_x^2 is calculated by

$$\sigma_x^2 = \frac{1}{N-1} \sum_{n=0}^{N-1} x^2[n]. \tag{2.4.3}$$

Particularly, the variance is a useful quantity to determine compressibility of waveform data and to choose suitable decorrelation method to be used. The coding gain is often used to measure the performance of both the predictive modeling and transform modeling,

$$G_c = \frac{\sigma_x^2}{\sigma_e^2}, \tag{2.4.4}$$

where σ_x^2 is the variance of the predictor input $x(n)$, σ_e^2 is the variance of the prediction error or the transform coefficient.

Another popular metric, the spectral flatness measure (sfm) gauges the flatness or whiteness of a power spectral density (psd). The sfm may take values between zero and unity where a unity spectral flatness measure represents a white process. The inverse of the sfm measures the waveform predictability of a process where the value for an unpredictable (white) process is unity and the value for a totally predictable process is infinity. For linear prediction, the prediction gain is upper bounded by the waveform predictability.

The complexity of an audio compression algorithm is measured by the number of arithmetic operations required to perform both the encoding and decoding processes. This is an important factor for applications involving online compression and decompression where speed is crucial.

2.5   Test Audio Materials

For our experiments in the thesis, nine audio materials are chosen; six materials from SQAM-CD [17][4] and two from published music CDs. All materials are sampled at 44.1kHz, 16 bits step size, and stereo channel (except for speech).

Nr.	Length	Description
1	1:11	SQAM Track 8, Violin, Arpeggio and Melodious Phrase
2	0:46	SQAM Track 13, Flute, Arpeggio and Melodious Phrase
3	0:21	SQAM Track 53, Female Speech, German (Mono)
4	1:32	SQAM Track 60, Piano, Schubert
5	1:22	SQAM Track 67, Wind Ensemble, Mozart
6	0:33	SQAM Track 69, ABBA, *Pop*
7	0:21	SQAM Track 70, Eddie Rabbitt, *Country*
8	0:29	Def Leppard "Adrenalize", Track 1 "Let's get rocked", Bludgeon Riffola Ltd, *Metal Rock*
9	0:29	Stan Getz "The Artistry of Stan Getz", Track 10 "Litha", Polygram Records, *Soft Jazz*

Table 2.4: Description of test audio materials

---

[4]SQAM(Sound Quality Assessment Material), European Broadcasting Union

Chapter 3

Predictive Modeling

Autoregressive (AR) modeling by linear prediction (LP) forms the basis of a wide variety of signal processing and communication systems including parametric spectral estimation and system identification. Perhaps the greatest success of linear prediction techniques is to be found in speech analysis and audio coding. It is also true that most of current lossless audio compression schemes employ a predictive coding approach or its hybrid algorithm combined with arithmetic coding.

The goal of the predictive modeling in audio compression is to reduce the sample magnitudes by making a prediction of the current sample on the basis of previous sample values and by transmitting just the difference between the current sample and the prediction. Several predictive methods exist for exploiting correlation between neighboring samples in a given data stream. The same decorrelation function is used in compression and reconstruction, and this function must take as input a delayed version of the input sequence.

In this chapter, we first review the general frameworks of predictive modeling and investigate various prediction filter structures including the modified linear predictor. We then compare the prediction filters by applying to the lossless audio compression system.

3.1 Theoretical Background

A discrete signal is a sequence of samples

$$x[n], \qquad n = \cdots, -2, -1, 0, 1, 2, \cdots \tag{3.1.1}$$

Signal $x[n]$ can be always be expressed as a linear combination of a set of some other sequences

$$x[n] = \sum_{k=1}^{N} c_k s_k[n] . \tag{3.1.2}$$

For example in

$$x[n] = \sum_{k=1}^{N} c_k e^{j2\pi kn/N} , \tag{3.1.3}$$

If $n = 1, 2, \cdots, N$, this is called the inverse discrete Fourier transform. The basis functions of this decomposition are defined by the following formula:

$$s_k[n] = e^{j\pi kn/N} . \tag{3.1.4}$$

There are a set of orthogonal functions and they form a *complete* basis. This means that a set of N basis functions can represent any signal of duration N exactly. In theory, signals which can be represented exactly by, e.g., a set of elementary functions, are called *singular* [18] or *deterministic* [19] signals.

The autocorrelation function of a discrete ergodic signal $s[n]$ is defined by

$$\forall k : \quad R_k = E(s[n]s[n-k]) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} s[n]s[n-k] . \tag{3.1.5}$$

White noise is a discrete stationary random signal $r[n]$ defined as a sequence with

$$\forall (n \neq k) : \quad R_k = E(r[n]r[n-k]) = 0 . \tag{3.1.6}$$

In classical literature [20], $r[n]$ is sometimes called a *fundamental sequence*. In practical applications, signals are of finite length, and therefore a signal may be called random only in respect to some signal model.

Signal $x[n]$ can always be written as a sum of a deterministic signal $s_d[n]$ and another signal $s_r[n] = x[n] - s_d[n]$. If $x[n]$ is a stationary signal and $s_d[n]$ and $s_r[n]$ are uncorrelated, it can be shown [18] that

$$x[n] = s_d[n] + s_r[n] = s_d[n] + \sum_{k=0}^{\infty} c_k r[n-k] , \tag{3.1.7}$$

where $r[n]$ is an uncorrelated white noise signal and $\sum_{k=0}^{\infty} |c_k|^2 < \infty$. This is called the *Wold decomposition theorem* for a stationary signal[1]. In classical terms [20], $s_r[n]$,

---

[1]Wold decomposition theorem was introduced in the first edition of Wold's book, his doctoral thesis, in 1938.

which is obtained from a fundamental sequence by sliding summation, is called a *regular* sequence. The Wold decomposition is of fundamental importance because it clearly divides the universe of linear spectral estimation methods into two main branches: deterministic and stochastic techniques. Deterministic techniques can be associated with non-parametric coding techniques such as transform coding. Similarly, parametric techniques are usually related to a stochastic signal modeling principle.

In the equation (3.1.7), if the coefficients c_k are fixed and $s_d[n] = 0$, $x[n]$ is a moving average model for the stochastic process $s_r[n]$ given by

$$s_r[n] = \sum_{k=0}^{\infty} c_k r[n-k] \,. \tag{3.1.8}$$

The z-transform of (3.1.8) is then given

$$S_r(z) = C(z)R(z) = \left[\sum_{k=0}^{\infty} c_k z^{-k}\right] R(z) \,, \tag{3.1.9}$$

and then

$$R(z) = \frac{S_r(z)}{C(z)} \tag{3.1.10}$$

which shows that the white noise excitation $r[n]$ is uniquely determined by the filter, its output $s_r[n]$, and the initial conditions at the filter's states.

Eq. (3.1.8) can be written as following

$$c_0 r[n] = s_r[n] - \sum_{k=1}^{\infty} c_k r[n-k] = s_r[n] - \tilde{s}_r[n] \,. \tag{3.1.11}$$

By denoting $r[n] = c_0 r[n]$ in (3.1.11) with assuming that $c_0 = 1$, this expression has two important aspects. Firstly, $r[n]$ obeys (3.1.6). Therefore, it also holds that $r[n]$ is uncorrelated with any linear combination of its past values $r[n-k]$, $k \geq 1$. That is,

$$E\left(r[n] \sum_{k=1}^{\infty} c_k r[n-k]\right) = E\left(r[n]\tilde{s}_r[n]\right) = 0 \,. \tag{3.1.12}$$

This is called the *orthogonality principle*. Secondly, as it was pointed out by Kolmogorov [20], $\tilde{s}_r[n]$, which is uniquely determined by the history of $s_r[n]$, can be seen as a *linear prediction* for $s_r[n]$. The prediction error is, by definition, a white noise signal $r[n]$. Therefore, (3.1.12) is an optimal solution to the *prediction problem* given by (3.1.11).

In parallel with Kolmogorov's work, Wiener [21] studied the prediction problem for *continuous signals* from a slightly different perspective[2]. Levinson [22] extended

---

[2]Wiener (1949) recognizes Kolmogorov's work with the same problem in his book and points out that: ... *the parallelism between them may be attributed to the simple fact that the theory of the stochastic processes had advanced to the point where the study of the prediction problem was the next thing on the agenda.*.

Wiener's theory for discrete-time signals. They started with minimization of the expectation of (3.1.11) by

$$\frac{\partial E[|r[n]|^2]}{\partial c_k} = 0, \qquad k = 1, 2, \cdots, \infty, \tag{3.1.13}$$

which leads to the same orthogonality condition given by (3.1.12) for an optimal set of coefficients c_k. It can be shown that this always gives the minimum of the expression. Basically, this is the classical least squares regression technique which was already used by Gauss and first published by Legendre in early 19th century, see, e.g., [23] [24] [25] for a historical survey. For time series, this technique was first applied by Yule [26] and Walker [27].

3.2   Linear Prediction System

3.2.1   General expressions

The basic idea behind linear prediction is that a sample of signal can be approximated as a linear combination of previous samples. By minimizing the sum of the squared differences between input samples and linearly predicted ones, a unique set of predictor coefficients can be determined.

From the discrete signal processing viewpoint, a digital filter $H(z)$ is assumed to have p poles and q zeros in the general *pole-zero* case, which means that given the input sequence $x[n]$, its approximation $\hat{x}[n]$ can be modeled by a combination of the q previous output samples and $p + 1$ previous input samples in a discrete filter system:

$$\hat{x}[n] = \sum_{k=0}^{p} b_k \, x[n-k] + \sum_{k=1}^{q} a_k \, \hat{x}[n-k], \tag{3.2.1}$$

which is equivalent to

$$H(z) = \frac{\hat{X}(z)}{X(z)} = \frac{\displaystyle\sum_{k=0}^{p} b_k \, z^{-k}}{1 - \displaystyle\sum_{k=1}^{q} a_k \, z^{-k}}. \tag{3.2.2}$$

This is the general description of pth-order *recursive* or *infinite impulse response* (IIR) system. Such model with both poles and zeros is also called an *autoregressive moving average* (ARMA) model. If we set the coefficients $a_1, a_2, \cdots, a_q$ equal to zero, (3.2.1)

becomes a *finite impulse response* (FIR) system, i.e.,

$$\hat{x}[n] = b_0 x[n] + b_1 x[n-1] + \cdots + b_p x[n-p] = \sum_{k=0}^{p} b_k\, x[n-k]\,. \qquad (3.2.3)$$

This is an allpole model, also known as an *autoregressive* (AR) model, while the allzero model ($p = 0$) is called a *moving average* (MA) model since the output is a weighted average of the q prior inputs. Figure 3.2.1 shows the general representation of these filter systems.
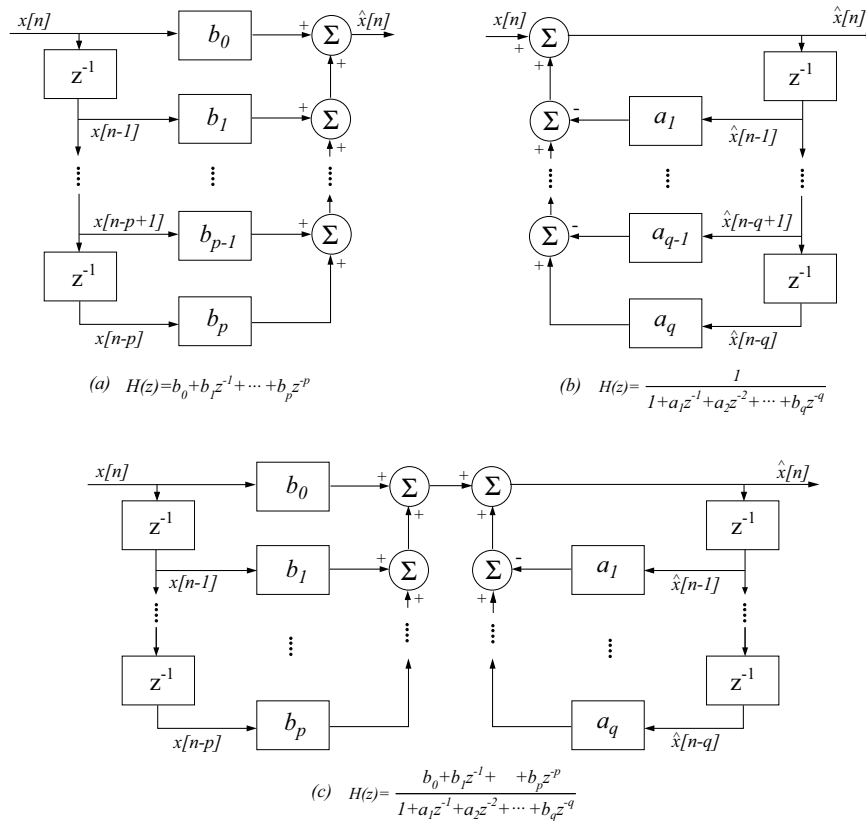
Figure 3.2.1: Block diagram representation for three types of filters

As showed in Figure 3.2.1, the prediction of present value in an FIR system is performed by combination of past values of the input signal, while for an IIR system the estimation of present value depends on the immediate past values of the output and the present and past values of the input. That is, an IIR system involves a recursive process in

order to update the output by the present and past values. As the result, the terms of FIR and IIR describe digital filters relative to the length of their sampled response sequences since it is possible to implement an FIR filter in a recursive fashion and an IIR filter in a nonrecursive manner.

Most of linear prediction system in literature assumes to use an allpole FIR filter because of its ability to provide extremely accurate estimate and its relative speed of computation. From (3.2.3) the system function of a pth-order FIR LP filter is the polynomial

$$P(z) = \sum_{k=1}^{p} b_k \, z^{-k} \,, \tag{3.2.4}$$

where any zeros at $z = 0$ are ignored because such zeros contribute nothing to the spectral magnitude and add only linear phase. Output of this predictive filtering is the difference between the original input signal $x[n]$ and its prediction $\hat{x}[n]$,

$$e[n] = x[n] - \hat{x}[n] = x[n] - \sum_{k=1}^{p} b_k \, x[n - k] \,, \tag{3.2.5}$$

where the difference signal $e[n]$ is called *residual* or *prediction error*[3]. From (3.2.5) it can be seen that the prediction error sequence is the output of a system whose transfer function is

$$A(z) = 1 - \sum_{k=1}^{p} b_k \, z^{-k} \,, \tag{3.2.6}$$

which is an inverse filter for the system $H(z)$, i.e., $H(z) = A(z)^{-1}$. The basic problem of LP is to determine a set of predictor coefficients $\{b_k\}$ directly from the signal in such a manner as to ontain a good estimate of the spectral properties of the signal through the use of (3.2.6).

The optimal predictor coefficients b_k are chosen to minimize the energy (average) in residual signal which is defined as

$$E = \sum_{n=-\infty}^{\infty} e^2[n] \tag{3.2.7a}$$

$$= \sum_{n=-\infty}^{\infty} (x[n] - \hat{x}[n])^2 \tag{3.2.7b}$$

$$= \sum_{n=-\infty}^{\infty} \left[x[n] - \sum_{k=1}^{p} b_k x[n - k] \right]^2 \,. \tag{3.2.7c}$$

---

[3]In the information theory $e(n)$ is often called the innovation sequence.

By setting $\partial E/\partial b_k = 0$, $k = 1, 2, 3, \ldots, p$, as a necessary condition for minimum energy, we obtain p linear equations

$$\sum_{n=-\infty}^{\infty} x[n-i]x[n] = \sum_{k=1}^{p} b_k \sum_{n=-\infty}^{\infty} x[n-i]x[n-k], \quad i = 1, 2, 3, \ldots, p\,. \qquad (3.2.8)$$

If we define

$$\phi(i, k) = \sum_{n=-\infty}^{\infty} x[n-i]x[n-k]\,, \qquad (3.2.9)$$

then the linear equations (3.2.8) can be written more compactly as

$$\sum_{k=1}^{p} b_k\,\phi(i, k) = \phi(i, 0), \quad i = 1, 2, 3, \ldots, p\,. \qquad (3.2.10)$$

From (3.2.7), (3.2.8) and (3.2.10) the minimum mean-squared prediction error can be shown to be

$$E = \sum_{-\infty}^{\infty} x^2[n] - \sum_{k=1}^{p} b_k \sum_{-\infty}^{\infty} x[n]x[n-k] \qquad (3.2.11\text{a})$$

$$= \phi(0, 0) - \sum_{k=1}^{p} b_k \phi(0, k) \qquad (3.2.11\text{b})$$

There are various algorithms to solve the linear equations (3.2.10), involving a special recursive algorithm for symmetric Toeplitz matrices. We review these algorithms in next section.

3.2.2   Estimation of the predictor coefficients

Autocorrelation method

To solve (3.2.10), the limits of summation in (3.2.7) and (3.2.8) must be over a finite interval. One of the basic approaches to determination of the limits is the well-known autocorrelation method. This approach assumes that the signal $x[n]$ has finite duration with length of N samples[4], i.e., $x[n] = 0$ outside the range $0 \leq n \leq N - 1$. In this case, the residual signal $e[n]$ obtained through pth-order LP filter will be nonzero over

---

[4]Normally, a Hamming or similar time window is used to segment the signal

the interval $0 \leq n \leq N - 1 + p$, and therefore the residual energy and (3.2.9) can be written as

$$E = \sum_{n=0}^{N+p-1} e^2[n] \,, \tag{3.2.12a}$$

$$\phi(i,k) = \sum_{n=0}^{N+p-1} x[n-i]x[n-k] \,, \tag{3.2.12b}$$

where $i = 1, 2, \ldots, p$ and $k = 0, 1, \ldots, p$. Since $\phi(i,k)$ of (3.2.12b) is identical to the autocorrelation function evaluated for $(i-k)$, the linear equations (3.2.10) can be expressed as a normal equations, called Wiener-Hopf equations[5] [28],

$$\sum_{k=1}^{p} a_k \, R(|i-k|) = R(i), \quad i = 1, 2, \ldots, p \,, \tag{3.2.13}$$

which is in matrix form,

$$\begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{bmatrix} = \begin{bmatrix} R(0) & R(1) & \ldots & R(p-1) \\ R(1) & R(0) & \ldots & R(p-2) \\ R(2) & R(1) & \ldots & R(p-3) \\ \vdots & \vdots & \vdots & \vdots \\ R(p-1) & R(p-2) & \ldots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} \tag{3.2.14}$$

Similarly to (3.2.11), the minimum mean squared prediction error becomes

$$E = R(0) - \sum_{k=1}^{p} a_k R(k) \,, \tag{3.2.15}$$

where $R(k)$ is the kth correlation of the source and $R(0)$ is equal to the source variance σ_x^2. In (3.2.14), the $p \times p$ matrix of autocorrelation values is a Toeplitz matrix; i.e., it is symmetric and all the elements along a given diagonal are equal. This special property will be exploited to obtain an efficient algorithm for the solution of (3.2.13).

The Wiener-Hopf equations can be efficiently solved by the well-known Levinson-Durbin recursive procedure [29], in which the following operations are performed recur-

---

[5]Note that (3.2.13) is identical to the Yule-Walker equations obtained for an AR(p) model. Thus the LP filter can be considered as a whitening filter. The residual sequence $e[n]$ is white and efficient codeable with lower entropy than the entropy of $x[n]$, only if the input sequence $x[n]$ is an AR(p) process and pth-order LP filter is optimal.

sively for $m = 1, 2, \ldots, p$:

$$k_m = \frac{\left[R(m) - \sum_{k=1}^{m-1} a_{m-1}[k]\, R(m - k) \right]}{E_{m-1}} \, , \tag{3.2.16a}$$

$$a_m[m] = k_m \, , \tag{3.2.16b}$$

$$a_m[k] = a_{m-1}[k] - k_m a_{m-1}[m - k], \qquad 1 \le k \le m - 1 \, , \tag{3.2.16c}$$

$$E_m = (1 - k_m^2) E_{m-1} \, , \tag{3.2.16d}$$

where initially $E_0 = R(0)$ and $a_0 = 0$. The *reflection coefficients* k_m guarantees a stable LP filter $H(z)$ so that at each cycle m the coefficients $a_m[k]$ describe the optimal mth-order linear predictor and the minimum error E_m is reduced by the factor $(1 - k_m^2)$. Therefore, the necessary condition for stable system is given by $|k_m| \le 1$. Particularly, the negatives of the k_m are called *partial correlation* (PARCOR) coefficients.

Covariance method

As mentioned in previous section, the autocorrelation method is usually preceded by a windowing of input signal, so that $x[n]$ values at the beginning and end of a block are tapered to zero. Thus, the autocorrelation solution introduces distortion into the spectral estimation procedure. The covariance method avoids this distortion since it does not need to window the input signal, but residual signal is weighted uniformly in time by a simple rectangular window of length N;

$$E = \sum_{n=0}^{N-1} e^2[n] \, . \tag{3.2.17}$$

Then the covariance function for $x[n]$ becomes

$$\phi(i, k) = \sum_{m=0}^{N-1} x[n - i] x[n - k] \tag{3.2.18a}$$

$$= \sum_{n=-k}^{N-k-1} x[n] x[n + k - i], \quad 1 \le i \le p, \ 0 \le k \le p \, . \tag{3.2.18b}$$

Setting $\partial E / \partial a_k = 0$ again to zero leads to p linear equations

$$\sum_{k=1}^{p} a_k \, \phi(i, k) = \phi(i, 0), \quad i = 1, 2, \ldots, p \, . \tag{3.2.19}$$

Since the evaluating $\phi(i, k)$ requires the input values of $x[n]$ in the interval $-p \leq n \leq N - 1$, it does not need to taper the segment of input signal to zero at the ends as in the autocorrelation method. In matrix form, (3.2.19) becomes

$$
\begin{bmatrix} \phi(1,0) \\ \phi(2,0) \\ \phi(3,0) \\ \vdots \\ \phi(p,0) \end{bmatrix} = \begin{bmatrix} \phi(1,1) & \phi(1,2) & \ldots & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \ldots & \phi(2,p) \\ \phi(3,1) & \phi(3,2) & \ldots & \phi(3,p) \\ \vdots & \vdots & \vdots & \vdots \\ \phi(p,1) & \phi(p,2) & \ldots & \phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} \tag{3.2.20}
$$

This covariance matrix is symmetric but not Toeplitz. As a result, the covariance method leads to a function that is not a true autocorrelation function, but rather, the cross-correlation between two very similar finite length segments of the input signal. From the difference of limitations in (3.2.12b) and (3.2.18a), the linear equations (3.2.19) have significantly different properties that strongly affect the method of solution and the properties of the resulting optimum predictor.

One of the well-known methods to solve the set of equations (3.2.19) is called the Cholesky decomposition (also referred as the square root method). In matrix notation, (3.2.19) can be written by

$$
\boldsymbol{\Phi}\mathbf{a} = \psi \,, \tag{3.2.21}
$$

where $\boldsymbol{\Phi}$ is a positive definite symmetric matrix with (i, j)th element $\phi(i, j)$, and $\mathbf{a}$ and ψ are column vectors with elements a_j, and $\phi(i, 0)$ respectively. For the Cholesky decomposition the matrix $\boldsymbol{\Phi}$ is given by

$$
\boldsymbol{\Phi} = \mathbf{VDV^T} \,, \tag{3.2.22}
$$

where $\boldsymbol{\Phi}$ is a lower triangular matrix whose main diagonal elements are all 1's, and $\mathbf{D}$ is a diagonal matrix. The elements of the matrices $\mathbf{V}$ and $\mathbf{d}$ are readily determined by

$$
V_{ij}d_j = \phi(i, j) - \sum_{k=1}^{j-1} V_{ik}d_k V_j k, \quad 1 \leq j \leq i - 1 \,, \tag{3.2.23a}
$$

$$
d_i = \phi(i, i) - \sum_{k=1}^{i-1} V_{ik}^2 d_k, \quad d_1 = \phi(1, 1), \quad i \geq 2 \,. \tag{3.2.23b}
$$

The column vector $\mathbf{a}$ can be solved by using a simple recursion of the forms

$$
Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij}Y_j, \quad p \geq i \geq 2 \,, \tag{3.2.24}
$$

where

$$
\mathbf{Y} = \mathbf{DV^T a}, \quad Y_1 = \psi_1 \,, \tag{3.2.25}
$$

and then

$$a_i = Y_i/d_i - \sum_{j=i+1}^{p} V_{ji}a_j, \quad 1 \le i \le p-1, \qquad (3.2.26)$$

with initial condition

$$a_p = Y_p/d_p. \qquad (3.2.27)$$

The index i in (3.2.26) proceeds backwards from $i = p-1$ down to $i = 1$.

3.2.3 Lattice structure of LP coefficients

It is possible to convert any digital filter to a corresponding lattice filter [30] [31]. The prediction in the LP systems discussed above is based on p previous samples of $x[n]$, i.e. *forward prediction*. In lattice LP model, the p ensuing samples are also used in a form of *backward prediction*. Figure 3.2.2 shows a general lattice structure which involves both the forward and backward prediction.
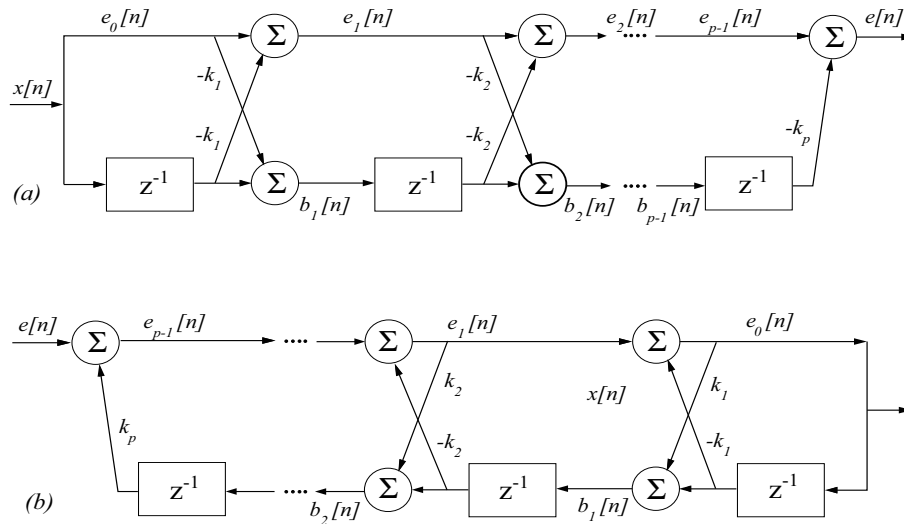
Figure 3.2.2: Lattice Filters: (a) inverse filter $A(z)$, which generates both forward and backward error signals at each stage of the lattice; (b) synthesis filter $1/A(z)$.

As showed in previous section, the lattice structure is a direct consequence of the Levinson-Durbin's recursion with the PACOR coefficients k_m. Applying (3.2.16c), we

see that the prediction error sequence $e_m[n]$ can be expressed as

$$e_m[n] = e[n] * a_{m-1}[n] - k_m x_m * a_{m-1}[m - n] , \qquad (3.2.28)$$

where * denotes convolution. The first term in (3.2.28) is the forward prediction error from an $(m - 1)$th predictor and the second term is a parallel backward error. Lattice methods of linear prediction are order-recursive. That is, the optimal coefficients are first solved for the first stage of the filter, then the prediction error signals are computed for the next stage and so on. Assigning $b_m[n]$ to this backward prediction error yields a recursion formula:

$$e_m[n] = e_{m-1}[n] - k_m b_{m-1}[n - 1] , \qquad (3.2.29)$$

where

$$b_m[n] = x[n] * a_m[m - n] = \sigma_{l=n-m}^n x[l] a_m[m - n + l] \qquad (3.2.30a)$$

$$= x[n - m] - \sum_{l=1}^m a_m[l] x[n - m + l] , \quad a_m[0] = 1 \qquad (3.2.30b)$$

$$= b_{m-1}[n - 1] - k_m e_{m-1}[n] . \qquad (3.2.30c)$$

(3.2.29) and (3.2.30c) define the forward and backward prediction error sequences for an mth order predictor in terms of the corresponding prediction errors of an $(m - 1)$th order predictor (see Figure 3.2.2, with initial conditions of $e_0[n] = b_0[n] = x[n]$). This block-based method is called *Burg Algorithm* [29]. In this method, the PACOR coefficients k_m can be directly related to the forward and backward prediction errors. The relationship in form of normalized cross-correlation function is

$$k_m = \frac{\displaystyle\sum_{n=0}^{N-1} e_{m-1}[n] b_{m-1}[n - 1]}{\left[\displaystyle\sum_{n=0}^{N-1} (e_{m-1}[n])^2 \sum_{n=0}^{N-1} (b_{m-1}[n - 1])^2 \right]^{1/2}} \qquad (3.2.31)$$

The reflection coefficients k_m have many interesting properties. In [32], these coefficients were derived directly from a non-uniform acoustic tube model, where the coefficients, as the name indicates, are reflection coefficients of individual tube elements. Therefore, the reflection coefficients and the lattice structure have firm physical interpretations. Their goal was to find a representation for LPC coefficients that is more robust to quantization. Reflection coefficients also act in a reasonable way in temporal interpolation of coefficients between frames. In addition, if all the reflection coefficients obey $|k_m| < 1, m = 1, 2, \ldots, p$, the synthesis filter is stable. Therefore, lattice methods of linear prediction also give direct means to check and guarantee the stability of the estimated model.

3.2.4   Determination of LP filter order

Finding an optimal LP filter order p is crucial to achieving optimal compression. It is not just a matter of decoder and encoder complexity because there is a tradeoff between the lower variance of the residual sequence and the increasing overhead due to larger predictor orders. On one hand, larger orders can capture the dynamics of a richer class of signals. On the other hand, larger orders also require proportionally larger data sets for the parameters to be accurately estimated and transmitted.

Obviously, determination of LP filter order is to find an order which minimizes the variance of zero-mean residual sequence, i.e.,

$$\sigma_e^2(p) = \frac{1}{N-p-1} \sum_{k=p}^{N-1} e^2[k] \,. \tag{3.2.32}$$

Simply, we might find an optimal LP filter order by incrementing p from $p = 1$ until the residual variance $\sigma_e^2(p)$ reaches a minimum. Another method, called the Akaike information criteria (AIC) [33], involves minimizing the following function

$$AIC(p) = N \ln \sigma_e^2(p) + 2p \,, \tag{3.2.33}$$

where $2p$ serves to penalize for unnecessarily high predictor orders. The AIC, however, has been shown to be statistically inconsistent, so the minimum description length (MDL) criterion has been formed [34] [35],

$$MDL(p) = N \ln \sigma_e^2(p) + p \ln N \,, \tag{3.2.34}$$

which implies to minimize the number of bits that would be required to describe the data. A method proposed by Tan [36] involves determining the optimal number of bits necessary to code each residual,

$$\mathcal{B}(p) = 2^{-1} \log_2 \sigma_e^2(p) \,. \tag{3.2.35}$$

In this case, p is increased until the following criterion is no longer true,

$$(N-p)\Delta B(p) > \Delta \beta(p) \,, \tag{3.2.36}$$

where $\Delta \mathcal{B}(p) = -[\mathcal{B}(p) - \mathcal{B}(p-1)]$ and $\Delta \beta(p)$ denotes the increase in overhead bits for each successive p. There are several other methods of order determination that are often used in practice include the Bayes information criterion (BIC) [37], which is equivalent to the MDL in many settings and the predictive least-squares (PLS) principle for sequential coding [38] [39].

3.2.5   Linear predictive coding (LPC)

Linear predictive coding(LPC) is an application of linear prediction modeling to signal encoding. LPC was first proposed for speech coding applications to make an accurate and economical representation of relevant speech parameters that capture information about the configuration of the vocal tract. The encoding process involves computation of filter coefficients a_k and the prediction error signal $e[n]$. The residual $e[n]$ and coefficients a_k are then quantized and transmitted to the decoder. In the decoder side, the original signal is reproduced using

$$X(z) = \frac{E(z)}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{E(z)}{A(z)} \, . \tag{3.2.37}$$

The quantization of the residual, or excitation for the synthesis filter, can be based on either scalar quantization or vector quantization. The quantization error $e_q[n] = e[n] - Q(e[n])$ is typically a nearly white noise process. Since $1/A(z)$ is a linear filter, its output for $E(z) + E_q(z)$ can be expressed as a sum of a clean signal $X(z)$ and an additive noise signal $X_q(z)$, that is,

$$X(z) + X_q(z) = \frac{E(z)}{A(z)} + \frac{E_p(z)}{A(z)} \, , \tag{3.2.38}$$

Equation (3.2.38) states that the coding error signal in prediction error coding has the z-transform, $X_q(z)$, characterized by the estimated allpole model $1/A(z)$. We have then a closed-loop encoder by reformulating (2.25) to

$$E(z) = X(z)A(z) - E_q(z)P(z) \, , \tag{3.2.39}$$

where $P(z) = 1 - \sum_{k=1}^{p} \gamma^k a_k z^{-k}$ with $0 < \gamma < 1$. If $\gamma = 1$, $X_p(z) = E_q(z)$, that is, coding error is approximately white noise. Other choices for $P(z)$ can be used to shape the spectrum of the coding error signal [40].

3.2.6   Polynomial LPC

With smooth signals, piecewise polynomial signal model can be employed. Filters with polynomial responses arise e.g. in unbiased extrapolation of polynomial signals with maximal noise attenuation. The goal of polynomial prediction filter design is to design such FIR coefficients $h(k)$, $k = 1, 2, \ldots, p$, where p is FIR length, that a piecewise polynomial input signal is exactly predicted. Thereafter, noise gain of the FIR is minimized.

$$NG = \sum_{k=1}^{p} h(k)^2 \tag{3.2.40}$$

With the last input sample taken at time $n - p$, prediction of an input signal $x[n]$ is generally given by

$$x[n + i] = \sum_{k=1}^{p} h[k]\, x[n - k]\,.$$ (3.2.41)

A polynomial with order p can be found that passes through the previous p data points $x[n-1], x[n-2], \ldots, x[n-p]$. This polynomial can be evaluated at the nth sample time in a restrictive form of the linear predictor to obtain the predicted value $\hat{x}[n]$. In this respect, a simple adaptive prediction method using only integer coefficients was first proposed for lossless audio compression in Shorten [41]. For example, we can obtain the

$$\hat{x}_0[n] = 0$$ (3.2.42a)
$$\hat{x}_1[n] = x[n - 1]$$ (3.2.42b)
$$\hat{x}_2[n] = 2x[n - 1] - x[n - 2]$$ (3.2.42c)
$$\hat{x}_3[n] = 3x[n - 1] - 3x[n - 2] + x[n - 3]$$ (3.2.42d)

These polynomials and the predicted values that they produce are illustrated in Fig. 3.2.3 for typical set of previous samples.
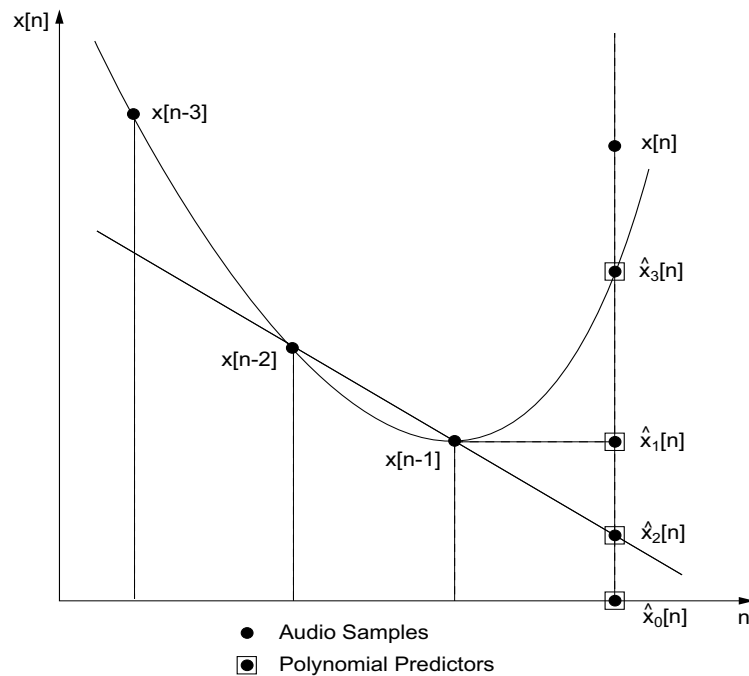
Figure 3.2.3: Four polynomial approximations of $x[n]$.(After [1])

An interesting property of these polynomial approximations is that the resulting residual signals, $e_p[n] = x[n] - \hat{x}_p[n]$, can be efficiently computed without any multiplications in the following recursive manner:

$$e_0[n] = x[n] \tag{3.2.43a}$$

$$\begin{aligned} e_1[n] &= e_0[n] - e_0[n-1] \\ &= x[n] - x[n-1] \end{aligned} \tag{3.2.43b}$$

$$\begin{aligned} e_2[n] &= e_1[n] - e_1[n-1] \\ &= (x[n] - x[n-1]) - (x[n-1] - x[n-2]) \\ &= x[n] - (2x[n-1] - x[n-2]) \end{aligned} \tag{3.2.43c}$$

$$\begin{aligned} e_3[n] &= e_2[n] - e_2[n-1] \\ &= (x[n] - 2x[n-1] + x[n-2]) \\ &\quad - (x[n-1] - 2x[n-2] + x[n-3]) \\ &= x[n] - (3x[n-1] - 3x[n-2] + x[n-3]) \end{aligned} \tag{3.2.43d}$$

The equations (3.2.43) mean also that the sum of absolute values is linearly related to the variance. This may be used as the basis of predictor selection. In our experiment (Section 3.6), the four prediction residuals are computed at every sample in the frame, and the absolute values of these residuals are averaged over the complete frame. The residual with the smallest sum of magnitudes over all samples in the frame is then defined as the best approximation for that frame, and that predictor is used for that frame. The information on which predictor was used can be coded with two bits of information, and that becomes part of the overhead information for the frame.

3.3   Nonstationary Signal Modeling

The techniques and concepts discussed in previous sections are all based on an assumption about stationarity of the input signal. In practical LPC algorithms, the filter coefficients are time-varying, i.e., parameters of a nonstationary signal model. Cremer [42] showed that Wold decomposition principle applies also to nonstationary signal models. It is possible to formulate an orthogonality condition of Eq. (3.1.12) for this signal model. However, there is no unique solution for the optimal time-varying coefficients $a_k[n]$. The coefficient evolutions must be restricted somehow in order to find one of the least-square optimal solutions to the coefficients. It is usually assumed that the signal is locally stationary [29], or the coefficients are smoothly time-varying [43]. These

are conceptually two different approaches. The local stationarity assumption is used in conventional frame-based and continuously adaptive techniques. Smooth coefficient evolution is assumed in smoothness priors techniques [44], and in techniques where the coefficient evolutions are restricted to a class of functions which can be expressed as linear combinations of predefined basis functions [45]. The latter approach is sometimes called deterministic regression approach of time-varying autoregressive modeling.

3.3.1 General approaches

Frame-based filter design

The basic technique to obtain a nonstationary signal model is to perform linear predictive analysis in frames such that the signal is assumed to be stationary within each analysis frame. In a long time scale, this means that the signal model for linear predictive coding is actually given by

$$x[n] = \sum_{k=1}^{p} a_k[n]x[n-k] + e[n] = \hat{x}[n] + e[n] \,, \qquad (3.3.1)$$

where filter coefficients $a_k[n]$ are now also functions of time n.

Audio and speech coding algorithms usually process the input signal in frames. For example, in LP-based speech coders, the frame-length is typically 10-20 ms. The frames are usually overlapping and, in the case of the autocorrelation method, some window function is applied to each signal frame before analysis. The filter coefficients corresponding to each frame are coded and transmitted along with excitation data. A direct application of this procedure would produce discontinuities to the coefficient trajectories in frame borders, which may produce unwanted artifacts. It is a common practice to interpolate filter coefficients smoothly from one frame to another. Therefore, the signal model which is considered in these algorithms is essentially given by (3.3.1) even if the spectral model is estimated in locally stationary frames.

It is possible to increase the amount of overlapping in the analysis so that the coefficients are estimated more frequently. An extreme example is a sliding window formulation of linear predictive modeling where coefficients are solved at each time instant. However, this is computationally expensive and leads to an increased number of filter coefficients to be transmitted. Barnwell [46] has introduced a computationally efficient method for computation of adaptive autocorrelation. In this method, the correlation terms in (3.2.14) are computed recursively using a leaky integrator. This is a version of the autocorrelation method of linear prediction where the window function is actually defined as an impulse response of a low-order IIR filter.

Deterministic regression time-varying LPC

It was proposed by Subba Rao [45] that the time-varying coefficient evolutions $a_k[n]$ could be expressed by

$$a_k[n] = \sum_{l=0}^{M} c_{kl}\phi_l[n] \,, \tag{3.3.2}$$

where $\phi_l[n]$ are a set of M predefined basis functions. For this system it is possible to formulate normal equations where the least squares optimal coefficients c_k can be solved directly. Typically, basis functions are some elementary mathematical functions such as the Fourier basis, Gaussian pulses, or prolate spheroidal sequences [47].

Adaptive filtering

While the method of frame-based linear prediction is effective, it suffers from the problem of finding a solution to the Yule-Walker equations, which becomes increasingly computationally expensive with large block sizes. Stochastic gradient methods for adaptive filtering also follow from a local formulation of the prediction problem. Here, the coefficients are not solved directly for a long signal frame but adjusted iteratively such that the filter coefficients converge towards optimal values. In this sense, these techniques are time-recursive. A classical example is least mean square (LSM) algorithm in which the coefficients of a direct form filter are adjusted using a simple gradient rule.

A backward adaptive formulation of linear predictive coding was introduced in [48] which is closely related to backward adaptive quantization methods presented in [49]. Here, the spectral model is not formed from the original input signal but from the already coded and transmitted signal. Since the same model can be computed at the decoder and the spectral model is completely estimated from the signal already transmitted, there is no need to code and transmit filter coefficients. Several different adaptive filtering techniques were compared in [50]. In next section we briefly describe two adaptive filtering methods.

3.3.2   Normalized least mean square algorithm

Adaptive FIR filters using normalized least mean square (NLMS) have been proposed and used successfully [51] [52]. From (3.3.1) the signal $x[n]$ and the time-varying filter coefficients $a[n]$ can be represented by the column vector, i.e., $\mathbf{x}[n] = (x[n-1], \cdots, x[n-N])^{\mathbf{t}}$ and $\mathbf{a}[n] = (a_0[n], \cdots, a_{N-1}[n])^{\mathbf{t}}$. Then a time-varying prediction error is given by

$$e[n] = x[n] - \mathbf{a}^T\mathbf{x}[n] \,, \tag{3.3.3}$$

If two fixed parameters, a smoothing parameter β and a convergence parameter u, are specified, then $\mathbf{a}[n]$ can be computed iteratively as following

$$\mathbf{a}[n+1] = \mathbf{b}[n] + \mu[n]e[n]\mathbf{x}[n] \,, \tag{3.3.4}$$

where

$$\mu[n] = \frac{u}{\sigma_N^2[n]} \tag{3.3.5}$$

$$\sigma_N^2[n] = \beta\sigma_N^2[n-1] + (1-\beta)(e^2[n-1]) \,. \tag{3.3.6}$$

Original signal can be exactly reconstructed by using the inverse of the algorithm

$$x[n] = e[n] + \mathbf{b^t x}[n] \,. \tag{3.3.7}$$

Eq. (3.3.7) shows that only $\mathbf{a}[0]$, $\mathbf{x}[0]$, and $e[n]$ are needed to reconstruct the original signal $x[n]$. This means that it is not necessary to transmit the coefficients $\mathbf{a}[n]$ and therefore to segment the signal in blocks. This algorithm requires less overhead than the standard LP method, but the coefficients should be updated very frequently.

3.3.3   Gradient adaptive lattice filter

In the gradient adaptive lattice (GAL) [53][54], the coefficients k_m are updated using the approximation of the error energy of the mth order predictor, i.e.,

$$k_m[n+1] = k_m[n] - \mu_m\frac{\partial\hat{E}_m(n)}{\partial k_m[n]} \,, \tag{3.3.8}$$

where μ_m are gradient weights and

$$\hat{E}_m(n) = e_m^2[n] + b_m^2[n] \,. \tag{3.3.9}$$

Applying the recursions after (3.2.29) (3.2.30c) for the current time index n leads to

$$\frac{\partial\hat{E}_m(n)}{\partial k_m[n]} = -2\big[e_m[n]b_{m-1}[n-1] + b_m[n]e_{m-1}[n]\big] \,. \tag{3.3.10}$$

From (3.3.8) and (3.3.10) we obtain a sample-by-sample update for the PACOR coefficients.

The gradient weights μ_m can be obtained by

$$2\mu_m = \frac{\alpha}{D_m(n)} \,, \tag{3.3.11}$$

where $D_m(n)$ is the expectation value of the sum of the forward and backward prediction error energies, i.e.,

$$D_m(n) = \lambda D_m(n-1) + (1-\lambda)\left(e_{m-1}^2[n] + b_{m-1}^2[n-1]\right), \text{ with } 0 < \lambda < 1, \quad (3.3.12)$$

and the constant value α is normally chosen to $\alpha = 1 - \lambda$.

Due to the cascaded structure of a lattice filter, the GAL algorithm is both time-recursive and order-recursive. In practice, GAL algorithm is significantly faster in convergence than the conventional LMS algorithm. In adaptive filtering techniques, the gradient update rule can also be interpreted as a method to produce a recursive window function for linear predictive analysis.

3.4 Modified Linear Prediction Filter

As showed in previous sections, in traditional one-step forward linear prediction an estimate for the current sample value is formed as a linear combination of previous sample values, and the filter is assumed to be a conventional allpole filter. There are infinitely many alternative ways to form a linear combination of signal history and use it to predict the next signal value. The selection for sampling of signal history is not based on any mathematical necessity. An example of a modified formulation of the prediction principle which has been used in speech and audio application is frequency-warped linear prediction [55] [56] [57] [58] [59]. Warped linear predictive coding is an alternative for conventional LPC in speech and audio coding applications, especially for the perceptual audio coding system.

3.4.1 Generalized form of linear prediction filter

Recall that the classical linear prediction for a sample value $x[n]$ is given by

$$\hat{x}[n] = \sum_{k=1}^{p} a_x\, x[n-k]. \quad (3.4.1)$$

The z-transform of (3.4.1) is

$$\widehat{X}(z) = \left[\sum_{k=1}^{p} a_k z^{-k}\right] X(z). \quad (3.4.2)$$

This scheme can be generalized by replacing the unit delay z^{-k} with allpass filter $D^k(z)$ to obtain

$$\widehat{X}(z) = \left[\sum_{k=1}^{p} a_k D^k(z) \right] X(z) \,. \tag{3.4.3}$$

In time domain one may write

$$\hat{x}[n] = \sum_{k=1}^{p} a_x d_k \big[x[n] \big] \,, \tag{3.4.4}$$

where $d_k\big[x[n]\big] = x[n-k]$. The mean square error of the estimate can now be written as

$$MSE = E \left\{ \left| x[n] - \sum_{k=1}^{p} a_k d_k \big[x[n] \big] \right|^2 \right\} \,, \tag{3.4.5}$$

where $E\{\cdot\}$ is expectation. A conventional minimization procedure leads to a system of normal equations

$$E\big\{ d_j\big[x[n]\big] d_0\big[x[n]\big] \big\} \sim \sum_{k=1}^{p} a_k\, E\big\{ d_k\big[x[n]\big] d_j\big[x[n]\big] \big\} = 0 \tag{3.4.6}$$

with $j = 0, \ldots, p-1$. Because $D(z)^{-1} = D(z)^{-1}$ for an allpass filter, Parceval's theorem can be applied so that

$$E\big\{ d_{j+l}\big[x[n]\big] d_{k+l}\big[x[n]\big] \big\} \sim \sum_{n=-\infty}^{\infty} d_{j+l}\big[x[n]\big] d_{k+l}\big[x[n]\big] \tag{3.4.7}$$

$$= \frac{1}{i2\pi} \oint_C D(z)^{j+l} X(z) D(z^{-1})^{k+l} X(z^{-1}) \frac{dz}{z}$$

$$= \frac{1}{i2\pi} \oint_C D(z)^{j+l-k-l} X(z) X(z^{-1}) \frac{dz}{z}$$

$$= \sum_{n=-\infty}^{\infty} d_j\big[x[n]\big] d_k\big[x[n]\big] \sim E\big\{ d_j\big[x[n]\big] d_k\big[x[n]\big] \big\} \,,$$

where k, j and l are any integers and $\sim$ indicates that the normalization of the expectation is omitted to simplify notation. The equation (3.4.7) states that the same correlation values appear in both terms of the left-hand side of (3.4.6). Therefore, (3.4.6) can be seen as a generalized form of the Wiener-Hopf equations, and the optimal coefficients a_k can be obtained by using, for example, the Levinson-Durbin algorithm just like in the conventional autocorrelation method of linear prediction.

3.4.2   Warped linear prediction (WLP)

A set of orthogonal polynomial functions are given by [60]

$$D(z)^k = \frac{\sqrt{1 - |\lambda_k|^2}}{1 - \lambda_k z^{-1}} \prod_{p=1}^{k} \frac{z^{-1} - \lambda_p}{1 - \lambda_p z^{-1}} \,. \tag{3.4.8}$$

In equation (3.4.8), if $\lambda_k = \lambda_p = 0$ $(\forall k, p)$, this reduces to a traditional FIR filter, and if $\lambda_k = \lambda_p$ $(\forall k, p)$, this is the *Laguerre model* [61] which is a long tradition in the theory of signal processing. A simplified version of (3.4.8) given by

$$D(z)^k = \prod_{p=1}^{k} \frac{z^{-1} - \lambda_p}{1 - \lambda_p z^{-1}} \tag{3.4.9}$$

is called a *frequency-warped filter*, and the corresponding modified LPC scheme is called warped linear predictive coding (WLPC). From (3.4.7), the filter coefficients can be computed by the autocorrelation method of LP. In Fourier domain the equation (3.4.8) can be written as

$$D(e^{-i\omega}) = \prod_{p=1}^{k} e^{-ip\left(\omega + 2\arctan\left(\frac{\lambda \sin(\omega)}{1 - \lambda \cos(\omega)}\right)\right)} \tag{3.4.10}$$

$$= \prod_{p=1}^{k} e^{-ip\nu(\omega)} \,,$$

where $\nu(\cdot)$ is the frequency warping function. Therefore the spectral mapping is determined by the phase function of $D(z)^k$ which can be controlled by the value of λ.

The transfer function of a warped linear predictor is then given by

$$A(z) = 1 - \sum_{k=1}^{p} a_k D(z)^k \,, \tag{3.4.11}$$

where $D(z)$ is an allpass filter

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \,. \tag{3.4.12}$$

The current sample is estimated from samples of the frequency warped signal [56], which is formed by the outputs of the allpass filter chain. Figure 3.4.1(a) gives the structure of an WLP synthesis filter with order 2.
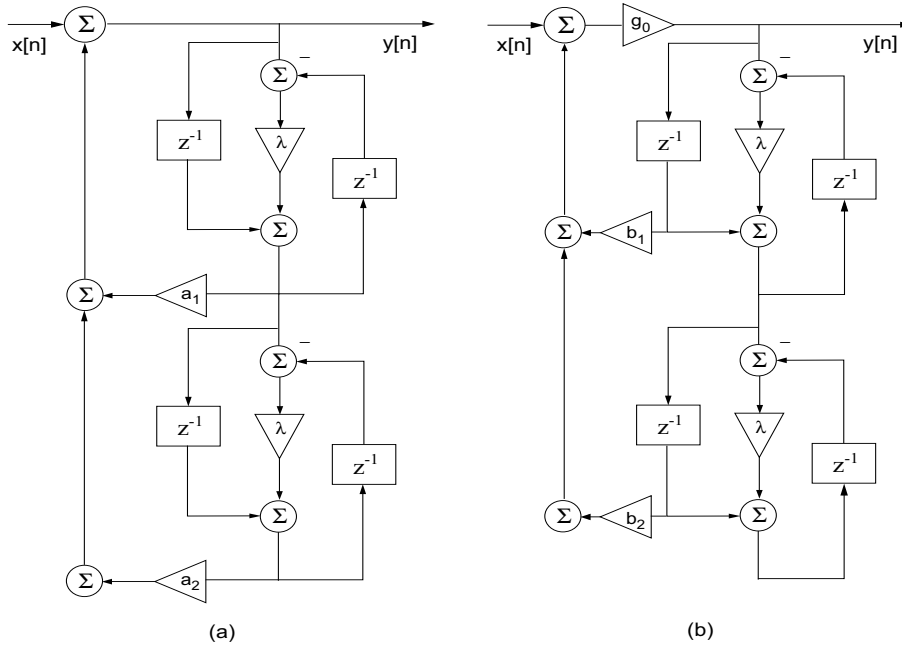
Figure 3.4.1: (a) WLP synthesis filter, (b) realizable WLP synthesis filter.

Note that this has a delay-free branch and is therefore not realizable in a practical implementation without modification. A possible modification is given Fig. 3.4.1(b) [58]. The coefficients g_0 and $b_1, \ldots, b_{p+1}$ can be calculated by using the following equations:

$$g_0 = \frac{1}{1 + \sum_{k=1}^{p} a_k(-\lambda)^k} \tag{3.4.13a}$$

$$b_1 = \sum_{k=1}^{p} a_k(-\lambda)^k \tag{3.4.13b}$$

$$b_k = \sum_{k=l}^{p} a_k(-\lambda)^{k-l} - \sum_{k=l-1}^{p} a_k(-\lambda)^{k-l+2} \tag{3.4.13c}$$

$$\text{with} \quad l = 1, \ldots, p+1$$

A warped FIR lattice filter may be derived directly by replacing the unit delays of the conventional lattice structure with first-order allpass elements. This leads to the structure shown in Figure 3.4.2. The reflection coefficient of the warped structure can be computed from the estimated coefficients of a warped direct-form filter as in the case of a conventional filter [62].
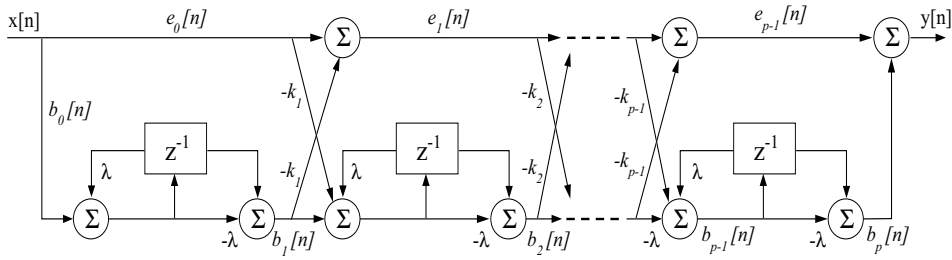
Figure 3.4.2: A warped FIR lattice filter structure

Since WLP performs linear prediction in the warped frequency domain, the resulting auto-regressive process has different resolutions in modeling the spectral peaks at different frequencies. The frequency resolution of a WLP model is controlled by the allpass coefficient λ. This is advantageous in speech and audio signal processing because the frequency resolution in the spectral estimate is relatively close to the frequency resolution of human hearing [63]. The performance of conventional and warped LPC algorithms is compared in a simulated coding system using listening tests and bit rate [59].

3.5   Context-based Modeling

Statistical modeling of the source being compressed plays a central role in any data compression systems. Given a finite source sequence with length N, $x_1, x_2, \ldots, x_N$, the optimal code length of the sequence in bits is

$$H_{opt} = -\log_2 \prod_{n=1}^{N} p(x_n | x^{n-1}) \tag{3.5.1}$$

where x^{n-1} denotes the sequence $x_{n-1}, x_{n-2}, \ldots, x_1$. The key issue in context modeling for compression of the signal x^N is to estimate $p(x_n | X^{n-1})$, where X^{n-1} denotes a subset of the available past sequence x^{n-1} (causal template), with the assumption that the random process producing x^N is Markovian. The probability of the current symbol is conditioned by the set of past observations X^{n-1} in the form of the conditional probability mass function (pmf), $\hat{p}(x_n | X^{n-1})$, that serves as a statistical model of the source. In classical universal context tree modeling [64] [65], for example, the different contexts of $\{X^{n-1}\}$ are grouped in a context tree, having as nodes as possible values of $\{X^{n-K}\}$, up to a depth level dictated by the affordable memory resources. The tree is grown as the

message is encoded, according to the contexts seen so far. Histogram tracking is used for modeling the pmf to each node. The context algorithm can also be used for prediction, by predicting $\hat{x}_n = \arg\max_x p(x_n|X^{n-1})$.

The algorithm is optimal in encoding Markovian sources, but becomes extremely complex for large size alphabets, such as the audio signal sampled at 16 bits/sample or higher. In the case of large size alphabets, the Markov finite state machine (FSM) becomes extremely large,[6] incurring two major problems; the estimating of the conditional probabilities for each context,[7] and the excessive memory requirement to store the large number of all possible states (contexts). As a result, the context model can not learn the source statistics fast enough to estimate accurately the conditional probability distribution, leading to the so called "context dilution" [66]. Hence, the main challenge in audio compression applications is to find a compromise between the context size and compression performance and to mimic the main principles of context algorithm for sources with large size of the alphabet. Several different approaches to the reduced context size are proposed such as context quantization [67], histogram bucketing [68] [69], or context tree modeling [64].

3.5.1 Design of statistical prediction filter

An adaptive context algorithm for prediction modeling is proposed [70], using the combination of L-predictor [71] and the FSM context modeler. The L-predictor is defined as a linear combination of the order statistics

$$\hat{x}_n = \sum_{k=1}^{K} b_k X_k \ , \tag{3.5.2}$$

where X_k denote the k-th order statistics in the set $\{X_1, ..., X_K\}$. The conditional information $X^{n-1} = [X_1, \ldots, X_K]$ is obtained by ordering the x^{n-1} increasingly. This means that the prediction is itself the main feature extracted from the contextual information. The prediction error in this case is defined as

$$e_n = x_n - \hat{x}(n|X_1, X_2, \ldots, X_K) \,. \tag{3.5.3}$$

For the given alphabet interval $[0, M-1]$, (3.5.2) can be parameterized by setting $X_0 = 0$ and $X_{K+1} = M - 1$;

$$\hat{x}_n = a_0 + \sum_{k=1}^{K+1} a_k(X_k - X_{k-1}) \,. \tag{3.5.4}$$

---

[6]e.g., given a sample value with Z bits resolution, there are 2^{ZK} different contexts.

[7]because even a very large signal does not provide sufficient samples to reach good estimates

This reparameterization in (3.5.4) has several advantages [70]; the flexibility of the prediction model, for example, that allows us to have a predictor with adapted properties at each different contexts. By selecting the parameters of the model at time n, (3.5.4) can be written as

$$\hat{x}_n = \hat{x}(n|w_n) = a_0(w_n) + \sum_{k=1}^{K+1} a_k(w_n)(X_k - X_{k-1}), \qquad (3.5.5)$$

where w_n denotes the suitable functions of the context at that time.

In [72], for example, two contexts are selected by context tree modeling and Hasse diagram. With these contexts the prediction value of current sample is computed using the adaptive linear prediction model similar to (3.5.5). First the contextual information is selected by a context mask containing the most recent N_c samples $x^{n-1} = [x_{n-1}, \ldots, x_{n-N_c}]$ and most recent N_e prediction errors $e^{n-1} = [e_{n-1}, \ldots, e_{n-N_e}]$, that is, the prediction errors from (3.5.3) are also used as part of the contexts. A recursive least squares (RLS) algorithm with forgetting factor w is used to perform the adaptive linear prediction;

$$\hat{x}_n = \sum_{k=1}^{p} w_{S_m^{(n-1)}}(k)\, x_{n-k}, \qquad (3.5.6)$$

where $S_m \in \{1, \ldots, N_m\}$ is the main context, obtained by using a tree classification procedure $S_m^{n-1} = \mathcal{T}(x^{n-1}, e^{n-1})$. A secondary context $S_s \in \{0, \ldots, N_s - 1\}$ is obtained by use of a Hasse diagram selection $S_s^{n-1} = \mathcal{H}(x^{n-1})$. The Hasse cube forms the state transition diagram of the finite state machine. Using the main and secondary contexts, the Eq. (3.5.6) can be extended with the intercept ρ,

$$\hat{x}_n = \rho_{S_m^{n-1}, S_s^{n-1}} + \sum_{k=1}^{p} w_{S_m^{n-1}}(k)\, x_{n-k}, \qquad (3.5.7)$$

where the intercept depends both on main and secondary contexts, while the parameters $w_{S_m^{n-1}}(k)$ depend only on the main context.

3.5.2   Context-based error modeling

To achieve high compression ratio by entropy coding, e.g., by arithmetic or Golomb-Rice coding, the residual signal consisting of prediction errors should be similar to stationary white noise as much as possible. As a matter of fact, however, the residual signal has ample redundancy and especially is still nonstationary. The better the prediction, however, the more noise like prediction errors will be. An improvement of coding rate can be

obtained by applying the context algorithm to make use of the dependency between the residuals. It is showed in [64] [73] that the context algorithm in conjunction with arithmetic or Huffman coding provides to produce good results in lossless image and audio compression.

In image compression, context algorithm for modeling of the prediction errors or the transform coefficients is highly studied in order to condition the distribution of the prediction errors [74] [75] [73]. In these cases, the prediction errors are arranged into a predefined number of statistically homogeneous classes based on their spatial *context*, i.e., prediction-based context modeling. If such classes are statistically discriminated, then the zeroth-order entropy of a *context-conditioned* model of prediction errors will be lower than that derived from a stationary memoryless model of the decorrelated source.

The cost of context modeling is proportional to the number of the parameters in the statistical model and could then offset the entropy savings. It means also that a direct implementation of the classical universal context algorithm to the high quality audio, such as DVD audio sampled by 96 kHz at 20 or 24 bits/sample, does not provide a good solution. Although the universal algorithms are proven to be asymptotically optimal for stationary sources, the complexity of their underlying models will add an extra term to the best achievable coding rate. Hence a key objective in a context modeling scheme for high quality audio compression is to reduce the number of parameters defining the coding distribution at each context and the number of contexts.

3.6 Experimental Comparison of Prediction Filters

A block diagram of the lossless predictive coding system for experimentation is shown in 3.6.1[8]. The linear prediction filters with different structures, i.e., FIR, IIR, lattice, and polynomial approximation (PAP) are tested with various block lengths and different filter orders. In addition, an efficient context-based error modeling (FIR-CM) is also experimented in conjunction with FIR linear predictor and Golomb-Rice coding. The audio materials used in the experiments are 9 files in CD-format, i.e., 44.1 kHz, Stereo, 16 bits, described in Section 2.5. There is provision to take advantage of dependency between the two stereo channels, but this does not provide a significant improvement except for the mono recorded test material Nr. 3 (female speech).

---

[8]This compression scheme is also used as a prototype system for all experiments in the thesis. In most cases, the decorrelation stage is replaced by each filter discussed in next chapters.
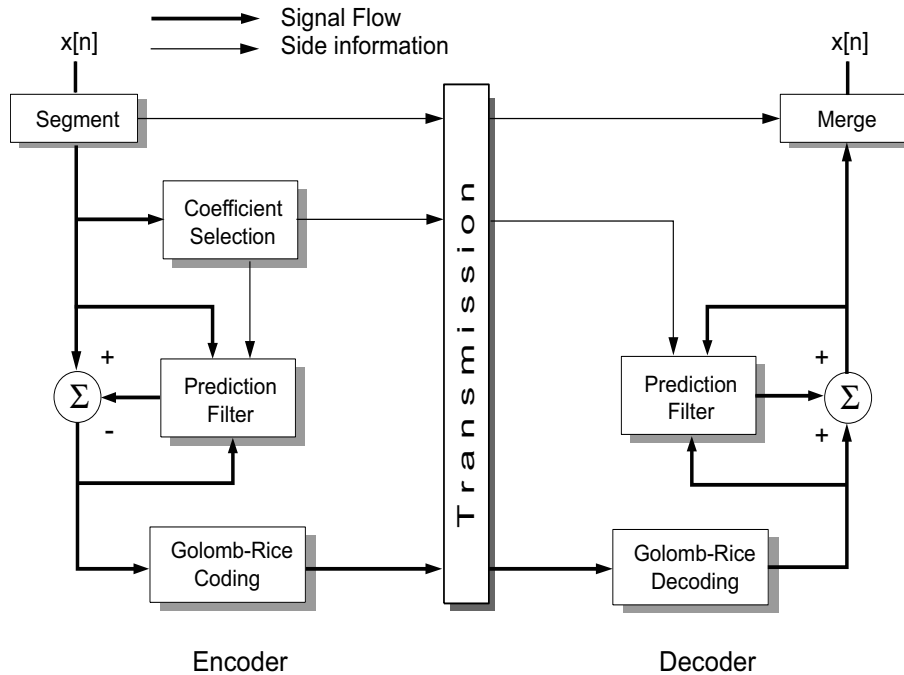
Figure 3.6.1: Block diagram of lossless audio coding system for the experiments.

3.6.1   Description of compression system

Segmentation

The input audio signal is divided into the blocks (frames), and each block is treated separately during the coding operation. A simple rectangular windowing is used for the segmentation without overlapping between frames and padding operation. We experimented the compression system with various block lengths being a power of two, 2^N, with N between 8 to 12, i.e., 256, 512, 1024, 2048, and 4096 samples.

Prediction filters

Four types of the linear prediction filters, i.e., FIR (Fig. 3.2.1.a), IIR (Fig. 3.2.1.b), lattice (Fig. 3.2.2), and polynomial approximation predictor (PAP) with order $p \le 4$ using Eq. (3.2.42), are tested to compare the efficiencies of the filters. Except for PAP, all filters are tested by the prediction order up to 10. The coefficients for FIR, IIR, and lattice filter are found using the standard Levinson-Durbin's recursion algorithm (3.2.16). First the autocorrelation coefficients of each block are computed and converted to prediction filter coefficients by using Levinson-Durbin algorithm.

Residual modeling and entropy coding

For entropy coding, we use Golomb-Rice code with fixed parameter k (see Section 2.2.3), since Golomb-Rice code is optimized for a block of signals having a Laplacian probability density (double-sided exponential distribution), which is found to be a good approximation for the distribution of the prediction residual samples resulted from the decorrelation operations. The estimation of optimal parameter k is linearly related to the variance of the signal. The Laplacian distribution is defined by

$$p\left(x\right) = \frac{1}{\sqrt{2}\sigma} e^{\frac{-\sqrt{2}}{\sigma}|x|} \,, \tag{3.6.1}$$

where σ^2 is the variance of the distribution. An expectation of the absolute value of x can be given as following

$$E(|x|) = \int_{-\infty}^{\infty} |x|\, p(x)\, dx \tag{3.6.2}$$

$$= \int_{0}^{\infty} x\, \frac{\sqrt{2}}{\sigma} e^{\frac{-\sqrt{2}}{\sigma}x}\, dx \tag{3.6.3}$$

$$= \int_{0}^{\infty} e^{\frac{-\sqrt{2}}{\sigma}x}\, dx - \left[x\, e^{\frac{-\sqrt{2}}{\sigma}x}\right]_{0}^{\infty} \tag{3.6.4}$$

$$= \frac{\sigma}{\sqrt{2}} \,. \tag{3.6.5}$$

Since the optimal parameter k means that half the samples lie in the range $\pm 2^k$, the code word length of integer n is optimal when $k + 1$ bits are for probability 0.5 and $k + n + 1$ bits for probability $2^{-(k+n)}$. It leads to that

$$\frac{1}{2} = \int_{-2^k}^{2^k} p(x)\, dx \tag{3.6.6}$$

$$= \int_{-2^k}^{2^k} \frac{1}{\sqrt{2}\sigma} e^{\frac{-\sqrt{2}}{\sigma}|x|}\, dx \tag{3.6.7}$$

$$= -e^{\frac{-\sqrt{2}}{\sigma}2^k} + 1 \,. \tag{3.6.8}$$

Therefore, the expectation function for the parameter k is given by

$$k = \log_2\left(\log_e(2)\frac{\sigma}{\sqrt{2}}\right) \tag{3.6.9}$$

$$= \log_2\left(\log_e(2)\, E(|x|)\right) \,. \tag{3.6.10}$$

In PAP implementation k is obtained using the Eq. (3.6.10). For the other schemes with FIR, IIR, and Lattice predictor, the expectation of residual samples $E(|e[n]|)$ is derived from σ, which is computed for the calculation of predictor coefficients.

Context-based error modeling (FIR-CM)

It is common that the residual signal has still redundancy even after the decorrelation by prediction. In FIR-CM scheme, a context algorithm for error modeling is embedded into the standard FIR prediction scheme in order to make use of the dependency between the residuals.

We used the similar context selection algorithm proposed in [72]. After the context indexing, the error remapping algorithm used in LOCO-I [73] is performed to improve the Golomb-Rice coding performance. The error signal $e[n]$ is remapped to positive integer using the revertible remapping;

$$e' = \begin{cases} 2e & \text{if } e \geq 0\,, \\ 2|e| - 1 & \text{otherwise}\,. \end{cases} \tag{3.6.11}$$

3.6.2   Test results and discussion

Table 3.1 shows the test results for comparison of different prediction filters. As showed in the test results, IIR prediction filter with large block length and higher prediction order has superior compression result for nearly all kind of audio samples over other prediction filters. Especially for the music samples with high treble energy and wide dynamic range (e.g., Nr. 6, 7, 8), IIR prediction scheme provides an acceptably stable compression ratio. It was clearly proven in our experiment that IIR prediction filter has more potential for improving the compression rate than FIR prediction filter, as discussed in [15]. As expected from its simple integer algorithm, PAP has lowest complexity with relative efficient compression performance. Using context-based error modeling, FIR-CM improved the compression rate 2-6% compared with that of FIR. In fact, this result of FIR-CM is less than we expected. Using all nodes in the Hasse diagram instead of middle layer nodes could have the potential of more improving the compression rate with several percents, but unfortunately with the cost of a twofold will increase in the overall complexity of the algorithm.

The choice of block length is an important consideration in implementation of most LP filter systems. In general, a longer block length may require an unreasonably high amount of computation and reduce the editibility. As showed in the results, for the non-stationary signal with highly localized statistical behavior like the speech (in our case Material Nr. 3) the block length should be small enough. Figure 3.6.2 shows the compression ratio with respect to the different block legnths. In our experiment, the efficient block length was between 2048 and 4096 samples for all LP filters.

Materials	PAP[1]	FIR	IIR	Lattice	FIR-CM	Gzip
Nr. 1 violin, solo	7.56 (4096,3)[2]	7.46 (4096,3)	**7.23** (4096,10)	7.59 (4096,3)	7.59 (2048,9)	12.05
Nr. 2 flute, solo	**5.76** (4096,3)	6.12 (4096,2)	5.77 (4096,10)	5.86 (4096,3)	6.11 (2048,2)	11.04
Nr. 3 speech, fem.	6.02 (256,2)	6.06 (1024,5)	5.96 (1024,9)	6.26 (512,2)	**5.92** (1024,4)	8.62
Nr. 4 piano, solo	4.23 (1024,3)	4.51 (4096,2)	**4.13** (4096,5)	4.45 (4096,3)	4.39 (2048,2)	10.43
Nr. 5 classic, orch.	**5.73** (1024,3)	6.07 (2048,3)	5.92 (4096,4)	5.90 (4096,3)	5.79 (2048,4)	12.79
Nr. 6 pop, abba	7.13 (512,2)	7.05 (2048,5)	**6.87** (2048,7)	7.21 (1024,2)	6.93 (2048,3)	11.76
Nr. 7 country	6.38 (4096,1)	6.35 (1024,5)	**6.28** (1024,4)	6.57 (1024,1)	6.37 (1024,3)	9.65
Nr. 8 rock, metal	12.16 (512,2)	11.40 (2048,10)	**11.31** (4096,10)	12.23 (4096,2)	11.49 (2048,9)	14.86
Nr. 9 jazz, soft	7.64 (512,2)	7.89 (4096,2)	7.73 (4096,10)	7.80 (2048,2)	**7.57** (2048,2)	13.93
Average	6.96 (1792,2.3)	6.99 (2731,4.1)	**6.80** (3186,7.7)	7.10 (2786,2.3)	6.91 (1820,4.2)	11.68

[1] prediction order $p \leq 4$, for the other filters $p \leq 10$

[2] (block length, prediction order)

Table 3.1: Test results with compression rates (bits/sample)

The order of the prediction filter is also set as a tradeoff. For large predictor orders, the number of bits necessary to send the side information will be large, but the quality of the prediction will be better, and therefore the residuals will be encoded with a small number of bits. For small predictor orders, the side information will be encoded using a small number of bits, but encoding of residuals will require a large amount of bits. However, there are reasons for keeping the filter order as low as possible. It is not just a matter of decoder and encoder complexity, as the coefficients of the filter have to be transmitted periodically, possibly with initialization data for the associated state variables, and the overheads of transmitting the side information can be significant and increase with filter order. Figure 3.6.3 shows the compression rate with respect to the prediction orders. Compared with the case of the block length in Figure 3.6.2, the efficient order of the prediction filter varies with the characteristics of each prediction filter. In our experiment, predictor order of 6 to 10 provided the best results from IIR filter, and 2 to 4 from the others.
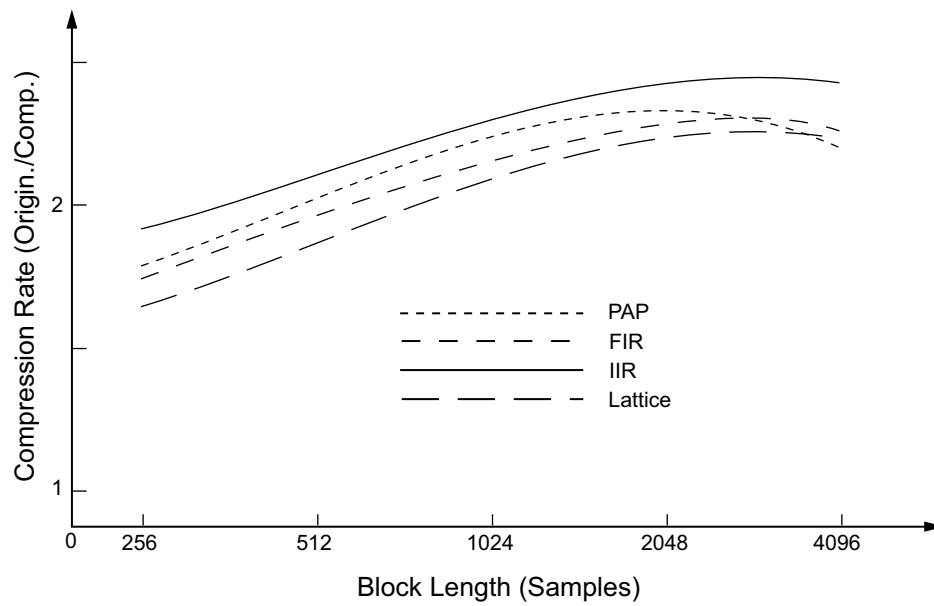
Figure 3.6.2: Comparison between block length and averaged compression ratios of all test materials
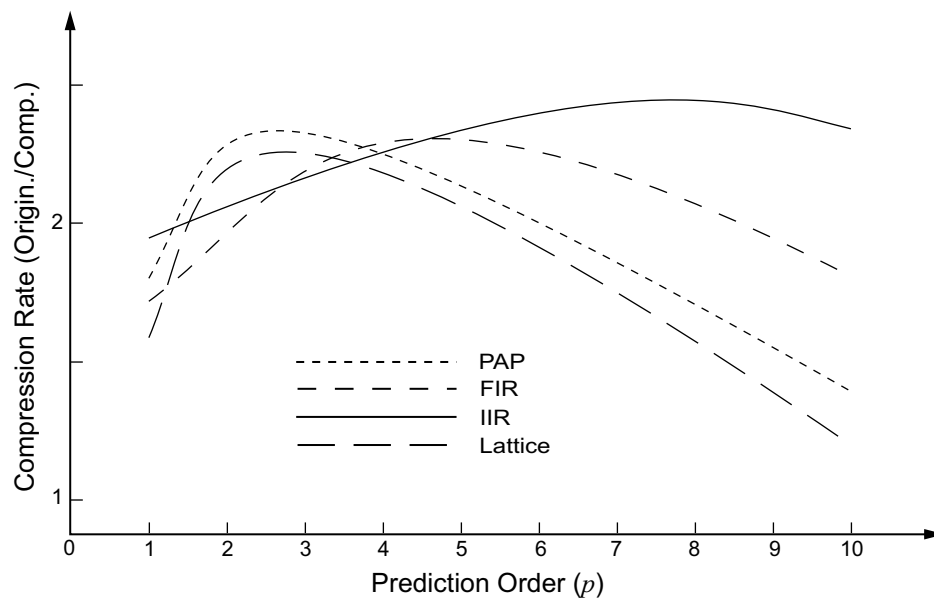
Figure 3.6.3: Comparison between prediction order and averaged compression rate of all test materials

Chapter 4

Filter Bank and Subband Transform

Subband transform is one of widely used signal decomposition techniques for data compression. In the subband transform the source spectrum is bandpass filtered and subsampled to obtain a time-frequency decomposition. Although the technique introduces delay into the system, this drawback is offset by the coding gains obtained from processing subband data rather than the fullband data. These observed gains have led to extensive use of subband decomposition for compression of video, image, audio and speech signal. Since the construction of the wavelet bases is equivalent to designing two-channel filter banks, the subband transform with octave band decomposition is also often referred to as discrete wavelet transform (DWT).

In this chapter we first review two well-known coding methods, Laplacian pyramid and subband coding scheme, and then describe the design methods of multirate filter banks. The efficiency of prediction from subbands and from fullband is formally compared and empirically investigated by designing the hybrid audio compression system.

4.1   Subband Coding Algorithms

4.1.1   The Laplacian pyramid

One of the first methods for hierarchical decomposition is formally proposed, especially for image coding, by Burt and Adelson [76]. They used a pyramid cascade of small Gaussian-like filters to create an overcomplete subband representation which they called a Laplacian pyramid. A signal will be decomposed in two pyramids. First, a *coarse approximation* of the original signal is divided using lowpass filtering with a halfband lowpass filter and subsampling by two (drop every other sample). Based on

this coarse approximation, the original signal is predicted using upsampling the coarse approximation and convolving with an interpolation filter and a *prediction error* is calculated. This prediction error represents the detail features that are missing in the coarse approximation. Note that the subsampling by two is equivalent to doubling the scale in the wavelet analysis. A schematic diagram of the one level pyramid decomposition is illustrated in Figure 4.1.1. The scheme can be iterated on $y[n]$ in order to create a hierarchy of lower resolution signals at lower scales. In the result, signals become shorter and shorter, therefore such schemes are called signal pyramid. Typically, the analysis filters and synthesis filters in the iterated full pyramid scheme are set to some common, compact lowpass filter, although better coding results are obtained by choosing the two filters independently.
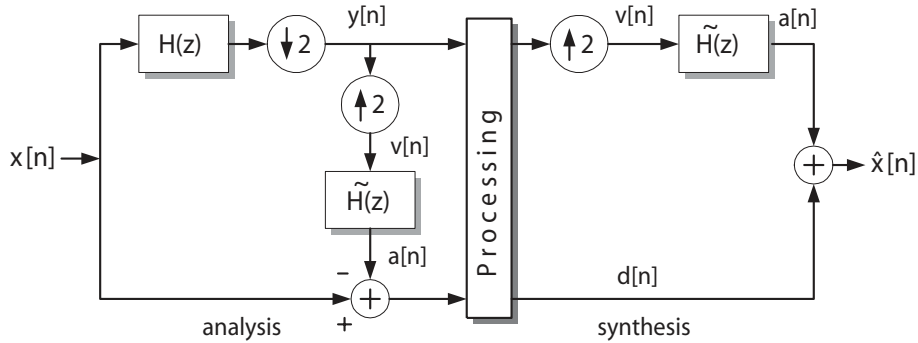
Figure 4.1.1: Construction of one level Laplacian pyramid decomposition: A full pyramid is recursively built by non-uniformly cascading this system to the lowpass subband.

Given an original sequence $x[n]_{n \in \mathbb{Z}}$, the coarse approximation $y[n]$ is given by

$$y[n] = \sum_k h[k] \, x[2n - k] \,, \tag{4.1.1}$$

where $h[n]$ is impulse response of lowpass filter. To find an approximation $a[n]$, $y[n]$ is upsampled by two, that is, inserting a zero between every sample, since we need a signal at the original scale for comparison. Then, the upsampled sequence $v[n]$ that has same length with original sequence $x[n]$ is interpolated with filter $\tilde{h}[n]$.

$$v[2n] = y[n], \quad v[2n + 1] = 0 \quad : \text{upsampling,} \tag{4.1.2}$$

$$a[n] = \sum_k \tilde{h}[k] \, v[n - k] \quad : \text{interpolation,} \tag{4.1.3}$$

$$d[n] = x[n] - a[n] \quad : \text{prediction error.} \tag{4.1.4}$$

It is obvious that $x[n]$ can be simply reconstructed by adding $d[n]$ and $a[n]$. This reconstruction is exact, regardless of the choice of the filters $h[n]$ and $\widetilde{h}[n]$. In the case of perfect halfband lowpass filter, $d[n]$ contains exactly the frequencies above $\pi/2$ of $x[n]$, and thus $d[n]$ can be subsampled by two as well without loss of information. This hints at the fact that critically sampled schemes must exist. It will be convenient to use z-transform of filters, i.e. $H(z) = \sum_{n\in\mathbb{Z}} h[n]\, z^{-n}$. (4.1.1) can be then written as follows,

$$Y(z) = \tfrac{1}{2}[H(z)X(z) + \underbrace{H(-z)X(-z)}_{\text{aliasing term}}]. \tag{4.1.5}$$

The prediction error is

$$D(z) = X(z) - \tfrac{1}{2}\widetilde{H}(z)[H(z)X(z) + H(-z)X(-z)], \tag{4.1.6}$$

and the input-output relation is

$$\begin{aligned}
\widehat{X}(z) &= \widetilde{H}(z)Y(z) + D(z) \\
&= \tfrac{1}{2}\widetilde{H}(z)[H(z)X(z) + H(-z)X(-z)] \\
&\quad + X(z) - \tfrac{1}{2}\widetilde{H}(z)[H(z)X(z) + H(-z)X(-z)] \\
&= X(z).
\end{aligned} \tag{4.1.7}$$

Formula (4.1.7) shows that the *aliasing* term can be eliminated so that the input is a delayed copy of the output, only if

$$\widetilde{H}(z)H(-z) = 0 \tag{4.1.8}$$

is satisfied for all frequency range. It means that the synthesis filters should be time-inverted version of the analysis filters. However, in practice, it is impossible to realize such a lowpass filter that satisfies this orthogonality condition. Most filter designs for pyramid decomposition attempt to minimize the aliasing resulting from the subsampling process in the pyramid schemes. An ideal subband system incorporates "brick-wall" bandpass filters which avoid the aliasing in process. Such filters, however, produce the ringing effect (Gibbs phenomenon) in the spatial domain which is perceptually undesirable.

In addition to suitability for data compression of the pyramid decomposition, the Laplacian pyramid code formed as multi-scale data construction is particularly useful for the progressive transmission. The progressive transmission can be achieved by sending the pyramid codes in order from lowest to highest resolution. At the receiving end, in result, an initial signal reconstructed from the code of lowest resolution becomes gradually a finer signal by adding the rest of codes with higher resolution. This progression allows also the receiver to choose the quality level of the signal.

The Laplacian pyramid has certain disadvantages, especially for *lossless* compression. The quantization errors from highpass subbands do not remain in these subbands. Instead, they appear in the reconstructed signal as broadband noise. Furthermore, as a result of the non-orthogonality of the pyramid scheme, the number of samples representing the signal is increased by a factor of 2 in one dimension and by a factor of $\frac{3}{4}$ in two dimensions. This is due to the correlation between the detail signals at different resolution. This correlation can be suppressed with the multiresolution analysis (MRA) described in section 5.6. In MRA, it is possible to extract the detail signal $d[n]$ as an exact difference of information between $x[n]$ and $y[n]$ by decomposing the signal into an orthonormal wavelet basis.

In two dimensions, the basis functions are not oriented, and thus will not extract the oriented structural redundancy. On the other hand, for motion-compensated video coding, this overcompleteness of the Laplacian pyramid has been effectively used to make the coding system robust against motion-compensation errors [77].

4.1.2 Subband coding scheme

As mentioned above, the Laplacian pyramid produces a overcomplete representation. One stage of a pyramid decomposition leads to both a half rate low resolution signal and a full rate difference signal, resulting in a redundant set of samples. Subband coding scheme, which is based on banks of *quadrature mirror filter* (QMF) and first developed in speech compression [78] [79] [80], provides no such redundancy. As illustrated in Figure 4.1.2, the lowpass, subsampled approximation is obtained exactly as the Laplacian pyramid, but instead of a difference signal, the "added detail" as a bandpass filtered version of input signal will be computed. It is then followed by subsampling by two as in lowpass process. It is obvious that the added detail to the lowpass approximation has to be a bandpass signal, and that if $h[n]$ is an ideal halfband lowpass filter, then an ideal bandpass or halfband highpass filter $g[n]$ will lead to a perfect representation of the original signal into two subsampled versions.

The input-output relation is

$$\widehat{X}(z) = T_d(z)X(z) + T_a(z)X(-z), \qquad (4.1.9)$$

where

$$T_d(z) = \frac{1}{2}[H(z)\widetilde{H}(z) + G(z)\widetilde{G}(z)],$$

$$T_a(z) = \frac{1}{2}[H(-z)\widetilde{H}(z) + G(-z)\widetilde{G}(z)]. \qquad (4.1.10)$$

The transfer function $T_d(z)$ is a linear shift-invariant (LSI) system response, and $T_a(z)$ is the system aliasing. In the subband decomposition, using a class of non-ideal FIR

bandpass filters allows to avoid the aliasing in the overall system output.
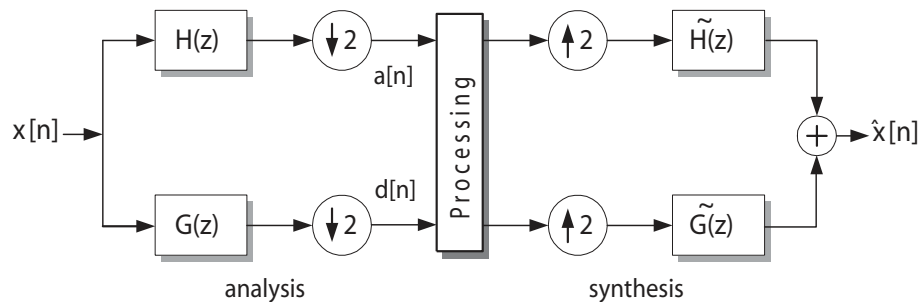
Figure 4.1.2: Subband decomposition. Two subsampled approximations are computed, one corresponding to low and the other to high frequencies. The reconstructed signal is obtained by re-interpolating the approximations and summing them.

As showed in Figure 4.1.2, the original signal $x[n]$ can be recovered from its two filtered subsampled versions, $a[n]$ and $d[n]$, which are upsampled and filtered by synthesis filters $\tilde{h}[n]$ and $\tilde{g}[n]$. Then we have the reconstructed signal $\hat{x}[n]$ by adding the two upsampled filtered versions. Till now, unlike the pyramid case, the reconstructed signal $\hat{x}[n]$ is not identical to the original signal $x[n]$ unless the filters satisfy some specific conditions, that is, *perfect reconstruction conditions* which will be discussed in following section.

4.2   Design of Multirate Filter Banks

Multirate filter banks are fundamental building blocks for subband decompositions and multiplexing on the frequency or time domain. Figure 4.2.1 illustrates a general structure of M-channel filter banks where $G_i(z)$ and $F_i(z)$ are the analysis and synthesis filters, respectively. A tree structure of the multirate filter banks can be obtained by uniformly or non-uniformly iterating on first output of the filter bank. This tree structure is a fundamental point in wavelet theory. Also, the tree leads to a multiresolution decomposition of signal. We will later investigate the convergence of the tree structure to the wavelet transform.
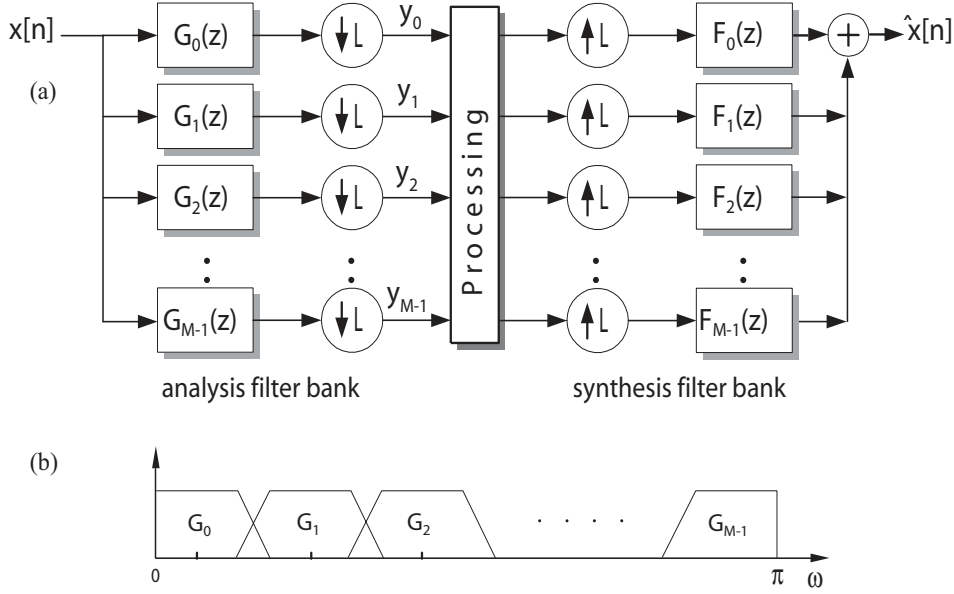
Figure 4.2.1: A general structure of M-channel multirate filter bank. (a) Spectrum division of the bandpass filters. (b) M-channel multirate filter banks with subsampling by factor L, $L \leq M$ (reconstruction condition).

In Figure 4.2.1, the system output in Fourier domain is given by

$$
\begin{aligned}
\widehat{X}(\omega) &= \frac{1}{L} \sum_{i=0}^{M-1} \left[\sum_{k=0}^{L-1} G_i\left(\omega + \tfrac{2\pi k}{L}\right) X\left(\omega + \tfrac{2\pi k}{L}\right) \right] F_i(\omega) \\
&= \frac{1}{L} \sum_{i=0}^{M-1} G_i(\omega) F_i(\omega) X(\omega) \\
&\quad + \frac{1}{L} \sum_{k=1}^{L-1} X\left(\omega + \tfrac{2\pi k}{L}\right) \sum_{i=0}^{M-1} G_i\left(\omega + \tfrac{2\pi k}{L}\right) F_i(\omega), \quad\quad (4.2.1)
\end{aligned}
$$

where the first sum corresponds to a linear shift-invariant system response and the second contains the system aliasing. Note that the reconstruction condition $L \leq M$ means that the filter banks constitute a frame so that any signal can be reconstructed from the subband signals. If $L = M$, it is critically sampled filter banks, whereas the term of oversampled filter banks refers to the case $L < M$. The oversampled filter banks are equivalent to a particular class of frame in $l^2(\mathbb{Z})$, and the necessary and sufficient conditions on a filter bank for implementing a frame or a tight frame expansion are investigated by Cvetković and Vetterli [81]. They focused also on nonsubsampled filter banks which implement transforms similar to continuous-time transforms and allow for very flexible design. In this section, we assume that all filter banks are critically subsampled, i.e. $L = M$, always.

Assume that we use a finite set of sequences $g_i[n]$, $i = 0 \ldots M - 1$, to obtain a structured orthonormal basis $\{\psi_i[n]\}_{i \in \mathbb{Z}}$ in $l^2(\mathbb{Z})$ such that

$$\langle \psi_i[n], \psi_j[n] \rangle = \sum_{n \in \mathbb{Z}} \psi_i^*[n]\, \psi_j[n] = \delta_{i,j}. \tag{4.2.2}$$

For example, it can be achieved through shifts by M,

$$\psi_{i+kM}[n] = g_i[n - kM], \quad k \in \mathbb{Z},\, i = 0, 1, \ldots M - 1\,. \tag{4.2.3}$$

In practice, this expansion can be computed using the M-channel multirate filter banks in Figure 4.2.1. First, the analysis filter bank decomposes the input signal into M sub-bands by convolving with parallel set of bandpass filters having impulse responses $g_i[n]$ followed by subsampling by $L = M$. To reconstruct the signal, the subband signals are upsampled, filtered with interpolation filters and then combined additively. Each subsampled subband signal encodes a particular portion of the frequency spectrum. For example, in signal compression, the spectral contents of the subband signals are coded depending on their energies whereas in radar system, the subband signals might be used to null out a narrow-band interference adaptively. Therefore, the structure of the filter bank depends on the practical applications.

Because of the subsampling process, the reconstructed signal has aliasing and distortion which is a main obstacle to designing the subband coding system. Perfect reconstruction filter banks provide the system that has no such errors, that is, the output as a time-delayed copy of the input signal. *Quadrature mirror filters*, developed by Croiser, Esteban and Galand [79], that are generally two channel filter bank, can be used to avoid aliasing in the overall system output. It is later discussed by Adelson [82] and Mallat [83] that these filters also form an orthogonal subband transform. However, a quadrature mirror filter can not have a finite impulse response, except for the simple Haar filter. To obtain perfect reconstruction orthogonal filters with a finite impulse response, necessary and sufficient conditions, called *conjugate mirror filters*, are found by Smith and Barnwell [84] [85] and Mintzer [86]. The generalization to the theory is completed by the biorthogonal equations of Vetterli [87] [88] and the general paraunitary matrix theory of Vaidyanathan [89]. M-channel cosine-modulated filter banks, which are first proposed by Nussbaumer [90] and Rothweiler [91] under the name of *pseudo* quadrature mirror filters, are used in MPEG standard for audio compression [92] [93] [94] [95]. In that case, the amount of aliasing is comparable to the stopband attenuation, and the distortion is minimized by an optimization process.

4.2.1 Perfect reconstruction conditions

We will concentrate on the critically sampled two-channel filter bank (Figure 4.2.2), because it is the simplest and most important case in practice and leads to wavelets.
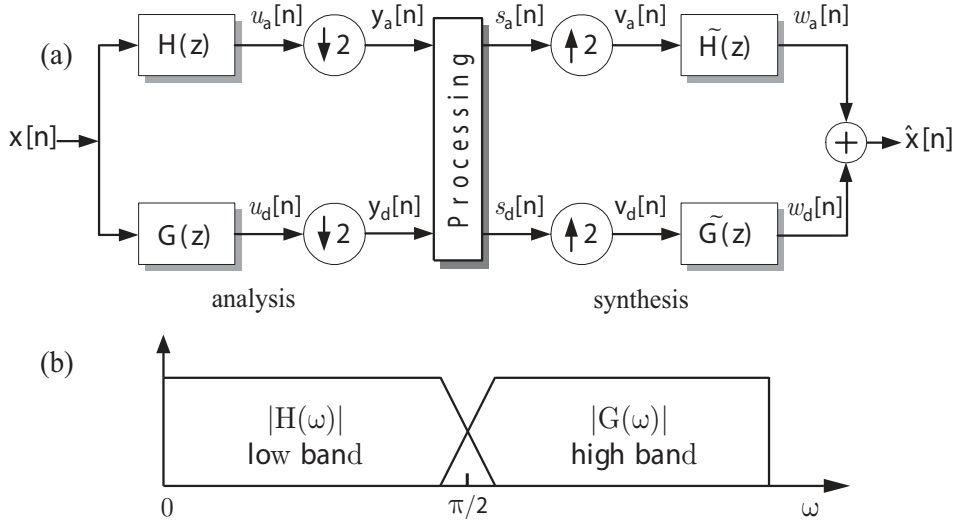
Figure 4.2.2: Two-channel filter bank. (a) The two-channels, with filtering using $h[n]$ and $g[n]$, subsampling, upsampling by factor 2 and interpolation filtering with $\tilde{h}[n]$ and $\tilde{g}[n]$. (b) Spectrum splitting between lowpass and bandpass filters.

The overall system response of the filter bank in Fourier domain is given by (4.2.1),

$$
\begin{aligned}
\widehat{X}(\omega) = {} & \frac{1}{2}\left[H(\omega)\widetilde{H}(\omega) + G(\omega)\widetilde{G}(\omega)\right]X(\omega) \\
& + \frac{1}{2}\left[H(\omega+\pi)\widetilde{H}(\omega) + G(\omega+\pi)\widetilde{G}(\omega)\right]X(\omega+\pi) \\
= {} & X(\omega)\,e^{-jl\omega},
\end{aligned}
\tag{4.2.4}
$$

where the first term is a linear shift-invariant (LSI) system response related to distortion and the second term with $(\omega+\pi)$ reflects the system aliasing[1]. For the PR filter bank, the filters have to satisfy the following two PR conditions:

- Distortion-free (DF) condition

$$
H(\omega)\widetilde{H}(\omega) + G(\omega)\widetilde{G}(\omega) = 2\,e^{-jl\omega}
$$

---

[1]It means that two inputs can give the same output, e.g. the constant vector $y[n] \equiv 1$ and the alternating vector $y[n] = (-1)^n$ have the same even components, and therefore they look the same after downsampling in Fig. 4.2.2 (a).

$$H(z)\widetilde{H}(z) + G(z)\widetilde{G}(z) = 2\,z^{-l}, \tag{4.2.5}$$

- Aliasing-free (AF) condition

$$H(\omega + \pi)\widetilde{H}(\omega) + G(\omega + \pi)\widetilde{G}(\omega) = 0$$

$$H(-z)\widetilde{H}(z) + G(-z)\widetilde{G}(z) = 0. \tag{4.2.6}$$

In matrix form, the two PR conditions can be written

$$\begin{bmatrix} H(z) & G(z) \\ H(-z) & G(-z) \end{bmatrix} \begin{bmatrix} \widetilde{H}(z) \\ \widetilde{G}(z) \end{bmatrix} = \begin{bmatrix} 2z^{-l} \\ 0 \end{bmatrix}. \tag{4.2.7}$$

Downsampling and upsampling

The two-fold subsampled signal $y[n] = u[2n]$ in figure 4.2.2 can be written in Fourier domain

$$Y(2\omega) = \sum_{n=-\infty}^{+\infty} u[2n]e^{-j2n\omega} = \tfrac{1}{2}\left(U(\omega) + U(\omega + \pi)\right). \tag{4.2.8}$$

The term $U(\omega + \pi)$ corresponds to a frequency folding leading to the system aliasing, which must be canceled at the reconstruction. The upsampling (zero insertion) in time domain defines

$$v[n] = s[n] = \begin{cases} s[p] & \text{if } n = 2p \\ 0 & \text{if } n = 2p + 1, \end{cases} \tag{4.2.9}$$

whose Fourier transform is

$$V(\omega) = \sum_{-\infty}^{+\infty} s[n]e^{-j2n\omega} = S(2\omega). \tag{4.2.10}$$

4.2.2 Quadrature mirror filter (QMF)

The term QMF refers to a particular choice of filters. The AF condition (4.2.6) leads to the "alternating sign" constructions. In terms of polynomials, this is

$$G(z) = \widetilde{H}(-z), \quad \widetilde{G}(z) = -H(-z) \tag{4.2.11}$$

With this choice of filters, the AF condition is automatically satisfied.[2] The remaining LSI system response is given,

$$\hat{X}(z) = \frac{1}{2} \left[H(z)\widetilde{H}(z) - H(-z)\widetilde{H}(-z) \right] X(z) = X(z)z^{-l}, \qquad (4.2.12)$$

which can be simplified using the product polynomial $P(z) = H(z)\widetilde{H}(z)$

$$P(z) - P(-z) = 2\,z^{-l}. \qquad (4.2.13)$$

The left side of (4.2.13) is an odd function, so l must be an odd integer. The equation means that the only odd terms in $P(z)$ is z^{-l} with coefficient 1, and is called *quadrature mirror condition*. Assuming that the filter lengths of $H(z)$ and $\widetilde{H}(z)$ are N_0 and N_1, respectively. One can verify that $P(z)$ is a halfband filter of the length $N_0 + N_1 - 1$.

Various designing methods and implementations of these filters are developed. For example, an even length filter, which can be designed by minimizing an error function containing a shift-invariant error term and a weighted stopband ripple term for a fixed number of filter taps, was investigated by Johnson [96]. Using a similar error criterion in the time domain, Jain [97] formulated an iterative design scheme in which each iteration requires the constrained minimization of a quadratic function.

4.2.3   Conjugate mirror filter (CMF)

The original QMFs were linear-phase and non-orthogonal. In the case of orthogonal (often called "paraunitary") filter bank, the filter lengths are equal, i.e., $N_0 = N_1 = N =$ *even integer*, and the synthesis filters are the time-inversed version of the analysis filters, i.e., $\widetilde{H}(z) = z^{-(N-1)}H(z^{-1})$. Then, the product filter $P(z) = H(z)\widetilde{H}(z)$ simplifies to $P(z) = z^{-(N-1)}H(z)H(z^{-1})$. Related to the subsampling formula (4.2.8), the orthonormality condition in z-domain is given

$$\langle\, h[n],\, h[n-2k]\,\rangle = \delta_k \iff \tfrac{1}{2}\left(P(z) + P(-z)\right) = 1 \qquad (4.2.14)$$

where $P(z) = H(z)H(z^{-1})$ is the deterministic autocorrelation of the sequence $h[n]$.

A case of great importance is when the filters have a finite impulse response (FIR) with length N which a QMF can not have. For FIR filters, there exist $a \in \mathbb{R}$ and $l \in \mathbb{Z}$ [98] such that

$$G(z) = az^{-(2l+1)}\,\widetilde{H}(-z)$$
$$\widetilde{G}(z) = a^{-1}z^{-(2l+1)}\,H(-z). \qquad (4.2.15)$$

---

[2]Note that it is the overall system aliasing that cancels, while the individual subbands contain aliasing.

The factor a is a gain which is inverse for the analysis and synthesis filters and l is a reverse shift. When we set $a = 1$ and $l = 0$, (4.2.15) becomes

$$g[n] = (-1)^{1-n} \tilde{h}[1 - n]$$
$$\tilde{g}[n] = (-1)^{1-n} h[1 - n] \qquad (4.2.16)$$

in time domain. The analysis and synthesis filter banks play a symmetric role and can be inverted.

If we impose that the analysis filter $H(z)$ is equal to the synthesis filter $\widetilde{H}(z)$, then (4.2.5) becomes

$$|H(e^{j\omega})|^2 + |H(e^{j(\omega+\pi)})|^2 = 2 \qquad (4.2.17)$$

on the unit circle, that is, the filter is power complementary [99]. (4.2.15)and (4.2.17) are conditions for obtaining perfect reconstruction orthogonal filters having a finite impulse response that are called *conjugate mirror filters* by Smith and Barnwell [85] and Mintzer [86]. Practically, they used following lowpass prototype $A(z)$ in first designs of orthogonal CMF banks with finite impulse response,

$$H(z) = A(z) \qquad\qquad \widetilde{H}(z) = A(z^{-1})z^{-(N-1)}$$
$$G(z) = A(-z^{-1})z^{-(N-1)} \qquad\qquad \widetilde{G}(z) = A(-z), \qquad (4.2.18)$$

where N must be an even integer in order to satisfy the AF condition. Apparently, if the filters are linear-phase, then the choice of filters (4.2.18) is equivalent with the case of QMF. However, for any non-trivial two-channel filter bank, linear-phase and orthogonality cannot be simultaneously imposed [100][3]. With $l = (N - 1)$, the DF condition (4.2.5) becomes

$$H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 2$$

$$\text{or} \quad P(z) + P(-z) = 2, \qquad (4.2.19)$$

which defines a halfband filter of length $2N - 1$. In order to satisfy the orthonormality condition (4.2.14), $P(z)$ must be a symmetric polynomial of the form

$$P(z) = 1 + \sum_{n=0}^{N-1} p_{2n+1}(z^{2n+1} + z^{-(2n+1)}). \qquad (4.2.20)$$

For dividing $P(z)$ into the mutually conjugated filters $H(z)$ and $H(z^{-1})$, there must exist the zeros z_0 inside the unit circle and the symmetric zeros z_0^{*-1} outside the unit circle. Therefore $P(z)$ as an autocorrelation function must satisfy

$$P(e^{j\omega}) \geq 0, \qquad (4.2.21)$$

---

[3]They can coexist for filter banks with more channels (e.g., discrete cosine transform and lapped orthogonal transform).

which assures of double zeros on the unit circle. Hence, one way to design an orthonormal FIR filter bank using CMF, which provides a systematic designing method as compared to QMF, is to find an autocorrelation sequence satisfying (4.2.14), and then to take its spectral factor, involving a choice of zeros and a possible delay factor.

4.2.4   Biorthogonal filters

By relaxing orthogonality, we can consider biorthogonal filter bank [101] [102]. In that case, the filters satisfy the PR conditions (4.2.5) and (4.2.6), but $\widetilde{H}(z)$ is not necessarily a time reversed version of $H(z)$. Thus, the product filter $P(z)$ in that case must be first considered as $P_0(z) = H(z)\widetilde{H}(z)$ instead of $P(z) = H(z)H(z^{-1})$. Namely, orthogonal filter banks obtained from CMFs are special cases of biorthogonal filter banks, and the design of the former is more restrictive than that of the latter. To design these filters, we choice the prototype of analysis and synthesis lowpass filters individually,

$$H(z) = A_1(z)z^{-K} \qquad\qquad \widetilde{H}(z) = A_2(z)z^{-(K+1)}$$
$$G(z) = A_2(-z)z^{-(K+1)} \qquad\qquad \widetilde{G}(z) = A_1(-z)z^{-K}. \qquad (4.2.22)$$

The filters are causal [4] with delay factor K and satisfy the AF condition. With total delay $l = K + 1$, DF condition becomes

$$H(z)\widetilde{H}(z) + H(-z)\widetilde{H}(-z) = 2$$

$$P_0(z) + P_0(-z) = 2, \qquad (4.2.23)$$

where $P_0(z) = H(z)\widetilde{H}(z)$. As mentioned, $H(z)$ is not necessarily a time reversed version of $\widetilde{H}(z)$.

In audio and image applications, one of the highly desirable filter characteristics is linear-phase, which corresponds to symmetry properties. A great interest of the biorthogonal solution is that construction of linear-phase filters or combination of symmetry and antisymmetry is possible, whereas it is excluded in the orthogonal case.[5] Recall that a linear-phase filter $H(z)$ of length N satisfies the self-symmetry property (compared with (4.2.16)), that is,

$$H(z) = \pm z^{-(N-1)}H(z^{-1}), \quad h[n] = \pm h(N-1-n). \qquad (4.2.24)$$

The linear-phase biorthogonal filters have become quite popular in image coding [103] [104], where a non-symmetric distortion in reconstruction is caused by using

---

[4]Causality can be ignored when the filters are FIR (or compactly supported

[5]except for the trivial Haar filters, where $\widetilde{H}(z) = 2^{-1/2}(1 + z^{-1})$

nonlinear-phase synthesis filters.

4.3 Subbands from Block Transforms

4.3.1 Discrete cosine transform

From a statistical signal processing standpoint, the DCT is a robust approximation to the optimal discrete-time Karhunen-Loeve transform (KLT) of a first-order Gauss-Markov process with a positive correlation coefficient ρ when $\rho \to 1$. The KLT is optimal in the energy compaction sense, i.e., among unitary transforms, the KLT packs signal energy into the fewest number of coefficients. However, the KLT is signal-dependent, therefore, computationally complex and expensive. DCT has proven to be a much better alternative in practice. It is signal independent and has linear phase, real coefficients, and fast algorithms. Four types of the DCT with block size N are commonly known

DCT-I:

$$X(k) = c(k)\sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} c(n)x(n) \cos\left(\frac{nk\pi}{N}\right) \tag{4.3.1}$$

$$x(n) = c(n)\sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} c(k)X(k) \cos\left(\frac{nk\pi}{N}\right) \tag{4.3.2}$$

DCT-II:

$$X(k) = c(k)\sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)k\pi}{2N}\right] \tag{4.3.3}$$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} c(k)X(k) \cos\left[\frac{(2n+1)k\pi}{2N}\right] \tag{4.3.4}$$

DCT-III:

$$X(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} c(n)x(n) \cos\left[\frac{n(2k+1)\pi}{2N}\right] \tag{4.3.5}$$

$$x(n) = c(n)\sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} X(k) \cos\left[\frac{n(2k+1)k\pi}{2N}\right] \tag{4.3.6}$$

DCT-IV:

$$X(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)(2k+1)\pi}{4N}\right] \tag{4.3.7}$$

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} X(k) \cos\left[\frac{(2n+1)(2k+1)\pi}{4N}\right] \tag{4.3.8}$$

where

$$c(k) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if} \quad k = 0 \\ 1, & \text{otherwise} \end{cases}$$

and $n, k = 0, 1, 2, \cdots, N-1$. All of these forms of the DCT follow from the generalized discrete Fourier transform defined as

$$X_g(k) = \sum_{n=0}^{N-1} x(k) W_N^{(k+k_0)(n+n_0)} \tag{4.3.9}$$

and

$$x(k) = \frac{1}{N} \sum_{k=0}^{N-1} X_g(k) W_N^{(k+k_0)(n+n_0)} \tag{4.3.10}$$

where

$$W_N = e^{2\pi j/N} \ . \tag{4.3.11}$$

Note that all of the DCT's basis functions have linear phase, and DCT places heavy emphasis on the lowpass frequency spectrum. Therefore, it provides superior energy compaction over other classical transforms such as FFT. The DCT-II is used in image and audio compression and has compaction properties very close to the KLT. The popular DCT in JPEG and MPEG is the 8-point DCT-II. When N is a power of 2, many efficient algorithms for computing the DCT are known [105], such as the sparse matrix factorizations. One factorization is even partly recursive, i.e., an N-point DCT-II can be implemented via an $N/2$-point DCT-II and an $N/2$-point DCT-IV,

$$\boldsymbol{C}_N^{II} = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{C}_{N/2}^{II} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{N/2}^{IV}\boldsymbol{J} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{J} \\ \boldsymbol{J} & -\boldsymbol{I} \end{bmatrix} , \tag{4.3.12}$$

where the symbols $\boldsymbol{C}_N^{II}, \boldsymbol{C}_N^{IV}, \boldsymbol{I}, \boldsymbol{J}$ denote, respectively, the $N \times N$ DCT-II matrix, DCT-IV matrix, the identity matrix, and the reversal (antidiagonal) matrix.

In practice, the computing a DCT on non-overlapping blocks is equivalent to convolving the signal with each of the block DCT basis functions and then subsampling by a factor equal to the block spacing. The Fourier transform of the basis functions indicates that each of the DCT functions is selective for a particular frequency subband,

although it is clear that the subband localization is rather poor. Thus, the DCT also qualifies as a subband transform.

4.3.2   Lapped orthogonal transform (LOT)

The block DCT is usually not computed globally, but is applied independently to non-overlapping sub-blocks of the signal. Therefore the resulting block DCT basis functions constitute a subband transform, but the subbands are not very well localized. Considered in the framework of the filtering system, the subsampled subband signal will contain severe amounts of aliasing. Since the transform is invertible (orthogonal), it should be clear that this subband aliasing is canceled in the synthesis stage. However, if the transform coefficients are quantized or discarded (e.g. in a coding system), the aliasing no longer cancels, and the errors appear as block edge artifacts in the reconstructed signal, called *blocking effect*. For images, this appears as a tiling effect, while for audio signals, periodic clicking is heard.

These problems motivate the development of an overlapped blocking scheme, resulting in the lapped orthogonal transform (LOT) popularized by Malvar [106] [107]. An M-band LOT is a linear transform that partitions the input signal into small overlapped blocks and then processes each block independently. In the resulting transform, the basis functions from adjacent blocks overlap each other, and their impulse responses are tapered at the edges.
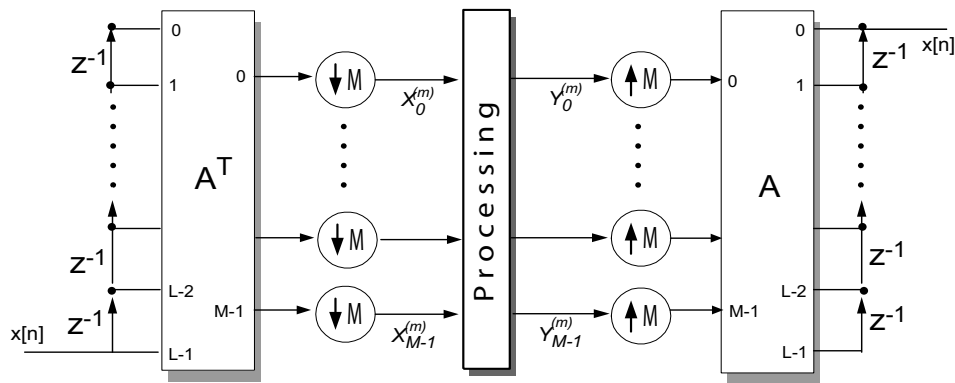
Figure 4.3.1: General structure of a lapped transform coding

The input signal $\mathbf{x}$ can be blocked into short sequences $\mathbf{x}_m$ of length L. The corresponding transform vector $\mathbf{X}_m$ of length N is obtained from $N \times L$ transform matrix $\boldsymbol{A}$

as $\mathbf{X}_m = \boldsymbol{A}^T \mathbf{x}_m$, which is in time domain

$$X(k) = \sum_{0}^{L-1} x(n)a_{n,k} \,. \tag{4.3.13}$$

For the LOT, we take L samples for each block, with $N < L \le 2N$, so that each block of input samples has an overlap of $(L - N)$ samples with each of its adjacent neighbors, $\mathbf{x}_{m-1}$ and $\mathbf{x}_{m+1}$. From now, we consider an uniform filterbank, so that the number of blocks (subbands) M is same with the size of output transform coefficients in each block N, i.e., $M = N$. Then this overlapped transform can be defined with $\boldsymbol{H}$ given by

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{D}_1 & & & & 0 \\ & \boldsymbol{D}_0 & & & \\ & & \ddots & & \\ & & & \boldsymbol{D}_0 & \\ 0 & & & & \boldsymbol{D}_2 \end{bmatrix} \tag{4.3.14}$$

where $\boldsymbol{D}_0$ is an $L \times M$ matrix that contains the LOT basis functions for each block. The matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are introduced because the first and last blocks of a segment have only one neighboring block, and thus the LOT for the first and last blocks must be defined in a slightly different way, to guarantee that none of the basis functions extends beyond the segment boundaries. Figure 4.3.1 shows a general structure of lapped transform coding, which is a family of perfect reconstruction QMF filter banks. The system is critically sampled, so that the outputs of the analysis filter bank are decimated by a factor M. Since the basis functions overlap the adjacent blocks, it is referred to as *lapped transform* and can be viewed as block transforms because of their relatively short lengths.

Each LOT basis function must be orthogonal not only to the other functions in same block but also to the functions in the two adjacent blocks. Therefore the following conditions must be hold,

$$\boldsymbol{D}^T \boldsymbol{D} = \boldsymbol{I} \,, \quad \text{and} \quad \boldsymbol{D}^T \boldsymbol{W} \boldsymbol{D} = \boldsymbol{0} \,, \tag{4.3.15}$$

where

$$\boldsymbol{W} \equiv \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{4.3.16}$$

It was shown that the orthogonality conditions were satisfied by the following condition:

$$\boldsymbol{D} = \frac{1}{2} \begin{bmatrix} \boldsymbol{C}_E^{II} - \boldsymbol{C}_O^{II} & \boldsymbol{C}_E^{II} - \boldsymbol{C}_O^{II} \\ \boldsymbol{J}(\boldsymbol{C}_E^{II} - \boldsymbol{C}_O^{II}) & -\boldsymbol{J}(\boldsymbol{C}_E^{II} - \boldsymbol{C}_O^{II}) \end{bmatrix} \boldsymbol{Z} \,, \tag{4.3.17}$$

where $\boldsymbol{Z}$ is an orthogonal matrix of order M, and $\boldsymbol{C}_E^{II}$ and $\boldsymbol{C}_O^{II}$ are the $M \times M/2$ matrices

containing the even and odd DCT-II, respectively, i.e., from Eqs. (4.3.3),

$$|\boldsymbol{C}_E^{II}|_{n,k} = c(k)\sqrt{\frac{2}{M}}\cos\left[\frac{\pi}{M}2k\left(n+\frac{1}{2}\right)\right], \tag{4.3.18}$$

$$|\boldsymbol{C}_O^{II}|_{n,k} = \sqrt{\frac{2}{M}}\cos\left[\frac{\pi}{M}(2k+1)\left(n+\frac{1}{2}\right)\right], \tag{4.3.19}$$

for $n = 0, 1, \ldots, M-1$, $k = 0, 1, \ldots, M/2-1$, where

$$c(k) = \begin{cases} \frac{1}{\sqrt{2}}, & k = 0 \\ 1, & \text{otherwise}. \end{cases}$$

There are several methods to obtain the matrix $\boldsymbol{Z}$. The optimal $\boldsymbol{Z}$ can be closely approximated by

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{Z}} \end{bmatrix} \tag{4.3.20}$$

where $\widetilde{\boldsymbol{Z}}$ is a cascade of $M/2-1$ plane rotations

$$\widetilde{\boldsymbol{Z}} = \boldsymbol{T}_1\boldsymbol{T}_2\cdots\boldsymbol{T}_{M/2-1} \tag{4.3.21}$$

where each plane rotation is defined by

$$\boldsymbol{T}_i = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}(\theta_i) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{4.3.22}$$

The matrix $\boldsymbol{Y}(\theta_i)$ is a 2×2 butterfly

$$\boldsymbol{Y}(\theta_i) = \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \tag{4.3.23}$$

With $\boldsymbol{Z}$ as in (4.3.20), the first $M/2$ columns of $\boldsymbol{D}$ will have even symmetry, and the last $M/2$ columns will have odd symmetry.

As pointed out in [107], the $M \times L$ lapped transform is simply the polyphase implementation of a maximally decimated M-channel L-tap filter banks. The precise relationship between the $M \times L$ lapped transform matrix $\boldsymbol{A} = [\boldsymbol{A}_0\boldsymbol{A}_1\cdots\boldsymbol{A}_{K-1}]$ and the polyphase matrix $\boldsymbol{P}$ is

$$\boldsymbol{P}(z) = \sum_{i=0}^{K-1} \boldsymbol{A}_i z^i \tag{4.3.24}$$

From the components with discrete sine transform (DST) and DCT, fast lapped transforms can be constructed in polyphase form, e.g., the type-II fast LOT is given (see Fig. 4.3.2),

$$\boldsymbol{P}(z) = \frac{1}{2} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{M/2}^{IV} \boldsymbol{C}_{M/2}^{II^T} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & z\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix} \boldsymbol{C}_M^{II} \boldsymbol{J}_M \qquad (4.3.25)$$

where $\boldsymbol{S}^{IV}$ denotes type-IV discrete sine transform (DST),

$$\boldsymbol{S}^{IV} = \sqrt{\frac{2}{M}} \sin \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right] . \qquad (4.3.26)$$
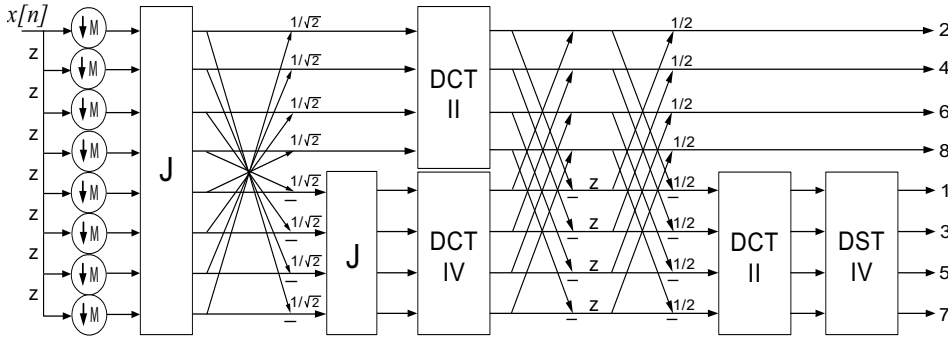
Figure 4.3.2: Fast lapped orthogonal transform.

There are many other fast solutions, and all of them involve replacing the product $\boldsymbol{S}_{M/2}^{IV} \boldsymbol{C}_{M/2}^{II^T}$ by different matrices $\boldsymbol{V}$, usually cascades of various 2×2 matrices along the diagonal axis. If a larger overlapping percentage is desired, more modules $\boldsymbol{U}_i(z)$ with different $\boldsymbol{V}_i$ can be added, i.e., $\boldsymbol{P}(z) = \boldsymbol{U}_{K-1}(z) \boldsymbol{U}_{K-2}(z) \cdots \boldsymbol{U}_1(z) \boldsymbol{C}_M^{II}$. This is known as the generalized LT [108].

4.3.3   Modulated lapped transform (MLT)

Another way to arrive at a good set of basis functions for a lapped transform is to define an appropriate lowpass filter prototype in a modulated filter bank structure. If the length of the lowpass prototype is chosen to be equal to $2M$, it is possible to achieve not only aliasing cancellation, but also perfect reconstruction with identical analysis and synthesis filters, as noted in [109]. In this case, a window of $2M$ samples from two consecutive

blocks undergoes a cosine transform, called modulated lapped transform (MLT). The MLT is defined as

$$X(k) = \sum_{n=0}^{2M-1} x(n)a_{n,k} \qquad (4.3.27)$$

for $k = 0, 1, \cdots, M - 1$, where

$$a_{n,k} = h(n)\sqrt{\frac{2}{M}} \cos\left[\frac{(2n + M + 1)(2k + 1)\pi}{4M}\right] \qquad (4.3.28)$$

and $h(n)$ is a lowpass half-band filter. The factor $\sqrt{2/M}$ is necessary to generate an orthogonal transform implementation. Note that only the first M transform coefficients are returned from $2M$ samples. The window is then shifted by M samples and the next set of M transform coefficients is computed. Thus, each window overlaps the last M samples of the previous window. This overlap ensures the continuity of the reconstructed samples despite the alteration of transform coefficients due to quantization.

The relationship of the MLT (4.3.28) to the DCT-IV (4.3.7) can be seen if we express the coefficients, $a_{n,k}$, in the following manner:

$$a_{n,k} = h(n)\sqrt{\frac{2}{M}} \cos\left[\frac{(2n + 1)(2k + 1)\pi}{4M} + (2k + 1)\frac{\pi}{2}\right] \qquad (4.3.29)$$

The phase shift $(2k + 1)(\pi/2)$ in the basis functions is important for ensuring aliasing cancellation and perfect reconstruction.

To recover $x(n)$, one requires not only $X(k)$ for the current block but also the previous block $X^P(k)$. Then

$$x(n) \simeq \sum_{k=0}^{M-1} \left[X(k)a_{n,k} + X^P(k)a_{n+M,k}\right]. \qquad (4.3.30)$$

The approximation symbol in Eq. (4.3.30) means that there is no guarantee that $x(n)$ can be recovered exactly unless $h(n)$ satisfies certain conditions, i.e.,

$$h(2M - 1 - n) = h(n), \quad \text{and} \quad h^2(n) + h^2(n + M) = c, \qquad (4.3.31)$$

where c is any constant. In order to keep orthogonality of the P matrix, we must choose $c = 1$. One possible choice for $h(n)$ which guarantees perfect reconstruction was given by

$$h(n) = \sin\left[\left(n + \frac{1}{2}\right)\left(\frac{\pi}{2M}\right)\right] \qquad (4.3.32)$$

where $0 \le n \le 2M$ and $h(n) = 0$, otherwise.

CHAPTER 4. FILTER BANK AND SUBBAND TRANSFORM 85

As the case of the fast LOT, MLT can be efficiently implemented by means of a fast transform of length M [107], i.e.,

$$X_k(m) = \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} y_m(n) \sin\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n+\frac{1}{2}\right)\right], \qquad (4.3.33)$$

where

$$y_m(n) = \begin{cases} x_m(n+M/2)h(n+M/2) - x_m(M/2-n-1)h(M/2-n-1) \\ \quad \text{for} \quad n = 0, \cdots, M/2-1 \\ x_m(n+M/2)h(n+M/2) + x_m(5M/2-n-1)h(5M/2-n-1) \\ \quad \text{for} \quad n = M/2, \cdots, M-1 \,. \end{cases}$$

$$(4.3.34)$$

Note that Eq. (4.3.33) can be obtained as the orthogonal DST-IV of $y_m(n)$.

MLT was extensively used in audio compression schemes (MPEG 1, MPEG 2, and AC-3) to transform the time representation, more suitable for perceptually masking and quantization. The MLT can be described from various viewpoints: filterbank theory, wavelet packets, and orthonormal transformations. All of these approaches are useful. In particular, the filterbank description leads to the set of conditions to be satisfied by the window function $h(n)$ for perfect reconstruction. The other viewpoints lead to a manner for implementing time varying filters.

4.3.4 Coding efficiency of the block transforms

To compare the transform efficiency, the coding gain G_{TC} can be employed which is equivalent to maximizing the energy compaction measure. The coding gain is defined by

$$G_{TC} = \frac{\frac{1}{M}\sum_{i=1}^{M}\sigma_i^2}{\left(\prod_{i=1}^{M}\sigma_i^2\right)^{1/M}}, \qquad (4.3.35)$$

where σ_i^2 is the variance of the output of the ith transform coefficient and also the ith diagonal entry of the matrix

$$\boldsymbol{R}_0 = \boldsymbol{P}^T \boldsymbol{R}_{xx} \boldsymbol{P}, \qquad (4.3.36)$$

where R_{xx} is the covariance matrix of a block of $2M$ samples of signal $x[n]$. The importance is of the G_{TC} measure is that it indicates the factor by which the mean-square reconstruction error is reduced, compared to direct quantization of the signal. The relative

efficiency of transforms is generally evaluated for a first-order autoregressive process, for which

$$[\boldsymbol{R_{xx}}]_{i,j} = \rho^{|i-j|} \,, \tag{4.3.37}$$

where ρ is the intersample correlation coefficient. The empirical comparison for energy compaction performance of several LOTs is presented in [110].

4.4 Linear Prediction from Subband Signals

Subband decomposition of a signal is shown to result in a whitening of the composite subband spectrum. This implies that, for any stationary source, a pth order prediction filter can be found that is better than the pth order prediction filter obtained by solving the Yule-Walker equations resulting from the fullband signal. It has been shown [111] [112] that when $p \to \infty$, there is no gain to be obtained by coding subband samples. In practice, however, when the order of prediction is finite, considerable gains have been obtained from using the linear prediction in subbands over in the fullband. Moreover, it is shown that the prediction error variance of the fullband always exceeds the total prediction error variance of the combined subbands, for a given prediction order p. This result is instrumental in proving that pth order entropy of the combined subbands is closer to the entropy rate of the source than the pth order entropy of the fullband, for any finite p, where p is the size of the block of source samples. In this section, we provide the theoretical approach to prove the gain of the prediction in subband that have so far been less focused.

4.4.1 pth order entropies

For a given Gaussian source x, the pth order per letter differential entropy of a block of p source samples is given by

$$H_p(x) = \frac{1}{2} \ln 2\pi e |\boldsymbol{\Phi}_p|^{\frac{1}{p}} \,, \tag{4.4.1}$$

where $|\boldsymbol{\Phi}_p|$ is the pth order covariance matrix of the source. For $p = 1$, we get the first-order entropy and for $p \to \infty$, we arrive at the entropy rate of the source. the entropy rate of the Gaussian source is defined as

$$H_\infty(x) = \frac{1}{2} \ln 2\pi e Q_x \,, \tag{4.4.2}$$

where Q_x is the source entropy power,

$$Q_x = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln P_{xx}(\omega)d\omega\right\} . \tag{4.4.3}$$

In Eq. (4.4.3) P_{xx} denotes the power spectral density (PSD). Similarly, the pth order per letter differential entropy for the mth subband is

$$H_p^m(x_m) = \frac{1}{2}\ln 2\pi e|\mathbf{\Phi}_p^m|^{\frac{1}{p}} . \tag{4.4.4}$$

Hence the entropy rate of the mth subband can be written as

$$H_\infty^m(x_m) = \frac{1}{2}\ln 2\pi e Q_{x_m}^m . \tag{4.4.5}$$

Since the subbands are subsampled by a factor of L_m, the contributions of their per-letter entropies to the combined per-letter entropy must each be reduced by this factor. Therefore the combined per-letter pth order subband entropy is given by

$$H_p(x_m) = \sum_{m=1}^{M}\frac{1}{L_m}H_p^m(x_m) , \tag{4.4.6}$$

for all non-negative integers p.

The spectral flatness measure (SFM) ξ_x of the source is defined as the following ratio

$$\xi_x = \frac{\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln P_{xx}(\omega)d\omega\right\}}{\frac{1}{2\pi}\int_{-\pi}^{\pi}P_{xx}(\omega)d\omega} = \frac{Q_x}{\sigma_x^2} . \tag{4.4.7}$$

The SFM is a real number between 0 and 1. When the source spectrum is white, the SFM is 1 and when the source spectrum is maximally correlated, the SFM is 0. For the SFM to be 0, the integral $\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln P_{xx}(\omega)d\omega$ must be $-\infty$. From the Paley-Wiener criterion, this means that the source is perfectly predictable. The inverse of the SFM is a measure of the predictability of the source. The distortion-rate function for the Gaussian source x is equal to $D(R) = 2^{2R}\xi_x\sigma_x^2$, for mean squared error $D(R)$ and $D(R) \leq \min P_{xx}(\omega)$. Thus smaller values of the SFM mean lesser entropy, more predictability and a smaller theoretical mean-square distortion D for a given rate R bits/sample.

The difference between the entropy rate and the pth order entropy is the amount of information per letter that the rest of the sequence contains about the first p samples, i.e.,

$$H_p(x) - H_\infty(x) = \lim_{N\to\infty}\left[\frac{1}{p}I(x_1, x_2, \cdots, x_p; x_{p+1}, \cdots, x_{p+N})\right] . \tag{4.4.8}$$

This difference, which we call the pth order memory, overbounds the rate difference between coding p-tuples of the source and optimally coding the source and, for small distortions, equals this rate difference. In terms of the rate-distortion function, the pth order memory of the source can be written as

$$H_p(x) - H_\infty(x) = R_p^x(D) - R_\infty^x(D), \tag{4.4.9}$$

where $R_p^x(D)$ is the pth order rate distortion function for the source x. The pth order memory of the combined subbands can similarly be expressed in terms of the rate-distortion function as

$$H_p(x_m) - H_\infty(x_m) = R_p^{x_m}(D) - R_\infty^{x_m}(D). \tag{4.4.10}$$

If this rate difference is smaller for subbands than for a fullband source, then there is rate advantage in coding subband p-tuples as opposed to encoding fullband p-tuples.

4.4.2   Prediction error variance in subbands

The kth correlation of the source, $r_{xx}(k)$, is given by

$$r_{xx}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{xx}(\omega) e^{j\omega k} d\omega \tag{4.4.11}$$

$$= \frac{1}{2\pi} \sum_{m=1}^{M} \int_{I_m} P_{xx}(\omega) e^{j\omega k} d\omega$$

At the output of the filter, the kth correlation is given by

$$r_{xx}^m(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H^m(\omega)|^2 P_{xx}(\omega) e^{j\omega k} d\omega \tag{4.4.12}$$

$$= \frac{1}{2\pi} \int_{I_m} P_{xx}(\omega) e^{j\omega k} d\omega,$$

where $H^m(\omega)$ denotes the ideal bandpass filters. From Eqs. (4.4.11) and (4.4.12) it is clear that

$$r_{xx}(k) = \sum_{m=1}^{M} r_{xx}^m(k), \quad \text{and} \quad \mathbf{\Phi}_p = \sum_{m=1}^{M} \mathbf{\Phi}_p^m, \tag{4.4.13}$$

where $\mathbf{\Phi}_p$ and $\mathbf{\Phi}_p^m$ are the $p \times p$ autocorrelation matrices of the source and bandpass filtered source, respectively.

The prediction error spectrum before subsampling is

$$P_{ee}^m(\omega) = |\tilde{A}(\omega)|^2 P_{xx}^m(\omega), \tag{4.4.14}$$

where $\tilde{A}(\omega) = 1 + \sum_{k=1}^{p} a(k)e^{-j\omega k}$,i.e., optimal subband prediction filter. The variance of the prediction error is then

$$\rho_p^m = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{ee}^m(\omega)d\omega \qquad (4.4.15)$$

After subsampling by L_m it is

$$P_{ee}^{(m)}(\omega) = \frac{1}{L_m}P_{ee}^m\left(\frac{\omega}{L_m} + \omega_m \mathrm{sgn}(\omega)\right). \qquad (4.4.16)$$

Integration over $P_{ee}^{(m)}(\omega)$ yields the prediction error variance after subsampling,

$$\begin{aligned}
\rho_p^{(m)} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{ee}^{(m)}(\omega)d\omega \qquad\qquad (4.4.17)\\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{L_m}P_{ee}^m\left(\frac{\omega}{L_m} + w_m\mathrm{sgn}(\omega)\right)d\omega\\
&= \frac{1}{2\pi} \int_{I_m} P_{ee}^m(\omega)d\omega\\
&= \rho_p^m
\end{aligned}$$

Eq. (4.4.17) shows that the prediction error variance before and after subsampling are equal, as claimed. One should notice that the subsampled error process is not a linear combination of the subsampled subband elements $x^{(m)}(n)$, but is comprised of elements which are linear combinations of subband elements $x^m(n)$, some of which are no longer present in $x^{(m)}(n)$ after subsampling. However, it is evident that since these latter elements can be reconstructed with zero mean-squared error from linear combinations of subband elements. When one first forms the prediction form the decimated subband samples, the variance of the error also remains the same after upsampling and ideal bandpass filtering with gain factor L_m in reconstruction of the bandpass subband process.

With the optimal subband prediction filters $\tilde{A}^{(m)}(\omega)$ and the optimal prediction filter for the fullband $\tilde{A}(\omega)$, the minimum prediction error can be written as

$$\begin{aligned}
\rho_p &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{A}(\omega)|^2 P_{xx}(\omega)d\omega \qquad\qquad (4.4.18)\\
&\geq \frac{1}{2\pi} \sum_{m=1}^{M} \int_{-\pi}^{\pi} |\tilde{A}^{(m)}(\omega)|^2 P_{xx}^{(m)}(\omega)d\omega\\
&= \sum_{m=1}^{M} \rho_p^{(m)}
\end{aligned}$$

It is also proven in [113] that the prediction error spectrum obtained by combining the individual subband prediction error spectra is flatter than the fullband prediction error spectrum, i.e.,

$$\xi_e^{SB} \geq \xi_e^{FB} \ . \tag{4.4.19}$$

This inequality in Eqs. (4.4.18) and (4.4.19) states that linear prediction from subbands is superior to linear prediction from the fullband.

Since each subband has a different SFM, one would intuitively choose a lower order of prediction for the flatter subbands, and conversely. Some algorithms to find the optimal assignment of orders p_i, $\sum_{i=1}^{M} p_i = p$, are proposed in [114] such that the composite SFM of the error spectrum is closed to 1.

4.4.3   Empirical comparison of prediction from Subbands

Audio Materials	Fullband LP Entropy ($p = 12$)	Subband LP Entropy ($\sum p_i = 12$)	Compressed Bit rates
Violin	11.56	10.92	10.01
Flute	11.39	10.84	9.89
Speech	11.20	10.45	9.67
Piano	12.26	11.90	11.02
Classic	11.48	10.53	9.74
Pop Abba	11.27	10.45	9.49
Country	11.75	11.08	10.29
Metal Rock	12.49	11.92	11.22
Soft Jazz	11.38	10.64	9.81
Average	11.64	10.97	10.12

Table 4.1: Test results for comparing the prediction error entropy from fullband and subbands

The performance of subband prediction versus fullband prediction for lossless audio compression applications is tested by entropy comparison in Table 4.1. For the experiment, the fast LOT (see Figure 4.3.2) with $L = 2M$ is chosen to converse the audio signal from time to a subband representation with $M = 4$.

By exploiting the time correlation of samples in each subband, we use the FIR linear predictor used for the prediction experiment in Chapter 3. The prediction value in mth

subband is computed as

$$\hat{X}_m(k) = \sum_{j=0}^{p_i-1} A_j^m X_m(k-j) \tag{4.4.20}$$

We have chosen the prediction order $p = \sum_{i=1}^{4} p_i = 12$. A different LP filter order was used in each subband, i.e., $p_1 = 2$, $p_2 = 3$, $p_3 = 3$, $p_4 = 4$. The last column in the Table 4.1 As we have shown in Chapter 3, the order of LP filter is set as a trade off. In practice, the prediction order over $p = 12$ does not provide a proportional improvement of compression ratio. Therefore we have chosen the order $p = 12 = p_{SB1} + p_{SB2} + p_{SB3} + p_{SB4}$.

Chapter 5

Wavelet Transform and Filter Bank

Transform coding systems are usually based on orthogonal linear transforms. The classic example of such a transform is the discrete Fourier transform (DFT), the discrete cosine transform (DCT) and the Karhunen-Loeve transform (KLT). Other choice for the transform is the wavelet transform (WT) that has achieved significantly improved performance over other transforms and represents the current state-of-the-art in image compression and audio compression as well. In addition to the compression performance, encoding of wavelet coefficients provides more interesting editibility, since the decomposition of the signal into a time-frequency representation allows to cope separately with different audio features and events.

From the signal processing viewpoint, it is clear that all block transforms and all subband transforms including wavelet transform can be defined and implemented by the filter design and windowing theory, as we mentioned in previous chapter. It is also the fact that the terms in the literature, e.g., subband coding, filter bank method, and multiresolution transform, that are used to eventually mean the wavelet transform, are defined ambiguously and used heedlessly. However, the wavelet transform, that is one of the most important new phenomenons in signal processing area since the Fourier transform has long ruled over the area, has obviously different history and characteristics. Hence, we study the wavelet transform in depth by beginning to review the time-frequency representation and describe the theoretical relationship between the wavelet and the multirate filter banks.

5.1   Introduction

The most important and fundamental variables in nature of signals are *time* and *frequency*. While the time domain function indicates how a signal's amplitude changes over time, the frequency domain function tells how often such changes take place. The bridge between time and frequency is the *Fourier transform*, which has long ruled over signal processing. Fourier series separate the "whole" signal into pure harmonics and are ideal for analyzing periodic signals. However, since the periodic sinusoidal functions as a basis function of Fourier transform are used to expand non-periodic signals and stretch out to infinite time, the Fourier integral transform is a far less natural tool for the real signals, especially for approximating a transient signal. This means that Fourier expansion has only frequency resolution, no spatial resolution, and is not associated with a particular instant in space. As a result, we might be able to determine all the frequencies present in a signal but do not know when, or where they are present.

To overcome this problem, in the past decades several solutions have been developed, such as short time Fourier transform (STFT), Gabor expansion, Wigner-Ville distribution (WVD) and wavelet analysis, which are able to represent a signal in the time and frequency domain at the same time, called time-frequency joint representation (TFR). The wavelet analysis or wavelet transform is most recent solution for efficient TFR. Not only for signal analysis and signal synthesis, the multi-resolution properties of wavelet transform can be effectively used for data compression with several advantages over the other traditional transforms for coding system, e.g., discrete cosine transform (DCT), lapped orthogonal transform (LOT) and Karhunen-Loeve transform (KLT).

5.2   Time-Frequency Distribution

5.2.1   Linear expansions of signals

In signal processing and numerical analysis, a preliminary task is to find an efficient representation of the signal that is particularly suitable for a problem. Signal expansions are central to such representation. It is methods of decomposing a signal as a linear combination of elementary "atoms" or building blocks. That is, we write a signal f as a linear combination

$$f = \sum_i \alpha_i \, \psi_i \,, \tag{5.2.1}$$

where the set of elementary signals $\{\psi_i\}_{i \in \mathbb{Z}}$ is complete for a space $\mathcal{S}$, if all signals $f \in \mathcal{S}$ can be expanded as in (5.2.1). Space $\mathcal{S}$ can be finite-dimensional (for example, $\mathcal{R}^n$, $\mathcal{C}^n$)

or infinite-dimensional (for example, $l^2(\mathbb{Z})$, $L^2(\mathbb{R})$). In general we will be concerned with signals that belong to the space of square integrable functions ($L^2(\mathbb{R})$) or the space of square summable sequences ($l^2(\mathbb{Z})$). In representation (5.2.1) that is conceptually a change of point of view, the set is actually a *basis*, if the vectors ψ_i are linearly independent. In that case, there exists a dual set $\{\tilde{\psi}_i\}_{i \in \mathbb{Z}}$ such that the expansion coefficients can be computed as follows,

$$\alpha_i = \sum_n \tilde{\psi}_i[n] \, f[n] \,, \tag{5.2.2}$$

where it is assumed that f and $\tilde{\psi}$ are sequences in $l^2(\mathbb{Z})$. If the set $\{\psi_i\}_{i \in \mathbb{Z}}$ is orthonormal and complete, then we have an *orthonormal basis* for $\mathcal{S}$, and the basis and its dual are the same, i.e. $\psi_i = \tilde{\psi}_i$. Then

$$\langle \psi_i, \psi_j \rangle = \delta_{i,j} \,, \tag{5.2.3}$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product and $\delta_{i,j}$, called *Kronecker's symbol*, is defined as 1 if $i = j$ and 0 otherwise. In such case of orthonormal set, each member of the set is orthogonal to every other member of the set and is also normalized. This is a great interest and utility both from theoretical and applied viewpoints.
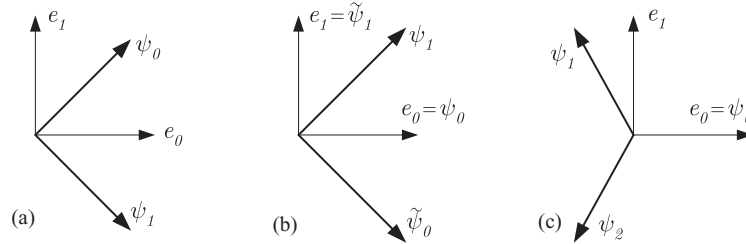
Figure 5.2.1: The sets of vectors for the expansion. The standard Euclidean basis is given by e_0 and e_1. (a) orthonormal basis. (b) biorthogonal basis. (c) overcomplete set.

The equation (5.2.1) becomes

$$f = \sum_i \langle \psi_i, f \rangle \, \psi_i \,, \tag{5.2.4}$$

where the convergence is in the square norm sense. If the set is complete and the vectors ψ_i are linearly independent but not orthonormal, then we have a *biorthogonal basis*, and the basis and its dual satisfy

$$\langle \tilde{\psi}_i, \psi_j \rangle = \delta_{i,j} \,, \tag{5.2.5}$$

and (5.2.1) becomes

$$f = \sum_i \langle \tilde{\psi}, f \rangle \, \psi_i = \sum_i \langle \psi, f \rangle \, \tilde{\psi}_i \,. \tag{5.2.6}$$

If the set is complete but the vectors are not linearly independent (i.e., *redundant*), then we do not have a basis but an overcomplete representation called a *frame*. In that case, this set behaves as an orthonormal basis, even though the vectors are linearly dependent (see section 5.5.2). Figure 5.2.1 illustrates the possible sets of vectors for the expansion.

For analytical and computational reasons, we will be interested only in structured bases, that is, the basis functions that are obtained through a small set of operations from a prototype function. These operations typically include the shifting in time, the scaling and the modulation which is shifting in frequency. In this case, the atoms or the basis vectors used to write the function f are closely related to each other.

Historically, the Fourier series for periodic signals is the first example of a structured signal expansion. Despite its obvious limitation to periodic signals, the Fourier transform has useful properties, such as the convolution property which comes from the fact that the basis functions are eigenfunctions of linear time-invariant systems. For non-periodic signals, a smooth window function is used to segment the signal and a piecewise Fourier expansion is applied to each segment of the signal. The Gabor expansion and short time Fourier transform are well-known examples of such local Fourier expansion. This localization, that is a key concept in signal analysis, leads to a time-frequency representation.

5.2.2   Time-frequency representation

The need for a time-frequency joint representation stemmed from the inadequacy of either time domain or frequency domain analysis to fully describe the nature of non-stationary signals. A time-frequency distribution of a signal provides information about how the spectral content of the signal evolves with time, thus providing an ideal tool to dissect, analyze and interpret non-stationary signals. This is performed by mapping a one dimensional signal in time domain into a two dimensional time-frequency plane so as to illuminate two important properties, localization in time of transient phenomena and presence of specific frequencies.

The fundamental idea behind Fourier's original work was to decompose a signal as the sum of weighted sinusoidal functions. These sinusoids are very well localized in the frequency domain, but not in time or space, since their support has an infinite length. Indeed, the Fourier transform is defined through an integral which covers the whole time domain, and does therefore not provide any information about the time dependency of the signal. This is problematic when analyzing signals of a non-stationary nature where it is often beneficial to be able to acquire a correlation between the time and frequency domains of a signal. Another deficiency of Fourier transform, because of its assumption that

the signal to be transformed is periodic in nature, is its inability to accurately represent a non-periodic signal without any specific parity properties, such as transient impulses.

In order to overcome these shortcomings of Fourier transform, in the past decades several solutions have been devised, most notably the short time Fourier transform (STFT), Wigner-Ville distribution and the wavelet transform. They provide a representation intermediating between a spatial and a Fourier representation. To represent the frequency behavior of a signal locally in time, the central task is to cut the signal of interest into several pieces, called *time-frequency atoms* which are localized both in time and frequency, and draw idealized representations of these atoms in the time-frequency plane and then analyze them separately. Note that a main idea behind the time-frequency analysis is thus to express the signal as a linear combination of the time-frequency atoms that should be functions constrained by the definition, to be zero outside of a small interval, formally called "*compact support*".

One of main purposes of the time-frequency representation is to explore *instantaneous* frequencies in signal. Let $f(t) = A(t)e^{j\phi(t)}$, where both $A(t)$ and $\phi(t)$ are real-valued time functions. Then, the first derivate of the phase $\phi(t)$ represents the weighted average of the momentary frequency. This $\phi^{'}(x)$ is traditionally called *instantaneous frequency*, which is actually incorrect for several reasons. For instance, $\phi^{'}(x)$ is a single-valued function, whereas at each time instant, generally there are multiple frequencies. It is caused also by the fact that the time and frequency measurements are limited by uncertainty relation in which Heisenberg's inequality[1](see Appendix) prevents us from making the product of the uncertainties smaller than a fixed constant. If a nonzero function f is small outside a time-interval of length T and its Fourier transform is small outside a frequency band of width Ω, then an inequality of the type $\Omega T \geq c$ must hold for some positive constant $c \sim 1$. The precise value of c depends on how the width T and Ω of the signal in time and frequency are measured.

5.2.3   Approaches to time-frequency analysis

The algorithms for TFR often fall into two categories, linear and quadratic. The first uses the inner product and the expansion, like the classical Fourier analysis, while the last uses the time-dependent spectrum of signal whose power spectra change with time. The Wigner-Ville distribution is one of the quadratic analysis. In addition, as Jean Ville(1947) proposed, there are two possible types of time-frequency analysis. The first approach is to appropriately cut the signal into slices in time and then pass these different slices through a system of filters to examine their frequency content. The other approach is to

---

[1] *"The more precisely the position is determined, the less precisely the momentum is known in this instant, and vice versa"* -Heisenberg, uncertainty paper, 1927- see Appendix

first filter different frequency bands and then carve up these bands individually to study their energy variations. The first approach leads us to construction of short time Fourier transform and Gabor expansion, and second to wavelet and wavelet packet transform, which leads to focus of this chapter. We briefly review these approaches:

Short Time Fourier Transform (STFT)
The most intuitive way to resolve the obstacle of Fourier transform is to localize the sinusoids in Fourier representation by windowing. Instead of processing the entire signal at once, STFT takes the Fourier transform on a block-by-block basis. Therefore, the resulting Fourier transform can then be thought of as a signal's frequency behavior during the time period covered by the windowed data block. For a given signal $f(t)$ in $L^2(\mathbb{R})$, the procedure can be described by

$$STFT_f(\omega, \tau) = \int_{-\infty}^{\infty} f(t)\, \gamma(t - \tau)e^{-j\omega t}dt, \quad (\omega, \tau) \in \mathbb{R}^2. \tag{5.2.7}$$

It is even more familiar to signal analysis in its discrete version, where τ and ω are assigned regularly values $\tau = mT$, $\omega = n\Omega$. Then (5.2.7) becomes

$$STFT_f^D(m, n) = \int_{-\infty}^{\infty} f(t)\, \gamma(t - mT)e^{-jn\Omega t}\, dt, \quad (m, n) \in \mathbb{Z} \tag{5.2.8}$$

The analysis window function $\gamma(t)$ balances the time and frequency resolution. The smaller the time duration of $\gamma(t)$, the better the time resolution (poorer frequency resolution). Therefore, there is a trade-off between time accuracy and frequency accuracy. The blocks could be overlapped or disjointed. The percentage of overlap between each block is determined by the time sampling step T and the length of the analysis window function $\gamma(t)$.

  The main disadvantage of STFT, especially on STFT-based spectrogram, is that there are sudden breaks between segments, called *window effect* or *blocking effect*. Hence, although the STFT can be simple and easily implemented, it has been found inadequate for the applications where both high time and frequency resolutions are required, and where frequency contents of a signal change dramatically over time.

Gabor expansion
Dennis Gabor, in 1946, proposed expanding a signal into a set of functions that are concentrated in both the time and frequency domains, then using the coefficients as the description of the signal's local property. The resulting representation is now known as the Gabor expansion.

$$f(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{m,n}\, h_{m,n}(t) \tag{5.2.9}$$
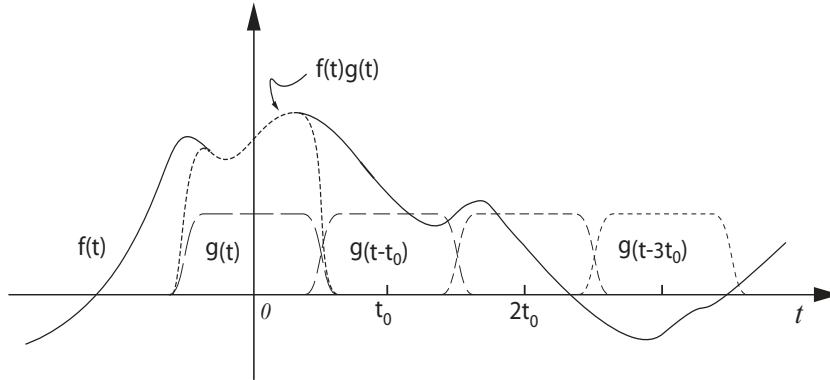
Figure 5.2.2: Short-time Fourier transformation

where $C_{m,n}$ are called the Gabor coefficients. The set of elementary functions $\{h_{m,n}(t)\}$ consists of a time-frequency-shifted window function $h(t)$ which allows the signal $f(t)$ to be cut into segments that are regularly spaced in time, $h(t - mT)f(t)$, $m = 0, \pm1, \pm2..$, i.e.,

$$h_{m,n}(t) = h(t - mT)e^{jn\Omega t}. \tag{5.2.10}$$

This is known as the first time-frequency wavelets or *Gabor wavelets*. In Gabor's original work, the parameter $T > 0$ is the time sampling step, and $h(t)$ is the normalized Gaussian function,

$$g(t) = (\alpha/\pi)^{1/4}e^{-\alpha t^2} \tag{5.2.11}$$

because it is optimally concentrated in the time-frequency domain, in terms of the uncertainty principle. The product of T and Ω determines the density of the sampling grid (however, Gabor limited $T\Omega = 2\pi$). The smaller the product, the denser the sampling (length of segments).

The time and frequency resolutions of $h_{m,n}(t)$ can be adjusted by the parameter α in (5.2.11). The smaller the value of α, the better the frequency resolution (poorer time resolution), vice versa. In two dimension, the Gabor wavelets are directional sinusoids weighted by Gaussian windows. Daugman [115] [116] has used two-dimensional Gabor transforms for image compression.

Note that the windows function in (5.2.8) and (5.2.10) have exactly the same form. Both are time- and frequency-shifted versions of a single prototype function. Let $C_{m,n} = STFT(m, n)$, then (5.2.9) and (5.2.7) form a pair of Gabor expansions. This indicates that the STFT is, in fact, the Gabor coefficient, and conversely the Gabor expansion can be thought of as the inverse of the STFT. However, the Gabor wavelets lead to serious algorithmic difficulties;

- The primary difficulty with Gabor expansion is that it is strongly non-orthogonal. If $h(t)$ is sufficiently regular and well localized, more precisely if the two inte-

grals $\int_{-\infty}^{\infty}(1+|t|)^2|h(t)|^2dt$ and $\int_{-\infty}^{\infty}(1+|\omega|)^2|\hat{h}(\omega)|^2d\omega$ are both finite, then the functions $h_{m,n}(t)$, $m,n \in \mathbb{Z}$, can never be an orthonormal basis for $L^2(\mathbb{R})$.

- By the choosing $\Omega = 1$ and $T = 2\pi$, the windowed analysis consists in restraining the signal to each interval $[2m\pi, 2(m+1)\pi)$ and using Fourier series to analyze each of the corresponding functions. But the functions obtained the crude segmentation, $h(t)$=1 on $[0,2\pi)$ and $h(t)$=0 elsewhere, are not 2π-periodic, and the Fourier analysis will highlight this lack of periodicity and interpret it as a discontinuity or an abrupt variation in the signal. One way to attenuate these numerical artifacts is to use the discrete cosine transform (DCT), which does not eliminate them nevertheless [117].

In conclusion, Gabor expansion can be used as if they formed an orthonormal basis, as long as one does only continuous decompositions. But the corresponding discrete algorithms do not exist, in general, or they require so much tinkering that they become too complicated. It is only very recently that some discrete versions of Gabor expansion are discovered [118] [119] [120] [121], but they are not motivated by expansion and inner product operations.

Granular Representation

In computer music, there is an approach to the representation and generation of music sounds, called "granular synthesis", which was first proposed also by Gabor [122]. The basic idea is that any sound can be considered as a sequence of elementary acoustic quanta called *grains* that correspond to a local time-frequency components of the sound. The features of the grains and their temporal location determine the sound's timbre. Furthermore, such descriptions are psychoacoustically appropriate given the time-frequency resolution tradeoffs and limitations observed in the auditory system.

Various types of grains and a high-level organization of grains based on the concept of tendency masks in the time-frequency plane are suggested by Roads [123]. Some simple analysis methods for deriving grains from real sounds were proposed in [124] [125]. The objective of such granulation approaches is to derive a representation of natural sounds that enables modifications such as time-scaling or pitch-shifting prior to resynthesis. The basic idea of these analysis methods is to extract grains by applying time-domain windows to the signal. Each windowed portion of the signal is treated as a grain, and parameterized by its window function and time location. These grains can be realigned in time or resampled in various ways to achieve desirable signal modifications. Similar ideas have been developed in the speech processing community [126]. In addition, while the granulation by time-windowing is effective for modification of signal, these grains are disparate from the elementary acoustic quanta suggested by Gabor and thus not appropriate analysis for Gabor's time-frequency representation. This fact motivates the technique called "sinusoidal modeling" as an alternative granular

analysis-synthesis, which is widely researched recently in acoustic signal analysis and speech audio coding [127] [128] [129].

Wigner-Ville distribution (WVD)

Formula (5.2.12) is known as the Wigner-Ville distribution, which was originally developed in the area of quantum mechanics by an American physicist, Eugene P. Wigner, in 1932. Fifteen years later, it was introduced into the signal processing area by a French scientist J. Ville.

$$WVD\,(t, \omega) = \int_{-\infty}^{\infty} R(t, \tau)\, e^{-j\omega\tau}\, dt, \tag{5.2.12}$$

where the correlation function $R(\tau)$ is the average of the instantaneous correlation $f(t)f(t - \tau)$, i.e.,

$$R(\tau) = \int_{-\infty}^{\infty} R(t, \tau)\, dt = \int_{-\infty}^{\infty} f(t + \tau/2)\, f^*(t + \tau/2)\, dt. \tag{5.2.13}$$

The mean frequency of the WVD at time t is equal to the signal's weighted average instantaneous frequency, i.e.,

$$\frac{\int_{-\infty}^{\infty} WVD(t, \omega)\, \omega\, d\omega}{\int_{-\infty}^{\infty} WVD(t, \omega)\, d\omega} = \phi^{'}(t)\ . \tag{5.2.14}$$

It is thus true that the WVD describes the signal's time-frequency behavior, and that the energy of the WVD is the same as the energy content in the signal $f(t)$, i.e.,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} WVD(t, \omega)\, dt d\omega = \int_{-\infty}^{\infty} |f(t)|^2\, dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |F(\omega)|^2\, d\omega\ . \tag{5.2.15}$$

As a result of (5.2.15), the WVD is often thought of as energy distribution of a signal in the time-frequency plane and does not introduce a reference function, such as window function or wavelet, against which the signal has to be integrated. This is the advantage of WVD by contrast with the STFT or the wavelet transform. The disadvantage is that the signal enters in the WVD in as quadratic rather than linear way, which is the cause of many interference phenomena. One of them is so-called "crossterm interference", in Figure 5.2.3, that reflects the correlation of signal components in vicinity of a certain moment. This artifact in the crossterm and the fact that there does not exist an algorithm to find an atomic decomposition of a signal by it's transform, prevent the WVD from being used for practical applications to TFR, even though it possesses many desirable properties for signal analysis and has practically much better time-frequency resolution compared with STFT, especially for signals which have a very short time duration, [130] [131].
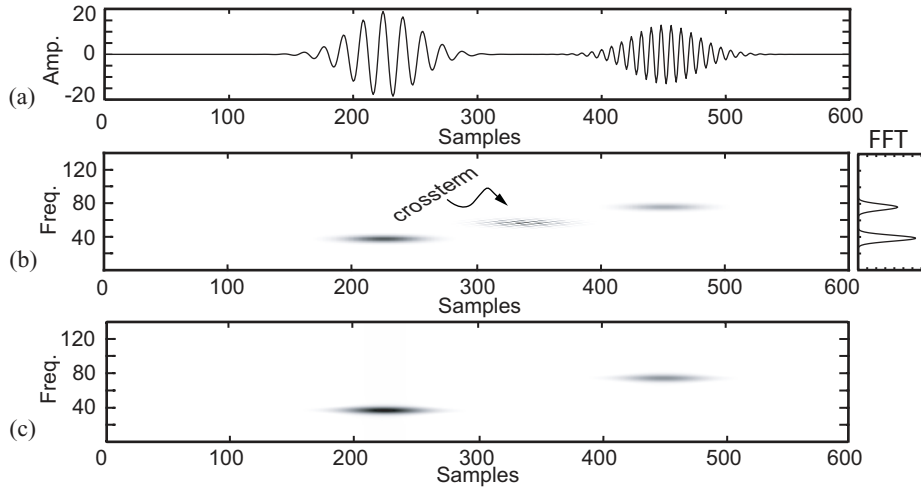
Figure 5.2.3: Wigner-Ville distribution of two Gaussian function. (a) input signal: sum of two Gaussian functions. (b) Wigner-Ville distribution. (c) Choi-Williams distribution

To reduce the crossterm interference with limited affects on the useful properties, an intuitive approach is to apply a 2-D low-pass filter to the WVD, i.e.,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} WVD(t-\mu, \omega-\tau)\, \phi(\mu,\tau)\, d\tau\, d\mu\,. \qquad (5.2.16)$$

Expanding the term of WVD in (5.2.16) yields

$$C(t,\omega) = \int_{-\infty}^{\infty} e^{-j\omega\tau} \int_{-\infty}^{\infty} \Phi(t-\mu,\tau)\, s(\mu+\tau/2)\, s^*(\mu-\tau/2)\, d\mu\, d\tau\,, \qquad (5.2.17)$$

where the function $\Phi(t,\tau)$ denotes the Fourier transform of $\phi(\mu,\tau)$. Formula (5.2.17) is called as "Cohen's class" developed in the field of the quantum mechanics by Leon Cohen [132] [133] [134]. Figure 5.2.3.(c) shows the alias-free WVD of the signal (a) filtered by Choi-Williams filter [135], which is one of the prominent members of Cohen's class.

Wavelet transform (WT)

The wavelet transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the Fourier transform. The main idea behind wavelet analysis is to decompose a signal f into a basis of windows function Ψ_i;

$$f = \sum_i a_i \Psi_i\,. \qquad (5.2.18)$$

The wavelet transform provides a similar time-frequency description to STFT (5.2.7), with a few important differences. The continuous wavelet transform formula is

$$\mathcal{W}_f(a,b) = |a|^{-1/2} \int f(t)\, \psi\left(\frac{(t-b)}{a}\right) dt, \quad (a,b) \in \mathbb{R}^2\,, \tag{5.2.19}$$

and by restricting a, b in (5.2.19) to only discrete values, $a = a_0^m$, $b = nb_0 a_0^m$, the discrete wavelet transform is obtained,

$$\mathcal{W}_f^D(m,n) = a_0^{-m/2} \int f(t)\, \psi(a_0^{-m}t - nb_0)\, dt, \quad (m,n) \in \mathbb{Z}\,, \tag{5.2.20}$$

where $a_0 > 1$, $b_0 > 0$ fixed. In both cases we assume that ψ satisfies

$$\int \psi(t)\, dt = 0. \tag{5.2.21}$$

One similarity between the wavelet transform and STFT is clear that both (5.2.7) and (5.2.19) take the inner products of f with a family of analyzing functions indexed by two labels, $\gamma_{\omega,\tau}(t) = \gamma(t-\tau)e^{-j\omega t}$ in (5.2.7) and $\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right)$ called "wavelets" in (5.2.19). The beginning atom function $\psi(t)$ is called *mother wavelet* or *analyzing wavelet*. The important difference that distinguishes the wavelet transform from the STFT lies in the shapes of their analyzing functions. While the frequency localization width of $\gamma_{\omega,\tau}$ in STFT does not depend on ω_0 and the resulting decomposition is consequently of $\Delta\omega = $ constant type, the analyzing functions $\psi_{a,b}$ of wavelet transform are shifted along the signal with the parameter of translation b in time and dilated with scaling parameter a, which covers different frequency ranges, i.e., large values of the scaling parameter $|a|$ correspond to small frequencies and small values of $|a|$ to high frequencies. Hence, in the case of wavelets we normally do not speak about time-frequency representations but about time-scale representations, scale being in a way the opposite of frequency, because the term frequency is reserved for the Fourier transform.

The localization of wavelets depends on the parameter a and the resulting decomposition will be a collection of time-frequency representation at $\Delta\omega/\omega = $ constant of the signal, all with different resolutions. Wavelets can be hereby interpreted as impulse responses of constant-$\mathcal{Q}$ filters, and because of this collection of representations we can speak of a multiresolution analysis. There exist many different types of wavelet transform, but they start from the basic formulas (5.2.19), (5.2.20).

5.2.4   Comparison by time-frequency tiling

In previous section, we have showed some expansions for time-frequency localization. An easy way to compare such expansions is in terms of their *time-frequency tiling*.

Each atom in an expansion has a particular localization characterized by energy concentration of the atoms. Furthermore, in the case of a structured expansion, these localizations are related to each other through simple transformations. This results in a tiling of time-frequency plane, which depends on the set of transformations. Figure 5.2.4 shows schematically the different localizations of the expansions in time-frequency plane.
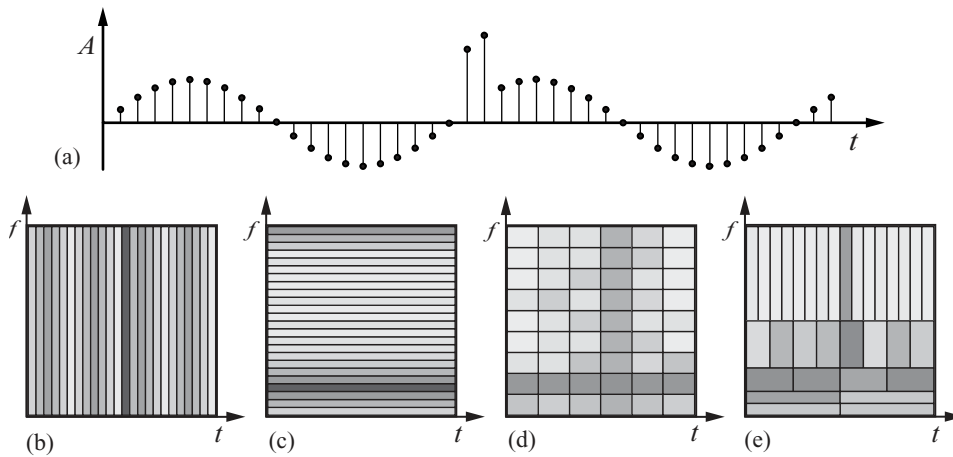
Figure 5.2.4: Time-frequency tiling. (a) input signal, sine wave plus impulse. (b) identity transform (expansion on impulse basis). (c) discrete-time Fourier transform. (d) short-time Fourier transform (Gabor expansion). (e) discrete-time wavelet transform.

Given the input signal (a) in Figure 5.2.4, it would be desirable to have an expansion (or transform) that captures both the isolated impulse and the isolated frequency component. The impulse basis in part (b) locates the impulse of the input signal very well, but its representation of the sinusoid is very redundant. By contrast, the Fourier transform (c) isolates the sinusoid frequency well, but its localization of impulse is not clear. The short-time Fourier transform (d) and wavelet transform (e) provide a compromise between the two former transforms. Typically, the short-time Fourier Transform leads to a rectangular tiling, while the wavelet transform leads to a dyadic tiling. The wavelet transform (e) achieves better localization of the time-domain impulse, without sacrificing too much of the frequency localization, however, a high-frequency sinusoid would not be well localized.

The wavelet transform operates as a microscope with scalable window function which focuses on smaller time phenomenon as the scale becomes small, whereas the STFT decomposes a signal with the window function that is constant at all frequencies, as shown in the figure. Namely, the wavelet permits a local characterization of signal, which the Fourier transform does not, and it is one of most important advantages of

wavelet transform in signal analysis. For the purpose of signal compression, it is clear that a transform whose basis function most closely resembles the input signal would provide the greatest efficiency and compression ratio. With some constraints one can constructs the wavelets that are appropriate to each class of input signals. Because of this adaptability of wavelet basis, the wavelet transform can be an efficient method for transform coding in signal compression system.

5.3 Wavelets In General

5.3.1 Wavelets

Wavelets are building block functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. In real world we usually have the signals that are limited in time or in frequency (band-limited). The time-limited signals can be efficiently represented using a basis of block functions, e.g. Dirac delta functions. However these block functions are not limited in frequency. For band-limited signals we intuitively use the Fourier basis of which sines and cosines are not limited in time. A compromise between the pure time-limited and band-limited basis functions is *wavelets*, namely, the elementary constituents used signal analysis and synthesis, and is also called as *time-frequency atoms* or *wavelet packets*, depending on the circumstances. The wavelets are obtained by scaling and translation of the single atom function called *mother wavelet* or *analyzing wavelet*. The mother wavelet $\psi(t)$ satisfies the conditions

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0, \quad \int_{-\infty}^{\infty} |\psi(t)|\, dt < \infty,$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$. A typical choice for the mother wavelet ψ is so-called "mexican hat" function as follows

$$\psi(t) = (1 - t^2)e^{-t^2/2},$$

which is second derivative of the Gaussian $e^{-t^2/2}$, as shown in Figure 5.3.1.

The mexican hat function is well localized in both time and frequency and satisfies the conditions mentioned above. Given a mother wavelet $\psi(t)$, we construct a sequence
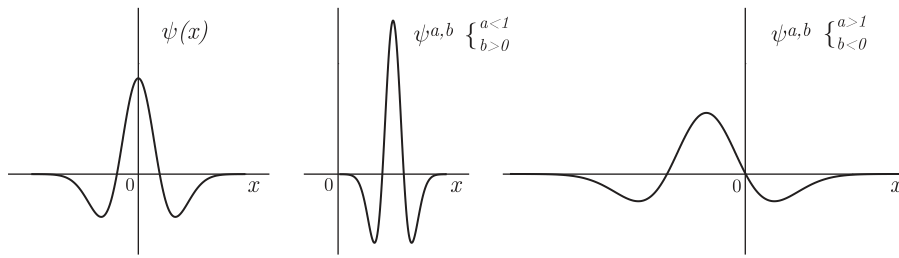
Figure 5.3.1: Mexican hat function (second derivative of the Gaussian probability density function) and its scaled and translated wavelets

of wavelets by dilation parameter a and translation parameter b of $\psi(t)$,

$$\psi_{a,b}(t) = |a|^{-1/2}\,\psi\left(\frac{t-b}{a}\right) \quad a > 0, \ b \in \mathbb{R}\,.$$

This is the Grossmann-Morlet wavelets.

In wavelet analysis, temporal analysis is performed with a contracted high-frequency version (small value of a) of the mother wavelet ψ, while frequency analysis is performed with a dilated low-frequency version (high value of a) of the same wavelet. This process is nearly identical with work of our auditory system, which consists of the representation of the ears as a bank of filters with constant $\delta f / f$ for frequencies higher than about 500Hz, so it suggests that wavelet-based methods for acoustical analysis have a better chance than other methods. Because the original signal or function can be represented in terms of wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients.

One of the main advantages of wavelet transform is the fact that one can choose the mother wavelet depending on the characteristics of information in the signal which has to be pointed out. The choice of mother wavelet is subject only to some mathematical admissibility conditions that are not very restrictive in practice (see the admissibility conditions in next subsection). By choosing the best wavelets adapted to our signal, or truncating the coefficients below a threshold, the signal is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression.

Since the wavelet transform associates to a function of two variables, time and scale, it generates a two-dimensional representation from a one-dimensional signal. This double representation allows to interpret all the specific information carried by the signal, and establishes a relation between the result of the wavelet transform and physical parameters associated with the signal. Consequently, the wavelet transform can be used to localize discontinuities and to extract modulation laws, and the phase of the wavelet transform can be used to define an instantaneous frequency at a given scale.

5.3.2 Properties of wavelet transform

Linearity

The Wavelet transform is linear. This property is very useful. It means that the transform of the sum of signals is the sum of their transforms, which is convenient for the analysis of polyphonic signals. It should be remarked that the TFR by Wigner-Ville distribution is not linear but rather bilinear or quadratic, as we showed in the previous section.

Admissibility Conditions

Since the wavelet transform is linear, we may ask for an inversion formula by resummation of wavelets weighted by the coefficients. From the Formula (5.2.19) it is given by

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty}\!\!\int_{-\infty}^{\infty} \mathcal{W}_f(a,b)\ \psi_{a,b}(t)\ \frac{da\ db}{a^2}, \tag{5.3.1}$$

where C_ψ is a constant,

$$C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{\omega}\ d\omega\ < \infty, \tag{5.3.2}$$

that depends on the choice of the mother wavelet ψ. From the formulas (5.3.1) and (5.3.2), we notice that the lossless reconstruction of the function f from the wavelet transformation is possible only if C_ψ exists, that is, if the integral in (5.3.2) converges. This condition that a mother wavelet has to satisfy is called the *admissibility condition*. It implies also that the Fourier transform of $\psi(t)$ *vanishes* at the zero frequency. In practice this condition therefore relaxes into

$$-E_\psi = \int |\psi(t)|^2\ dt\ < +\infty, \quad -\hat{\psi}(0) = \int \psi(t)\ dt =\ 0. \tag{5.3.3}$$

They mean that $\psi(t)$ has finite energy, that is, a bandpass like spectrum and zero mean value in time domain, respectively. So $\psi(t)$ has to oscillate, and in other words it must be a ***wave***. This is a very important observation, which we will use later on to build an efficient wavelet transform.

Regularity (decay property)

For practical applications that seek a signal description with as few components as possible, it is not a desirable property that the time-bandwidth product of wavelet transform is the square of the input signal. It can be tackled by making a quickly decreasing wavelet transform with corresponding scale a. The regularity conditions state that the wavelet function should have some smoothness and concentration in both time and frequency domains.

The term of regularity seems to be derived from the definition that a filter is regular with k if its z-transform has k zeros at $z = e^{i\pi}$. In wavelet theory, it is related with scale function, which is possible only if all wavelet moments up to $k-1$ vanish. The regularity condition can be also explained using the concept of *vanishing moments*. We expand the wavelet transform into the Taylor series at $t = 0$ [136], for simplicity $b = 0$, and it is given by

$$\mathcal{W}_f(a, b_{=0}) = \frac{1}{\sqrt{a}} \left[\sum_{p=0}^{n} f^{(p)}(0) \int \frac{t^p}{p!} \, \psi\left(\frac{t}{a}\right) \, dt \; + \; O(n+1) \right], \qquad (5.3.4)$$

where $f^{(p)}$ stands for the pth derivative of f, and $O(n+1)$ denotes the rest of the expansion. We define the moments of the wavelet by $\mathcal{M}_p$,

$$\mathcal{M}_p = \int t^p \, \psi(t) \, dt, \qquad (5.3.5)$$

then we can rewrite (5.3.4) with the finite development as follows,

$$\mathcal{W}(a, 0) = \frac{1}{\sqrt{a}} \left[f(0)a\mathcal{M}_0 + \frac{f^{(1)}(0)}{1!}a^2\mathcal{M}_1 + \frac{f^{(2)}(0)}{2!}a^3\mathcal{M}_2 + ... \right.$$
$$\left. ... + \frac{f^{(n)}(0)}{n!}a^{n+1}\mathcal{M}_n + O(a^{n+2}) \right]. \qquad (5.3.6)$$

From the admissibility condition, $\mathcal{M}_0 = 0$, therefore, the first term in the right-hand side of the (5.3.6) is zero. If the other moments $\mathcal{M}_n = 0$ also, then the coefficients of $\mathcal{W}(a, b)$ will decay as fast as a^{n+2} for a smooth signal $f(t)$. This is called *vanishing moments* or *approximation order*. N vanishing moments corresponds to N approximation order of the wavelet transform. These moments do not need to be exactly zero, but small value is often enough. In practice, the number of vanishing moments required depends heavily on the application [137]. With the admissibility conditions, the regularity is one of the most important properties of wavelet transform. For more detail about it, we refer to [138] [100].

As mentioned in previous section, we precedently have the ***wave*** from the admissibility condition. The regularity and vanishing moments give us the property of fast decay, that is, ***let***. Now they together give us the term ***wavelet***.

Energy Conservation
The total energy of the signal can be expressed in term of the values of the transforms by

$$E_f = \int |f(t)|^2 \, dt = \frac{1}{C_\psi} \int |\mathcal{W}_f(a, b)|^2 \, \frac{da \, db}{a^2} . \qquad (5.3.7)$$

This expression allow us to interpret the square of the modulus of the transforms as a density of energy distributed over the domains of representation.

Constraints on Coefficients

The wavelets are, in general, not orthogonal:

$$\langle \psi_{a,b}, \psi_{a',b'} \rangle = \int \bar{\psi}_{a,b}(t)\, \psi_{a',b'}(t)\, dt \;\neq\; 0. \tag{5.3.8}$$

The coefficients of the transform are constrained to satisfy

$$\mathcal{W}_f(a,b) = \int P_\psi(a,b;a',b')\, \mathcal{W}_f(a',b')\, \frac{da'\, db'}{a'^2}\,, \tag{5.3.9}$$

where $P_\psi(a,b;a',b') = C_\psi^{-1}\langle \psi_{a,b},\, \psi_{a',b'}\rangle$. Formula (5.3.9) means that it is possible to find a grid such that the analysis coefficients at the points of the grid allow an arbitrarily accurate reconstruction of the coefficients on the representation plane. Figure 5.5.1 illustrates the points form of such a grid compared with the case of STFT.

5.4 The Continuous Wavelet Transform

5.4.1 Definition of CWT

As mentioned previously, a wavelet expansion uses dilations and translations of one fixed function, the mother wavelet $\psi \in L^2(\mathbb{R})$. In the case of continuous wavelet transform (CWT), the dilation and translation parameters a,b vary continuously over $\mathbb{R}$, with the constraint $a \neq 0$. In result, a doubly-indexed family of wavelets from ψ by dilating and translating is generated as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\, \psi\!\left(\frac{x-b}{a}\right) \quad \text{with} \quad a,b \in \mathbb{R},\; a \neq 0\,. \tag{5.4.1}$$

The variable x usually denotes either time or space. These functions are scaled so that their $L^2(\mathbb{R})$ norms are independent of a, i.e., for all $a,b\; \|\psi_{a,b}\| = \|\psi\|$ with assuming $\|\psi\| = 1$. The CWT of a function $f \in L^2(\mathbb{R})$ is now defined by

$$\mathcal{W}_f(a,b) = \langle\, f,\, \psi_{a,b}\,\rangle \tag{5.4.2}$$

$$= \int f(x)\, |a|^{-1/2}\, \overline{\psi\!\left(\tfrac{x-b}{a}\right)}\, dx\,. \tag{5.4.3}$$

Using Parseval's identity, we can also write this as

$$2\pi \, \mathcal{W}_f(a, b) = \langle \, \hat{f}, \, \hat{\psi}_{a,b} \, \rangle, \tag{5.4.4}$$

where

$$\hat{\psi}_{a,b}(\omega) = \frac{a}{\sqrt{|a|}} \, e^{-i\omega b} \, \hat{\psi}(a\omega). \tag{5.4.5}$$

Suppose that the wavelet ψ satisfies the admissibility condition mentioned in the section 5.3.2, and we recall it,

$$C_\psi = 2\pi \int |\omega|^{-1} \, |\hat{\psi}(\omega)|^2 \, d\omega \, < \infty. \tag{5.4.6}$$

Then, the resolution of the identity holds for all $f, g \in L^2(\mathbb{R})$ as follows,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}_f(a, b) \, \overline{\mathcal{W}_g(a, b)} \, \frac{da \, db}{a^2} \, = C_\psi \langle f, g \rangle. \tag{5.4.7}$$

Now the continuous wavelet transform $\mathcal{W}(a, b)$ is invertible on its range, and an inverse transform is given by the relation

$$f(x) = C_\psi{}^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}_f(a, b) \, \psi_{a,b}(x) \, \frac{da \, db}{a^2}. \tag{5.4.8}$$

In formula (5.4.2), the continuous wavelet transform is calculated by continuously shifting a continuously scalable function over a signal and calculation the correlation between the two signals. The correspondence $f(x) \rightarrow \mathcal{W}_f(a, b)$ represents a one-variable function by a function of two variables, into which lots of correlations are built in, and it will be also clear that the scaled functions will be nowhere near an orthogonal basis. Therefore, the obtained wavelet coefficients will be highly *redundant*. In fact, orthogonality is one of most important properties in signal analysis, but it is often not desirable, for example, when one wants to explore the characteristics of signal with well correlated fine representation [139] [140].

In some applications the redundancy can help to reduce the sensitivity to noise [136]or improve the shift invariance of the transform [138]. However, for most practical applications, especially for the data compression, in where require a perfect reconstruction property and a signal description with as a few components as possible, we would like to remove this redundancy. This links to the orthonormal wavelet basis, the perfect reconstruction filter banks and the multiresolution analysis in discrete-time wavelet transform, which are discussed in section 5.6 and 5.7.

5.4.2 The CWT in higher dimensions

We consider the CWT that can be made in dimensions lager than 1, i.e., $L^2(\mathbb{R}^n)$ with $n > 1$. One possibility is to choose the wavelet $\psi \in L^2(\mathbb{R}^n)$ so that it is spherically symmetric, as well as $\hat{\psi}(\omega) = \eta(|\omega|)$ [100]. Then the admissibility condition becomes

$$C_\psi = (2\pi)^n \int_0^\infty |\eta(t)|^2 \, \frac{dt}{t} \; < \infty \,. \tag{5.4.9}$$

The resolution of the identity in (5.4.7) is supplemented for all $f, g \in L^2(\mathbb{R}^n)$ as follows,

$$\int_0^\infty \int_{-\infty}^\infty \mathcal{W}_f(a,b) \, \overline{\mathcal{W}_g(a,b)} \; db \, \frac{da}{a^{n+1}} = C_\psi \langle f, g \rangle \,, \tag{5.4.10}$$

where $\mathcal{W}_f(a,b) = \langle f, \psi_{a,b} \rangle$, as before, and $\psi_{a,b}(x) = a^{-n/2} \psi(\frac{x-b}{a})$, with $a \in \mathbb{R}_+$, $a \neq 0$, and $b \in \mathbb{R}^n$. Consequently, the reconstruction version of high-dimensional CWT is given by

$$f = C_\psi^{-1} \int_0^\infty \int_{\mathbb{R}^n} \mathcal{W}_f(a,b) \, \psi_{a,b} \; db \, \frac{da}{a^{n+1}} \,. \tag{5.4.11}$$

Other approach is to choose a ψ that is not spherically symmetric, and to introduce rotation as well as dilations and translations [141] [142]. In two dimensions, for instance, we then define

$$\psi_{a,b,\theta}(x) = a^{-1} \, \psi\left(R_\theta^{-1} \left(\frac{x-b}{a} \right) \right), \tag{5.4.12}$$

where $a > 0$, $b \in \mathbb{R}^2$, and where R_θ is the matrix $\left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix} \right)$. Then the admissibility condition becomes

$$C_\psi = (2\pi)^2 \int_0^\infty \int_0^{2\pi} |\hat{\psi}(r\cos\theta, r\sin\theta)|^2 \, d\theta \, \frac{dr}{r} < \infty \,, \tag{5.4.13}$$

and the corresponding resolution of the identity is

$$f = C_\psi^{-1} \int_0^\infty \int_{\mathbb{R}^2} \int_0^{2\pi} \mathcal{W}_f(a,b,\theta) \, \psi_{a,b,\theta} \; d\theta \, db \, \frac{da}{a^3} \,. \tag{5.4.14}$$

5.5 Discreet Wavelets: Wavelet Series

In CWT we do not put any constraints on the choice of the two coordinates a and b, and they can therefore map the whole (a, b) plane, in principle. In discrete wavelets, however, we restrict the choice of possible (a, b) values as $a = a_0^m$ and $b = nb_0a_0^m$. Compared with *discrete-time* wavelet transform, the discrete wavelets or wavelet series are merely the discrete representation of the both time and scale axes of CWT. This discretization allows the wavelet transform to be numerically computed as a series of discrete convolutions in time at discrete intervals of scale.

5.5.1 Discretization of parameters

In practical applications, the CWT can only be computed on a discrete grid of points $(a_n, b_n)_{n \in \mathbb{Z}}$. As in the case of the STFT, the most intuitive way of doing this is simply uniform sampling the time-frequency plane. However, in the case of wavelet transform, the scale change can be used to reduce the sampling rate. At higher scales (lower frequencies), the sampling rate can be decreased, according to Nyquist's rule. In other words, if the time-scale plane needs to be sampled with a sampling rate of N_1 at scale s_1, the same plane can be sampled with a sampling rate of N_2, at scale s_2, where $s_1 < s_2$ and $N_2 < N_1$. The actual relationship between N_1 and N_2 is given by

$$N_2 = \frac{s_1}{s_2}N_1 \,. \tag{5.5.1}$$

If synthesis is not required, the Nyquist criteria does not need to be satisfied. The restrictions on the discretization and the sampling rate become important if, and only of, the signal reconstruction is desired.

Now we describe the discretization procedure in mathematical terms. Suppose that ψ is simply defined by

$$\hat{\psi}(\omega) = \begin{cases} 1 & |\omega| \in [1, a_0], \\ 0 & \text{elsewhere}, \end{cases} \tag{5.5.2}$$

for some $a_0 > 1$. Since, in that particular case, the scaled function f_a is band-limited by $\hat{\psi}(\omega)$ to $[-\frac{a_0}{a}, -\frac{1}{a}] \cup [\frac{1}{a}, \frac{a_0}{a}]$, it is completely determined by its samples $f_a(\frac{ma}{\pi(a_0-1)})$, $m \in \mathbb{Z}$. Moreover, the frequency axis is tiled by the intervals $[-a_0^m, -a_0^{m-1}] \cup [a_0^{m-1}, a_0^m]$, $m \in \mathbb{Z}$, so that f can be recovered form the data of $f_m = f_{a_0^m}$, $m \in \mathbb{Z}$, where the different values of m correspond to wavelets of different widths. Therefore, the discretization of the translation parameter b should depend on m, that is, high frequency wavelets are

translated by small dilation steps in order to cover the whole time range, while lower frequency wavelets are translated by larger dilation steps[2].

To discretize b we therefore choose $b = nb_0 a_0^m$, $m, n \in \mathbb{Z}$, for some fixed $b_0 > 0$. With formula (5.4.1) given before, we can now define the *discrete wavelets* as follows,

$$\psi_{m,n}(x) = a_0^{-m/2}\, \psi(a_0^{-m}x - nb_0), \quad m, n \in \mathbb{Z}, \tag{5.5.3}$$

where fixed $a_0 > 1$ and $b_0 > 0$. Although (5.5.3) is called a discrete wavelet, it is piecewise a continuous function normally. Consequently, the effect of discretizing the wavelet is that the time-scale plane is now sampled at discrete intervals, as shown in Figure 5.5.1.
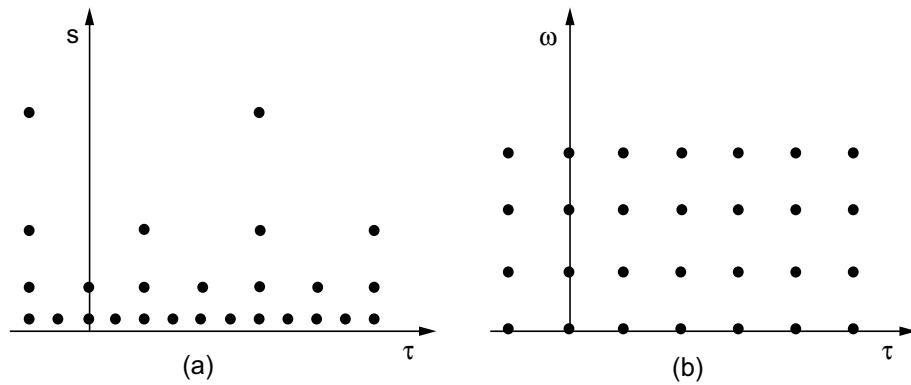
Figure 5.5.1: Comparison of analysis grids for wavelet transform (a), and the STFT (b).

We usually choose $s_0 = 2$ so that the sampling of the frequency axis corresponds to *dyadic sampling* [143] [144]. This is very natural choice for computers, the human ear and music, for instance. For the translation factor we usually choose $b_0 = 1$ so that we also have dyadic sampling grid of the time axis. In this case, the sampling rate is reduced for the time axis by a factor of 2 at every scale, since the discrete scale changes by factors

---

[2]The following analogy may clear this concept. Consider the whole process as looking at a particular object. The human eyes first determine the coarse view which depends on the distance of the eyes to the object. This corresponds to adjusting the scale parameter a_0^{-m}. When looking at a very close object, with great detail, m is negative and large (low scale, high frequency, analyses the detail in the signal). Moving the head (or eyes) very slowly and with very small increments (of angle, of distance, depending on the object that is being viewed), corresponds to small values of $b = na_0^m b_0$. Note that when m is negative and large, it corresponds to small changes in time, b, (high sampling rate) and large changes in s_0^{-m} (low scale, high frequencies, where the sampling rate is high). The scale parameter can be thought of as magnification too.

of 2. This choice of $a_0 = 2$ and $b_0 = 1$ is also used to construct an orthonormal wavelet basis which is a very important class of discrete wavelets (see section 5.5.3). It has been also covered to sample with more irregular sets of a, b [145] [146] [147].

Now the wavelet transformation using discrete wavelets is given by

$$\mathcal{W}_f(m,n) = a_0^{-m/2} \int f(t)\,\psi(a_0^{-m}t - nb_0)\,dt, \quad (m,n) \in \mathbb{Z}\,. \tag{5.5.4}$$

Formula (5.5.4) implies also that the transformation of a continuous signal using the discrete wavelets results in a *series* of wavelet coefficients. Generally, the transform that only uses the dyadic values of a and b as (5.5.4) was called the *discrete wavelet transform*. However, at this time, this term is ambiguous, since the discrete wavelets are not time-discrete actually, only the translation and scale steps are discretized, as described above. Furthermore, the term DWT is also used to denote the transform from the sequence of scaling function coefficients of a function to its wavelet coefficients (see section 5.6). Therefore, in the discrete wavelets case, it is referred to as the *wavelet series* or *wavelet series decomposition* instead of DWT, analogous to the Fourier series and discrete Fourier transform (DFT).

In such a decomposition scheme, an important issue is naturally the question of reconstruction. In the case of the CWT, this question was answered immediately by the resolution of the identity (5.4.7), which does not exist in this discrete case. The reconstruction of a signal from its wavelet series is only possible if the $\psi_{m,n}$ constitute either a *frame*, an *orthonormal basis* or a *biorthogonal basis*. Let us start with the *frame* that we give some detail in the next section. The relationship between the frame and numerically stable reconstructions from discretized wavelets was first pointed out by A. Grossmann (1985). In [100], it is proven that the necessary and sufficient condition for stable reconstruction is that the energy of the wavelet coefficients must lie between two positive bounds, i.e.

$$A\|f\|^2 \le \sum_{m,n} |\langle f, \psi_{m,n}\rangle|^2 \le B\|f\|^2\,, \tag{5.5.5}$$

where $\|f\|^2$ is the energy of $f(t)$, $A > 0$, $B < \infty$, and A, B are independent of $f(t)$. The family of basis functions $\psi_{m,n}(t)$ with $m, n \in \mathbb{Z}$ is referred to as a *frame* with bounds A and B, if, and only if, for all $f \in Ł^2(\mathbb{R})$, (5.5.5) is satisfied. When $A = B$, the frame is *tight* and the discrete wavelets behave exactly like an orthonormal basis. When $A \ne B$, exact reconstruction is still possible at the expense of a *dual frame*. In a dual frame the decomposition wavelet is different from the reconstruction wavelet. Now, if $\psi_{m,n}$ satisfy (5.5.5), we can recover f by summing the wavelet series, weighted by the wavelet transform coefficients:

$$f(t) = \sum_{m,n} \mathcal{W}_f(m,n)\,\psi_{m,n}(t). \tag{5.5.6}$$

If $\psi_{m,n}$ are orthonormal or biorthonormal, the transform will be non-redundant, whereas if they form a frame, the transform will be redundant. On the other hand, it is much easier to find frames than to find orthonormal or biorthonormal bases. One of the main disadvantages of discrete wavelets is that the resulting wavelet transform is no longer shift-invariant (or time-invariant), which means that the wavelet transforms of a signal and of a time-shifted version of the same signal are not simply shifted versions of each other.

5.5.2 Frames of wavelets

A frame is a family of vectors which can represent any finite energy signal by the sequence of its inner products with the vectors of the family. It is a generalization of the concept of basis in a linear space. Note that not all sequences of values may represent a sequence of samples. In general, frames are a stable and redundant representation of signals. We briefly review here this frames introduced by Duffin and Schaffer, [148].

Let $\Phi = \{\varphi_n\}_{n \in \mathbb{Z}}$ in a Hilbert space $\mathcal{H}$. Φ is called a *frame* if there exist $A > 0$, $B < \infty$ so that

$$A\|f\|^2 \leq \sum_{n \in \mathbb{Z}} |\langle f, \varphi_n \rangle|^2 \leq B\|f\|^2, \quad \text{for all } f \in \mathcal{H}. \tag{5.5.7}$$

It is called *frame condition*. A and B are called the *frame bounds*. A frame Φ is called a *tight frame* if the frame bounds can be taken to be equal. To such a sequence in (5.5.7), we associate the frame operator U that transforms any $f \in \mathcal{H}$ into the square-summable sequence of its frame inner products, i.e. $(\langle f, \varphi_n \rangle)_{n \in \mathbb{Z}}$ in $l^2(\mathbb{Z})$. The adjoint of U is U^* : $l^2 \to \mathcal{H}$. If we define that $G \equiv U^*U : \mathcal{H} \to \mathcal{H}$ is any positive Hermitian operator such that $f * Gf \geq 0$ for all $f \in \mathcal{H}$, then we can simply rewrite the frame condition (5.5.7) with the identity operator $I : C^n \to C^n$ (defined by $Iu \equiv u$ for all $u \in C^n$) as follows,

$$AI \leq G \leq BI, \tag{5.5.8}$$

in the sense that $A\langle f, f \rangle \leq \langle U^*Uf, f \rangle \leq B\langle f, f \rangle$ for all $f \in \mathcal{H}$. From (5.5.8) it can be immediately concluded that the eigenvalues of G lie in the interval $[A, B]$, and are equal in the tight frame case.

With traditional assuming that A is chosen as large as possible and B as small as possible, in particular, the frame bounds will be often taken as the minimum and maximum eigenvalues of G. Because all eigenvalues of G are nonzero it follows from (5.5.8) that G is invertible so that we can define a new sequence $\tilde{\Phi} = G^{-1}\Phi$. Thus it turns out that $\tilde{\Phi}$ is itself a frame with bounds $A^{-1} > 0$, $B^{-1} < \infty$,

$$\frac{1}{B}\|f\|^2 \leq \sum_{n \in \mathbb{Z}} |\langle f, \tilde{\varphi}_n \rangle|^2 \leq \frac{1}{A}\|f\|^2, \tag{5.5.9}$$

since we have

$$\langle f, \tilde{\varphi}_n \rangle = \langle f, G^{-1} \varphi_n \rangle = \langle G^{-1} f, \varphi_n \rangle . \tag{5.5.10}$$

The frame $\tilde{\Phi}$, which is an image of the frame Φ through G^{-1}, is called the dual frame of the Φ. It can be easily checked that the associated frame operator $\tilde{U}$ satisfies $\tilde{U} = U(U^*U)^{-1}$ so that we have

$$\tilde{U}^*U = [UG^{-1}] * U = G^{-1}G = I ,$$
$$U^*\tilde{U} = U^*UG^{-1} = I . \tag{5.5.11}$$

This shows that any $f \in \mathcal{H}$ can be expressed as

$$f = \sum_{n \in \mathbb{Z}} \langle f, \varphi_n \rangle \tilde{\varphi}_n = \sum_{n \in \mathbb{Z}} \langle f, \tilde{\varphi}_n \rangle \varphi_n . \tag{5.5.12}$$

Thus, it is true that the reconstruction of an arbitrary signal is also possible by applying the dual frame $\tilde{\Phi}$ or using inverse of G ($\equiv U^*U$). This can be done by a fast algorithm given as follows,

$$G = \frac{A+B}{2} (I - R) , \tag{5.5.13}$$

where the residual operator $R = I - \frac{2}{A+B} G$, which satisfies

$$\|R\| \leq \frac{B-A}{A+B} < 1 , \tag{5.5.14}$$

according to (5.5.8). Hence the series $\sum_{k=0}^{\infty} R^k$ converges in norm, and its limit is $(I - R)^{-1}$. Consequently, we have

$$\tilde{\varphi}_n = G^{-1} \varphi_n = \frac{2}{A+B} \sum_{k=0}^{\infty} R^k \varphi_n , \tag{5.5.15}$$

which means that we can approximate G^{-1} by truncating the series of R^k at a certain order depending on the desired accuracy. Note that the convergence of (5.5.15) is very fast when the frame bounds are *close* in the sense $A - B \ll A + B$.

In the tight frame case, i.e., $A = B$, we simply have $\tilde{\Phi} = A^{-1}\Phi$ and rewrite (5.5.7), for all $f \in \mathcal{H}$,

$$\sum_{n \in \mathbb{Z}} |\langle f, \varphi_n \rangle|^2 = A\|f\|^2 ,$$

and it follows

$$f = A^{-1} \sum_{n} \langle f, \varphi_n \rangle \varphi_n , \tag{5.5.16}$$

which gives a trivial way to recover f from the $\langle f, \varphi_n \rangle$. Although the formula (5.5.16) is very similar to the expansion of f in an orthonormal basis, it is redundant and also not

orthonormal basis. However if the frame bounds $A = B = 1$, and if $\|\Phi\| = 1$ for all $n \in \mathbb{Z}$, then the Φ constitute an orthonormal basis, and its transform is not redundant. In other words, the discrete wavelets, for instance, can be made orthogonal to their own dilations and translations by special choice of the mother wavelet, e.g.

$$\int \psi_{m,n}(t)\, \psi^*_{m',n'}(t)\, dt = \begin{cases} 1 & \text{if } m = m' \text{ and } n = n'\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{5.5.17}$$

Then the reconstruction formula (5.5.6) will be the expansion of orthonormal bases. In next section we discuss more about the orthonormal basis.

In the case of wavelet family $\psi_{m,n}$, a criterion for the existence of the frame bounds was derived by Daubechies [100] under the assumptions that

$$0 < c_1 \le \sum_{n \in \mathbb{Z}} |\hat{\psi}(a_0^m \omega)|^2 \le c_2 < \infty \tag{5.5.18}$$

and

$$\beta(s) = \sup_{\omega \in \mathbb{R}} \sum_{n \in \mathbb{Z}} |\hat{\psi}(a_0^m \omega)\hat{\psi}(a_0^m \omega + s)| \le C(1 + |s|)^{-(1+\epsilon)}\,, \tag{5.5.19}$$

for some $\epsilon > 0$. Note that the dual frame $\tilde{\psi}_{m,n} = G^{-1}\psi_{m,n}$ will not be generated by the translations and dilations of its mother function $\tilde{\psi}$. The frame bounds are given by

$$A = \frac{2\pi}{b_0} \left\{ c_1 - \sum_{\substack{k \in \mathbb{Z} \\ k \ne 0}} \left[\beta\Big(\frac{2\pi}{b_0}k\Big)\, \beta\Big(-\frac{2\pi}{b_0}k\Big) \right]^{1/2} \right\},$$

$$B = \frac{2\pi}{b_0} \left\{ c_2 - \sum_{\substack{k \in \mathbb{Z} \\ k \ne 0}} \left[\beta\Big(\frac{2\pi}{b_0}k\Big)\, \beta\Big(-\frac{2\pi}{b_0}k\Big) \right]^{1/2} \right\}, \tag{5.5.20}$$

under the condition that A is strictly positive. It is easy to verify that this is always the case when b_0 is smaller than a certain threshold b_t, recall that $c_1 > 0$. Moreover, if we choose a_0 and b_0 close to 1 and 0, respectively, then the ratio $\frac{A-B}{A+B}$ tends to 0, hence, A and B arbitrarily close to 1 after renormalization. In the result, the $\psi_{m,n}$ will be closely a tight frame and an orthonormal basis. This process is quite equivalent to oversampling the CWT on a very dense grid. For this reasons, the wavelet series decomposition is often referred to as the *oversampled* wavelet transform.

5.5.3 Orthonormal wavelet bases

As the case of CWT, it is clear that the wavelet series decomposition is also redundant representation of a signal, since it is merely to discretize the scale and translation steps

of CWT. In the case of a tight frame, if all wavelets $\psi_{m,n}(x)$ in (5.5.3) are necessary to reconstruct a general signal, then the wavelets form an *orthonormal* basis of the space of signals with finite energy [149]. That is, by considering more restrictions on the sampling parameters a_0 and b_0, it is possible to remove the redundancy in the reconstruction formula (5.4.8), so that it may be regarded as the expansion of f in a basis.

Meyer [150] and Strömberg [151] showed that if we choose $a_0 = 2$, $b_0 = 1$, then there exist some wavelets $\psi(x) \in L^2(\mathbb{R})$ with good time-frequency localization properties, such that

$$\psi_{m,n}(x) = 2^{-m/2}\,\psi(2^{-m}x - n), \quad m, n \in \mathbb{Z} \tag{5.5.21}$$

constitute an orthonormal basis for $L^2(\mathbb{R})$. These particular wavelets are called *orthogonal wavelets*. In this case, an arbitrary signal can be represented exactly as a weighted sum of basis functions obtained by scaling and translation of a single prototype function $\psi(x)$,

$$f = \sum_{m,n\in\mathbb{Z}} \langle f, \psi_{m,n} \rangle \, \psi_{m,n}. \tag{5.5.22}$$

It is true that the orthonormal wavelet bases can be built for sequences of scales other than $(2^m)_{m\in\mathbb{Z}}$, but we will concentrate on dyadic scales which lead to simpler decomposition algorithms.
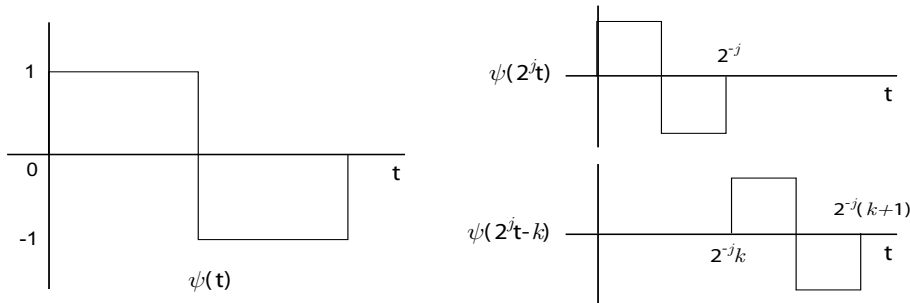
Figure 5.5.2: Haar wavelet and its scaled and shifted version.

The Haar basis is the oldest example of a function ψ for which the $\psi_{m,n}$ defined by (5.5.21) constitute an orthonormal basis for $L^2(\mathbb{R})$. The Haar basis is a piecewise function such that

$$\psi(x) = \begin{cases} 1 & 0 \le x < \frac{1}{2} \\ -1 & \frac{1}{2} \le x < 1 \\ 0 & \text{otherwise}. \end{cases} \tag{5.5.23}$$

Note that the Haar wavelet is not continuous (see Figure 5.5.2), and its Fourier transform decays only like $|\omega|^{-1}$. These correspond to inefficient approximation for smooth

functions and bad frequency localization, respectively, and are therefore a major inconvenience of the Haar wavelet for many applications. On the other hand, the Haar basis is an unconditional basis for $L^p(\mathbb{R})$, $1 < p < \infty$, and it is advantage of the Haar basis in comparison with the short-time Fourier basis [100].

Another most important wavelet example is the sinc basis, which is well known as a key function for the reconstruction of special band-limited signals in signal processing, that is,

$$\hat{\psi}(\omega) = \begin{cases} 1 & \pi < |\omega| \le 2\pi, \\ 0 & \text{otherwise.} \end{cases} \tag{5.5.24}$$

They are limit cases of orthonormal wavelet bases with good time resolution (Haar basis) and good frequency resolution (sinc basis).[3] Namely, they have complemental time-frequency properties to each other.
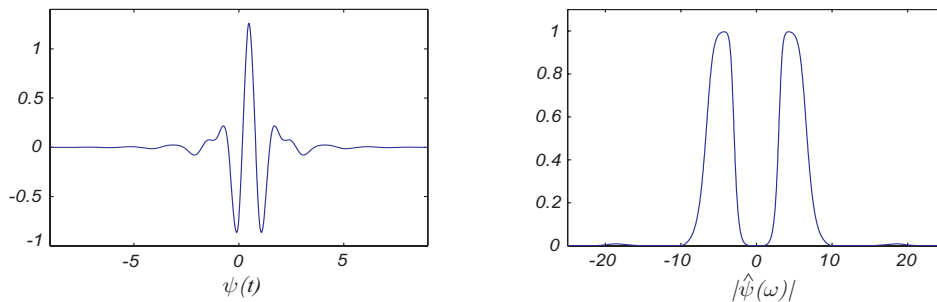
Figure 5.5.3: Example of orthogonal wavelet. (a) Cubic spline by Lemarié and Battle. (b) Modulus of its Fourier transform. This wavelet $\psi(x)$ can be interpreted as the impulse response of a band-pass filter.

The methods for orthogonal wavelet construction can be divided into two classes. First, there are direct constructions by Strömberg [151], Meyer [152], Battle [153] and Lemarié [154], which consist in "smoothing" the sinc function in the Fourier domain so that the function in time domain has faster decay. Figure 5.5.3 shows an example of a particular orthogonal wavelet which is a cubic spline studied independently by Lemarié and Battle. Another method, which is pioneered by Daubechies [155] and seems more relevant for numerical applications, is to construct indirectly the wavelets from iterated discrete-time filter banks in time domain, and this study of the iteration leads to the notion of regularity of the discrete-time filter. These constructions can be considered as smoothed versions of the Haar function, leading to better frequency resolution. Hence,

---

[3]In addition to sinc basis, the Littlewood-Paley basis, with $\psi(x) = (\pi x)^{-1}(\sin 2\pi x - \sin \pi x)$ or $\hat{\psi}(\xi) = (2\pi)^{-1/2}$ if $\pi \le |\xi| \le 2\pi$, 0 otherwise, is also an orthonormal wavelet basis with good frequency localization.

it can be seen that one of the main goals of the wavelet bases theory is to construct the wavelet bases that have the time-frequency properties of the two extreme cases of Haar and sinc bases, simultaneously. In the STFT case, according to the Balian-Low Theorem [156] [98], it is impossible to construct an orthonormal windowed Fourier basis with such a function well localized in time and frequency.
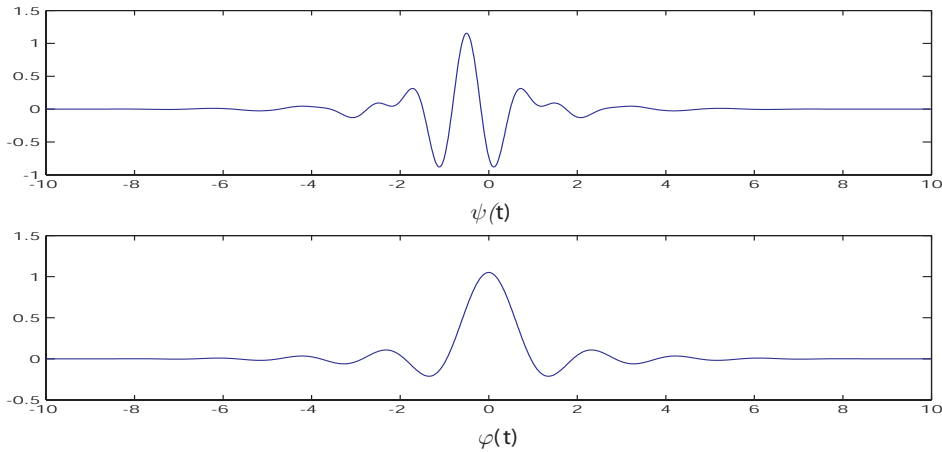
Figure 5.5.4: Meyer wavelet and scaling function.

As an example of the direct constructions of orthonormal wavelet basis, the Meyer wavelet basis is given by

$$\hat{\psi} = \begin{cases} 2^{-1/2}\, e^{i\omega/2}\, \sin[\frac{\pi}{2}\, \nu\,(\frac{3|\omega|}{2\pi} - 1)] & \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ 2^{-1/2}\, e^{i\omega/2}\, \cos[\frac{\pi}{2}\, \nu\,(\frac{3|\omega|}{4\pi} - 1)] & \frac{4\pi}{3} \leq |\omega| \leq \frac{8\pi}{3} \\ 0 & \text{otherwise,} \end{cases} \qquad (5.5.25)$$

where ν is a C^n function that is on the interval $[0,1]$ and satisfies

$$\forall x \in [0,1], \quad \nu(x) + \nu(1-x) = 1. \qquad (5.5.26)$$

Meyer wavelet is a tight frame, and its transform is generally implemented in the Fourier domain. The construction of (5.5.25) is a frequency band-limited function whose Fourier transform is smooth, and this smoothness provides a much faster asymptotic decay in time. As a result, the combining positive and negative frequencies gets rid of the redundancy of the frame (i.e., tight frame with frame bound 1), and the orthonormality is achieved by the smooth function $\nu(x)$. We refer to [100] for proving that the wavelets constitute an orthonormal basis. Figure 5.5.4 shows an example of the $\psi(x)$, with the choice $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$.

In fact, the simple statement of (5.5.21) opened an important exchange between harmonic analysis and discrete signal processing. For example, orthogonal wavelets dilated by 2^m carry signal variations at the resolution 2^{-m}, and the decomposition of a function in such an orthonormal wavelet basis can be computed with a *quadrature mirror filter bank*. It means that the construction of the orthonormal bases is directly related to the *multiresolution analysis* and the *multirate filter banks* which are discussed in following sections. Practically, using the multiresolution analysis developed by S. Mallat and Y. Meyer, all the constructions of the orthonormal wavelet bases referred in this section can be more satisfyingly explained.

5.6 Multiresolution Analysis and Wavelets

Stephane Mallat (1985) discovered some close connections among the quadrature mirror filters, the pyramid algorithms and the orthonormal wavelet bases. With the concept of orthonormal wavelet bases as a tool, he tried to describe mathematically the "increment in information" needed to go from a coarse resolution to a higher resolution approximation. On the basis of this insight, Mallat [157] and Meyer [152] formulated such successive approximations and created a framework for wavelet expansions called *multiresolution analysis (MRA)*, which provides a natural framework to understand wavelet decompositions and new construction methods of orthonormal wavelet bases.

One of the most important contributions of MRA was to establish links with the ideas and methods that are related to the wavelet analysis and independently developed in various fields, for example, in quantum mechanics, mathematics and signal processing. Indeed, the MRA provided a formalized bridge between the ideas of other disciplines and then catalyzed the rediscovery of wavelets with comprehensive studies and the proliferation of its application areas, indeed.

5.6.1 Multiresolution approximations of closed subspaces

A multiresolution analysis (MRA) of $L^2(\mathbb{R})$ is defined as a sequence of closed subspaces V_j of $L^2(\mathbb{R})$, $j \in \mathbb{Z}$, with the following properties [157] [155] [100]:

$$f(t) \in V_j \iff f(t - 2^j k) \in V_j \,, \tag{5.6.1}$$

$$V_{j+1} \subset V_j \,,^4 \tag{5.6.2}$$

$$f(t) \in V_j \iff f(2^j t) \in V_0 \,, \tag{5.6.3}$$

$$\lim_{j \to +\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\} \,, \tag{5.6.4}$$

$$\lim_{j \to -\infty} V_j = \overline{\bigcup_{j=-\infty}^{+\infty} V_j} = L^2(\mathbb{R}) \,, \tag{5.6.5}$$

and there exist a scaling function $\varphi \in V_0$ such that

the collection $\{\varphi(t - n)\}_{n \in \mathbb{Z}}$ is an orthonormal[5]basis of V_0 . (5.6.6)

The space V_j can be viewed as the set of all possible approximations of functions at the resolution 2^{-j}. MRA is then obtained by computing the approximation of signals at various resolutions with orthogonal projections onto different spaces $\{V_j\}_{j \in \mathbb{Z}}$. In order to calculate the approximation, the orthogonal basis of each space V_j is generated by dilating and translating a single function φ called *scaling function*, i.e.,

$$\varphi_{j,n}(t) = 2^{-j/2} \varphi(2^{-j} t - n), \quad n \in \mathbb{Z}. \tag{5.6.7}$$

When the resolution 2^{-j} goes to 0, the property (5.6.4) implies that we lose all the details of f and

$$\lim_{j \to +\infty} \|\mathcal{P}_j f\| = 0, \tag{5.6.8}$$

where $\mathcal{P}_j$ denotes the orthogonal projection operator onto V_j. When the resolution 2^{-j} goes $+\infty$, the completeness property (5.6.5) ensures that

$$\lim_{j \to -\infty} \mathcal{P}_j f = f \text{ in } L^2(\mathbb{R}), \quad \lim_{j \to -\infty} \|f - \mathcal{P}_j f\| = 0. \tag{5.6.9}$$

When the resolution 2^{-j} increases, the decay rate of the approximation error $\|f - \mathcal{P}_{V_j} f\|$ depends on the regularity of f.

The inner products $\langle f, \varphi_{j,k} \rangle$, which provide a discrete approximation at the scale 2^j, can be rewritten by convolution product as following,

$$\langle f, \varphi_{j,n} \rangle = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} f(t)\, \varphi(2^{-j} t - n)\, dt = f * \bar{\varphi}_j(2^j n) \,, \tag{5.6.10}$$

---

[4]As in the ladder of Sobolev spaces, we use also this nesting order which is the most practical for numerical analysis, even though the reverse ordering (i.e. $V_j \subset V_{j+1}$) is used in many literatures.

[5]Since an orthogonal MRA is of interest for most practical applications, we use only orthonormal basis instead of Riesz basis to define the MRA

where $\bar{\varphi}_j(t) = 2^{-j/2}\varphi(2^{-j}t)$. Since the frequency contents of φ are typically concentrated in $[-\pi, \pi]$, the calculation of the inner products $\langle f, \varphi_{j,n} \rangle$ is a lowpass filtering of signal f sampled at interval 2^j.

From the multiresolution causality property (5.6.2) with (5.6.6), it is immediate that the scaling function φ may be expressed as a linear combination of the basis function of V_{-1}. Since $\varphi \in V_0 \subset V_{-1}$, and the $\varphi_{-1,n}$ are an orthonormal basis in V_{-1}, a sequence $h[n] \in l^2(\mathbb{Z})$ exists such that φ satisfies

$$\varphi(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} h[n] \, \varphi(2t - n), \qquad (5.6.11)$$

where

$$h[n] = \langle \varphi(t), \sqrt{2}\,\varphi(2t - n) \rangle \Rightarrow \sum_{n \in \mathbb{Z}} h[n] = \sqrt{2}. \qquad (5.6.12)$$

The equation (5.6.11) is called *refinement equation*[6]. The sequence $h[n]$ will be interpreted as a discrete filter. The Fourier transform of (5.6.11) yields

$$\Phi(\omega) = \frac{1}{\sqrt{2}} H(\omega/2)\Phi(\omega/2), \qquad (5.6.13)$$

where H is a 2π-periodic function defined by $H(\omega) = \sum_{n=-\infty}^{+\infty} h[n]e^{-jn\omega}$. The periodicity gives $H(n\pi) = 0$ for all odd integer n. Then the equation (5.6.13) yields

$$\Phi(2k\pi) = \delta_k, \qquad k \in \mathbb{Z}. \qquad (5.6.14)$$

Practically, this is the first of the so-called "Strang-Fix conditions"[7]. To be able to use the collection $\{\varphi(t - n)\}$ to approximate even the simplest functions such as $f(t) \equiv 1$, it is natural to assume that the scaling function and its integer translates form a *partition of unity*, which we may normalize to 1:

$$\sum_{n \in \mathbb{Z}} \varphi(t - n) = \sum \varphi(t) = \Phi(0) = 1. \qquad (5.6.15)$$

By Poisson's summation formula, (5.6.15) is equivalent to (5.6.14) essentially. The orthogonality of the collection $\{\varphi(t - n)\}_{n \in \mathbb{Z}}$ in property (5.6.6) means that

$$\langle \varphi(t), \varphi(t - n) \rangle = \delta_n. \qquad (5.6.16)$$

---

[6]It goes by several different names: *dilation equation, scaling equation or two-scale equation*

[7]The Strang and Fix conditions characterize the approximating properties of a shift-invariant localized operator by its ability to reconstruct polynomials. In MRA, it is used to relate the number of vanishing moments of a wavelet to the order of multiresolution approximation. [158] [159] [160]

As the relation between (5.6.14) and (5.6.15), the orthogonality (5.6.16) is equivalent to

$$\sum_{k \in \mathbb{Z}} |\Phi(\omega + 2k\pi)|^2 = 1. \tag{5.6.17}$$

When we apply (5.6.13) recursively, we obtain the infinite product formula:

$$\Phi(\omega) = \Phi(2^{-j}\omega) \prod_{j=1}^{+\infty} \frac{H(2^{-j}\omega)}{\sqrt{2}} . \tag{5.6.18}$$

If $\varphi(t)$ is integrable, then $\Phi(\omega)$ is continuous for all ω. It means also that $\lim_{j \to +\infty} \Phi(2^{-j}\omega) = \Phi(0)$. Since $\Phi(0) = 1$, (5.6.18) can be simplified such that

$$\Phi(\omega) = \prod_{j=1}^{+\infty} \frac{H(2^{-j}\omega)}{\sqrt{2}}. \tag{5.6.19}$$

(5.6.19)[8] shows that scaling function $\varphi(t)$ can be constructed directly from the $h[n]$. In many applications, we never need the explicit expression of the scaling function, indeed. However, there are fast algorithms that use the refinement equation (5.6.11) to evaluate the scaling function φ at dyadic points [155][161][159][162]. To guarantee that (5.6.19) is the Fourier transform of a scaling function, the Fourier series of $h[n] = \langle \varphi(t), \sqrt{2}\varphi(2t - n) \rangle$ satisfies

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 2 \tag{5.6.20}$$

$$H(0) = \sqrt{2}. \tag{5.6.21}$$

The relation between the two necessary conditions is that

$$H(\pi) = 0 \quad \Rightarrow \quad \sum_{n \in \mathbb{Z}} (-1)^n h[n] = 0. \tag{5.6.22}$$

Since the condition (5.6.20) is identical to the quadrature condition (4.2.17) of discrete filter in section 4.2.3, it is proved that any scaling function is specified by the CMF.

---

[8]The convergence of this product is proved in [100] [98].

5.6.2   Orthogonal wavelet function and detail spaces

By considering a space W_j to be the orthogonal complement of V_j in V_{j-1}, we can construct wavelets from MRA. The space W_j satisfies

$$V_{j-1} = V_j \oplus W_j \,, \tag{5.6.23}$$

where the symbol $\oplus$ stands for direct sum. Because $V_j \subset V_{j-1}$, the orthogonal projection of f on V_{j-1} can be decomposed as

$$\mathcal{P}_{j-1}f = \mathcal{P}_j f + \mathcal{Q}_j f \,, \tag{5.6.24}$$

where $\mathcal{Q}_j$ denotes the orthogonal projection operator onto W_j. The space W_j contains the "detail" information of f that is needed to go from a coarser approximation at resolution 2^{-j} to a finer approximation at resolution 2^{-j-1}. Since $W_j \subset V_{j-1} \perp W_{j-1}$, it follows that W_j is also orthogonal to W_{j-1}. It is therefore immediate that all the subspaces W_j are mutually orthogonal unlike the subspaces V_j. Consequently, it implies that we can decompose

$$L^2(\mathbb{R}) = \bigoplus_j W_j \,. \tag{5.6.25}$$

Similarly to the definition of scaling function φ, if a collection of closed subspaces satisfies (5.6.1)-(5.6.6), there exist wavelet function ψ such that $\{\psi(t-n)\}_n \in \mathbb{Z}$ is an orthonormal basis of W_0. For all scales, the whole collection $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}}$, $\psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t - n)$, is then an orthonormal basis of $L^2(\mathbb{R})$. Since $\psi_{0,n} \in W_0 \subset V_{-1}$, a sequence $g[n] \in l^2(\mathbb{Z})$ exists such that

$$\psi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} g[n]\varphi(2t - n) \,. \tag{5.6.26}$$

The Fourier transform of (5.6.26) is given by

$$\Psi(\omega) = \frac{1}{\sqrt{2}}G(\omega/2)\Phi(\omega/2) \,, \tag{5.6.27}$$

where G is a 2π-periodic function, i.e., $G(\omega) = \sum_{n=-\infty}^{+\infty} g[n]e^{-jn\omega}$. If the sequence $h[n]$ is the corresponding CMF, the sequence $g[n]$ can be directly obtained from $h[n]$, i.e.,

$$G(\omega) = e^{-j\omega}H^*(\omega + \pi) \,, \tag{5.6.28}$$

and its inverse Fourier transform is

$$g[n] = (-1)^{1-n}h[1 - n] \,. \tag{5.6.29}$$

This mirror relation enables us to consider a fast wavelet transform.

The sufficient conditions for a orthogonal MRA and for a orthogonal wavelet, i.e., $\langle \psi(t), \varphi(t-n)\rangle = 0$ and $\langle \psi(t), \psi(t-n)\rangle = \delta_n$, are equivalent to

$$\sum_{k\in\mathbb{Z}} \Psi(\omega + 2k\pi)\Phi^*(\omega + 2k\pi) = 0\,, \tag{5.6.30}$$

$$\sum_{k\in\mathbb{Z}} |\Psi(\omega + 2k\pi)|^2 = 1\,, \tag{5.6.31}$$

respectively. From these conditions, the necessary and sufficient conditions on $G(\omega)$ for designing an orthogonal wavelet are given as

$$|G(\omega)|^2 + |G(\omega + \pi)|^2 = 2, \tag{5.6.32}$$

$$G(\omega)H^*(\omega) + G(\omega + \pi)H^*(\omega + \pi) = 0. \tag{5.6.33}$$

We observe that once the filter H has been chosen, it completely determines the functions φ and ψ and therefore, the multiresolution analysis. Moreover, in properly constructed algorithms, the values of the functions φ and ψ are usually never computed. Due to the recursive definition of the wavelet bases, all the manipulations are performed with the CMFs H and G, even if they involve quantities associated with φ and ψ.

5.6.3   Relation between regularity and vanishing moments

Regularity condition states that the wavelet function should have some smoothness, which plays an important role in compression applications, and concentration in both time and frequency domains. A higher degree of smoothness corresponds to better frequency localization of the filters, and is strongly desired in numerical analysis applications. From the refinement equation it is clear that the regularity of φ and ψ is the same. Tchamitchian [163] related the uniform Lipschitz regularity of φ and ψ to the number of zeros of $H(\omega)$ at $\omega = \pi$. To make ψ and φ reasonably regular, $H(\omega)$ should be of the form

$$H(\omega) = \sqrt{2}\left(\frac{1 + e^{-j\omega}}{2}\right)^M L(\omega)\,, \tag{5.6.34}$$

with $M \geq 1$. $L(\omega)$ is a 2π-periodic smooth function with $L(0) = 1$ and $L(1) \neq 0$, since $H(\omega)$ has a root of multiplicity M at $\omega = \pi$. This factorization of $H(\omega)$ implies that the regularity of φ and ψ increases with M, which is equal to the number of vanishing moments of ψ. Note that when $L(\omega) = 1$, the choice $H(\omega) = \left(\frac{1 + e^{-j\omega}}{2}\right)^M$ corresponds to the basic spline scaling function. The moments of φ and ψ are defined by

$$\mathcal{M}_\varphi = \int_{-\infty}^{\infty} t^m \varphi(t)\,dt\,, \quad \mathcal{M}_\psi = \int_{-\infty}^{\infty} t^m \psi(t)\,dt\,, \tag{5.6.35}$$

with $0 \le m < M$. The vanishing moments property means that the basis functions are chosen to be orthogonal to any polynomial of degree less than M, namely, if the collection of functions $\{\psi(t-k)\}$ is an orthonormal basis of W_0, then

$$\int_{-\infty}^{\infty} t^m \, \psi(t) \, dt = 0, \quad m = 0, \dots, M-1. \tag{5.6.36}$$

If a signal f is reasonably regular and ψ has enough vanishing moments then the wavelet coefficients $|\langle f, \psi_{j,k}\rangle|$ are small at the fine scales 2^j. If $m = 0$ in (5.6.36), then ψ has a vanishing integral (admissibility condition of CWT) which allows us to give a precise characterization of the functions with a certain smoothness in terms of the decay of the CWT. Similarly, to obtain the characterization of functions with more smoothness, the ψ needs to have more vanishing moments. According to the Strang-Fix conditions [158], this is closely related to the property that the scaling function and its translates can be used to represent polynomials. By using (5.6.13) and (5.6.14), the fact that $H(\omega)$ having number of zeros at $\omega = \pi$ implies that

$$i^m \, \Phi^m(2k\pi) = \delta_k \, \mathcal{M}_\varphi, \qquad 0 \le m < M. \tag{5.6.37}$$

It follows that

$$\mathcal{M}_\varphi = \sum_{n \in \mathbb{Z}} (t-n)^m \, \varphi(t-n), \qquad 0 \le m < M. \tag{5.6.38}$$

Thus this proves that ψ has M vanishing moments if and only if any polynomial with degree smaller than M can be written as a linear expansion of $\{\varphi(t-n)\}_{n \in \mathbb{Z}}$.

The decay of $|\varphi(t)|$ and $|\psi(t)|$ implies that $\Phi(\omega)$ and $\Psi(\omega)$ are M times continuously differentiable. Hence the rate of decay depends also on the number of vanishing moments of the functions of the basis. The Haar functions have only one vanishing moment, and for this reason the gain in the decay is insufficient to make computing in the Haar basis practical. To have a faster decay, it is necessary to use basis functions with several vanishing moments. The vanishing moments are responsible for attaining practical algorithms, i.e. controlling the constants in the complexity estimates of the fast algorithms.

5.6.4   Compactly supported wavelets

If a signal f has an isolated singularity at the inside of the support of $\psi_{j,n}(t)$, then the coefficients may have a large amplitude. To minimize the number of high amplitude coefficients we must reduce the support size of ψ. To minimize the size of the support, we must synthesize CMFs with as few non-zero coefficients as possible. Form the definition

of the $h[n]$, $h[n] = \langle \varphi(t), \sqrt{2}\psi(2t - n)\rangle$, it is immediate that if $h[n]$ has finite impulse response (FIR) CMFs then φ and ψ have also compact support. For compactly supported φ the $H(\omega)$ becomes a trigonometric polynomial:

$$H(\omega) = \sum_{n=0}^{N-1} h[n]\, e^{-jn\omega}. \tag{5.6.39}$$

We recall the factorization formula (5.6.34),

$$H(\omega) = \sqrt{2}\left(\frac{1 + e^{-j\omega}}{2}\right)^M L(e^{-j\omega}), \tag{5.6.40}$$

where $(1 + e^{-j\omega})^M$ is a minimum size polynomial having M zeros at $\omega = \pi$. Since $|H(\omega)|^2$ and $|L(e^{-j\omega})|^2$ are polynomial in $\cos\omega$, we can rewrite (5.6.40) as following

$$|H(\omega)|^2 = 2\left(\cos\frac{\omega}{2}\right)^{2M} P\left(\sin^2\frac{\omega}{2}\right). \tag{5.6.41}$$

In terms of P, the orthonormal condition (5.6.20) is equivalent to

$$(1 - y)^M P(y) + y^M P(1 - y) = 1, \tag{5.6.42}$$

for any $y = \sin^2(\omega/2) \in [0, 1]$.

By using the Bezout theorem [100] we can obtain the solution $P(y) \geq 0$ of minimum degree, which enables to minimize the number of non-zero terms of the finite Fourier series $H(\omega)$. Let us study this theorem in some detail. If $Q_1(y)$ and $Q_2(y)$ are two polynomials of degrees p_1 and p_2 with no common zeros, then there exist unique polynomials $P_1(y)$ and $P_2(y)$ of degrees $p_2 - 1$ and $p_1 - 1$ so that

$$P_1(y)Q_1(y) + P_2(y)Q_2(y) = 1. \tag{5.6.43}$$

By applying this to (5.6.42), there exist two unique polynomials $P_1(y)$ and $P_2(y)$ such that

$$(1 - y)^M P_1(y) + y^M P_2(y) = 1. \tag{5.6.44}$$

Substituting $P_2(y) = P_1(1-y) = P(1-y)$ leads to the unique minimum degree solution of (5.6.42) as

$$P(y) = \sum_{k=0}^{M-1} \binom{M - 1 + k}{k} y^k, \tag{5.6.45}$$

with $P(y) \geq 0$ for $y \in [0, 1]$. Higher degree solution $P_+ \geq 0$ can also be obtained by the factorization of

$$P_+(y) = P(y) + y^M R(\tfrac{1}{2} - y), \tag{5.6.46}$$

where R is an odd polynomial. To construct a minimum degree polynomial given by

$$L(e^{-j\omega}) = \sum_{k=0}^{p} l_k e^{-jk\omega} = l_0 \prod_{k=0}^{p} (1 - a_k e^{-j\omega}) \tag{5.6.47}$$

with $|L(e^{-j\omega})|^2 = P(\sin^2(\omega/2))$, we factor the $|L(e^{-j\omega})|^2$ as

$$|L(e^{-j\omega})|^2 = L(e^{-j\omega})L(e^{j\omega}) \tag{5.6.48}$$

$$= P\left(\frac{2 - e^{j\omega} - e^{-j\omega}}{4}\right) = Q(e^{-j\omega}), \tag{5.6.49}$$

which can then be solved by extending it to the whole complex plane with the variable $z = e^{j\omega}$:

$$L(z)L(z^{-1}) = l_0^2 \prod_{k=0}^{p} (1 - a_k z)(1 - a_k z^{-1}) \tag{5.6.50}$$

$$= P\left(\frac{2 - z - z^{-1}}{4}\right) = Q(z). \tag{5.6.51}$$
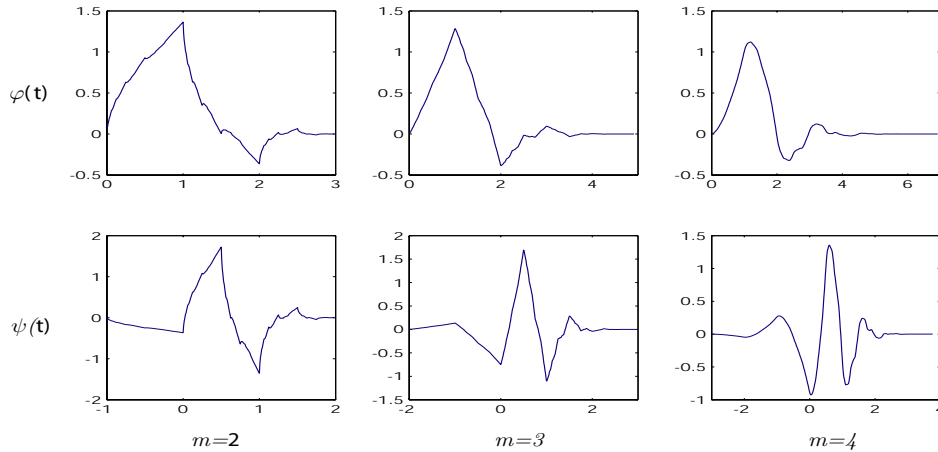
Figure 5.6.1: Daubechies orthogonal wavelets ψ and scaling function φ with m vanishing moments.

In order to obtain the polynomial $L(z)$ of the form (5.6.40), it remains only to find the square root $L(z)$ of $Q(z)$, referred to as *spectral factorization*. To prove the existence of $L(z)$ of $Q(z)$, we refer to [100] [164]. In the result, we have the polynomial $L(z)$ with the minimum degree $p = M - 1$, with $l_0^2 = Q(0) = P(1/2) = 2^{M-1}$. It leads to the filter

h of minimum size $N = M + p + 1 = 2M$. All the works till now can then be seen as to synthesize a FIR CMF h of size $2M$.

Daubechies compactly supported orthonormal wavelets are constructed by the spectral factorization of (5.6.46), with different choices M and R. Figure 5.6.1 displays φ and ψ, for $M = 2, 3, 4, 5$ and $R \equiv 0$, i.e. minimum size support. Due to the choice of minimum phase square root of $Q(e^{-j\omega})$, Daubechies wavelets are highly non-symmetric as showed in Figure 5.6.1. A more symmetric wavelets, called *symmlet*, can be constructed by obtaining an almost linear complex phase [155]. Lina and Mayrand [165] constructed the complex CMFs with a compact support and a linear phase, but they have redundant coefficients when the signal is real. Determining the regularity of compactly supported orthonormal wavelets is harder than, e.g. the case of the Meyer or the Battle-Lemarié wavelets. It is due to the partly different degree of regularity of them. Several methods for the problem are suggested in [155] [161] [162] [166].

5.6.5   Biorthogonal wavelet bases

Note again that the Haar wavelet is the only real-valued wavelet that satisfies the linear-phase (symmetry) and orthogonality (paraunitarity) simultaneously. By relaxing the orthogonality conditions to biorthogonal properties, we can obtain wavelets with more flexibility that are limited in the case of orthogonal wavelets. In the biorthogonal case, a dual scaling function $\tilde{\varphi}$ and a dual wavelet $\tilde{\psi}$ exist that generate a dual multiresolution analysis with subspaces $\widetilde{V}_j$ and $\widetilde{W}_j$, such that

$$\widetilde{V}_j \perp W_j \,, \quad V_j \perp \widetilde{W}_j \,, \quad \text{and} \quad \widetilde{W}_j \perp W_{j'} \quad (j \neq j')\,. \tag{5.6.52}$$

Namely the biorthogonality of the decomposition and reconstruction wavelets implies that W_j is not orthogonal to V_j but is to $\widetilde{V}_j$ whereas $\widetilde{W}_j$ is not orthogonal to $\widetilde{V}_j$ but is to V_j. The biorthogonality of (5.6.52) is equivalent with

$$\langle \tilde{\varphi}(t), \psi(t - n) \rangle = \langle \tilde{\psi}(t), \varphi(t - n) \rangle = 0\,, \tag{5.6.53}$$

$$\langle \tilde{\varphi}(t), \varphi(t - n) \rangle = \langle \tilde{\psi}(t), \psi(t - n) \rangle = \delta_n\,, \tag{5.6.54}$$

$$\langle \tilde{\varphi}_{j,n}, \varphi_{j,n'} \rangle = \delta_{n-n'}\,, \tag{5.6.55}$$

$$\langle \tilde{\psi}_{j,n}, \psi_{j',n'} \rangle = \delta_{j-j'}\delta_{n-n'}\,, \tag{5.6.56}$$

for any $(j, j', n, n') \in \mathbb{Z}$. Since the dual functions define also a MRA, they have to satisfy

$$\tilde{\varphi}(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} \tilde{h}[n]\, \tilde{\varphi}(2t - n)\,, \tag{5.6.57}$$

$$\tilde{\psi}(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} \tilde{g}[n]\, \tilde{\varphi}(2t - n)\,. \tag{5.6.58}$$

Using the same fashion as (5.6.28), (5.6.30) and (5.6.31) the necessary conditions are given by

$$H^*(\omega)\widetilde{H}(\omega) + H^*(\omega + \pi)\widetilde{H}(\omega + \pi) = 2\,, \tag{5.6.59}$$

$$G^*(\omega)\widetilde{G}(\omega) + G^*(\omega + \pi)\widetilde{G}(\omega + \pi) = 2\,, \tag{5.6.60}$$

$$H^*(\omega)\widetilde{G}(\omega) + H^*(\omega + \pi)\widetilde{G}(\omega + \pi) = 0\,, \tag{5.6.61}$$

$$G^*(\omega)\widetilde{H}(\omega) + G^*(\omega + \pi)\widetilde{H}(\omega + \pi) = 0\,, \tag{5.6.62}$$

and

$$G(\omega) = e^{-j\omega}\widetilde{H}^*(\omega + \pi)\,, \quad \widetilde{G}(\omega) = e^{-j\omega}H^*(\omega + \pi). \tag{5.6.63}$$

The projection operators take the form

$$\mathcal{P}_j f = \sum_{n \in \mathbb{Z}} \langle f, \tilde{\varphi}_{j,n} \rangle \, \varphi_{j,n} \quad \text{and} \quad \mathcal{Q}_j f = \sum_{n \in \mathbb{Z}} \langle f, \tilde{\psi}_{j,n} \rangle \, \psi_{j,n}\,. \tag{5.6.64}$$

and

$$f = \sum_{j,n \in \mathbb{Z}} \langle f, \tilde{\psi}_{j,n} \rangle \, \psi_{j,n} = \sum_{j,n \in \mathbb{Z}} \langle f, \psi_{j,n} \rangle \, \tilde{\psi}_{j,n}\,. \tag{5.6.65}$$

If $H(\omega)$ and $G(\omega)$ are FIR CMFs, φ and ψ are compactly supported, whereas the dual functions are, in general, not compactly supported. The number of vanishing moments of ψ and $\tilde{\psi}$ depends on the number of zeros at $\omega = \pi$ of $H(\omega)$ and $\widetilde{H}(\omega)$. Since $G(\omega) = e^{-j\omega}\widetilde{H}^*(\omega + \pi)$, the number of vanishing moments of $\tilde{\psi}$ is equal to the number M of zeros at $\omega = \pi$ of $H(\omega)$.

As mentioned, an advantage of the biorthogonal wavelets is that the construction of linear-phase filters or combination of symmetry and antisymmetry is possible. If $h[n]$ and $\tilde{h}[n]$ have an odd number of non-zero samples and are symmetric about $n = 0$, then $\varphi(t)$ and $\tilde{\varphi}(t)$ are symmetric about $t = 0$, while $\psi(t)$ and $\tilde{\psi}(t)$ are symmetric with respect to a shifted center. If $h[n]$ and $\tilde{h}[n]$ have an even number of non-zero samples and are symmetric about $n = 1/2$, then $\varphi(t)$ and $\tilde{\varphi}(t)$ are symmetric about $t = 1/2$, while $\psi(t)$ and $\tilde{\psi}(t)$ are antisymmetric with respect to a shifted center.

5.6.6 Discrete-time wavelet transform (DWT)

From the equations (5.6.11), (5.6.16) and (5.6.26) we see that

$$h[n - 2k] = \langle \varphi(t - k), \varphi(2t - n) \rangle\,, \tag{5.6.66}$$

$$g[n - 2k] = \langle \psi(t - k), \varphi(2t - n) \rangle\,. \tag{5.6.67}$$

This leads to a form of multirate decomposition as a discrete wavelet transform (DWT), i.e.

$$\varphi(2t - n) = \sum_{k \in \mathbb{Z}} h[n - 2k]\,\varphi(t - k) + \sum_{k \in \mathbb{Z}} g[n - 2k]\,\psi(t - k)\,. \tag{5.6.68}$$

Since $V_j = V_{j+1} \oplus W_{j+1}$, we can express a function $v_j \in V_j$ as the sum of a function $v_{j+1} \in V_{j+1}$ and a function $w_{j+1} \in W_{j+1}$:

$$v_j(t) = v_{j+1}(t) + w_{j+1}(t) = \sum_{n \in \mathbb{Z}} a_j[n]\,\varphi_{j,n} \tag{5.6.69}$$

$$= \sum_{n \in \mathbb{Z}} a_{j+1}[n]\,\varphi_{j+1,n} + \sum_{n \in \mathbb{Z}} d_{j+1}[n]\psi_{j+1,n}\,, \tag{5.6.70}$$

where $a_j[n] = \langle f, \varphi_{j,n} \rangle$ and $d_j[n] = \langle f, \psi_{j,n} \rangle$. The coefficients $a_j[n]$ and $d_j[n]$ can then be calculated as

$$a_{j+1}[k] = \sum_{n \in \mathbb{Z}} h[n - 2k]\,a_j[n]\,, \tag{5.6.71}$$

$$d_{j+1}[k] = \sum_{n \in \mathbb{Z}} g[n - 2k]\,d_j[n]\,, \tag{5.6.72}$$

for the decomposition, and

$$a_j[k] = \sum_{n \in \mathbb{Z}} h[k - 2n]\,a_{j+1}[n] + \sum_{n \in \mathbb{Z}} g[k - 2n]\,d_{j+1}[n]\,, \tag{5.6.73}$$

for the reconstruction. It is easy to see that the same algorithm can be applied to the dual functions. This process is equivalent to the cascade discrete filtering with subsampling and upsampling, as illustrated in Figure 5.6.2. It is often called Mallat's algorithm [83]. Since the complexity of this operation is $\mathcal{O}(n)$[9], i.e. the amount of work is proportional to the signal length, this algorithm is also known as the *fast wavelet transform* (FWT).

If we iterate the decomposition up to the largest scale 2^J, we obtain an multirate orthogonal wavelet representation of a signal f with resolutions $2^{-J} < 2^{-j} \leq 2^{-L}$. The finest approximation a_L can then be composed by following

$$a_L = a_J \bigoplus_{j}^{L < j \leq J} d_j. \tag{5.6.74}$$

---

[9]as a comparison, the fast Fourier transform (FFT) has a complexity of $\mathcal{O}(n \log n)$
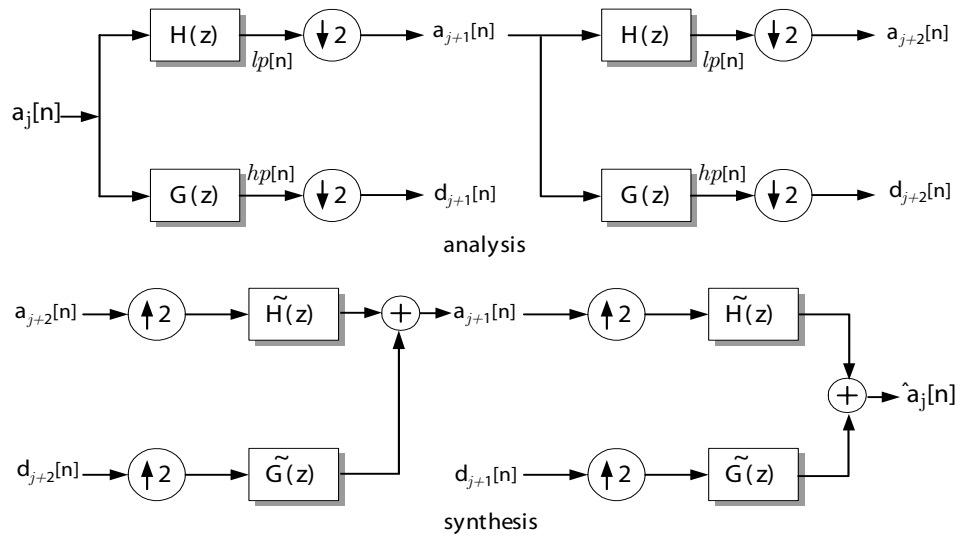
Figure 5.6.2: Cascade structure for fast wavelet transform.

The economy of the wavelet transform stems from the fact that the DWT tends to compress real-world signals into just a few coefficients of large magnitude. Compression follows form the vanishing moments property of wavelets, which guarantees that the wavelet coefficients of low-order polynomial signals are zero. Thus, if a signal is exactly polynomial, it can be completely described using scaling coefficients alone. In more realistic situations, the signal will not be polynomial, but may be well-approximated by a piecewise polynomial function. Because wavelet functions also have localized support, most of the wavelet coefficients of such a signal will be zero except those corresponding to wavelets having support near the breakpoints of the polynomial segments.

5.7   Filter Banks and Wavelets

Paralleling the advances of wavelet analysis in pure and applied mathematics were those of multirate filter bank in signal processing, but in the context of discrete-time signals. Discrete-time wavelet transform (DWT),[10] which we use practically in applications, can be computed by using multirate filter banks. The construction of the wavelet bases is equivalent to designing two-channel filter banks. In computer vision, the filter bank

---

[10]It must be cleared that discrete wavelets or wavelet series are not discrete-time processing, but they are merely discrete representation of the time-scale plane of CWT

has been used to obtain the successive approximations of images starting from a coarse resolution and going to finer resolution. This successive approximation view is closely related to the multiresolution analysis (MRA), which is a natural framework for the construction of the orthonormal wavelet bases.

5.7.1 QMF and wavelets

We have seen with (5.3.3) that the wavelet has a bandpass like spectrum. If wavelet can be considered as a bandpass filter, then a series of dilated wavelets can be seen as a bandpass filter bank. In a particular configuration, that is, when the filter bank has octave bands, one obtains a discrete wavelet series. In signal processing, such an octave band or logarithmic spectrum has been popular less for its mathematical properties than because it is more nature for certain applications such as audio compression since it emulates the hearing process of the ears.

To accomplish the octave division of the frequency space, a standard way is using the QMF such as in Figure 4.2.2(b), which consists of a pair of lowpass and highpass filter. The symmetry of the QMFs enables to facilitate minimizing leakage between filters and to minimize phase distortion. If we can create a wavelet function whose spectral content mimics the QMFs and which also satisfy the additional requirements for valid wavelet functions, then we can use these functions as a basis set for the wavelet transform[11]. Thus one way to understand the need for several types of wavelet functions is to realize that they correspond to different shapes of the QMFs in the Fourier domain, i.e. different ways to bandlimit the signal. In result, if one wavelet can be seen as a bandpass filter and a scaling function is a lowpass filter, then a series of dilated wavelets together with the scaling function can be seen as a filter bank.

5.7.2 Scaling function and lowpass filter

Even with discrete wavelets (wavelet series) in previous section 5.5, there still exist the problem of an infinite number of scalings and translations to calculate the wavelet transform. This problem can be solved by using scaling function, introduced by Mallat [83]. Scaling function sets lower bound for the wavelets as a lowpass filter[12] in Figure 5.7.1.

---

[11]The STFT can also be described in terms of lowpass and highpass filters but here they divide the frequency axis equally and not proportionally as for the wavelet QMFs.

[12]Because of the lowpass nature of the scaling function spectrum, it is sometimes referred to as the averaging filter

If we analyze a signal using the combination of scaling function and wavelets as in Figure 5.7.1, the scaling function undertakes the low spectrum band which otherwise should be covered by all the wavelets up to unlimited scale, while the rest is done by the wavelets. Consequently we can limit the number of wavelets from an infinite number to a finite number. In that case, the highpass filter corresponds to the wavelet function whereas lowpass filter corresponds to the scaling function. Scaling function is defined by

$$\varphi_{m,n}(x) = 2^{-m/2}\varphi(2^{-m}x - n). \tag{5.7.1}$$

Its Fourier transform is

$$\Phi(\omega) = \frac{\Theta(\omega)}{\left(\sum_{k=-\infty}^{\infty} |\Theta(\omega + 2k\pi)|^2\right)^{1/2}}, \tag{5.7.2}$$

where the function $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ is a orthonormal Riesz basis.
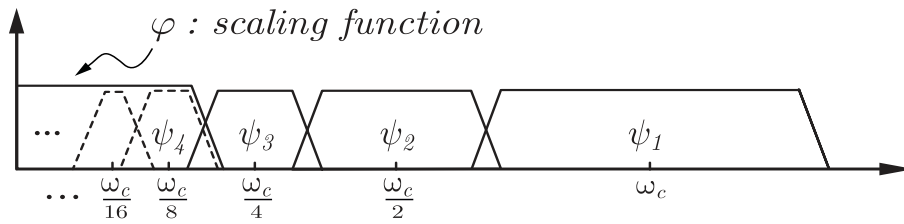
Figure 5.7.1: Lower bounding of an infinite set of wavelets by scaling function.

For a lowpass filter, the polynomial $H(z) = \sum h[n]e^{-jn\omega}$ is near zero at the $\omega = \pi$ ($z = -1$ in z-plane). The lowest frequency $\omega = 0$ ($z = 1$) passes through an ideal filter provided $H(1) = 1$. The condition $H(-1) = 0$ is fundamental in wavelet theory. It must hold exactly to have any chance of continuous scaling functions and wavelets. Therefore the scaling function has to satisfy admissibility condition similar to (5.3.3),

$$\int \varphi(x)\,dx = 1, \tag{5.7.3}$$

which means that the 0th moment of the scaling function can not vanish.

The scaling function $\varphi(x)$ can be written in terms of itself on a smaller scale $\varphi(2x)$,

$$\varphi(x) = \sum_{k=0}^{M-1} c_k\,\varphi(2x - k), \tag{5.7.4}$$

where the c_k are numerical constants that define the scaling function. This functional equation is called *refinement equation* or *dilation equation*. The range of the summation

is determined by the specified number of nonzero coefficients M. (5.7.4) is orthogonal to its translation, i.e., $\int \varphi(x)\varphi(x - k)\, dx = 0$. The corresponding wavelet function $\psi(x)$ which is orthogonal to its dilations, i.e., $\int \psi(x)\psi(2x - k)\, dx = 0$, has a recursive relation to the scaling function with reversed order of the coefficients,

$$\psi(x) = \sum_k (-1)^k\, c_{1-k}\, \varphi(2x - k)\,. \tag{5.7.5}$$

The wavelet function ψ is dependent upon the solution of scaling function φ. The value of the coefficients is determined by constraints of orthogonality and normalization [161]. Normalization requires that

$$\sum_k c_k c_{k-2l} = 2\delta_{0l}\,, \tag{5.7.6}$$

which means that the sum of the squares of all coefficients is two. From the above conditions,

$$\sum_k (-1)^k\, c_{1-k} c_{k-2l} = 0\,. \tag{5.7.7}$$

Hence, it turns out that the above conditions of scaling function correspond to the conditions of perfect reconstruction filter bank, and that the c_k are the coefficients of lowpass filter $H(z)$ for the CMF.

Because of the recursive relations between the scaling and wavelet function, it is possible to represent a signal gradually, i.e., coarse-to-fine representation. This "increment in information" leads to the MRA. Mallat proved that a multiresolution approximation is entirely characterized by the scaling function which is specified by a CMF.

When we use scaling function instead of wavelets, we lose information. That is to say, from a signal representation point of view we do not loose any information, since it will still be possible to reconstruct the original signal, but from a wavelet analysis point of view we discard possible valuable scale information. The width of the scaling function spectrum is therefore an important parameter for using the wavelet transform. The shorter its spectrum the more wavelet coefficients we will have and the more scale information. But, as always, there will be practical limitations on the number of wavelet coefficients we can handle.

Data compression using wavelet transform will be achieved by its ability to efficiently approximate a signal with few non-zero wavelet coefficients. The design of wavelets must therefore be optimized to produce a maximum number of wavelet coefficients that are close to zero. A signal f has few non-negligible wavelet coefficients if most of the fine-scale wavelet coefficients are small. This depends mostly on the regularity of f, the number of vanishing moments of ψ and the size of its support.

Chapter 6

Reversible Design of Subband and Wavelet Transform

Theoretically the wavelet transformation and PR filter banks are losslessly invertible. However, this invertibility often depends on the fact that the transform is calculated using exact arithmetic. For lossless compression, however, it should be considered that the resulting output data of most non-singular linear transforms consist of rational or real numbers because of their floating point coefficients. In practice, a finite-precision arithmetic is employed to perform the transforms, and such arithmetic is inherently inexact due to errors introduced by rounding. Therefore it is to say, in the strict sense, that the transforms that are lossless invertible in exact arithmetic are *lossy* in finite-precision arithmetic, e.g., when using computers in practice. Even when the input data consist of sequences of integer samples (this is the case for audio, image, and video signal in digital world), the outputs no longer consist of integers. Moreover the loss of the information would increase still more in the stage of the quantization of the transform coefficients in signal compression system.

This chapter begins by introducing the concept of reversibility of the transforms and the several reversible design methods that are efficiently used in signal compression. Next, the lifting scheme is presented as a general framework for the design and implementation of wavelet transform, and we show how the lifting scheme can be modified to a transform that allows an integer version of every wavelet transform. We then examine how lifting can be used in conjunction with M-channel subband transform. Finally, an experimental investigation of the proposed reversible transforms is presented by testing them in decorrelation stage on the lossless audio compression system.

6.1   The concept of reversible transform

It is important to clearly define the terms of *perfect (exact) reconstruction system* and *reversible system* in order to avoid the possible ambiguity in both terms;

Perfect reconstruction system: if the signals $x(n)$ and $\hat{x}(n)$ are identical up to multiplicative constant and a delay term, the system is called PR system.

Reversible System: a reversible system is an implementation of a PR system, in integer arithmetic, so that a signal with integer coefficients can be losslessly recovered. An efficient reversible system is a reversible system with transform matrix of determinant $\approx \pm 1$.

In principle, the construction of reversible transforms is simple. Through the appropriate use of quantization or rounding, any PR linear transform can be modified so that it can be computed using finite-precision arithmetic while preserving lossless invertibility. Because of the use of quantization, the resulting reversible transform is generally nonlinear and only serves to approximate the linear transform from which it was derived. If the reversible transform fails to mimic the behavior of its parent transform, the desirable properties of the parent transform will likely be lost and poor results will be obtained. Hence, the key consideration in the design process is to construct *efficient* reversible transforms that successfully approximate their parent transform. Hence, this efficient construction is by no means a straightforward task, although the idea behind generating reversible transform is easily stated.

In the past, reversible transforms have been developed largely by ad hoc methods that are difficult to generalize. Some good transforms were developed, e.g., S-transform, RTS transform, and S+P transform, that are all nonlinear approximation to well-known linear transforms, and were devised using ad hoc methods. Fortunately, a symmetric method for construction reversible transform based on *lifting scheme* has been recently proposed. The lifting scheme enables us to construct the wavelets without the translation and dilation of one fixed function (wavelet). More interestingly, it can be easily modified to implement reversible wavelet transform that maps integers to integers, and that is also reversible. On the other hand, integer computations are much faster than floating point for computers and much easier to implement in hardware which is more important in some applications.

6.2 Reversible subband transforms

6.2.1 S-Transform

A classic example of a reversible transform is the *S-transform* (Sequential transform) proposed in [167] [168] which has become quite popular for lossless signal compression (especially image compression). A sequence of random integers $x[n]$ with length N, can be perfectly represented by the two sequences with length $N/2$ defined by

$$lp[n] = \lfloor (x[2n] + x[2n+1])/2 \rfloor \,, \tag{6.2.1}$$
$$hp[n] = x[2n] - x[2n+1] \,,$$

where the floor $\lfloor \cdot \rfloor$ represents a maximum integer not exceeding a real number x. This is so called (2×2)[1] S-transform. Several slightly different definitions of this transform exist in the literature. In fact, the S-transform is a nonlinear approximation to a scaled version of the Haar transform. The Haar transform itself is one of the simplest two-channel subband transforms, as shown in Chapter 5. Again that the transfer functions of Haar transform are defined as,

$$H(z) = (1+z)/2 \,, \qquad\qquad G(z) = 1 - z \,,$$
$$\widetilde{H}(z) = 1 + z \,, \qquad\qquad \widetilde{G}(z) = (1-z)/2 \,.$$

This transform is a scaled version of an orthogonal transform. Therefore, the coefficients of the S-transform must be weighted on a per subband basis in order to approximate an orthogonal transform. The inverse transformation of (6.2.1) is

$$x[2n] = lp[n] + \lfloor (hp[n]+1)/2 \rfloor \,, \tag{6.2.2}$$
$$x[2n+1] = x[2n] - hp[n] \,. \tag{6.2.3}$$

The idea behind the reversibility of the S-transform is the observation of the facts; the sum and the difference of two integers are sufficient knowledge to recover the numbers and have the same parity, i.e., they share the same least significant bit. Therefore the division by 2 (or a shift right by 1) in Eq. (6.2.1) eliminates a redundant least significant bit. Figure 6.2.1 shows a block diagram of the transform.

In fact, this transformation is equal to subband decomposition, except for the truncation procedure. Therefore $lp[n]$ and $hp[n]$ are the lowpass and highpass components, respectively. The main idea behind this representation is that, if the correlation coefficient of $x[2n]$ and $x[2n+1]$ is larger than $1/3$, then the average variance of $lp[n]$ and $hp[n]$ is smaller than the variance of $x[n]$. In this case, $hp[n]$ normally has small variance, while the variance of $lp[n]$ is approximately equal to the variance of $x[n]$. The main advantage

---

[1]Let the numbers of taps of the lowpass filter and that of the highpass filter be $(n \times m)$.

of S-transform is also that it is easy to find the truncation allowing the PR system. Moreover, there is no data expansion, i.e., it uses use the same number of samples of the original signal.
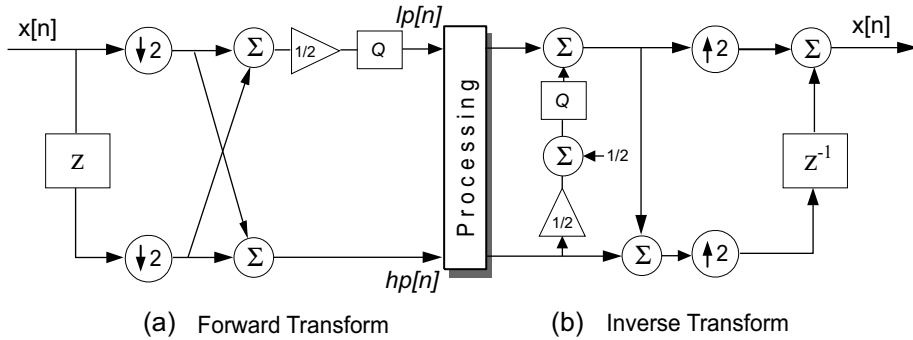
Figure 6.2.1: Filtering structure of S-transform ($Q(x) = \lfloor x \rfloor$).

Unfortunately, the ideas on which the S-transform is based do not generalize to transforms using more complicated relationships that simple pairwise sums and differences. Consequently the S-transform does not provide any further insight into how other classes of reversible transforms might be constructed.

6.2.2 Reversible TS-Transform

To design a symmetric short kernel filter (SSKF), Gall and Tabatabei in [169] used a factorization of a product filter into two linear phase low-pass components. These correspond to the lowpass analysis and synthesis filters. By using the QMF properties the highpass filters are derived. In their most important example, the following product filter is factorized

$$P(z) = \tfrac{1}{16}(1 + z^{-1})^3(-1 + 3z^{-1} + 3z^{-2} - z^{-3}) \,. \tag{6.2.4}$$

Its two factorized versions are given in [169]

$$
\begin{aligned}
P_1(z) &= [\tfrac{1}{4}(1 + z^{-1})^3] \times [\tfrac{1}{4}(-1 + 3z^{-1} + 3z^{-2} - z^{-3})] \\
P_2(z) &= [\tfrac{1}{2}(1 + z^{-1})^3] \times [\tfrac{1}{8}(1 + z^{-1})(-1 + 3z^{-1} + 3z^{-2} - z^{-3})] \,.
\end{aligned}
\tag{6.2.5}
$$

Using this factorization method, another version is considered in [170],

$$P_3(z) = [\tfrac{1}{2}(1 + z^{-1})] \times [\tfrac{1}{8}(1 + z^{-1})^2(-1 + 3z^{-1} + 3z^{-2} - z^{-3})] \,. \tag{6.2.6}$$

From the third version, a (2×6) PR subband filter can be defined with following filter coefficients

$$h = \frac{1}{\sqrt{2}}(1, 1)$$

$$g = \frac{1}{8\sqrt{2}}(-1, -1, 8, -8, 1, 1). \qquad (6.2.7)$$

This transform is called *TS(two-six)-transform* in the literature [171]. A reversible version (RTS) of Eq. (6.2.7) is proposed in [172],

$$lp[n] = \left\lfloor \frac{x[2n] + x[2n+1]}{2} \right\rfloor \qquad (6.2.8)$$

$$hp[n] = \left\lfloor \frac{1}{4} \left(- \left\lfloor \frac{(x[2n] + x[2n+1])}{2} \right\rfloor + 4(x[2n+2] - x[2n+3]) \right.\right.$$
$$\left.\left. + \left\lfloor \frac{(x[2n+4] + x[2n+5])}{2} \right\rfloor \right) \right\rfloor \qquad (6.2.9)$$

The expression for hp can be simplified and written with the use of lp, and the integer division by 4 can be rounded by adding a 2 to the numerator. These result in,

$$lp[n] = \left\lfloor \frac{x[2n] + x[2n+1]}{2} \right\rfloor \qquad (6.2.10)$$

$$hp[n] = x[2n+2] - x[2n+3] + \left\lfloor \frac{-lp[n] + lp[n+2] + 2}{4} \right\rfloor \qquad (6.2.11)$$

The inverse transform of the RTS-transform is quite simple,

$$x[2n] = lp[n] + \left\lfloor \frac{s[n] + 1}{2} \right\rfloor \qquad (6.2.12)$$

$$x[2n+1] = lp[n] - \left\lfloor \frac{s[n]}{2} \right\rfloor \qquad (6.2.13)$$

where

$$s[n] = hp[n-1] - \left\lfloor \frac{-lp[n-1] + lp[n+1] + 2}{4} \right\rfloor \qquad (6.2.14)$$

As the case of S-transform, the lowpass signal of RTS-transform has the same range of values as the input signal. This property is especially important in a pyramid system where the lowpass signal is successively decomposed. There is no systemic error due to rounding in the integer implementation of the transform, so all error in a lossy system can be controlled by quantization.

6.2.3   S+P Transform

Initially proposed by Said and Perlman [173], the S+P transform (S-transform + prediction) is a reversible transform that maps integers to integers and is parameterized by two sets of filter coefficients. This transform is a further refinement of the S-transform where the S-transformed highpass output hp_o is replaced by the difference between the $hp_o[n]$ and the estimate $\hat{hp}[n]$ obtained using the prediction;

$$lp[n] = \lfloor \tfrac{1}{2}(x[2n] + x[2n+1]) \rfloor \qquad (6.2.15)$$

$$hp[n] = hp_o[n] - \lfloor \hat{hp}[n] + 1/2 \rfloor \qquad (6.2.16)$$

$$(6.2.17)$$

where

$$hp_o[n] = x[2n] - x[2n-1] \qquad (6.2.18)$$

$$\hat{hp}[n] = \sum_{i=L_0}^{L_1} \alpha_i \Delta lp[n+i] - \sum_{j=1}^{H} \beta_j \, hp_o[n+j] \,, \qquad (6.2.19)$$

where $\Delta lp[n] = lp[n-1] - lp[n]$. The use of $\Delta lp[n]$ instead of $lp[n]$ allows to have zero-mean estimation terms, and thus there is no need to subtract the mean from $x[n]$. Note that the index i can be negative because $lp[n]$ is not replaced by a prediction error. The optimal predictor coefficients α and β can be found by solving the Yule-Walker equations (see Chapter 3). The inverse transform uses $lp[n]$ and $hp[n]$ to reconstruct the original input signal $x[n]$ as given by

$$x[2n] = lp[n] + \lfloor \tfrac{1}{2}(hp_o[n] + 1) \rfloor \qquad (6.2.20)$$

$$x[2n+1] = x[2n] - hp_o[n] \,, \qquad (6.2.21)$$

where

$$hp_o[n] = hp[n] + \lfloor \hat{hp}[n] + \tfrac{1}{2} \rfloor \,, \qquad (6.2.22)$$

and $\hat{hp}[n]$ is as given above.

Type	α_{-1}	α_0	α_1	β_1
A	0	1/4	1/4	0
B	0	2/8	3/8	2/8
C	-1/16	4/16	8/16	6/16

Table 6.1: S+P transform predictor coefficients

In Table 6.1^2, three sets of predictor coefficients are listed that have been suggested in [174]. The predictor A in the table has the smallest computational complexity and yields a reversible version of the TS-transform. In the degenerate case where all of the predictor coefficients are zero ($\alpha_i = \beta_i = 0$), the S-transform is obtained.

(a) Forward Transform

(b) Inverse Transform
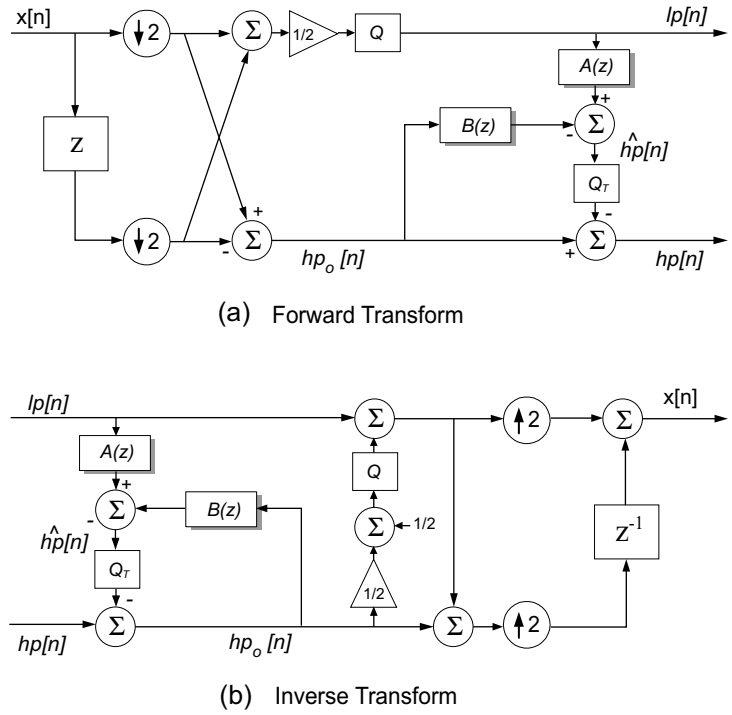
Figure 6.2.2: Structure of S+P transform. (a) Forward transform, (b) Inverse transform

The filtering structure for the S+P transform is shown in Figure 6.2.2, where $A(z)$, $B(z)$, $Q_T(x)$, and $Q(x)$ are defined as

$$A(z) = (1 + z^{-1}) \sum_{i=L_0}^{L_1} \alpha_i z^{-1}, \quad B(z) = \sum_{j=1}^{H} \beta_j z^{-j}, \tag{6.2.23}$$

$$Q(x) = \lfloor x \rfloor, \quad Q_T(x) = \lfloor x + \tfrac{1}{2} \rfloor. \tag{6.2.24}$$

Note that the first part of the forward transform structure and last part of the inverse transform structure are nothing more than the S-transform. If we disregard the effects of

---

^{2}In image compression applications, the predictor B is the best suited for natural images and the predictor C is for very smooth medical images.

the truncation in the S+P transform, we have a linear subband transform that corresponds to a QMF bank having analysis filters with transfer functions,

$$H(z) = \tfrac{1}{2}(1 + z),$$
$$G(z) = -\tfrac{1}{2}(1 + z)A(z^2) + (1 - z)[1 + B(z^2)] \qquad (6.2.25)$$

As we have seen that the S+P transform does not directly approximate an orthogonal or near-orthogonal transform. By weighting the transform coefficients associated with each subband by an appropriate constant, a near-orthogonal transform is obtained. To handle finite-length signal, it should be assumed that the signal is defined for $n = 0, 1, 2, \ldots, N - 1$ where N is even. This assumption is required so that when the input signal is split into two polyphase components. If the signal is not of even length, the simple solution is to pad the signal by one sample. The S+P transform is typically applied in a pyramid fashion such that the lowpass signal is successively decomposed. In this case, we have a reversible wavelet transform.

6.3   Lifting Scheme

Lifting is an iterative procedure for constructing biorthogonal wavelet bases. It was first developed by Swelden [175] [176] as a method to improve the general wavelet properties. The procedure takes a simple initial basis and fine-tunes such properties as the number of vanishing moments and the order of approximation, through successive modifications of the basis functions. The canonical lifting step involves either modifying the wavelets while holding the scaling functions fixed or modifying the dual wavelets while holding the dual scaling functions fixed. This cycle of modifications is repeated until the desired basis properties are obtained.

6.3.1   Characterization of biorthogonal wavelet bases

The lifting scheme is inspired from a relationship among all biorthogonal wavelets that share the same scaling function such that one can construct the desired wavelet from a simple one. The lifting scheme is available only if an initial set of biorthogonal filters having a finite impulse response is defined. We assume that the scaling function, wavelet and their duals are compactly supported. As shown in section 5.6.5, the necessary conditions

for the finite biorthogonal filters $\{h, g, \tilde{h}, \tilde{g}\}$ in matrix form are given by

$$\boldsymbol{M}(z^{-1})^T \widetilde{\boldsymbol{M}}(z) = 2\boldsymbol{I} \,, \tag{6.3.1}$$

with the modulation matrix $\boldsymbol{M}$,

$$\boldsymbol{M}(z) = \begin{bmatrix} H(z) & H(-z) \\ G(z) & G(-z) \end{bmatrix} \,, \tag{6.3.2}$$

and

$$G(z) = z^{-1}\widetilde{H}(-z^{-1}) \,, \quad \widetilde{G}(z) = z^{-1}H(-z^{-1}) \,. \tag{6.3.3}$$

Note that the orthogonal case corresponds to $\boldsymbol{M}$ being a unitary matrix.

To design the biorthogonal filters, there are various methods for deriving the highpass filter $G(z)$ from the lowpass filter $H(z)$. For example, Chui [177] showed that if, given the biorthogonal filters h and $\tilde{h}$, g and $\tilde{g}$ are chosen such that

$$G(z) = z^{-1}\widetilde{H}(-z^{-1})R^{-1}(z^{-2}) \,, \tag{6.3.4}$$

$$\widetilde{G}(z) = z^{-1}H(-z^{-1})R(z^2) \,, \tag{6.3.5}$$

then the filters g and $\tilde{g}$ satisfy the condition (6.3.1). $R(z)$ belongs to the Wiener class if $R(z) \neq 0$ and $r[n] \in l^1$. In this case, the highpass filters g and $\tilde{g}$ are finite only if the $R(z)$ is a monomial.

According to the theorem of Chui, it is proven by Herley, Vetterli [178] and Sweldens [175] that if there exists a finite filter s such that

$$H^S(z) = H(z) + z^{-1}\widetilde{H}(-z^{-1})S(z^{-2}), \tag{6.3.6}$$

then the finite filter h^s is also biorthogonal to $\tilde{h}$. From using (6.3.3) it then turns out that if $\{h, g, \tilde{h}, \tilde{g}\}$ are biorthogonal then we can construct a new set of finite biorthogonal filters $\{h^s, g, \tilde{h}, \tilde{g}^s\}$ with

$$H^S(z) = H(z) + G(z)S(z^{-2}) \tag{6.3.7}$$

$$\widetilde{G}^S(z) = \widetilde{G}(z) - \widetilde{H}(z)S(z^2) \,, \tag{6.3.8}$$

where $S(z)$ is a trigonometric polynomial. In the time domain it becomes

$$h^s[n] = h[n] + \sum_{k\in\mathbb{Z}} g[n - 2k]\, s[-k] \tag{6.3.9}$$

$$\tilde{g}^s[n] = \tilde{g}[n] - \sum_{k\in\mathbb{Z}} \tilde{h}[n - 2k]\, s[k] \,, \tag{6.3.10}$$

and these new filters are said to be *lifted*. In result the lowpass filter h is lifted with the help of the highpass filter g. It is called *primal lifting*.

Given an initial set of biorthogonal scaling functions and wavelets $\{\varphi, \psi, \tilde{\varphi}, \tilde{\psi}\}$, the new set $\{\varphi^s, \psi^s, \tilde{\varphi}, \tilde{\psi}^s\}$ can be derived by inserting (6.3.9) (6.3.10) in the refinement equation (5.6.57) (5.6.58):

$$\varphi^s(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h[k]\, \varphi^s(2t - k) + \sum_{k \in \mathbb{Z}} s[-k]\, \psi^s(t - k) \qquad (6.3.11)$$

$$\psi^s(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} g[k]\, \varphi^s(2t - k) \qquad (6.3.12)$$

$$\tilde{\psi}^s(t) = \tilde{\psi}(t) - \sum_{k \in \mathbb{Z}} s[k]\, \tilde{\varphi}(t - k)\,, \qquad (6.3.13)$$

where the coefficients of finite filter $s[n]$ can be freely chosen. In practice the lifting filter s is designed to produce the new filter $\widetilde{G}^S(z)$ with more zeros at $z = 1$. Since the support size of ψ and $\tilde{\psi}$ increases correspondingly to the length of the filter s, construction for a minimum size of s should be considered in order to achieve specific properties. For each choice of s, however, the biorthogonality conditions (5.6.53) and (5.6.54) have to be verified.

To improve the properties of ψ and $\tilde{\varphi}$ we use a *dual lifting* which modifies highpass filter g with the help of the lowpass filter $\tilde{h}$. By inverting h with g and g with $\tilde{g}$ in (6.3.9) and (6.3.10), we obtain the new set $\{h, g^u, \tilde{h}^u, \tilde{g}\}$ from an initial set $\{h, g, \tilde{h}, \tilde{g}\}$:

$$G^U(z) = G(z) + H(z)U(z^{-2})\,, \qquad (6.3.14)$$

$$\widetilde{H}^U(z) = \widetilde{H}(z) - \widetilde{G}(z)U(z^2)\,, \qquad (6.3.15)$$

and in the time domain

$$g^u[n] = g[n] + \sum_{k \in \mathbb{Z}} h[n - 2k]\, u[-k]\,, \qquad (6.3.16)$$

$$\tilde{h}^u[n] = \tilde{h}[n] - \sum_{k \in \mathbb{Z}} \tilde{g}[n - 2k]\, u[k]\,. \qquad (6.3.17)$$

Similarly we have new set of formally biorthogonal scaling function and wavelets $\{\varphi, \psi^u, \tilde{\varphi}^u, \tilde{\psi}^u\}$:

$$\tilde{\varphi}^u(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \tilde{h}[k]\, \tilde{\varphi}^u(2t - k) - \sum_{k \in \mathbb{Z}} u[k]\, \tilde{\psi}^u(t - k) \qquad (6.3.18)$$

$$\tilde{\psi}^u(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \tilde{g}[k]\, \tilde{\varphi}^u(2t - k) \qquad (6.3.19)$$

$$\psi^u(t) = \psi(t) + \sum_{k \in \mathbb{Z}} u[-k]\, \varphi(t - k)\,, \qquad (6.3.20)$$
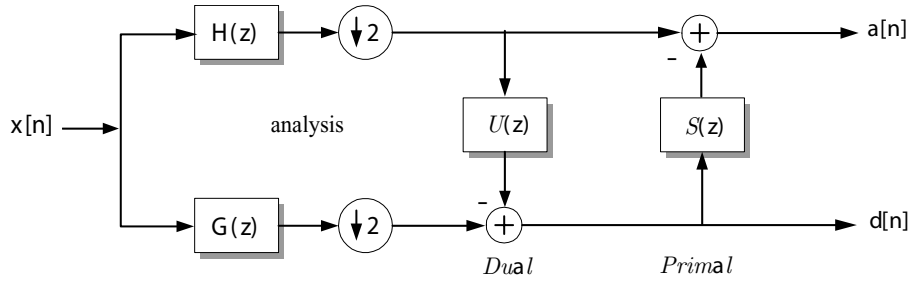
Figure 6.3.1: Primal lifting and dual lifting on classical subband scheme

Figure 6.3.1 displays the lifting and dual lifting based on the classical subband filter scheme. Primal lifting and dual lifting can be iterated to improve the regularity and vanishing moments of both ψ and $\tilde{\psi}$. It allows us to construct a multiresolution analysis with desired properties.

6.3.2 Polyphase representations

The polyphase representation provides a particularly convenient tool to describe the special structure of the modulation matrix [179] and the mechanics of lifting scheme. It is based on the fact that one can associate a unitary operator with an orthogonal filter bank.

We define the even and odd polyphase components h_e and h_o of the filter h by

$$h_e(z) = \sum_{k \in \mathbb{Z}} h_{2k} \, z^{-k} \quad \text{and} \quad h_o(z) = \sum_{k \in \mathbb{Z}} h_{2k+1} \, z^{-k} \,. \tag{6.3.21}$$

This can be written also as a discrete equivalent of Euler's formula:

$$h_e(z^2) = \frac{H(z) + H(-z)}{2} \,, \quad h_o(z^2) = \frac{H(z) - H(-z)}{2z^{-1}} \,. \tag{6.3.22}$$

The polyphase representation of the filter h is then given by

$$H(z) = \sum_k h_k \, z^{-k} = h_e(z^2) + z^{-1} h_o(z^2) \,. \tag{6.3.23}$$

Concerning to the modulation matrix $\boldsymbol{M}(z)$ (6.3.2) we define the polyphase matrix $\boldsymbol{P}(z)$ as following

$$\boldsymbol{P}(z) = \begin{bmatrix} h_e(z) & h_o(z) \\ g_e(z) & g_o(z) \end{bmatrix} = \sum_{k \in \mathbb{Z}} \begin{bmatrix} h_{2k} & h_{2k+1} \\ g_{2k} & g_{2k+1} \end{bmatrix} z^{-k} \tag{6.3.24}$$

so that

$$P(z^2) = 2^{-1} M(z) \begin{bmatrix} 1 & z \\ 1 & -z \end{bmatrix}. \tag{6.3.25}$$

$\widetilde{P}(z)$ is defined in similar way. The condition for perfect reconstruction now can be written in unitary form such that

$$\boldsymbol{P}(z) \, \widetilde{\boldsymbol{P}}(z^{-1})^T = \boldsymbol{I}. \tag{6.3.26}$$

Assuming that $\boldsymbol{P}(z)$ is of finite degree, i.e. Laurent polynomial[3], and has a monomial determinant. It implies that $\boldsymbol{P}(z)$ and $\widetilde{\boldsymbol{P}}(z)$ are invertible and their inverses are also of finite degree. Such matrices are often called *unimodular matrix*. From Cramer's rule with assuming, $\det \boldsymbol{P}(z) = 1$, and (6.3.26) we see that

$$\begin{aligned}
\tilde{h}_e(z) &= g_o(z^{-1}), \quad \tilde{h}_o(z) = -g_e(z^{-1}), \\
\tilde{g}_e(z) &= -h_o(z^{-1}), \quad \tilde{g}_o(z) = h_e(z^{-1}).
\end{aligned} \tag{6.3.27}$$

It shows that once we have such a matrix $\boldsymbol{P}(z)$, the dual matrix $\widetilde{\boldsymbol{P}}(z)$ follows immediately. If the polyphase matrix $\boldsymbol{P}(z)$ has determinant 1, then the filter pair (h, g) is called *complementary*. If the pair (h, g) is complementary, then the pair $(\tilde{h}, \tilde{g})$ is also so. Because of the perfect reconstruction condition the wavelet filter pairs are always in this case. If the polyphase matrix is the unit matrix, i.e. $\boldsymbol{P}(z) = \boldsymbol{I}$, then it results in $h(z) = \tilde{h}(z) = 1$ and $g(z) = \tilde{g}(z) = z^{-1}$. This transform is referred to as the *lazy wavelet transform* in the context of lifting. The lazy transform does nothing more than splitting the input signal into even and odd components.

6.3.3   DWT and MRA in polyphase representation

As we showed in previous chapter, the DWT is performed by lowpass, highpass filtering and subsampling as illustrated in Figure 6.3.2. Then we have a coarse approximation $a_{1,i}$ (lowpass filtered) and a detail information $d_{1,i}$ (highpass filtered) of input signal $x[n] = a_0[n]$. In order to obtain a successive multiresolution representation of the signal, we iterate this same processing on each coarse approximations $a_j[n]$.

---

[3]The z-transform of a FIR filter is a Laurent polynomial $h(z)$ given by $h(z) = \sum_{k=p}^{q} h_k z^{-k}$, A Laurent polynomial differs from a normal polynomial in that it can have negative exponents. The degree of a Laurent polynomial h is defined as $|h| = q - p$. Note that de Laurent polynomial z^p has degree zero.
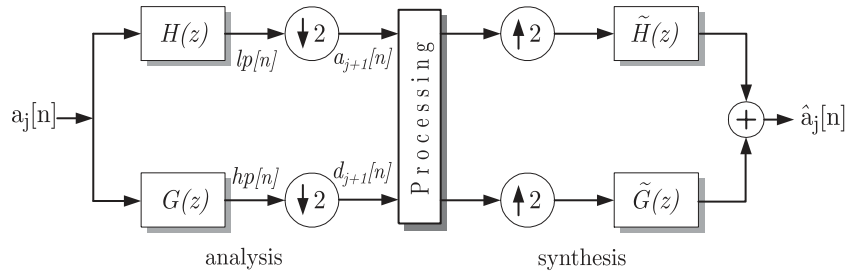
Figure 6.3.2: Discrete wavelet transform using 2-channel perfect reconstruction filter bank

We can describe the discrete wavelet transform by using the polyphase representation. The lowpassed signal $jp(z)$ and highpassed signal $hp(z)$ in Figure 6.3.2 can be written as

$$\begin{bmatrix} lp(z) \\ hp(z) \end{bmatrix} = \begin{bmatrix} H(z) \\ G(z) \end{bmatrix} X(z),$$ (6.3.28)

while the following subsampling part corresponds to

$$lp(z^2) = \frac{lp(z) + lp(-z)}{2} = lp_e(z^2),$$ (6.3.29)

$$hp(z^2) = \frac{hp(z) + hp(-z)}{2} = hp_e(z^2).$$ (6.3.30)

This can be combined as

$$\begin{bmatrix} lp(z^2) \\ hp(z^2) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} H(-z) & H(z) \\ G(-z) & G(z) \end{bmatrix} \begin{bmatrix} X(-z) \\ X(z) \end{bmatrix}.$$ (6.3.31)

However, by first filtering and then subsampling, we first compute all the coefficients and then throw away half of the work done. It would be more efficient if the subsampling step could be done before the filtering, which means that we only compute the even parts of lp and hp, i.e.,

$$lp_e(z) = [H(z)X(z)]_e = h_e(z)x_e(z) + z^{-1}h_o(z)x_o(z)$$ (6.3.32)

Thus $lp_e(z)$ and $hp_e(z)$ are obtained using the polyphase matrix as following

$$\begin{bmatrix} lp_e(z) \\ hp_e(z) \end{bmatrix} = \boldsymbol{P}(z) \begin{bmatrix} x_e(z) \\ z^{-1}x_o(z) \end{bmatrix}$$ (6.3.33)
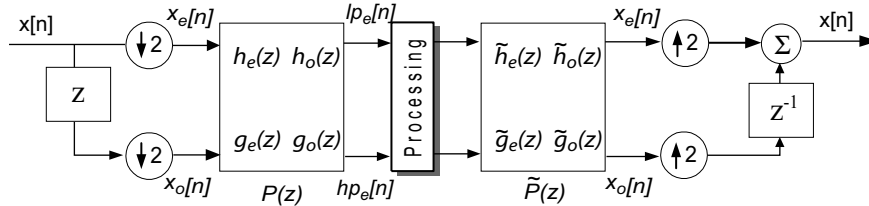
Figure 6.3.3: Polyphase representation for wavelet filter bank

Figure 6.3.3 illustrates the polyphase representation of the wavelet filter bank in Figure 6.3.2. The interest in such a polyphase representations is that paraunitary matrices possess complete factorizations in terms of elementary matrices of degree zero (rotations) and one (diagonal matrix with delays). Thus one can optimize such a factorized structure to find a filter bank meeting specific constraints. Note that the constraint of a large number of zeros at $\omega = \pi$ (as required for wavelet designs) is difficult to enforce in this form.

6.3.4   Lifting algorithm

In previous section we have shown through the biorthogonal wavelets how the lifting affects them. In this section we introduce the lifting scheme using a natural way, i.e. how the lifting affects the discrete wavelet transform and multiresolution representation. It is fruitful to view the DWT as a prediction-error decomposition. The scaling coefficients at a given scale j are "predictors" for the data at the next higher resolution or finer scale $j - 1$. The wavelet coefficients are simply the "prediction errors" between the scaling coefficients and the higher resolution data that they are attempting predict. In fact, this interpretation has led to lifting scheme as a new framework for DWT design.

Figure 6.3.4: General structure of lifting scheme.

The fundamental idea behind the lifting scheme is to efficiently implement all the decorrelation procedures in classical wavelet transform in time (spatial) domain. The corresponding tool in lifting scheme is comprised of three steps, i.e. split, predict, and update:

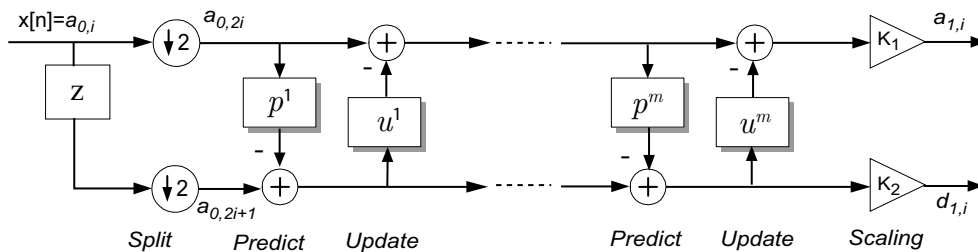**Split**: The first step of lifting algorithm separates the input signal $a_0[n]$ (from the MRA viewpoint, we set a signal $x[n] = a_{j=0}[n] = a_0[n]$) into its even and odd indexed subsequences, $a_{0,2i} = a_0[2n]$ and $a_{0,2i+1} = a_0[2n + 1]$.[4] Practically the splitting step is performed using the unit polyphase transform, called *lazy transform*. This is equivalent to the subsampling in wavelet transform. If the signal $a_0[n]$ is smooth and regular, then $a_{0,2i}$ and $a_{0,2i+1}$ are closely correlated. In this case it is intuitive that one can efficiently predict one of the sets from the other set.

Predict: Since the correlation structure is typically local, we predict the odd samples $a_{0,2i+1}$ (for example) from the neighboring even samples $a_{0,2i}$:

$$d_{1,i} = a_{0,2i+1} - \mathcal{P}(a_{0,2i}),\qquad(6.3.34)$$

where the prediction function $\mathcal{P}$ is a linear combination of neighboring even samples, i.e. $\mathcal{P}a_{0,2i} = \sum_{k\in\mathbb{Z}} p_k\, a_e[n + k]$. In many cases the order of $\mathcal{P}$ is up to and including $M - 1$, while the analysis wavelet has M zero moments. If the signal is locally smooth, the prediction errors $d_{1,i}$ will be small, and then its first order entropy is also smaller than that of $a_{0,2i+1}$. When we replace $a_{0,2i+1}$ by the detail $d_{1,i}$ (it is basic idea of the well-known DPCM methods [12]), we will lose no information of the signal $a_{0,2i+1}$ because it can easily be recovered by noting that

$$a_{0,2i+1} = d_{1,i} + \mathcal{P}(a_{0,2i}).\qquad(6.3.35)$$

As a result, this prediction procedure is equivalent to applying a highpass filter $g[n]$ to signal $x[n]$ in classical wavelet transform.

Update: In order to preserve the average of the $a_{0,2i}$ same as that of the input signal $a[n]$, we need a second lifting step, which replaces $a_{0,2i}$ with

$$a_{1,i} = a_{0,2i} + \mathcal{U}(d_{1,i}),\qquad(6.3.36)$$

where $\mathcal{U}$ is a linear combination of neighboring detail samples, i.e., $\mathcal{U}[n] = \sum_{k\in\mathbb{Z}} u_k\, d_1[n + k]$. Then $a_{1,i}$ corresponds to the lowpass output in wavelet transform. The updating step is also trivially invertible:

$$a_{0,2i} = a_{1,i} - \mathcal{U}(d_{1,i}).\qquad(6.3.37)$$

---

[4]It is also possible to split $a[n]$ into other non-overlapping sets [176]

CHAPTER 6. REVERSIBLE DESIGN OF SUBBAND AND WAVELET **151**
TRANSFORM

By iteration of these steps on the $a_j[n]$ we obtain a multiresolution decomposition of a signal. Consequently, no matter how $\mathcal{P}$ and $\mathcal{U}$ are chosen, the lifting scheme with these steps is always invertible and thus leads to critically sampled perfect reconstruction filter banks. It also leads to fast polyphase implementation of filter bank decomposition. Figure 6.3.4 illustrates the typical lifting stage.

The flexibility of the lifting scheme allows to adjust wavelet transforms to complex geometric situations and irregular sampling leading to so-called *second generation wavelets* [176] [180].

6.4   Factoring WT into the Lifting Scheme

It is proved in [181] that any wavelet with FIR filters can be factorized into a finite number of alternating lifting and dual lifting steps starting from the Lazy wavelet. This implies that all classical wavelet transforms can be implemented by factorization of the wavelets into lifting steps. The key problem is how to design these lifting and dual lifting steps to satisfy some constraints, e.g., the number of vanishing moments. For the factorization the Euclidean algorithm provides a bridge between the classical wavelet transforms and the lifting scheme.

6.4.1   Decomposition into lifting steps

Any polyphase matrix representing of a wavelet transform with finite filters can be factored in a finite product of unit upper and lower triangular 2×2 matrices, and a diagonal normalization matrix;

$$\boldsymbol{P}(z) = \underbrace{\begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}}_{normalization} \prod_{i=m}^{1} \left\{ \overbrace{\begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix}}^{primallifting} \underbrace{\begin{bmatrix} 1 & 0 \\ u_i(z) & 1 \end{bmatrix}}_{duallifting} \right\} , \qquad (6.4.1)$$

where $s_i(z)$ and $u_i(z)$ are Laurent polynomials and K_1 and K_2 are scaling constants unequal to zero. If scaling is not desired for some reason, it is even possible to factor the scaling matrix into four more lifting steps, one of which can be combined with the last real lifting step so that factoring a scaling matrix costs three extra lifting steps. From this

it immediately follows that the inverse transform can be written as

$$
\boldsymbol{P}^{-1}(z) = \prod_{i=1}^{m} \left\{ \begin{bmatrix} 1 & 0 \\ -u_i(z) & 1 \end{bmatrix} \begin{bmatrix} 1 & -s_i(z) \\ 0 & 1 \end{bmatrix} \right\} \begin{bmatrix} 1/K_1 & 0 \\ 0 & 1/K_2 \end{bmatrix} \tag{6.4.2}
$$

To find the $s_i(z)$ and $u_i(z)$, we first decompose the polyphase matrix using primal lifting as

$$
\boldsymbol{P}(z) = \begin{bmatrix} h_e(z) & h_o(z) \\ g_e(z) & g_o(z) \end{bmatrix} = \begin{bmatrix} h_e(z) & h_o^{new}(z) \\ g_e(z) & g_o^{new}(z) \end{bmatrix} \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix} \tag{6.4.3}
$$
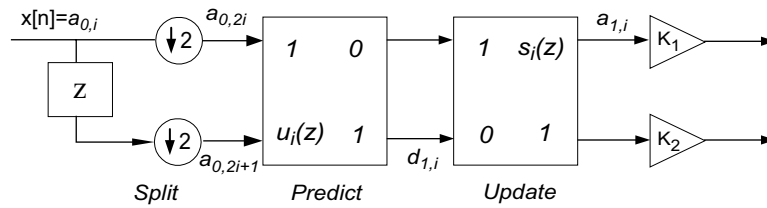
Figure 6.4.1: Polyphase representation of lifting process, analysis side.

We then have to find Laurent polynomials $s(z)$, h_o^{new}, and $g_o^{new}(z)$ such that

$$
h_o(z) = s(z)h_e(z) + h_o^{new}(z) , \tag{6.4.4}
$$
$$
g_o(z) = s(z)g_e(z) + g_o^{new}(z) ,
$$

with

$$
|h_o^{new}(z)| < |h_o(z)| \tag{6.4.5}
$$
$$
|g_o^{new}(z)| < |g_o(z)| ,
$$

where $|\cdot|$ means degree. This corresponds to two *long division with remainder* schemes of Laurent polynomials, with a common quotient $s(z)$. To perform this, we can use the Euclidean algorithm for Laurent polynomials, which was originally developed to find the greatest common divisor (gcd) of two natural numbers. In [181] it is given as:
Take two Laurent polynomials $a(z)$ and $b(z) \neq 0$ with $|a(z)| \geq |b(z)|$. Let $a_0(z) = a(z)$ and $b_0(z) = b(z)$ and iterate the following steps starting from $i = 0$,

$$
a_{i+1}(z) = b_i(z) \tag{6.4.6}
$$
$$
b_{i+1}(z) = a_i(z)\%b_i(z) \tag{6.4.7}
$$
$$
q_{i+1}(z) = a_i(z)/b_i(z) \tag{6.4.8}
$$

Then $a_n(z) = gcd(a(z), b(z))$ where n is the smallest number for which $b_n(z) = 0$.

The result of this algorithm can be written as

$$\begin{bmatrix} a(z) \\ b(z) \end{bmatrix} = \prod_{i=1}^{n} \begin{bmatrix} q_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_n(z) \\ 0 \end{bmatrix} , \qquad (6.4.9)$$

which looks very much like a series of lifting steps. The gcd found might not be unique since it is defined only up to a factor z^p, i.e. there are several factorizations possible[5]. This turns out to be an advantage because it allows us to select the factoring which best suits our needs.

As an example, we make a decomposition of the Cohen-Daubechies-Feauveau biorthogonal wavelets [101] with 2 vanishing moments. The CDF(2,2)[6] is given as

$$H(z) = -\frac{1}{8}z^{-2} + \frac{1}{4}z^{-1} + \frac{3}{4} + \frac{1}{4}z - \frac{1}{8}z^2 , \qquad (6.4.10)$$

$$G(z) = \frac{1}{4}z^{-2} - \frac{1}{2}z^{-1} + \frac{1}{4} . \qquad (6.4.11)$$

The corresponding polyphase matrix is

$$\boldsymbol{P}(z) = \begin{bmatrix} h_e^{new}(z) & \frac{1}{4} + \frac{1}{4}z \\ g_e^{new}(z) & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ u(z) & 1 \end{bmatrix} \qquad (6.4.12)$$

We have to find Laurent polynomials $u(z)$, $h_e^{new}(z)$, and $g_e^{new}(z)$ such that

$$-\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z = h_e^{new}(z) + u(z) \left(\frac{1}{4} + \frac{1}{4}z \right) , \qquad (6.4.13)$$

$$\frac{1}{4}z^{-1} + \frac{1}{4} = g_e^{new}(z) + u(z) \left(-\frac{1}{2} \right) . \qquad (6.4.14)$$

Thus we calculate the quotient $u(z)$ and the remainder $h_e^{new}(z)$ of a long division. This division is not unique, but there are three solutions:

$$-\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z = \begin{cases} \left(-\frac{1}{2}z^{-1} + \frac{7}{2} \right) \left(\frac{1}{4} + \frac{1}{4}z \right) & -z \\ \left(-\frac{1}{2}z^{-1} - \frac{1}{2} \right) \left(\frac{1}{4} + \frac{1}{4}z \right) & +1 \\ \left(-\frac{7}{2}z^{-1} - \frac{1}{2} \right) \left(\frac{1}{4} + \frac{1}{4}z \right) & -z^{-1} \end{cases} \qquad (6.4.15)$$

---

[5]We write the long division with remainder of a_0 and a_1 as $a_0 = q_1 \Delta a_1 + a_2$. But we can express a_1 in a similar way as $a_1 = q_2 \Delta a_2 + a_3$ and a_2 also, and so on. The row of remainders will eventually reach zero, $a_1 > a_2 > \cdots > a_n > a_{n+1} = 0$, and this is where it stops. The gcd of a_0 and a_1 is now an. However, we are more interested in the intermediate results $a_0 = q_1 \Delta (q_2 \Delta (\ldots (q_n \Delta a_n + a_{n+1}) + \ldots) + a_3) + a_2$ $(an + 1 = 0)$.

[6](number of vanishing moments of the analyzing high pass filter, number of vanishing moments of the synthesizing high pass filter)

If we choose the middle line of the solutions in (6.4.15) as the factorization, we have a symmetrical one, which goes nicely with $g(z)$ as well, i.e.,

$$u(z) = -\frac{1}{2}z^{-1} - \frac{1}{2}, \tag{6.4.16}$$

$$h_e^{new} = 1, \tag{6.4.17}$$

$$g_e^{new}(z) = 0, \tag{6.4.18}$$

and thus the decomposition is given by

$$\boldsymbol{P}(z) = \begin{bmatrix} 1 & \frac{1}{4} + \frac{1}{4}z \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}z^{-1} - \frac{1}{2} & 1 \end{bmatrix} \tag{6.4.19}$$

We can continue the decomposition to obtain

$$\boldsymbol{P}(z) = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{4} + \frac{1}{4}z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}z^{-1} - \frac{1}{2} & 1 \end{bmatrix} \tag{6.4.20}$$

Clearly the number of lifting steps is bounded by the length of the original filters. Figure 6.4.2 shows the practical implementation scheme with the factorized CDF(2,2) biorthogonal filter.
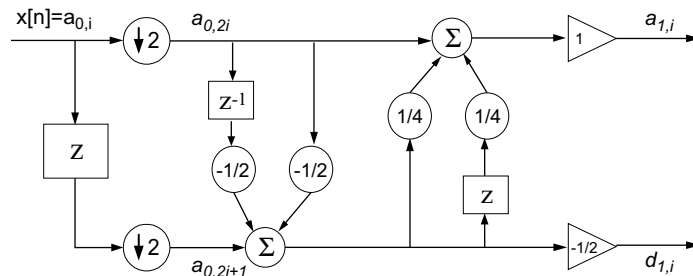

Figure 6.4.2: Implementation of lifting transform using CDF(2,2) biorthogonal filter

It is important to point out that the lifting factorization is highly non-unique process. For a given wavelet transform, we do not know exactly how many essentially different factorizations are possible and how they differ. One can choose the best one depending on the applications.

6.4.2   Normalization factor K

In practice, we can omit the normalization factors K_1 and K_2. However, for compression applications it is important to know the magnitudes of the normalized wavelet coefficients that indicate the importance of the corresponding wavelet components. Furthermore, if we want the energy of the signal to be retained in the wavelet coefficients (or at least an approximation in the non-orthogonal case) we have to use normalized filter coefficients that are $\sqrt{2}$ times as large as the filter coefficients we used. Thus the real normalization factors are $\sqrt{2}K_1$ for the lowpass band, and $\sqrt{2}K_2$ for the highpass band. Note that the normalization factors for the lowpass band accumulate, due to the iteration on the lowpass band.

It is also shown in [182] how, with at most 3 extra lifting steps, K can always be set to 1. One can then build the integer version of those extra lifting steps. This comes down to shifting one coefficient b bits to the left and the other b bits to the right. The b bits that get lost in the shift to the right end up in the b locations freed up by the shift to the left.

6.4.3   Lifting properties

The lifting scheme has some properties which are not found in many other transforms. First, the inverse transform is immediately as showed in Figure 6.3.4. This easy invertibility is always true for the lifting scheme. Secondly, lifting can be done in-place. It means that we never need samples other than the output of the previous lifting step and therefore we can replace the old stream by the new stream at every summation point. This allows also for adaptive wavelet transforms. One can start the analysis of a function from the coarsest levels and then build the finer levels by refining only in the areas of interests [180]. Thirdly, lifting is not causal. Usually this is not really a problem since we can always delay the signal enough to make it causal, but it will never be real-time. In some cases however it is possible to design a causal lifting transform. The last important property is the calculation complexity. In [181] it is proven that for long filters the lifting scheme cuts computation complexity in half, compared to the standard iterated FIR filter bank algorithm. This type of wavelet transform has already a complexity of $\mathcal{O}(N)$, in other words, much more efficient than the FFT with its complexity of $\mathcal{O}(N \log N)$ and lifting speeds things up with another factor of two.

6.5 Integer Reversible WT

In previous section we have seen that the lifting scheme leads to a fast in-place calculation of the wavelet transform that does not require auxiliary memory. The lifting scheme can be easily modified to implement integer reversible wavelet transform (IRWT) that maps integers to integers. Namely, the IRWT provides the decomposition of original signal into a set of integer coefficients and is also invertible. Since it allows perfect reconstruction [182], by inverse transform of IRWT the original signal can be reconstructed without any loss. Practically, non-integer transforms expand the input data (for example, 16 bit audio signal) to 32 bit wide floating point numbers in order to describe the real numbers of their coefficients. During the quantization or rounding process of these real numbers to low bit integers in a compression system, we will lose some corresponding information and thus can not reconstruct the original signal from the decoder side of the system. From a lossless compression point of view, it is thus very important that IRWT coefficients consist of the integers and have same dynamical range as the input signal. These discharge some from the consideration regarding the size of the variables to be used and the designing fast algorithms. The memory utilization of integers is also a positive consideration.

6.5.1 Rounded Lifting Scheme

From the Figure 6.3.4 and Eqs. (6.3.34) and (6.3.36), we consider the quantized version of $\mathcal{P}$ and $\mathcal{U}$, as following

$$d_{1,i} = a_{0,2i+1} - Int\{\mathcal{P}(a_{0,2i})\} \tag{6.5.1}$$
$$a_{1,i} = a_{0,2i} + Int\{\mathcal{U}(d_{1,i})\} \tag{6.5.2}$$

where the operation $Int\{\cdot\}$ denote the two possible quantization operations, either rounding or truncation, i.e.,

$$\left\lfloor x + \tfrac{1}{2} \right\rfloor , \quad \text{or} \quad \lfloor x \rfloor \tag{6.5.3}$$

Obviously, as long as the same rounding operation is employed on both the analysis and synthesis sides, PR is not affected. The original signal can always be reconstructed without any loss as

$$a_{0,2i} = a_{1,i} - Int\{\mathcal{U}(d_{1,i})\} \tag{6.5.4}$$
$$a_{0,2i+1} = d_{1,i} + Int\{\mathcal{P}(a_{0,2i})\} \tag{6.5.5}$$

It means that whatever deterministic rounding operation is used, the lifting scheme is always reversible. Of course, the resulting system is nonlinear, and the new subband

signals serve only to approximate the original subband signals.

6.5.2   Rounding the rational filter coefficients

As we discussed in Section 6.3.1, the lifting step can be written by using z-transform

$$H^S(z) = H(z) + S(z)G(z) \tag{6.5.6}$$

Because the signal part $G(z)$ is not changed by the lifting step, the result of the filter operation can be rounded

$$H^S(z) = H(z) + Int\{S(z)G(z)\} \tag{6.5.7}$$

while retaining a reversible operation

$$H(z) = H^S(z) - Int\{S(z)G(z)\} \tag{6.5.8}$$

In some cases, all filter coefficients of $S(z)$ are rational and the corresponding lifting step can be written as

$$h[n] = h[n] - \frac{1}{a}\sum_k b[k]g[k] \tag{6.5.9}$$

where a and $b[k]$ are integers. This is true for various wavelet transforms, among which the CDF biorthogonal wavelets. Such a lifting step can be modified in one of the following ways:

Full rounding The result of the division by a is rounded,

$$\dot{h}[n] = h[n] - Int\left\{\frac{\sum_k b[k]g[k]}{a}\right\}, \tag{6.5.10}$$

where $\dot{h}$ is an integer approximation to the h obtained with floating point lifting. In most cases, this rounding method is used, except when we want to control the dynamic range of the result.

No rounding We avoid the division by a by multiplying the other terms with a,

$$\dot{h}[n] = ah[n] - \sum_k b[k]g[k]. \tag{6.5.11}$$

In this case, the dynamic range of the coefficients will increase. This has to be taken into account in later steps of lifting scheme. Note that in this case no real rounding is performed, and thus this can be considered to be an exact implementation of the floating point version, yielding integers.

Mixed form We combine the rounding and multiplication steps of both methods above,

$$\dot{h}[n] = a_1 h[n] - Int\left\{\frac{\sum_k b[k]g[k]}{a_2}\right\}, \qquad (6.5.12)$$

with

$$a_1, a_2 \in \mathbb{Z},$$
$$a_1, a_2 = a.$$

This variant can be used if we want to have more control over the dynamic range of the resulting $\dot{h}$.

All of the three cases above the modified lifting step is still reversible, and thus PR property is still present.

6.5.3 Boundary treatment

The signals in practice do not extend infinitely in time or space, but are limited to a finite interval. However, filter bank algorithms assume infinite signal lengths. The difficulty is how to handle filtering at the boundaries of the signal. Since the outputs of the filters depend on past and/or future sample values, once we get close enough to a signal edge, the filters need sample values that are not defined. A general approaches that can be used to address this problem is to extend the signal so that it is defined for all possible sample indices.

Periodic extension

Periodic extension is the simplest approach to handling finite-length signals. We extend the finite-length signal by putting copies of itself in front of and behind the original signal. After a wavelet transform we can simply discard the coefficients that lie outside of the interval in which the original signal was defined, i.e., if the original signal is defined over N sample indices, periodic extension results in an infinite-length signal of period N. However, periodic extension has the potential disadvantage that unless the first and the last sample have the same value, the extended signal may possess some unexpected discontinuities at the boundaries of the original signal. These discontinuities will locally enlarge the wavelet coefficients and make compression of the signal more difficult. Of course, we can always pad the signal with extra samples so that divisibility is achieved, but doing this is not acceptable, since it would lead to much more wavelet coefficients than original samples.

Symmetric extension

A relatively better solution is the symmetric extension of the finite-length signal. We extend the signal by mirroring it around its endpoints so that it is both symmetric and periodic, which makes the discontinuities disappear. Thus the symmetric extension can be considered as periodic extension applied to a concatenation of the original signal and a mirrored copy of the original signal. After filtering one has to retain twice as much coefficient. Fortunately one can discard half of them if the filters are symmetric, yielding the same number of coefficients as the original signal length[7].
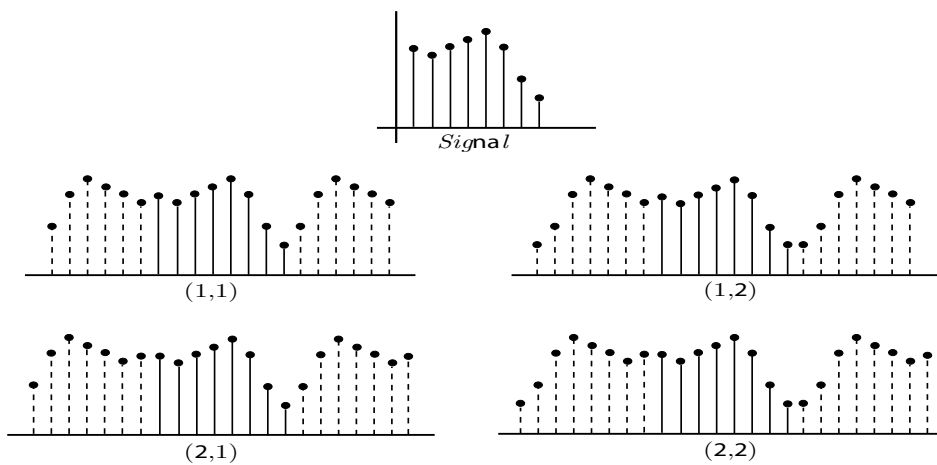
Figure 6.5.1: Symmetric signal extension. (first sample,last sample): 1 for undupli-cated and 2 for duplicated.

A signal can either be symmetric about one of its samples or about a point midway between two samples. These two cases are referred to as whole-sample symmetry and half-sample symmetry, respectively. The centers of symmetry of periodic signals always come in pairs. If the period of the signal is even, both symmetry centers will be whole-sample symmetry, while if the period is odd, one will be whole-sample symmetry and the other half-sample symmetry. There is more than one way in which to symmetrically extend a signal. For example, Figure 6.5.1 shows the four possible ways to symmetrically extend a discrete signal.

---

[7]That is why we use biorthogonal wavelets for lifting implementation instead of orthogonal wavelets, since they can be symmetric

6.6   Experiment on lossless audio compression

We tested various reversible transforms described in previous sections for lossless audio compression. For the decorrelation stage in our prototype compression system 3.6.1, the two symmetric biorthogonal interpolating wavelets, Deslauriers-Dubuc symmetric wavelet (4,2) and Cohen-Daubechies-Feauveau wavelet (2,2), and the S+P transform with the predictor type B (see Table 6.1) were investigated with various block lengths 2^9 to 2^{12} and decomposition levels 3 to 6 in successive form through the lowpass subbands.

6.6.1   Tested reversible filters

Deslauriers-Dubuc (4, 2): symmetric, biorthogonal, interpolating, $K = 1$
    impulse response:

$$h = \frac{1}{16} \left(1, 0, -9, 16, -9, 0, 1\right)$$
$$g = \frac{1}{16} \left(1, 0, -2, 4, \frac{23}{2}, 4, -2, 0, \frac{1}{4}\right) \qquad (6.6.1)$$

lifting implementation:

$$d_{1,i} = a_{0,2i+1} - \left\lfloor \frac{9}{16}(a_{0,2i} + a_{0,2i+2}) - \frac{1}{16}(a_{0,2i-2} + a_{0,2i+4}) + \frac{1}{2} \right\rfloor^{8},$$
$$a_{1,i} = a_{0,2i} + \left\lfloor \frac{1}{4}(d_{1,i-1} + d_{1,i}) + \frac{1}{2} \right\rfloor. \qquad (6.6.2)$$

Cohen-Daubechies-Feauveau (2, 2): symmetric, biorthogonal, interpolating, $K = 1$
    impulse response:

$$h = \frac{1}{8} \left(-1, \ 2, \ 6, \ 2, \ 1\right),$$
$$g = \frac{1}{4} \left(1, \ -2, \ 1\right). \qquad (6.6.3)$$

lifting implementation:

$$d_{1,i} = a_{0,2i+1} - \left\lfloor \frac{1}{2}(a_{0,2i} + a_{0,2i+2}) + \frac{1}{2} \right\rfloor,$$
$$a_{1,i} = a_{0,2i} + \left\lfloor \frac{1}{4}(d_{1,i-1} + d_{1,i}) + \frac{1}{2} \right\rfloor. \qquad (6.6.4)$$

---

[8]For the floor operations above, we used a biased floor function, i.e., $\lfloor \cdot \rfloor = \lfloor x + \frac{1}{2} \rfloor$.

S+P Transform: $\alpha_0 = \frac{2}{8}, \; \alpha_1 = \frac{3}{8}, \; \beta_1 = \frac{2}{8}$

$$\hat{hp}[n] = \frac{1}{8}\{2(\Delta lp[n] + \Delta lp[n+1] - hp_o[n+1]) + \Delta lp[n+1]\} \qquad (6.6.5)$$

6.6.2   Test results and discussion

The resulting bite rates are tabulated in Table 6.2. It is turned out that there is no single filter that has superior compression performance for all classes of audio materials, and that the transform effectiveness depends strongly on the audio signal content.

Materials	S	SPB	CDF(2,2)	DD(4,2)	LP-IIR
Nr. 1 violin, solo	9.12 (1.74)	8.92 (1.79)	**8.61** (1.86)	8.69 (1.84)	7.23 (2.21)
Nr. 2 flute, solo	7.54 (2.12)	**7.49** **(2.14)**	7.57 (2.11)	7.52 (2.13)	5.77 (2.77)
Nr. 3 speech, fem.	7.32 (2.19)	**7.30** **(2.19)**	7.37 (2.17)	7.42 (2.16)	5.92 (2.70)
Nr. 4 piano, solo	5.95 (2.69)	5.82 (2.75)	**5.73** (2.79)	5.81 (2.75)	4.13 (3.87)
Nr. 5 classic, orch.	7.50 (2.13)	**7.36** **(2.17)**	7.46 (2.14)	7.43 (2.15)	5.92 (2.70)
Nr. 6 pop, abba	8.35 (1.92)	8.33 (1.92)	**8.29** **(1.93)**	8.32 (1.92)	6.87 (2.33)
Nr. 7 country	**7.42** **(2.16)**	7.53 (2.12)	7.43 (2.15)	7.55 (2.12)	6.28 (2.55)
Nr. 8 rock, metal	12.44 (1.29)	12.49 (1.28)	**12.39** **(1.29)**	12.47 (1.28)	11.31 (1.41)
Nr. 9 jazz, soft	9.54 (1.68)	9.45 (1.69)	9.47 (1.69)	**9.44** **(1.69)**	7.73 (2.06)
Average	8.35 (1.92)	8.30 (1.93)	**8.26** **(1.94)**	8.29 (1.93)	6.91 (2.32)

Table 6.2: Test results with compressed bit rate (bits/sample) und compression ratio (orig./comp.)

For smooth audio signals, the DD(4,2) and the SPB are most effective, while the CDF(2,2) and S-transform perform best for audio signals with a large dynamic range and a strong treble energy with the lowest computational complexity. Consequently, the test

results suggest that the number of vanishing moments is not only the important factor which affects compression efficiency.

As illustrated in Figure 6.6.1, the compression ratio is nearly independent of the block length. It is because that the reversible transforms tested do not need to perform a block-based calculation for any correlation coefficient. Figure 6.6.2 shows the compression ratios with respect to different decomposition levels. Considering the whole complexity, a 6-level decomposition seems to be a good tradeoff for lossless audio compression.

The main advantage of using integer reversible wavelet transform to represent signal is multiresolution representation, which lossless compression methods based on time-domain prediction cannot offer. This easily allows the progressive transmission, lower resolution versions first, followed by transmissions of successive details. Such a mode of transmission is especially valuable in scenarios where bandwidth is limited, audio signal size are large and lossy compression is not acceptable.

In Table 6.2, the bit rates obtained by the IIR linear prediction filter in Chapter 3 is also tabulated in order to compare the compression performance between linear prediction in time domain and reversible transform. Obviously, the linear prediction methods outperform the reversible transform methods. However, the reversible transform possesses still big improvement potential, because of the highly non-unique factorization results, as mentioned before. A good choice of the factorization, which approximates better the characteristics of its parent linear transform and is more suitable for the lossless compression application, can improve its compression performance substantially more effective. Therefore it may be an interesting topic for future work.
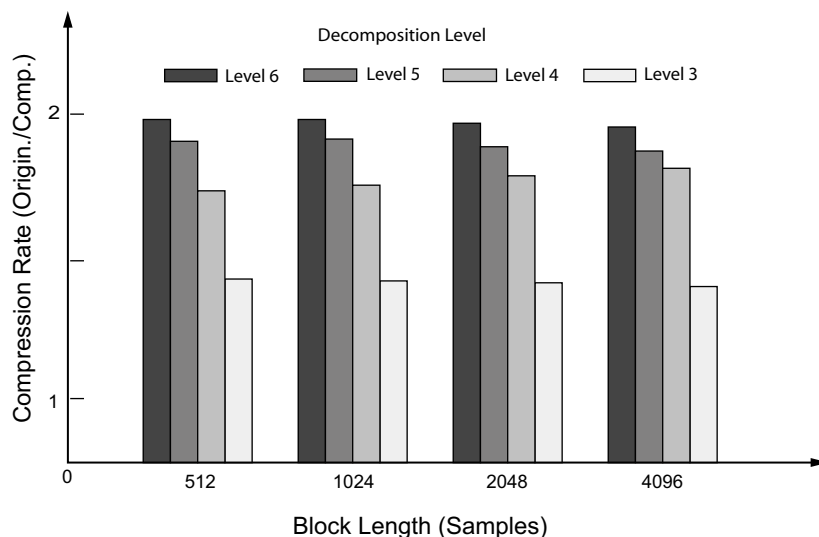

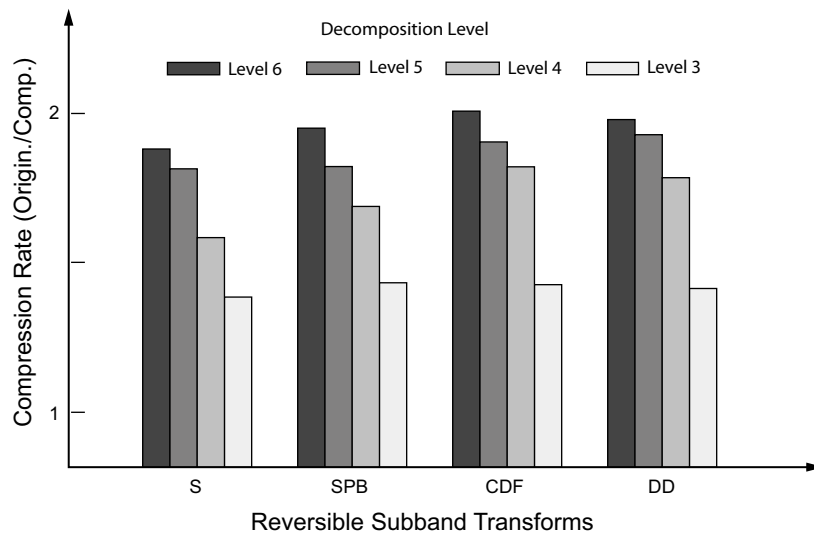**Figure 6.6.1:** Compression ratios in function of decomposition level and block length

Figure 6.6.2: Compression ratios in function of decomposition level and tested reversibel transforms

Chapter 7

Conclusion and Future Work

This thesis has touched on a range of issues in lossless audio compression. More precisely, its major focus is on the signal decorrelation methods in the compression scheme, predictive modeling and transform modeling. A variety of signal modeling techniques used for signal compression are explored and empirically validated, namely the linear prediction filter, subband decomposition, multirate filter banks, wavelet transform, reversible integer transforms, and lifting algorithms. In this Chapter, the key results of the thesis are summarized and directions for further research are discussed.

7.1   Prediction and Transform

Predictive modeling

In modeling a signal, it is of primary importance that the model be well approximated to the signal in question. Otherwise, the model will not necessarily provide a meaningful decorrelation performance in the compression scheme. The goal of the predictive modeling is to reduce the sample magnitudes by making a prediction of the current sample based on previous sample values.

Various prediction filters that are characterized by their prediction structures and coefficients are compared by applying to lossless audio compression. To explore the influence of the filter parameters on compression performance, various block lengths and prediction orders are tested for each prediction filter. Complexity is also an important factor for the evaluation of the prediction filters. With its lowest complexity, the polynomial predictor (PAP) provides efficient compression performance for certain of the audio materials. On the other side, applying the context-based error modeling into the compression scheme

does not specifically improve the compression ratio compared to its very high complexity. It is clearly shown through the test results that the selection of the prediction filter and the parameters depends strongly on the audio signal content.

In general, the lattice structure offers a powerful characterization in both filter design and implementation with the independently variable coefficients. However, for the compression applications, it may be not good choice because of the increasing side information and complexity. IIR prediction filter with large block length (2048 to 4096 samples) and high prediction order ($p = 10$) has superior compression result for nearly all kind of audio samples over other prediction filters. Furthermore, for the music samples with high treble energy and wide dynamic range, the IIR prediction filter provides an acceptably stable compression ratio. In contrast with the FIR filter, the IIR filter offers more possible ways to efficiently design or modify the filter structure because of its backward adaptation.

Transform Modeling

In transform modeling, the subband transform, wavelet transform, and reversible integer transform are theoretically and empirically investigated. In fact, from signal processing or implementation viewpoint, they converge onto almost same transform results. However, it is also the fact that they can be improved by understanding the particular characteristics of each transform. In the thesis, therefore, the wavelet transform is mathematically in depth explored in order to make clear its heterogeneous nature related to the other transforms and to avoid the ambiguous use of the term in literature.

With the design theory of multirate filter banks, the subband coding scheme and two-channel perfect reconstruction filter banks are explored. Parallel with the subband transform, the block transforms are discussed. Lapped transform takes account inter-block correlation and is capable of eliminating the discontinuities between block boundaries. The effectiveness of the linear prediction from subbands and fullband are first formally studied and then empirically compared by the entropies of prediction errors in both cases. Theoretically, it is shown that the prediction error variance of the fullband always exceeds the total prediction error variance of the combined subbands, for a given prediction order p. From the test results, it is also shown that the effectiveness of the subband prediction depends strongly on the used subband transform. For example, in the experiment, an adaptive inverse transform used to transform the subband signal into time domain has improved compressibility of the subband signals.

The lifting scheme is an efficient implementation of a wavelet transform algorithm. Furthermore, it can easily be converted to operate on integer values. The inverse transform is trivial to find and always provides perfect reconstruction of the input signal. All other reversible transforms discussed in the thesis, e.g., S-transform, RTS transform, and

S+P transform, can be shown as a special cases of the lifting scheme. We have worked out the details of integer wavelet transforms for the biorthogonal wavelets. A problem, that can also mean a potential for an improvement factor, is that the lifting factorization is highly non-unique process. For a given wavelet transform, we do not know exactly how many essentially different factorizations are possible and how they differ. One can choose the best one depending on the applications.

In conclusion, for the signal compression there is always trade-off between the decorrelation level and the side information. It means that the better decorrelation does not always provide a better compression result. It is also empirically shown that the predictive modeling methods discussed in the thesis outperform over the transform modeling methods in lossless audio compression. IIR prediction filter offers more potential for improving the compression ratio. Main advantage of the using transform modeling in compression scheme is multiresolution representation of the signal, which lossless compression methods based on time-domain prediction cannot offer. This allows the progressive transmission and editibility of the audio features and events.

7.2 Suggestions for Future Research

This thesis has made some meaningful contributions to broadly examining the lossless compression methods, from linear prediction to subband transform. In fact, the works done in the thesis and other state-of-the-art lossless audio coders show us that the lossless audio compression seems to be already reached its limit of compression ratio (i.e., max. 3 to 4), as often said in the literature, no matter whether using a simple prediction method or a complex transform to decorrelate the audio signal. However, there are still some points that could potentially pave the way to overcome the limit.

First, parametric signal segmentation will improve compression performance. The choice of effective block length is a critical factor for designing signal compression system. For both the prediction method and the transform method, the decorrelation procedure is performed in each block separately. In this case, a cross correlation between the blocks is generally ignored. The musical signal, for example, is quasi-periodic signal because of its repeating rhythm and harmonic progress. If we segment the signal according to the period of rhythm and beat and take a cross correlation between the blocks into account, it is feasible to enhance compression ratios. As a starting point, one might incorporate a beat tracking stage into the compression system to estimate the parameters for the effective block length.

Another possibility for improving precision of predictors is to develop prediction

method with blockwisely varying prediction order. As we know, efficient decorrelation of signal with high treble portion and wide dynamic range needs in general higher prediction order than the case of the smooth signal. Therefore it might be useful endeavor to estimate the prediction order for each block by calculating variance of the blocks, for example.

Discrete Wavelet transform in two-channel subband scheme have been exhaustively studied. There are, however, some good reasons to believe that M-channel ($M > 2$) transforms may be more effective for signal compression. One desirable feature of M-band subband systems is that it is possible to have orthogonality with symmetric, finitely-supported scaling and wavelet functions. This is not possible with two-channel systems except for the trivial case of the Haar and other Haar-like transforms. Further work on developing a hybrid compression system that combines the M-channel wavelet transform with linear prediction for each subband would be beneficial.

Histogram forgetting is another technique that can potentially improve the proposed context based entropy coder. Using this technique, a coefficient less than one is used to multiply the older histogram counts before it is incremented. The influence of the old symbols is gradually reduced. Entropy coders with modified histograms are more likely to have a higher compression ratio than the conventional coders with uniformly-initialized histograms.

Finally, a reversible transform scheme where the scaling is chosen based on the characteristics of a particular signal would need to be devised. Reversible transform is constructed to calculate a scaled version of its parent linear transform. It has been observed, however, that one particular scaling does not lead to the best compression results in all cases. In lifting scheme, since the factorization of biorthogonal wavelets is highly non-unique process, we could design further experiments to explore potential of the integer lifting scheme.

Appendix A

Heisenberg's Uncertainty Principle

In quantum mechanics, Heisenberg's derivation of the uncertainty relations showed that the simultaneous measurement of two *canonically conjugate* variables (such as the momentum(ξ) and position(x) or the energy(E) and time(t) for a moving particle) entails a limitation on the precision (standard deviation) of each measurement. In the most extreme case, absolute precision of one variable would entail absolute imprecision regarding the other. This uncertainty relation can be written in the shorthand by,

$$\Delta x \Delta \xi \geq h/4\pi, \quad \Delta E \Delta t \geq h/4\pi, \qquad (A.0.1)$$

where h denotes Planck's constant. Suppose that ψ is a modulated waveform of finite total energy, and that both the position and momentum uncertainties of ψ are finite;

$$\Delta x(\psi) \stackrel{def}{=} \inf_{x_0} \left(\frac{\| (x - x_0)\, u(x)\, \|}{\| \psi(x)\, \|} \right) < \infty, \qquad (A.0.2)$$

$$\Delta \xi(\psi) \stackrel{def}{=} \inf_{\xi_0} \left(\frac{\| (\xi - \xi_0)\, \hat{u}(\xi)\, \|}{\| \hat{\psi}(\xi)\, \|} \right) < \infty. \qquad (A.0.3)$$

Finite Δx requires that on average $\psi(x)$ decays faster than $|x|^{-3/2}$ as $|x| \to \infty$. Finite $\Delta \xi$ requires that ψ is smooth, in the sense that ψ' must also have finite energy. Note that every function ψ belonging to the Schwartz class $\mathcal{S}$ satisfies (A.0.2).

In signal processing terms, if the window function ψ gives the instantaneous value of a time-varying signal, then it is reasonable to speak of *time t* and *frequency ω* rather than position x and momentum ξ, especially since both pairs of quantities are related by the Fourier transform. We will say then that ψ is well *localized* in both time and frequency if the product of its time and frequency uncertainties is small enough. Heisenberg's uncertainty principle states that it is impossible to know the exact frequency and the exact time of occurrence of this frequency in a signal, $\Delta t \Delta \omega \geq 1/4\pi$. In other words, if f is non zero with a compact support, then its Fourier transform cannot be zero on a whole

interval. Similarly, if its Fourier transform is compactly supported, then it cannot be zero on a time interval, i.e. in the limit case of a sinusoid, Δt is zero and $\Delta \omega$ is infinite. Hence, even if the Heisenberg constrains are verified, it is impossible to have a function in $L^2(\mathbb{R})$ which is compactly supported both in the time and Fourier domains. In particular, this means that there is no instantaneous frequency analysis for finite energy signals. The time-frequency localization is thus achievable only in the mean squares sense, i.e.,

$$\Delta t^2 = \frac{\int (t - t_0)^2 \, |\psi(t)|^2 \, dt}{\int |\psi(t)|^2 \, dt} \qquad\qquad t_0 = \frac{\int t \, |\psi(t)|^2 \, dt}{\int |\psi(t)|^2 \, dt} \qquad (A.0.4)$$

$$\Delta \omega^2 = \frac{\int (\omega - \omega_0)^2 \, |\hat{\psi}(\omega)|^2 \, d\omega}{\int |\hat{\psi}(\omega)|^2 \, d\omega} \qquad\qquad \omega_0 = \frac{\int \omega \, |\hat{\psi}(\omega)|^2 \, d\omega}{\int |\hat{\psi}(\omega)|^2 \, d\omega} \, , \qquad (A.0.5)$$

and is represented not as a "point" in the time frequency plane but as a rectangle or an ellipse (Figure A.0.1), whose position indicates the nominal time and frequency and whose shape suggests the relative uncertainties of the two quantities. The amplitude of a waveform may be indicated by darkening the rectangle in proportion to its energy. Consequently, a balance has to be reached between time and frequency resolution for efficient TFR, and it is very important how one cuts a signal.
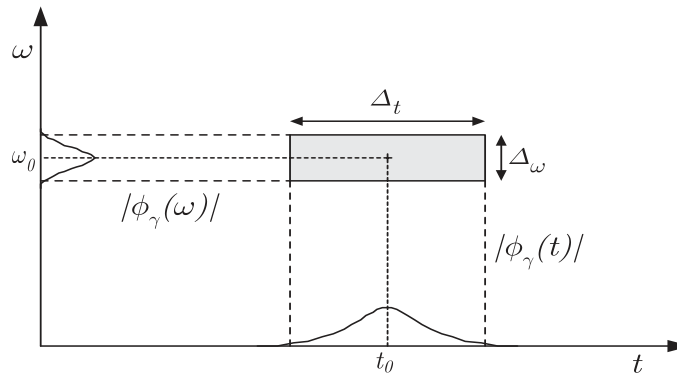
Figure A.0.1: Heisenberg's cell in time-frequency plane

In Figure A.0.1, the uncertainties in time and in frequency are given by the width and hieght of the cell, respectively. The product of the uncertainties is the area of the cell, which can never be made smaller than the lower bound $1/4\pi$ given by Heisenberg's inequality. Only for Gaussian function $g(t) = e^{-\pi t^2/2}$, the Heisenberg inequality becomes an equality, i.e. minimal Heisenberg cell area, $4\pi \Omega T = 1$. However, Gaussian function is not causal because it does not vanish for $t > 0$. In practice, we will avoid the many restrictions of the Gaussian by relaxing the minimality condition in order to perform

the approximate time-frequency analysis. For instance, the concept of multiresolution analysis (MRA) allows us to estimate the approxiamate atoms in a natural way.

APPENDIX A. HEISENBERG'S UNCERTAINTY PRINCIPLE   **171**

Appendix B

A Short History of Wavelets

The term *wavelets* or *ondelettes* (French) was coined originally in France, the so called "French school" lead by J. Morlet, A. Grossmann and Y. Meyer. The concept of wavelets can be viewed as a synthesis of ideas which indepandently originated in engineering (quadrature mirror filters, subband coding), physics (coherent states, renormalization group), and pure mathematics (study of Calderón-Zygmund operators). In looking back over the history of mathematics, for instance, we will uncover several different origins of wavelet analysis [117]. Therefore, for a conceptional understanding about wavelets, it is not vain to survey a flow of the researches which correspond to a specific point of view and a particular techniques in each area. As the history of wavelets could be the topic of a separate paper, we are going to give a short, subjective account.

Although the wavelet transform has come into prominence during the last decade, the founding principles behind wavelets can be traced back as far as 1909 when Alfred Haar [183] discovered the first orthonormal system of basis functions, at the same time, opened one of the routes leading to wavelets. For any continuous function $f(x)$, the series

$$f(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} a_{2^j+k} \, h(2^j x - k), \quad \text{with} \quad 0 \le x < 1 \tag{B.0.1}$$

converges to $f(x)$ uniformly over the interval [0,1]. This Haar's work led to the simplest of the orthogonal wavelets. However, the Haar basis function has a lack of coherence since the atoms $h_n(x)$ used to construct the continuous function $f(x)$ are not themselves continuous functions and its Fourier transform decays only like $|\xi|^{-1}$, corresponding to bad frequency localization. The theories expounded by Haar extended to Lévy's study of Brownian motion, a type of random signal, in the 1930s. He used the Schauder basis to examine local regularity properties that were not accessible through the trigonometric system. A similar difficulties were encountered when Littlewood and Paley attempted to localize the energy of function, the integral $\frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 dx$. They were interested in whether the energy of a function was spread evenly over its entire interval or concentrated

about a few points. To reveal this information that is hidden in the Fourier series of $f(x)$, Littlewood and Paley discovered a series manipulated using the "dyadic blocks" $\Delta_j f(x)$, which is a sequence of operators that act essentially as a bank of band pass filters with an interval of separation of approximately an octave. The dyadic block was defined by

$$\Delta_j f(x) = \sum_{2^j \leq k < 2^{j+1}} (a_k \cos kx + b_k \sin kx), \tag{B.0.2}$$

where $a_0 + \sum_1^\infty (a_k \cos kx + b_x sinkx)$ denotes the Fourier series of $f(x)$. Then we have a effective substitution for Fourier series,

$$f(x) = a_0 + \sum_0^\infty \Delta_j f(x), \tag{B.0.3}$$

which later provides an effective algorithm for numerical image processing, thanks to the work of Marr [184] and Mallat [185] [186].

In the 1940s, as we indicated in the preceding section, the problem of developing a mixed signal representation in term of a double sequence of a elementary signals in time-frequency plane was addressed by Dennis Gabor (1946) who introduced the first time-frequency wavelets (Gabor wavelets) in communication theory, and by Jean Ville (1947) who proposed another approach which was tied into the research of Hermann Wigner (1932), a physicist working in the field of quantum mechanics, and led to the development of the Wigner-Ville transform.

In the 1950s and 60s, the Littlewood-Paley techniques were brought in to relief as powerful tools for studying other things, such as solutions of partial differential equations and integral equations. They were then efficiently engrafted into *Calderón-Zygmund theory* (1960) in area of harmonic analysis that is still very heavily researched. It was at this point that the *mother wavelet $\psi(x)$* appeared. One of the standard approaches in that time, not only in Calderón-Zygmund theory, but in analysis in general, is to break up a complicated phenomenon into many simple pieces and study each of the pieces separately.

Guido Weiss and Ronald R. Coifman, in 1975, interpreted Lusin's Hardy spaces $H^p(\mathbb{R})$ in terms of *atoms* and *atomic decompositions* [187]. The purpose of their study is to find what we now consider the "atoms" (simplest elements of a function space) and "molecules" (assembly rules of theirs) which allow the reconstruction of all the elements of the function space using these atoms. One of the approaches to decomposition is given also by Calderón formula,

$$f(x) = \int_0^\infty \int_{-\infty}^\infty (\psi_t * f)(y)\, \tilde{\psi}_t(x - y)\, dy \frac{dt}{t}, \tag{B.0.4}$$

where $*$ denotes convolution, and $\psi_t(x) = t^{-1}\psi(x/t)$. $\tilde{\psi}_t(x)$ is defined similarly for appropriate fixed functions ψ and $\tilde{\psi}$. The representation (B.0.4) is , in fact, an example

of a continuous wavelet transform. In mathematical physics the Aslaksen-Klauder construction of the $(ax + b)$-coherent states can be seen as another independent derivation of the Calderón formula [188] [189]. With further understanding the functions in the Hardy spaces $H(\mathbb{R})$ which play now a fundamental role in signal processing, Strömberg discovered the first orthogonal wavelets [151] in the early 1980s. He showed also that the orthonormal wavelet basis for $L^2(\mathbb{R})$, defined $\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n)$, is an unconditional basis for the real Hardy space $\mathcal{H}(\mathbb{R})$.

Independent from these developments in harmonic analysis, Alex Grossmann (theoretical physicist) and Jean Morlet (geophysicist) [190] [191] [192] studied the wavelet transform in its continuous form using theory of *frames* introduced by Duffin and Schaeffer (1952) [148] [193] [194], and employed the wavelets to analyse earthquakes and model the process of sound waves travelling through the Earth's crust [195]. Indeed, their research gave an impetus to the enormous interest in wavelet theory. Later they suggested the word "wavelet" for the building blocks, and what earlier had been referred to as Littlewood-Paley theory, now started to be called wavelet theory. In terms of Grossmann-Morlet theory, which is identical to Calderón's theory, a wavelet is defined as a function in $L(\mathbb{R})$ whose Fourier transform $\psi(\xi)$ satisfies the criterion $\int_0^\infty \left|\tilde{\psi}(t,\xi)\right|^2 \frac{dt}{t} = 1$ almost everywhere.

Paralleling the advances in pure and applied mathematics were those in signal processing, but in the context of discrete-time signals. Driven by applications such as speech and image compression, a method called subband coding was proposed by Croisier, Esteban, and Galand [79] using a special class of filters called quadrature mirror filters (QMF) in the late 1970's, and by Crochiere, Webber and Flanagan [78]. This led to the study of perfect reconstruction filter banks, a problem solved in the 1980's by several people, including Smith and Barnwell [84] [85], Mintzer [86], Vetterli [88], and Vaidyanathan [196]. In a particular configuration, namely when the filter bank has octave bands, one obtains a discrete-time wavelet series. Such a configuration has been popular in signal processing less for its mathematical properties than because an octave band or logarithmic spectrum.

In the early to mid 80's, several groups, perhaps most notably the one associated with Yves Meyer and his collaborators, independently realized, with some excitement, that tools from Calderón-Zygmund theory, in particular the Littlewood-Paley representations, had discrete analogs and could give a unified view of many of the results in harmonic analysis. Also, one started to understand that these techniques could be effective substitutes for Fourier series in numerical applications. In 1985, Y. Meyer found a family of wavelets that he showed to be the most efficient for modelling complex phenomena [150] [197]. He was actually unaware at that time of Strömberg's construction of the orthonormal wavelets. Soon after, Tchamitchian constructed the first example of biorthogonal wavelet bases [163]. Battle (1987) and Lemarié (1988) constructed new orthogonal wavelet expansions using some of Tchmitchian's computations [153] [154]. In 1986,

Stephane Mallat discovered some relationships between quadrature mirror filters, pyramid algoritms and orthonormal wavelet bases through his work in numerical image processing. In result, Mallat and Meyer developed a systematic framework "multiresolution analysis" [157] [83] [152], which gave a satisfactory explanation for all constructions of orthogonal expansions, and provided a tool for the construction of yet other bases. Soon, Ingrid Daubechies gave a construction of wavelets, non-zero only on a finite interval and with arbitrarily high, but fixed, regularity [155] [198].

Since this final transition from continuous signal processing to discrete signal processing was achieved by Mallat and Daubechies, there has been a proliferation of activity with comprehensive studies expanding on the wavelet transform and its implementation into many fields of endeavour. Applications that have been explored include multiresolution signal processing, image and data compression, telecommunications, fingerprint analysis, numerical analysis and speech processing.

Bibliography

[1] M. Hans and R. W. Schafer, "Lossless compression of digital audio," *IEEE Sig. Proc. Mag.*, vol. 18, pp. 21–32, July 2001.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 398–403, July 1948.

[3] J. Rissanen and G. G. L. Jr., "Arithmetic coding," *IBM journal of Research and Development*, vol. 23, pp. 149–162, Mar. 1979.

[4] I. H. Witten, R. R. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communication of the ACM*, vol. 30, pp. 520–540, June 1987.

[5] S. D. Stearns, "Arithmetic coding in lossless waveform compresssion," *IEEE Trans. Signal Proc.*, vol. 43, pp. 1874–1879, Aug. 1995.

[6] S. W. Golomb, "Run-length encodings," *IEEE Trans. Inform. Th.*, vol. IT-12, pp. 399–401, July 1966.

[7] R. G. Gallager and D. C. V. Voorhis, "Optimal source codes for geometrically distributed integer alphabets," *IEEE Trans. Inform. Th.*, vol. IT-21, pp. 228–230, Mar. 1975.

[8] R. F. Rice, "Some practical universal noiseless coding techniques," *JPL Publication*, vol. 79-22, Mar. 1979.

[9] R. F. Rice, "Some practical universal noiseless coding techniques-part ii," *JPL Publication*, vol. 83-17, Mar. 1983.

[10] R. F. Rice, "Some practical universal noiseless coding techniques-part III," *JPL Publication*, vol. 91-3, Nov. 1991.

[11] J. Venbrux, N. Liu, K. Liu, P. Vincent, and R. Merrell, "A very high speed lossless compression/decompression chip set," *JPL Publication*, vol. 91-13, July 1991.

[12] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Prentice-Hall Signal Processing Series, 1984.

[13] J. S. Tobias, ed., *Foundation of Modern Auditory Theory*. New York and London: Academic Press, 1970.

[14] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding og High-Quality Digital Audio," *Journal of Audio Eng. Soc.*, vol. 42, pp. 780–792, Oct. 1994.

[15] P. Craven and M. J. Law, "Lossless Compression Using IIR Prediction Filters," in *102nd Convention of AES*, (Munich), p. Preprint 4415, Mar. 1997.

[16] P. Craven and M. Gerzon, "Lossless coding method for waveform data." International Patent Applicatin no. PCT/GB96/01164, May 1996.

[17] "Sound Quality Assessment Material CD." Technical Centre of the European Broadcasting Union (EBU), 1988. 4222042.

[18] H. Wold, *A Study in the Analysis of Stationary Time Series*. Stockholm, Sweden: Almquist and Wiksell, 2 ed., 1954.

[19] J. L. Doob, "The elementary gaussian processes," *Ann. Math. Statist.*, vol. 15, pp. 229–282, 1944.

[20] A. N. Kolmogorov, "Stationary sequences in Hilbert space," *Bull. Math. Univ. Moscow.*, 1941. English translaton available in [23].

[21] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationalry Time Series with Engineering Applications*. New York: Technology Press and John Wiley & Sons, Inc., 1949.

[22] N. Levinson, "The Wiener RMS (root mean square) error criterion in filter design and prediction," *Journ. of Math. Phys.*, vol. 25, no. 4, pp. 261–278, 1947.

[23] T. Kailath, "A view of three decades of linear filtering theory," *Trans. Information Theory*, vol. 20, no. 2, pp. 146–181, 1974.

[24] H. W. Sorenson, *Parameter Estimation: Principles and Problems*, vol. Vol.9 of Control and Systems Theory. New York: Marcel Dekker, Inc., 1980.

[25] E. A. Robinson, "A historical perspective of spectrum estimation," in *Proc. IEEE*, vol. 70, pp. 885–907, 1982.

BIBLIOGRAPHY 177

[26] G. U. Yule, "On a method of investigating periodicities in disturbed series with special references to Wolfer's sunspot numbers," *Philos. Trans. Roy. Soc. London, Series A*, vol. 226, no. 267-298, 1927.

[27] G. T. Walker, "On periodicity in series of related terms," in *Proc. Roy. Soc. London, Series A*, vol. 131, pp. 518–532, 1931.

[28] A. Papoulis, "Maximun entropy and apectral estimation: A review," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 29, pp. 1176–1186, Dec. 1991.

[29] J. Makhoul, "Linear prediction: A tutorial review," in *Proceedings of the IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[30] F. Itakura and S. Saito, "On the optimum quantization of feature parameters in the PACOR speech synthesizer," in *Proc. Conf. Speech Commun. and Processing*, pp. 434–437, 1972.

[31] R. W. Schafer and J. D. Markel, *Speech Anaysis*. New York: Selected Reprints Series, IEEE Press, 1979.

[32] B. Atal and S. L. Hanauer, "Speech anaysis and synthesis by linear prediction of the speech wave," *Journal of Audio Eng. Soc.*, vol. 50, pp. 637–655, 1971.

[33] H. Akaike, "A new look at the statistical model identification," *IEEE Tran. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

[34] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[35] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inform. Th.*, vol. 44, pp. 2743–2760, Oct. 1998.

[36] L.-Z. Tan, *Theory and Techniques for Lossless Waveform Data Compression*. PhD thesis, The University of New Mexico, 1992.

[37] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[38] J. Rissanen, "A predictive least squares principle," *IMA J. Math. Contr. Inform.*, vol. 3, no. 2-3, pp. 221–222, 1986.

[39] M. Wax, "Order selection for ar models by predictive least squares," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, pp. 581–588, Apr. 1988.

[40] J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-27, no. 1, pp. 63–73, 1979.

[41] T. Robinson, "Shorten: Simple lossless and near-lossless waveform compression," Tech. Rep. CUED/F-INFENG/TR.156, Cambridge University, UK, Dec. 1994.

[42] H. Cremer, "On some classes of nonstattionalry processees," in *Proc. 4th Berkeley Symp. Math. Statist. Probability*, vol. 2, (Berkeley, CA, USA), pp. 57–78, Univ. California Press, 1961.

[43] M. B. Priestley, *Spectral Analysis and Time Series*. Vol. 2 of *Probability and Mathematical Statistics*, London: Academic Press Inc., 1981.

[44] J. P. Kaipio and M. Juntunen, "Deterministic regression smoothness prior $t$var modelling," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, vol. III, pp. 1693–1696, IEEE, Phoenix, Arizona, USA, 1999.

[45] T. Subba Rao, "The fitting of non-stationary signals," *J. R. Statis. Soc.*, vol. B32, pp. 312–322, 1970.

[46] T. Barnwell, "Recursive autocorrelation computation for (lpc) analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, (Hartford), pp. 1–4, 1977.

[47] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-V: The discrete case," *Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, 1978.

[48] J. D. Gibson, S. K. Jones, and J. L. Melsa, "Sequentially adaptive prediction and coding of speech signals," *IEEE Trans. Commun.*, vol. 22, no. 11, pp. 1789–1797, 1974.

[49] N. S. Jayant, "Adaptive quantization with one word memory," *Bell System Technical Journal*, pp. 1119–1144, 1973.

[50] J. D. Gibson, Y. C. Cheong, W.-W. Chang, and H. C. Woo, "A comparison of backward adaptive prediction algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, vol. 1, (Albuquerque, New Mexico), pp. 237–240, 1990.

[51] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ:Printice-Hall, 1996.

[52] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Printice-Hall, 1985.

BIBLIOGRAPHY

[53] L. J. Griffiths, "A continuously-adaptive filter implemented as a lattice structure," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, (Hartford, USA), pp. 683–686, 1977.

[54] S. J. Orfanidis, *Optimum Signal Processing, An Introduction*. Signapore: McGraw-Hill, second edition ed., 1990.

[55] H. W. Strube, "Linear prediction on a warped frequency scale," *Journ. of Acoust. Soc. Am.*, vol. 68, pp. 1071–1076, October 1980.

[56] U. K. Laine, M. Karjalainen, and T. Altosaar, "WLP in speech and audio processing," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, vol. 3, pp. 349–352, 1994.

[57] A. Härmä, U. K. Laine, and M. Karjalainen, "Warped linear prediction in audio coding," in *Proc. NORSIG'96*, 1996.

[58] A. Härmä, U. K. Laine, and M. Karjalainen, "WLPAC-a perceptual audio codec in a nutshell," in *AES 102nd Conv.*, (Munich, Germany), p. Preprint 4420, 1997.

[59] A. Härmä and U. K. Laine, "A comparison between frequency-warped and conventional linear predictive coding," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 9, pp. 579–588, July 2001.

[60] B. Ninness and F. Gustafsson, "A unifying construction of orthogonal bases for system identification," *IEEE Trans. Automatic Control*, vol. 42, no. 4, pp. 515–521, 1997.

[61] Y. W. Lee, *Statistical Theory of Communication*. New York: Wiley, 1960.

[62] D. G. Messerschmitt, "A class of generalized lattice filters," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-28, pp. 198–204, Apr. 1980.

[63] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transform," *IEEE Trans. Speech, Audio Proc.*, vol. 7, pp. 697–708, Nov. 1999.

[64] J. Rissanen, "A universal data compression system," *IEEE Trans. Inform. Th.*, vol. IT-29, pp. 656–664, Sept. 1983.

[65] M. Weinberger, J. Rissanen, and M. Feder, "A universal finite memory source," *IEEE Trans. Inform. Th.*, vol. IT-3, pp. 643–652, May 1995.

[66] M. Weinberger, J. Rissanen, and R. Arps, "Applications of universal context modeling to lossless compression of gray-scale images," *IEEE Trans. Image Proc.*, vol. IP-5, pp. 575–586, Apr. 1996.

[67] X. Wu, "Lossless compression of continuous-tone images via context selection, quantization, and modeling," *IEEE Trans. Image Proc.*, vol. 6, pp. 656–664, May 1997.

[68] S. Todd, G. Langdon, and J. Rissanen, "Parameter reduction and context selection for compressing of grey-scale images," *IBM J. Res. Develop*, pp. 188–193, Mar. 1985.

[69] I. Tăbuş and J. Astola, "Adaptive Boolean predictive modeling with application to lossless image coding," in *SPIE-Statistical and Stochastic Methods for Image Processing II*, (San Diego, Califonia), pp. 234–245, Oct. 1997.

[70] I. Tăbuş, J. Rissanen, and J. Astola, "Adaptive L-predictors based in finite state machine context selection," in *ICIP'97 International Conference on Image Processing*, (Santa Barbara), pp. 401–404, Oct 1997.

[71] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[72] C. D. Giurcăneanu, I. Tăbuş, and J. Astola, "Adaptive context based sequential prediction for lossless audio compression," in *Proc. Eusipco-98, IX European Signal Processing Conference*, (Rhodes, Greece), Sept. 1998.

[73] M. J. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A Low Complexity, Context-based Lossless Image Compression Algorithm," in *Proc. Data Compr. Conf.*, vol. 40, (Snowbird, Utah, USA), pp. 140–149, Oct. 1996.

[74] P. G. Howard and J. S. Vitter, "Fast and efficient lossless image compression," in *Proc. Data Compr. Conf.*, (Snowbird, Utah, USA), pp. 351–360, Mar. 1993.

[75] X. Wu and N. Memon, "Context-based adaptive lossless image coding," *IEEE Trans. Commun.*, vol. 45, pp. 437–444, April 1997.

[76] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532–540, Apr. 1983.

[77] R. Schäfer, P. Kauft, and U. Gölz, "On the application of spatio-temporal contrast sensitivity functions to HDTV," in *Conference on Applied Vision*, (San Francisco), pp. 118–121, Optical Society of America, July 1989.

[78] R. E. Crochiere, S. A. Weber, and J. L. Flanagan, "Digital coding of speech in subbands," *Bell System Technical Journal*, vol. 55, pp. 1069–1085, Oct. 1976.

[79] A. Croisier, D. Esteban, and C. Galand, "Perfect channel spliting by use of interpolation, decimation, tree decomposition techniques," in *Int. Conf. on Information Sciences and Systems*, (Patras, Greece), pp. 443–446, Aug. 1976.

[80] D. Esteban and C. Galand, "Application of quadrature mirror filters to split-band voice coding schemes," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, (Hartford, Connecticut), pp. 191–195, May 1977.

[81] Z. Cvetković and M. Vetterli, "Oversampled filter banks," *IEEE Trans. Signal Proc.*, vol. 46, pp. 1245–1255, May 1998.

[82] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transforms for image coding," in *Proc. SPIE*, vol. 845, (Cambridge, MA), pp. 50–58, Oct. 1987.

[83] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," in *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 11, pp. 674–693, July 1989.

[84] M. J. Smith and T. P. Barnwell III, "A procedure for designing exact reconstruction filter banks for tree-structured subband coders," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, (San Diego, CA), pp. 27.1.1–27.1.4, March 1984.

[85] M. J. Smith and T. P. Barnwell III, "Exact reconstruction for tree-structured subband coders," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 34, pp. 431–441, June 1986.

[86] F. Mintzer, "Filters for distortion-free two-band multirate filter banks," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 33, pp. 626–630, June 1985.

[87] M. Vetterli, "Splitting a signal into subsampled channels allowing perfect reconstruction," in *Proc. IASTED Conf. on Appl. Sig. Proc. and Dig. Filt.*, (Paris), June 1985.

[88] M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Proc.*, vol. 10, pp. 219–244, April 1986.

[89] P. P. Vaidyanathan, "Quadrature mirror filter banks, m-band extensions and perfect reconstruction techniques," *IEEE ASSP Mag.*, vol. 4, pp. 4–20, July 1987.

[90] H. J. Nussbaumer, "Pseuso qmf filter bank," in *IBM Tech. disclosure bulletin*, vol. 24, pp. 3081–3087, Nov. 1981.

[91] J. H. Rothweiler, "Polyphase quadrature filters, a new subband coding technique," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, vol. 24, pp. 1980–1983, Apr. 1983.

[92] G. Stoll and F. Dehéry, "High quality audio bit rate reduction family for different applications," in *Proc. IEEE Int. Conf. Commun.*, pp. 937–941, Apr. 1990.

[93] Y. F. Dehéry, G. Stoll, and L. Kerkof, "MUSICAM: Source Coding for Digital Sound," in *17th Int. Television Symp.*, (Montreux, Switzerland), p. Preprint 2937, May 1990.

[94] Y. F. Dehéry, M. Lever, and P. Urcum, "A MUSICAM source codec for digital audio broadcasting and storage," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, (Toronto, Canada), pp. 3605–3608, May 1991.

[95] G. Theile and G. Stoll, "MUSICAM-Surround: A Universal Multi-Channel Coding System Compatible with ISO 11172-3," in *93rd Convention of AES*, (San Frrancisco), p. Preprint 3403, Oct. 1992.

[96] J. D. Johnson, "A filter family ddesigned for use in quadrature mirror filter banks," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pp. 291–294, 1980.

[97] V. K. Jain and R. E. Crochiere, "Quadrature mirror filter design in the time domain," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 32, pp. 353–360, April 1984.

[98] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[99] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. NJ:Prentice Hall, 1993.

[100] I. Daubechies, *Ten Lectures on Wavelets*. no. 61 in CBMS-NSF Series in Applied Mathematics, Philadelphia: SIAM, 1992.

[101] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. on Pure and Appl. Math.*, vol. 45, pp. 485–560, 1992.

[102] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Proc.*, vol. 40, pp. 2207–2232, Sept. 1992.

[103] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Proc.*, vol. 1, pp. 205–220, Apr. 1992.

[104] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Proc.*, vol. 4, pp. 1053–1060, Aug. 1995.

[105] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. NY: Academic Press, 1990.

[106] H. S. Malvar and D. H. Staelin, "Lapped transforms for efficient transform/subband coding," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 38, pp. 969–978, June 1990.

BIBLIOGRAPHY

[107] H. S. Malvar, *Signal Processing with Lapped Transform*. Nordwood: Artech House, 1992.

[108] R. L. de Queiroz, T. Q. Nguyen, and K. R. Rao, "The genlot: generalized linear-phase lapped orthogonal transform," *IEEE Trans. Signal Proc.*, vol. 40, pp. 497–507, Mar. 1996.

[109] M. Vetterli and D. L. Gall, "Perfect reconstruction FIR filter banks: Lapped transforms, pseudo-QMF's and paraunitary matrices," in *Proc. IEEE Int. Symp. Circ. and Syst.*, (Espoo, Finland), pp. 2249–2253, June 1988.

[110] A. N. Akansu and F. E. Wadas, "On lapped orthogonal transforms," *Signal Proc.*, vol. 40, Feb. 1992.

[111] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-34, pp. 1278–1286, Oct. 1986.

[112] J. W. Woods, ed., *Subband Image Coding*. Kluwer Academic Publishers, 1991.

[113] S. Rao and W. A. Pearlman, "On the superiority of coding and estimation from subbands," in *Proc. 1992 Conf. on Information Sciences and Systems*, (Princeton, N.J.), pp. 890–895, Mar. 1992.

[114] S. Rao, *Analysis of Linear Prediction and Spectral Estimation from Subbands with Applications*. PhD thesis, Electrical, Computer and Systems Engineering Dept., Rensselaer Polytechnic Institute, Dec. 1993.

[115] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, pp. 1169–1179, 1988.

[116] J. G. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," *IEEE Trans. Biomedical Engineering*, vol. 36, no. 1, pp. 107–114, 1989.

[117] Y. Meyer, *Wavelets: Algorithms and Applications*. Philadelphia: SIAM, 1993.

[118] S. Li, "A general theory of discrete Gabor expansion," in *Proc. SPIE'94 Mathematical Imaging: Wavelet Applications*, (San Diago, CA), July 1994.

[119] S. Li and D. M. Healy Jr., "A parametric Class of discrete Gabor Expansions," *IEEE Trans. Signal Proc.*, vol. 44, pp. 201–211, Feb. 1996.

[120] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Trans. Signal Proc.*, vol. 41, pp. 2429–2439, July 1993.

[121] J. Yao, "Complete Gabor trnaformation for signal representation," *IEEE Trans. Image Proc.*, vol. 2, pp. 152–159, Apr. 1993.

[122] D. Gabor, "Acoustical quanta and the theory of hearing," *Natur*, vol. 159(1044), pp. 591–594, May 1947.

[123] C. Roads, "Automated granular synthesis of sound," *Computer Music Journal*, vol. 2(2), pp. 61–62, 1978.

[124] D. Jones and T. Parks, "Generation and combination of grains for music synthesis," *Computer Music Journal*, vol. 12(2), pp. 27–34, 1988.

[125] B. Truax, "Discovering inner complexity: Time shifting and transposition with a real-time granulation technique," *Computer Music Journal*, vol. 18(2), pp. 38–48, 1994.

[126] E. Mouines and J. Laroche, "Non-parametric techniques for pitch-scale and time-scale modification of speech," *Speech Communication*, vol. 16, pp. 175–205, 1995.

[127] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 34, Aug. 1986.

[128] X. Serra and J. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14(4), pp. 12–24, 1990.

[129] P. Prandoni, M. goodwin, and M. Vetterli, "Optimal time segmentation for signal modeling and compression," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, 1997.

[130] B. Boashash, *Time-Frequency Signal Analysis*, ch. in Advances in Spectrum Analysis and Array Processing, pp. 418–517. Prentice-Hall, Englewood Cliffs, S. Haykin ed., 1990.

[131] C. P. Janse and A. Kaiser, "Time-frequency distributions of loud-speakers: The application of the wigner distribution," *Journal of Audio Eng. Soc.*, vol. 37, pp. 198–223, 1983.

[132] L. Cohen, "Generalized phase-space distribution functions," *Journ. of Math. Phys.*, vol. 7, pp. 781–806, 1966.

[133] L. Cohen, "Time-frequency distribution - a review," in *Proc. IEEE*, vol. 77, pp. 941–981, July 1989.

[134] L. Cohen, *Time-Frequency Anaysis*. Englewood Cliffs, NJ: Prentice Hall, 1995.

BIBLIOGRAPHY

[135] H. Choi and W. J. Williams, "Improved time-frequency representation of multi-component signals using exponential kernels," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 37, pp. 862–871, June 1989.

[136] Y. Sheng, "Wavelet transform," in *The Transforms and Applications Handbook* (A. D. Poularikas, ed.), The Electrical Engineering Handbook Series, pp. 747–827, Boca Raton, FL(USA): CRC Press, 1996.

[137] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proc. IEEE Conf. on Image Proc.*, vol. 1, (Santa Barbara, California), pp. 596–599, Oct. 1997.

[138] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms, A Primer.* Upper Saddle River, NJ, USA: Pretice Hall, 1998.

[139] M. Frazier and B. Jawerth, "Decomposition of Besov spaces," *Indiana Univ. Math. J.*, vol. 34, pp. 777–799, 1985.

[140] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Th.*, vol. 38, pp. 617–643, 1992.

[141] R. Murenzi, "Wavelet transforms associated to the n-dimensional Euclidean group with dilations: signals in more than one dimension," in *Wavelets* (J. M. Combes, A. Grossmann, and P. Tchamitchian, eds.), pp. 239–246, Berlin: Springer-Verlag, 1989.

[142] F. Argoul, A. Arnéodo, J. Elezgaray, G. Grasseau, and R. Murenzi, "Wavelet transform of two-dimensional fractal aggregates," *Phys. Lett. A*, vol. 135, pp. 327–336, 1989.

[143] M. Frazier, "A discrete transform and decompositions of distribution spaces," *J. Func. Anal., 93*, pp. 34–170, 1990.

[144] I. Daubechies, "The wavelet transformation, time-frequency localization and signal analysis," *IEEE Trans. Inform. Th.*, vol. 36, pp. 961–1005, Sept. 1990.

[145] J. J. Benedetto, "Irregular sampling and frames," in *Wavelets: A Tutorial in Theory and Applications* (C. K. Chui, ed.), pp. 445–507, San Diego: Academic Press, 1992.

[146] H. G. Feichtinger and K. Gröchenig, "Irregular sampling theorems and series expansions of band-limited functions," *SIAM Journ. on Math. Analysis*, vol. 23, 1992.

BIBLIOGRAPHY

[147] H. G. Feichtinger and K. Gröchenig, "Theory and practice of irregular sampling," in *Wavelets: Mathematics and Applications* (J. Benedetto and M. Frazier, eds.), pp. 305–363, Boca Raton, FL, USA: CRC Press, 1993.

[148] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic Fourier series," *Trans. Amer. Math. Soc.*, vol. 72, pp. 341–366, 1952.

[149] C. E. Heil, "Wavelets and frames," in *Signal Processing, Part 1: Signal Processing Theory* (L. Auslander, et al., ed.), vol. 22 of *IMA*, pp. 147–160, New York: Springer, 1990.

[150] Y. Meyer, "*Principe d'incertitude, bases hilbertiennes et algébres d'opérateurs.*" Séminaire Bourbaki, no. 662, 1985-1986.

[151] J. O. Strömberg, "A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces," in *Conference on Harmonic Analysis in Honor of Antoni Zygmund* (W. Beckner et al., ed.), vol. II, pp. 475–493, Wadsworth math. series, 1982.

[152] Y. Meyer, *Ondelettes et Opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs multilinéaires*. Paris: Hermann, 1990.

[153] G. Battle, "A block spin construction of odelettes. Part I: Lemarié functions," *Comm. Math. Phys.*, vol. 110, pp. 601–615, 1987.

[154] P. G. Lemarié, "Ondelettes à localisation exponentielle," *Journ. de Math. Pures et Appl.*, vol. 67, pp. 227–236, 1988.

[155] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. on Pure and Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.

[156] R. Balian, *Un principe d'incertitude en théorie du signal ou en mécanique quantique*. Série 2, C.R. Acad. Sci. Paris, 292, 1981.

[157] S. Mallat, "Multiresolution approximations and Wavelet Orthonormal Bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, Sept. 1989.

[158] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," *Construct. Aspects of Funct. Anal.*, pp. 796–830, 1971.

[159] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Review*, vol. 31, no. 4, pp. 614–627, 1989.

[160] G. Fix and G. Strang, "Fourier analysis of the finite element method on Ritz-Galerkin theory," *Stud. Appl. Math*, vol. 48, pp. 265–273, 1969.

BIBLIOGRAPHY

[161] I. Daubechies and J. C. Lagarias, "Two-scale difference equattions: I. Existence and global regularity of solutions," *SIAM Journ. on Math. Analysis*, vol. 22, pp. 1388–1410, 1991.

[162] I. Daubechies and J. C. Lagarias, "Two-scale defference equations: II. Local regularity, infinite products of matrices and fractals," *SIAM Journ. on Math. Analysis*, vol. 23, pp. 1031–1079, 1992.

[163] P. Tchamitchian, "Biorthogonalité et théorie des opérateurs," *Revista Mathemática Iberoamericana*, vol. 3, no. 2, pp. 163–189, 1987.

[164] G. Polya and G. Szegö, *Aufgaben and Lehrsätze aus der Analysis*, vol. Vol. II. Berlin: Springer Verlag, 1971.

[165] J. M. Lina and M. Mayrand, "Complex daubechies wavelets," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 2, pp. 219–229, 1995.

[166] O. Rioul, "Simple regularity criteria for subdivision schemes," *SIAM Journ. on Math. Analysis*, no. 23, pp. 1544–1576, 1992.

[167] K. Irie and R. Kishimoto, "A study on the perfect reconstructive subband coding," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 1, pp. 42–48, Mar. 1991.

[168] I. Shah, O. Akiwumi-Assani, and B. Johnson, "A chip set for lossless image compression," *IEEE Journal Solid-State Circuits*, vol. 26, no. 3, pp. 237–244, 1991.

[169] D. L. Gall and A. Tabatabai, "Subband coding of digital images using symmetric short kernel filters and arithmetic coding techniques," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, (New York), pp. 761–765, 1988.

[170] D. Speck, "Low-complexity subband coding for image compression," tech. rep., Computer Research Laboratory, University of California at Santa Cruz, May 1993.

[171] J. Villasenor, B. Belzer, and J. Liao, "Filter evaluation and selection in wavelet image compression," in *IEEE Data Compression Conference*, (Snowbird, Utah), 1994.

[172] A. Zandi, J. Allen, E. Schwartz, and M. Boliek, "CREW: Compression with reversible embedded wavelets," in *IEEE Data Compression Conference*, (Snowbird, Utah), Mar. 1995.

[173] A. Said and W. Pearlman, "Reversible image compression via multiresolution representation and predictive coding," in *Proc. SPIE in Visual Commun. and Image Proc.*, vol. 2094, pp. 664–674, Nov. 1993.

[174] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy compression," *IEEE Trans. on Image Processing*, vol. 5, pp. 1303–1310, Sept. 1996.

[175] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 3, no. 2, pp. 186–200, 1996.

[176] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journ. on Math. Analysis*, vol. 29, no. 2, pp. 511–546, 1997.

[177] C. K. Chui, *An Introduction to Wavelets*. San Diego, CA: Academic Press, 1992.

[178] C. Herley and M. Vetterli, "Wavelets and recursive filter banks," *IEEE Trans. Signal Proc.*, vol. 41, pp. 2536–2556, Aug. 1993.

[179] M. G. Bellanger and J. L. Daguet, "TDM-FDM transmultiplexer: Digital polyphase and FFT," *IEEE Trans. Commun.*, vol. 22, no. 9, pp. 1199–1204, 1974.

[180] P. Schröder and W. Sweldens, "Spherical wavelets: Efficiently representing functions on the sphere," in *SIGGRAPH '95 Proceedings (Computer Graphics)*, pp. 161–172, 1995.

[181] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *Journal of Fourier Analysis and Applications*, vol. 4, no. 3, pp. 245–267, 1998.

[182] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 5, no. 3, pp. 332–369, 1998.

[183] A. Haar, "Zur Theorie der Orthogonalen Funktion-systeme," *Mathematics Analysis*, vol. 69, pp. 331–371, 1910.

[184] D. Marr, *Vision*. New York: W. H. Freeman and Co., 1982.

[185] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 37, pp. 2091–2110, Dec. 1989.

[186] S. Mallat and S. Zhong, "Complete signal representation with multiscale edges," Robotics Report 219, Courant Institute of Mathematical Sciences, New York, Dec. 1989.

[187] R. R. Coifman, *A real variable characterization of H^p*. 51, Studia Math, 1974.

BIBLIOGRAPHY

[188] E. W. Aslaksen and J. R. Klauder, "Unitary representation of the affined group," *Journ. of Math. Phys.*, vol. 9, pp. 2267–2275, 1968.

[189] J. R. Klauder and B. S. Skagerstam, *Coherent States*. World Scientific, Signapore, 1985.

[190] A. Grossman and J. Molet, "Decomposition of Hardy Function into Square Integrable Wavelets of Constant Shape," *SIAM Journ. on Math. Analysis*, vol. 15, pp. 723–736, July 1984.

[191] A. Grossmann and J. Morlet, "Decomposition of functions into wavelets of constant shape, and related transforms," in *Mathematics and Physics, Lectures on Recent Results* (L. Streit, ed.), World Scientific Publishing, Signapore, 1985.

[192] A. Grossmann, J. Morlet, and T. Paul, "Transforms associated to square integrable group representations I. General results," *Journ. of Math. Phys.*, vol. 26, pp. 2473–2479, 1985.

[193] R. M. Young, *An Introduction to Nonharmonic Fourier Series*. New York: Academic Press, 1980.

[194] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *Journ. of Math. Phys.*, vol. 27, pp. 1271–1283, 1986.

[195] A. Grossmann, M. Holschneider, R. Kronland-Martinet, and J. Morlet, "Detection of abrupt changes in sound signals with the help of wavelet transforms," in *Inverse Problem: An Interdisxiplinary Study; Advances in Electronics and Electron Physics, Supplement 19*, (New York), pp. 298–306, Academic Press, 1987.

[196] P. P. Vaidyanathan, "Theory and design of m-channel maximally decimated quadrature mirror filter with arbitrary m, having perfect reconstruction property," *IEEE Trans. Acoust., Speech, and Signal Proc.*, pp. 476–492, Apr. 1987.

[197] Y. Meyer, "De la recherche pétrolière à la géometrie des espaces de Banach en passant par les paraproduits," tech. rep., École Polytechnique, Palaiseau, 1985-1986.

[198] I. Daubechies, "Orthonormal basess of compactly supported wavelets II. Variations on a theme," *SIAM Journ. on Math. Analysis*, vol. 24, pp. 499–519, 1993.

Curriculum Vitae

Name: Jong-Hwa Kim
Date of Birth: May 5. 1966 in Seoul / Republic of Korea

EDUCATION

Ph.D. **Candidate in Communication Sciences**, degree expected in 2003
Institute for Communication Sciences, Technical University Berlin, Germany
Dissertation title: *Lossless Wideband Audio Compression: Prediction and Transform*
Supervisor: Prof. Dr. Manfred Krause, Prof. Dr. Helga de la Motte-Haber
Completion of Course: **Diplom-Toningenieur**, summer 1997 - winter 1998
Communication Engineering I II, Sound Studio Engineering I II, Digital Signal
Processing, Lab. and Studio Practices

M.Sc. **Electronic Engineering**, February 1994
DSP Laboratory, Graduate School of Kyungwon University, Seongnam, Korea
Thesis: *A study on signal reconstruction using adapted iterative estimation algorithm*
Supervisor: Prof. Dr. Sung-Il Kim

B.A. **Electronic Engineering**, February 1992
Department of Electronic Engineering, Kyungwon University, Seongnam, Korea

POSITIONS HELD AND PROJECTS

Research Assistant, July 2002-present, Lab for Multimedia Concepts and Applications,
Dept. of Computer Sciences, University Augsburg, Germany.

PhD Work, 1998-present, supported by NaFöG Scholarship from Berliner Senate.
Institute for Communication Sciences, Technical University Berlin, Germany.
Multimedia Signal Processing: Video/image/audio coding (source coding and source-
channel joint coding), multichannel sound system (surround sound, ambisonics) and ap-
plication of wavelet transform.

Software Engineer, 2000-2002, Inscribe GmbH, Berlin, Germany.
Virtual Reality: Programming of real-time 3D-View/Sound engine for virtual-reality simu-
lation and DVD authoring tool.

Project Scientist, 1998-1999, Cegelec AEG GmbH, Berlin, Germany.
Control Simulation: Design and development of simulation system for flux control of asyn-
chronous motor using adapted Kalman filter.

Research Assistant, 1993-1994, Samkyung Electronics Co., Seoul, Korea.
Sound Chip Design: Design and development of real-time sound processing system for
musical instruments.

Student Employee, 1992-1994, DSP Lab, Kyungwon University, Seongnam, Korea.
Assisted teaching and laboratory in electronics.