Multirate Statistical Signal Processing

Springer

MULTIRATE STATISTICAL SIGNAL PROCESSING

Omid S. Jahromi

Multirate Statistical Signal Processing

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-5316-9 (HB)
ISBN-13 978-1-4020-5316-0 (HB)
ISBN-10 1-4020-5317-7 (e-book)
ISBN-13 978-1-4020-5317-7 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

"If you want to build a ship don't herd people together to collect wood and don't assign them tasks and work, but rather teach them to long for the endless immensity of the sea."

The Little Prince by Antoine de Saint-Exupry, 1943

Preface

The field of multirate signal processing has witnessed a great deal of progress and an increasingly wide range of applications since the publication of the first textbook by Crochiere and Rabiner (1983). However, this progress has been mainly in the area of deterministic systems with emphasis on perfect-reconstruction and/or orthogonal systems.

This book introduces a statistical theory for extracting information from signals that have different sampling rates. This new theory generalizes the conventional (deterministic) theory of multirate systems beyond many of its constraints. Furthermore, it allows for the formulation of several new problems such as spectrum estimation, time-delay estimation and sensor fusion in the realm of multirate signal processing.

I have arrived at the theory presented here by integrating concepts from diverse areas such as information theory, inverse problems and theory of inequalities. The process of merging a variety of concepts of different origin results in both merits and shortcomings. The former include the fresh and undifferentiated view of an amateur, providing scope of application. The latter include a lack of in-depth experience in each of the original fields. Granted, this may lead to gaps in continuity, however it goes without saying that a complete theory can seldom be achieved by one person and in a short time.

My goal in writing this book has been to inspire the reader to initiate his own research and add to the theory of multirate statistical signal processing. I have tried to present background material, key principles, potential applications and open research problems while striking the appropriate balance between clarity and brevity. I hope you find it informative, useful and above all interesting!

Acknowledgments

It is a great pleasure to me to acknowledge those many people who have influenced my thinking and contributed to my knowledge. The now-classic

book by Prof. Vaidyanathan (1993) was the initial source that triggered my interest in multirate signal processing back in 1995. I am very indebted to my M.Sc. thesis advisor, Prof. M. A. Masnadi-Shirazi, who encouraged me and supported my research on filter bank theory at Shiraz University during the period 1995–97.

I would like to express my deepest gratitude to my Ph.D. thesis supervisors, Profs. Bruce Francis and Raymond Kwong at the University of Toronto. Bruce in particular provided me with unprecedented freedom to spend my time on almost any topic that stimulated my curiosity. His liberal supervision combined with his strict emphasis on clarity were fundamental in shaping the research effort that led to the theory presented in this book. Moreover, Bruce and Raymond both supported me financially throughout the entire period of my Ph.D. study from 1998 to 2002.

I also wish to thank the support, friendship and encouragement I received from Prof. Parham Aarabi while working as a postdoctoral fellow at his Artificial Perception laboratory (APL) at the University of Toronto. The stimulating, youthful and yet relaxed environment of APL was fundamental in shaping the research that led to the material in Chapters 5 and 8. Furthermore, Parham has remained a true friend and a constant source of encouragement far beyond what I can possibly acknowledge here.

I admit that the style and the presentation of this book have been greatly influenced by my acquaintance over the years with Dr. David Smith of Toronto. Whether we got together for a morning coffee at the Arbor Room[1] or shared a melancholic evening at the Rebel House[2], David always managed to teach me something new about life!

Finally, I would like to thank Mr. Mark de Jongh, my editor at Springer, for his patience and understanding during the long and treacherous process of writing this book.

Palm Beach, Florida, January 2007 Omid Jahromi

[1] `http://www.harthouse.utoronto.ca/`
[2] `http://www.rebelhouse.ca/`

Contents

1

Introduction

1.1 Digital signal processing

Digital signal processing (DSP) is one of the great technological innovations of the twentieth century. DSP is the art of analysis, manipulation and interpretation of signals using digital computers. This includes a wide variety of goals: noise reduction, classification, enhancement, compression, filtering and much more. With the increasing use of general-purpose and embedded computers the usage and need of digital signal processing has increased dramatically. DSP has revolutionized not only industrial applications such as radar, communications receivers and medical imaging, but also our everyday life. Today, advanced DSP computations are done routinely in mobile phones, in entertainment appliances such as CD/DVD/MP3-players, in high-definition television (HDTV), in car safety systems and so on. DSP will remain as one of the most powerful technologies to shape science and engineering in the twenty-first century as well.

The popularity of digital signal processing stems from certain advantages: DSP filters do not need tuning and may be exactly duplicated from unit to unit; temperature variations are virtually non-existent; and DSP represents the ultimate in flexibility, since general-purpose DSP hardware can be programmed to perform many different functions, often eliminating other hardware.

The world of science and engineering is filled with signals: images captured by remote space probes, voltages generated by the heart and brain, radar and sonar echoes, seismic vibrations, and countless other natural and artificial sources. Virtually all natural signals are *analog*, which means they are real-valued quantities varying continuously in time. In order to process an analog signal on a digital computer the signal must be digitized. An analog signal $x(t)$ can be digitized using a process involving sampling and quantization. Sampling simply means taking samples of the signal at discrete times $t = nT_s$, $n = 0, 1, 2, 3$, etc. The rate at which the signal is sampled is called the *sampling rate* or *sampling frequency* and is denoted f_s. Clearly, $f_s = 1/T_s$.

One of the main consequences of sampling an analog signal is a phenomenon known as *aliasing*. Aliasing is an effect that causes different continuous signals to become indistinguishable (or aliases of one another) when sampled. When this happens, the original signal cannot be uniquely reconstructed from the sampled signal. If the effective bandwidth of an analog signal $x(t)$ is W Hz, the famous Nyquist Sampling Theorem states that the minimum sampling frequency needed for aliasing-free discretization of $x(t)$ is $f_s = 2W$.

Aliasing is a major concern in the analog-to-digital conversion of audio/visual signals: improper sampling of the analog signal might cause high-frequency components to be aliased with genuine low-frequency ones, and be incorrectly reconstructed as such during the subsequent digital-to-analog conversion. To prevent this problem, the signals must be appropriately filtered before sampling.

Once a signal has been sampled, one needs to convert the samples into a binary (digital) number that is directly proportional to the real-valued input sample. This process is called *quantization*. Obviously, this is a process of approximation since the number of bits in the binary output limits the number of discrete signal levels that can be represented. Therefore, some information is lost.

Remark 1.1. In this book we assume that sufficiently long binary numbers are used during analog to digital conversion so that the error caused by quantization (or *quantization noise* as it is often called) is negligible. In view of this, we use the terms *discrete-time signal* and *digital signal* interchangeably unless otherwise stated.

In addition to quantization noise, noise is introduced in practical analog to digital converters (ADCs) by slight variations in the exact times of sampling. Phase noise in the ADC's clock source, as well as other inaccuracies in the sampling mechanisms contribute to this *aperture jitter noise*.

The nature of the above mentioned noise sources is such that if we could increase the sampling rate by a factor of N, then digitally filter the output and convert it back down to a lower rate, we could improve the signal to noise ratio by almost the factor N. This is because both the quantization noise and the aperture jitter noise would be spread over a larger band width thus much of the high-frequency noise would be eliminated by the digital filter. This technique is called *over sampling*[1]. Sampling at high rates is beneficial during digital to analog conversion as well. It reduces the effect of *zero-order sample-and-hold distortion* used in practical digital to analog converters (DACs) and simplifies the design of the filters used to avoid aliasing.

[1] So-called *sigma-delta converters* use this method to achieve the best possible dynamic range. They use one-bit quantizers at a very high sampling rate and digital decimation techniques (described later in this chapter) to reduce the sampling frequency, thus improving signal to noise ratio. Sigma-delta converters represent the state of the art in ADC technology at the time of this writing.

While an artificial increase in sampling rate is beneficial both during analog to digital conversion and converting back to analog, lower internal sampling rates help reduce the computational burden in DSP systems. In addition, when digitally filtering some signals, making the filters' bandwidth a large fraction of the sampling frequency makes it easier to build sharp-skirted filters — exactly what DSP is famous for. In view of these considerations, there is a clear need for sampling rate conversion in DSP systems. To meet this need and to provide a deeper understanding of how to process digital signals in systems that require more than one sampling rate, an entire subfield of digital signal processing known as *multirate signal processing* has been developed.

1.2 Multirate signal processing

Conceptually, there is a very simple and straightforward approach to changing the sampling rate of a digital signal. In this approach one merely reconstructs the continuous-time signal from the original set of samples and then resamples the signal at the new rate (assuming that no additional anti-aliasing filtering is required). This approach, however, is not without problems in practice. A major practical problem is that the ideal operations required to reconstruct the continuous-time signal from the original samples and to resample the signal at the new rate cannot be implemented exactly. Due to limitations of practical ADC and DAC devices, the resulting signal will have additive noise (due to resampling), signal-dependent distortions (due to non-ideal samplers), and frequency distortions (due to non-ideal frequency response of the analog filters used). These practical distortions can be minimized by a careful design of the individual components used but cannot be eliminated entirely.

Multirate signal processing techniques provide a very attractive alternative to the above *analog approach* to sampling rate conversion. The following basic operations are essential in multirate signal processing:

(i) Decimation (sampling rate decrease) by an integer factor M
(ii) Interpolation (sampling rate increase) by an integer factor L

1.2.1 Decimation by an integer factor M

Consider a discrete-time signal $x(n)$ which has been obtained by sampling an analog signal $x(t)$. Assume that the sampling frequency is f_s and that $x(n)$ is a full-band signal, that is, its spectrum is nonzero for all frequencies in the interval $[-f_s/2\ f_s/2]$, except possibly at an isolated set of points. To convert the signal $x(n)$ to a signal $y(m)$ corresponding to a lower sampling rate $f_s^{'} = f_s/M$, one must first filter $x(n)$ with a digital low-pass filter that approximates the ideal characteristics

$$H(e^{j\omega}) = \begin{cases} 1, & |\omega| < \pi/M \\ 0, & \text{otherwise} \end{cases} \tag{1.1}$$

The sampling rate reduction is then achieved by keeping only every Mth sample of the filtered output. Let $v(n)$ be the low-pass filter's output. Then, we have

$$y(m) = v(Mm). \tag{1.2}$$

This process is illustrated in Fig. 1.1. The box in this figure containing a down arrow followed by an integer is called a *decimator* or *down sampler* and corresponds to the resampling operation described by (1.2).

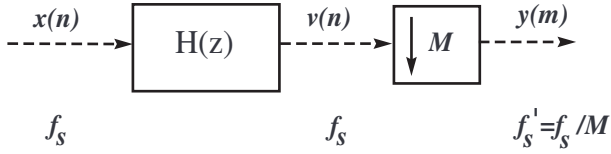


Fig. 1.1. Reducing the sampling rate of a signal by an integer factor M.

The purpose of the lowpass filter $H(z)$ is to filter $x(n)$ sufficiently so that its energy above the frequency $\omega = \pi/M$ is negligible. Thus it serves as an antialiasing filter. If the frequency response $H(e^{j\omega})$ of this filter closely approximates the ideal response of (1.1), then

$$Y(e^{j\omega}) = \frac{1}{M} X(e^{j\omega/M}), \ |\omega| < \pi. \tag{1.3}$$

1.2.2 Interpolation by an integer factor L

Consider the process of increasing the sampling rate (interpolation or upsampling) of a discrete-time signal $x(n)$ by an integer factor L. This implies that we must interpolate $L-1$ new sample values between each pair of sample values of $x(n)$. This is done by inserting $L-1$ zeros between each pair of samples of $x(n)$ resulting to the signal

$$u(m) = \begin{cases} x(m/L), & m = 0, \pm L, \pm 2L, \dots \\ 0, & \text{otherwise} \end{cases} \tag{1.4}$$

Inserting zero samples in between the samples of $x(n)$ results in unwanted spectral image components. To eliminate this artifact, it is necessary to filter the signal $u(m)$ with a digital lowpass (anti-imaging) filter that approximates the ideal characteristics

$$F(e^{j\omega}) = \begin{cases} L, & |\omega| < \pi/L \\ 0, & \text{otherwise} \end{cases} \tag{1.5}$$

Note that the filter has a gain L in the passband. This is necessary to ensure that the amplitude of the final interpolated signal $y(m)$ is correct.

The process of interpolating a digital signal by a factor of L is illustrated in Fig. 1.2. As with the resampling operation, the block diagram symbol of an up arrow with an integer corresponds to increasing the sampling rate as specified by (1.4), and is referred to as *expander* or *upsampler*.

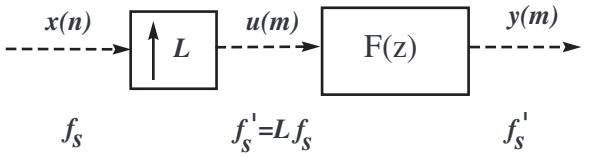


Fig. 1.2. Increasing the sampling rate of a signal by an integer factor L.

Remark 1.2. From now on, we will use the same variable name, say n, to denote the time index for the signals before and after a decimator or expander block. This is a common practice and is done to simplify notation. It is implicitly understood that the indices before and after down/up sampling pertain to different time instances.

The reader is referred to the excellent texts Crochiere and Rabiner (1983), Vaidyanathan (1993), Fliege (1994) and Mertinz (1999) for further reading on multirate signal processing theory and techniques.

1.3 Applications of multirate signal processing

The basic multirate signal processing building blocks introduced above are key to many signal compression and communication applications. In this section, we mention two classic applications where multirate signal processing techniques have proved viable. We will also introduce a new third area where we believe multirate techniques can have significant potential impact.

1.3.1 Scalable representation of multimedia signals

In many multimedia applications, it's very desirable to store or broadcast a signal in *multiple resolutions*. This is to say, it is desirable to have a signal representation where various "coarse" or "fine" approximations to the signal are readily retrieved.

In principle, a multi-resolution representation can be obtained by decomposing the original signal into M components. The components are formed such that the first component represents the signal at the lowest resolution,

the second component, when added to the first one, provides a better approximation; the third component, once added to the first two, provides an even better approximation and so on. Finally, the original signal (i.e. highest quality approximation) is retrieved when all the components are put together.

A very interesting question is whether it is possible to obtain an "optimal" multi-resolution representation where every approximation provided by the representation is as good as possible at it's respective resolution. When the answer is yes, the resulting decomposition is called *scalable* or *optimal in the sense of scalability.*

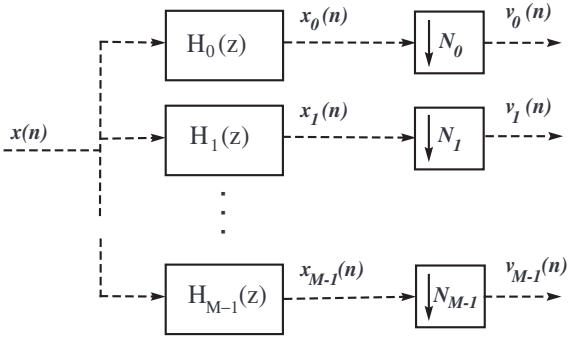


Fig. 1.3. An M-channel multirate analysis filter bank.

Scalable representation of audio-visual signals is of paramount importance in multicasting applications where a group of users, each with different demands for quality of service, are to be addressed by a single source through a communication network. In the literature, the broad concept of scalability is also described by such terms as *progressive transmission*, *multi-resolution coding*, *successively refinable coding*, *multi-scale representation* and so on. See, for example, Bull et al. (1999) and Goyal (2001).

A natural way to obtain a multi-resolution representation of a signal is to use a *multirate analysis filter bank* shown in Fig. 1.3. This system decomposes the input signal $x(n)$ into M low-rate components $v_0(n)$, $v_1(n)$, ..., $v_{M-1}(n)$ called *subband components* or *subband signals.* At the receiving end or during retrieval, some or all of the subband components are passed through a *multirate synthesis filter bank* (Fig. 1.4) to reconstruct an approximation $y(n)$ to the original signal $x(n)$.

Let $y_0(n)$ denote the output of the synthesis filet bank when only the subband component $v_0(n)$ is present. Then, define $y_1(n)$ as the approximation obtained by feeding both subband signals $v_0(n)$ and $v_1(n)$ to the synthesis filter bank. Similarly, define $y_2(n)$, $y_3(n)$, ..., $y_{M-1}(n)$ to denote the approximations obtained by including successively more subbands.

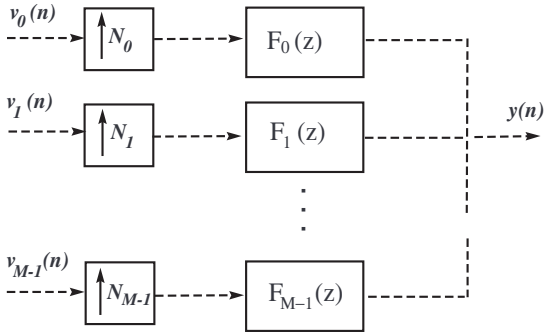


Fig. 1.4. An M-channel multirate synthesis filter bank.

To obtain a scalable representation, we have to design the analysis and synthesis filter banks such that all of the approximations $y_0(n)$, $y_1(n)$, $\cdots$, $y_{M-1}(n)$ are as close to the original measurement signal $x(n)$ as possible. This is a multi-objective optimization problem. Therefore, it is possible that the objective functions to be minimized have no common solution. This means, in general, a scalable decomposition of an arbitrary signal might not be be possible. We will give an extensive analysis of the problem of scalable decomposition of signals in Chapter 6.

1.3.2 Subband coding

Most current image/video compression standards such as JPEG, MPEG-2 and H.264 (a.k.a. MPEG-4, part 10) are based on block-based discrete cosine transform (DCT). However, a significant amount of research work has demonstrated the benefits of more elaborate *subband coding* techniques that use multirate filter banks instead of block transforms. This is reflected in the upcoming JPEG-2000 image compression standard as well as modern audio compression techniques such as MP3, ATRAC3plus[2] and AAC[3].

[2] This is a proprietary audio compression technique used in SONY's HiMD players, PSP console and ATRAC CD players. It is thought to be a hybrid subband/DCT codec where the signal is split into 16 sub-bands before MDCT and bit allocation, though not much information has been released.

[3] Advanced Audio Coding (AAC) is a lossy audio compression scheme developed with the contributions of several companies including Dolby, AT&T, Sony and Nokia. It was officially declared an international standard by the Moving Pictures Experts Group in April of 1997. As a result, it is also known as MPEG-2 Part 7 and MPEG-4 Part 3 depending on its implementation. The popularity of this format is maintained by it being the default codec on iTunes, the jukebox which powers Apple's iPod, the most popular digital audio player on the market as of 2006.

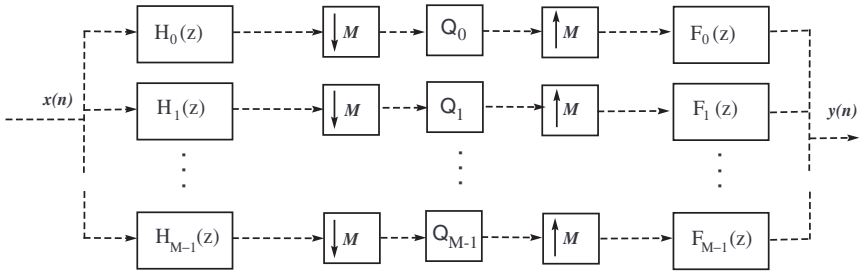


Fig. 1.5. An M-channel subband coding system.

Subband coding is a classic method for compression of raw audio/visual signals (Jayant and Noll, 1984). In this method, the input signal $x(n)$ is decomposed into M subbands which are then decimated and quantized by a set of M memoryless quantizers Q_0 to Q_{M-1} (Fig. 1.5). An approximation to the original input is then synthesized by up-sampling the quantized subbands and combining them using a set of synthesis filters. The following problem naturally arises in the context of designing an efficient subband coding system:

Problem 1.1 (Optimal Subband Coding). Assume that the input signal $x(n)$ is a random signal with known statistical properties. Assume, also that the analysis and synthesis filter banks used belong to a specific class $\mathcal{L}$. Given these information, specify

(i) a way to distribute b_{total} bits among the quantizers
(ii) a filter bank in $\mathcal{L}$

such that the expected difference between the input $x(n)$ and output $y(n)$ is minimized.

The problems of scalable signal decomposition and optimal subband coding are closely connected. We will discuss the precise mathematical connection between these two problems in Chapter 6 as well.

1.3.3 Distributed measurement and sensor networks

In recent years, a new information collection paradigm which advocates connecting a large number of inexpensive and small sensors in a *sensor network* has emerged. In defence applications, sensor networks can provide enhanced battlefield situational awareness which can revolutionize a wide variety of operations from armored assault on open terrain to urban warfare. Sensor networks have many potential applications in environmental monitoring, biomedicine, factory automation and control of transportation systems as well[4].

[4] The reader is referred to the IEEE Signal Processing Magazine, special issue on sensor networks (Vol. 23, No. 4, July 2006) for pointers to further applications.

The trend to network many sensors together has been reinforced by the widespread availability of cheap embedded processors and easily accessible wireless networks. The building blocks of a sensor network, often called "Motes", are self-contained, battery-powered computers that measure light, sound, temperature, humidity, and other environmental variables (Fig. 1.6). Motes can be deployed in large numbers providing enhanced spatio-temporal sensing coverage in ways that are either prohibitively expensive or impossible using conventional techniques.

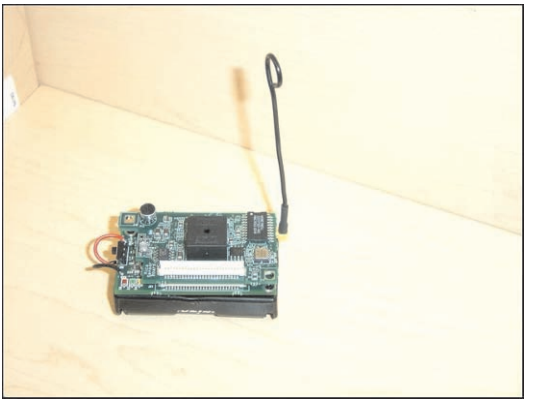


Fig. 1.6. A wireless sensor node or "Mote" made by Crossbow Technology, Inc. in San Jose, California.

In this book we focus on sensor network applications where a collection of Motes placed at various locations observe a single information source. It is assumed that this source produces some kind of information-bearing signal such as sound, speech, electromagnetic wave, etc. The goal is to measure the information-bearing signal produced by the source as accurately as possible using distorted, low-resolution and possibly noisy measurements made by the Motes (Fig. 1.7).

In principle, a distributed network of sensors can be very flexible, cost effective, and robust with respect to individual Mote's failure. However, there are many technological hurdles that must be overcome for sensor networks to become viable. For instance, Motes are inevitably constrained in processing speed, storage capacity, and communication bandwidth. Additionally, their lifetime is determined by their ability to conserve power. These constraints require new hardware designs, novel network architectures and special information processing algorithms.

The design of information processing algorithms under energy, bandwidth and other application-specific constraints results in theoretical challenges that span all layers of the network's protocol stack. However, there are some

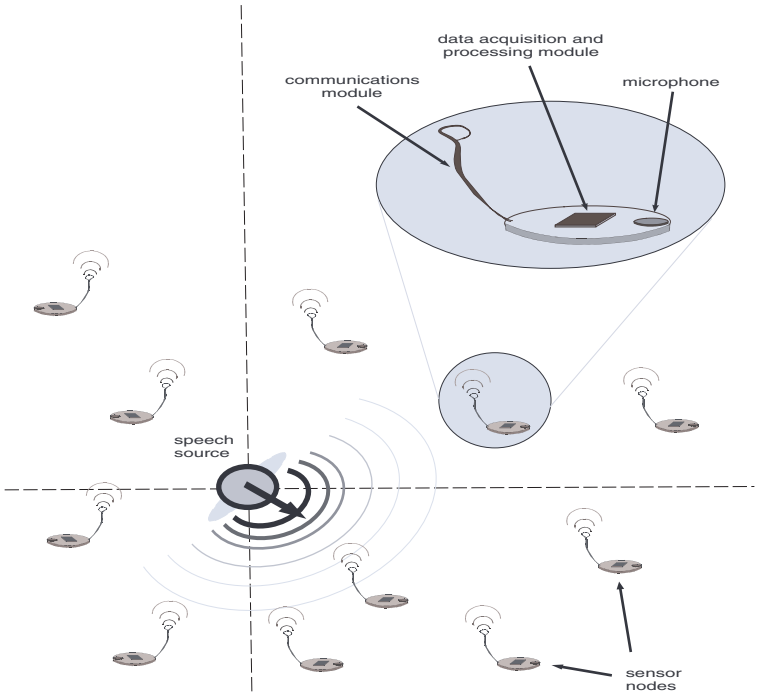


Fig. 1.7. A sensor network monitoring a speech source in a room.

fundamental challenges that clearly fall in the realm of digital signal processing. Central among these is the establishment of models, metrics and algorithms for efficient fusion of sensory information on a common mathematical framework.

In this book we advocate the use of multirate signal processing building blocks for modelling sensor network information fusion problems. To give the reader an idea how signal fusion in a sensor network can be cast as a multirate signal processing problem, we consider a simple two-sensor measurement scenario in the example below.

Example 1.1. Consider the two-sensor distributed measurement scenario shown in Fig. 1.8(a). In this figure, $x(t)$ denotes the source signal arriving at the reference sensor node. This signal is the object of measurement.

Assume that attenuation is negligible and the environment is non-dispersive. In this case, the signal received by the second sensor is $x(t-\Delta)$ where Δ represents the unknown time-delay of arrival (TDOA). Each sensor node samples and communicates its measured data at only half of the Niquist rate required to discretize $x(t)$ faithfully.

The objective is to design the sensor nodes and the reconstruction system at the receiving end such that the signal $y(n)$ reconstructed at the receiver is equivalent to the direct high-sampling-rate measurement $x(n)$ obtained as shown in Fig. 1.8(b).

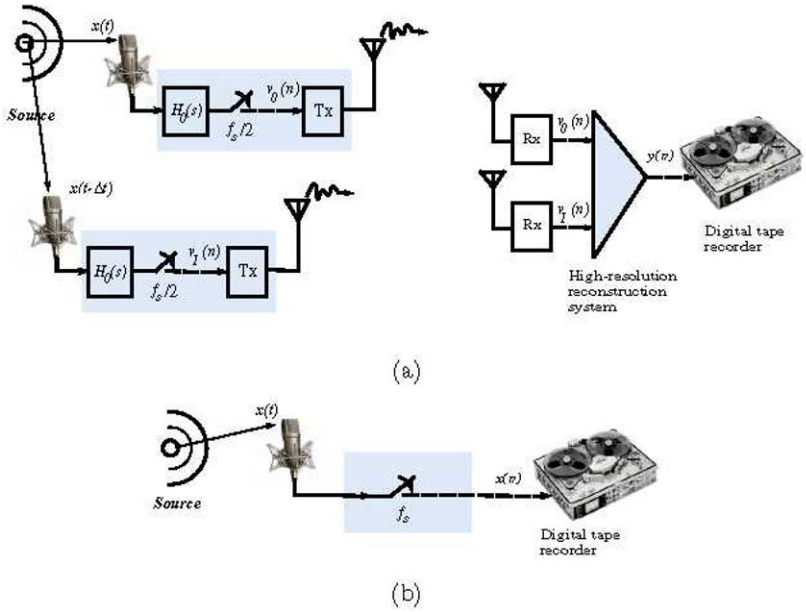


Fig. 1.8. (a) A two-node multirate sensor array system. Each sensor node samples and communicates data at only half of the Niquist rate required to discretize the signal $x(t)$ faithfully. The objective is to design the sensor nodes and the reconstruction system at the receiving end such that the signal $y(n)$ reconstructed at the receiver is a replica of the direct high-sampling-rate measurement $x(n)$ shown in (b).

In the special case that the time-delay of arrival Δ is an integer multiple of the Nyquist sampling period (i.e. if $\Delta \triangleq D/f_s$, $D \in \mathbb{Z}$) the observation model shown in Fig. 1.8(a) is easily discretized, leading to the multirate filter bank model shown in Fig. 1.9(a). In the general case that Δ is not an integer multiple of the Nyquist sampling period $1/f_s$, the transfer function z^{-D} in Fig. 1.8 has no formal meaning. However, if we assume that $x(t)$ is bandlimited to $W = f_s/2$ Hz, z^{-D} can be interpreted "symbolically" in light of the following generalized interpolation formula (Oppenheim and Schafer, 1989, Sec 3.5), (Meyr et al., 1998, Sec 4.2.2):

$$Y(z) = z^{-D}X(z) \;\Leftrightarrow\; y(n) = \sum_{k=-\infty}^{\infty} x(k)\frac{\sin(\pi(n-D-k))}{\pi(n-D-k)}.$$

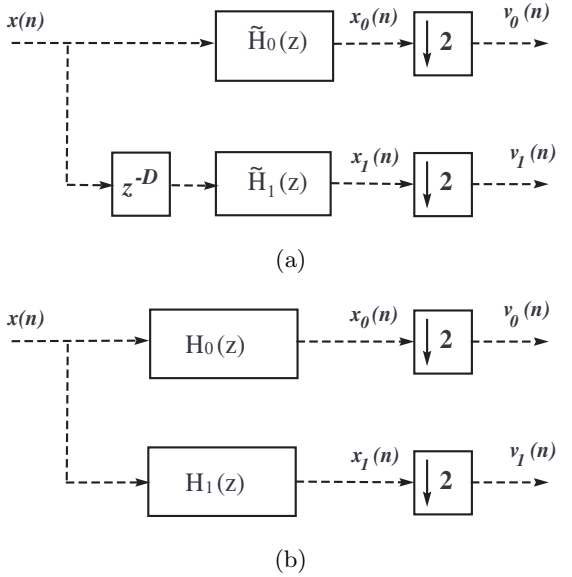


Fig. 1.9. (a) Discrete-time filter bank model for the measurement scenario shown in Fig. 1.8. (b) Simplified model where the delay block in the lower channel is integrated with the linear filter in that channel. Here, $H_0(z) = \widetilde{H}_0(z)$ and $H_1(z) = z^{-D}\widetilde{H}_1(z)$.

Therefore, when $x(t)$ is bandlimited to $W = f_s/2$ Hz, the relationship between the desired full-rate measurement $x(n)$ and the low-rate measurements $v_i(n)$ can be modelled by linear filters and decimators as shown in Fig. 1.9(b). ♢

The above example suggests that a multirate analysis filter bank such as the one shown in Fig. 1.3 can be a useful model for posing signal fusion problems in sensor networks. The filters $H_i(z)$ in the filter bank can represent several effects at the same time:

(i) the frequency response of the actual sensory device or transducer used by a Mote,
(ii) signal propagation effects such as attenuation, time delay of arrival, reverberation, frequency-selective fading and so on,
(iii) any filtering or other linear operation deliberately performed on the measured signal by Mote's signal processing hardware.

The decimator blocks in the filter bank can represent one or more of the following artifacts:

(i) the difference between the actual sampling rate of the ADC used by a Mote and the sampling rate with which we would like to monitor the information baring signal,

(ii) the difference between symbol rate by which a Mote can transmit its measured data and and the sampling rate with which we would like to monitor the information baring signal,
(iii) the difference between symbol rate by which a Mote's DSP hardware can process it's measured data and the sampling rate with which we would like to monitor the information baring signal.

In future chapters, we will address several key problems in distributed measurement by reference to the filter bank model shown in Fig. 1.3.

1.4 Multirate statistical signal processing

Statistical signal processing is a subfield of digital signal processing which deals with random processes and their manipulation using digital computers. Classic problems in statistical signal processing include power spectrum estimation, parametric signal modelling, detection and estimation of signals contaminated by noise and adaptive filtering.

In this book we discuss the use of statistical modelling as a powerful and unifying tool for processing multirate signals. We assume that a signal of interest, denoted $x(n)$, can be modelled as stationary random process. Then, we consider a measurement model where $x(n)$ is observed or measured indirectly through a multirate analysis filter bank of the type shown in Figs. 1.3 or 1.9. The goal is to make inferences about the non-observable signal $x(n)$ or other quantities of interest such as the time-delay D in Fig. 1.9 using the observable low-rate signals $v_i(n)$.

The theory of multirate statistical signal processing as presented in this book consists of several interconnected sub-theories. Each sub-theory deals with a specific problem as described below.

Multirate Spectrum Estimation (Chapter 3) : Consider the multirate system shown in Fig. 1.3. Estimate the statistical properties of $x(n)$ given statistical properties of $v_i(n)$.

Multirate Signal Estimation (Chapter 4): Consider the multirate system shown in Fig. 1.3. Assume that the statistical properties of $x(n)$ are known. Estimate some sample values of $x(n)$ given some sample values of $v_i(n)$.

Multirate Time-Delay Estimation (Chapter 5): Consider the multirate system shown in Fig. 1.9(a). Estimate the time-delay D given the observable low-rate measurements $v_i(n)$. Also, specify conditions on the filters $H_i(z)$ such that the time-delay D can be reliably estimated from the low-rate measurements $v_i(n)$.

Optimal Multirate Decomposition of Signals (Chapter 6): Consider the multirate system shown in Fig. 1.3. Assume that the statistical properties of $x(n)$ are known. Characterize a filter bank within a given class of filter banks such that it decomposes $x(n)$ in a way which is more scalable than all other filter banks in that class. Also, discuss the existence and uniqueness of such a filter bank.

Information in Multirate Systems (Chapter 7): Consider the multirate system shown in Fig. 1.3. Assume that the statistical properties of $x(n)$ are unknown. Quantify the amount of information gained about statistical properties of $x(n)$ if we specify statistical properties of one or more low-rate signal $v_i(n)$.

The statistical approach to multirate signal processing presented in this book has several advantages over classical (deterministic) formulations developed in previous works. Some of these advantages are as follows:

(i) The statistical theory is applicable to a very broad class of analysis system models regardless of the sampling rates used or the types of filters used. In contrast, previous material on multirate systems and filter banks covered only FIR, Perfect Reconstruction and/or orthogonal filter banks and often required uniform sampling rate in all the channels.
(ii) Statistical algorithms can estimate an arbitrarily large number of samples of the original signal and provide a bound on the expected error for each and every one of the samples estimated.
(iii) When it is possible to uniquely specify some samples of the original signal, our estimator will do that too! This is because the statistical idea of *estimation* used in this book is much more powerful than the deterministic idea of *reconstruction* used in classical filter bank literature.
(iv) The concept of "aliasing" which is a key ingredient of the classical theory of multirate systems is almost nonexistent in the statistical theory!

Remark 1.3. Since mid 1980's, there has been considerable interest and activity in signal processing, applied mathematics and statistical research communities in developing multi-resolution data models and algorithms. Basseville, Benveniste, Willsky and their coworkers have put considerable effort into the development of a theory of multi-resolution stochastic modeling and associated techniques for optimal "multi-scale statistical signal analysis" (Basseville et al., 1992a), (Basseville et al., 1992b), (Basseville et al., 1992c), (Chou et al., 1994a), (Chou et al., 1994b). These authors consider processes indexed by nodes on binary (or n-ary) lattices (or trees) in which different depths in the lattice or tree correspond to different spatial scales in representing a signal or an image. It is important to note that this paradigm is conceptually different from our theory which uses the classical wide-sense stationary model for random processes.

1.5 Notation

We use capital letters or boldface lowercase letters to represent vectors. Boldface capital letters are reserved for matrices. Elements of a matrix $\mathbf{A}$ are referred to as $[\mathbf{A}]_{ij}$. We denote the set of real M-tuples by $\mathbb{R}^M$ and use the notation $\mathbb{R}_+$ for positive real numbers. The expected value of a random variable x is denoted by $E\{x\}$. The end of an example is indicated using the symbol $\diamondsuit$. The linear convolution operator is denoted by $\star$.

The spaces of Lebesgue-measurable[5] functions are represented by $L^1(a,b)$, $L^2(a,b)$, etc. For $f(x) \in L^p(a,b), p < \infty$, the L^p norm is defined by

$$\|f\|_p \triangleq \frac{1}{b-a} \sqrt[p]{\int_a^b |f(x)|^p dx}.$$

The L^∞ norm is defined by

$$\|f\|_\infty \triangleq \sup_{x \in (a,b)} |f(x)|.$$

[5] See Jones (1993) for an excellent introduction to the theory of Lebesgue measures.

2

Background

This chapter is intended to provide some background material which is fundamental in construction of the theory of statistical multirate signal processing. It also serves as a means to introduce certain notation and conventions which will be needed in later chapters. It is assumed that the reader is already familiar with elementary notions of probability theory, linear algebra and digital signal processing.

2.1 Inverse and ill-posed problems

When using the term *inverse problem* one would naturally ask "inverse to what?" In mathematical physics, two problems are called inverse to each other if the formulation of one problem involves the other one. For mostly historic reasons, one might call one of these problems (usually the simpler one or the one which was studied earlier) the *direct problem*, the other one the *inverse problem*. However, if there is a real-world problem behind the mathematical problem studied, there is, in most cases, a quite natural distinction between the direct and the inverse problem. For example, if one wants to predict the future behaviour of a physical system from knowledge of its present state and the physical laws, one will call this the direct problem. Possible inverse problems are the determination of the present state of the system from future observations (i.e., the calculation of the evolution of the system backwards in time) or the identification of physical parameters from observations of the evolution of the system. Thus, one might say *the inverse problems are concerned with determining causes for a desired or an observed effect.*

Most often, inverse problems are much more difficult to deal with (from a mathematical point of view) than their direct counterparts. This is because they might not have a solution in the strict sense or solutions might not be unique and/or might not depend continuously on data. Mathematical problems having such undesirable properties are called *ill-posed problems* and

cause (mostly because of the discontinuous dependence of solutions on the data) severe numerical difficulties.

The study of inverse problems has been one of the fastest-growing areas in applied mathematics in the last two decades. This growth has largely been driven by the needs of applications in both natural sciences (e.g. inverse scattering theory, astronomical image restoration and statistical learning theory) and industry (e.g. computerized tomography). See e.g. Tikhonov and Arsenin (1977), Vasin and Ageev (1995), Engl et al. (1996) and Tikhonov et al. (1998).

The study of concrete inverse problems often involves the question "how can one enforce uniqueness by additional information or assumptions?" Not much can be said about this in a general context. However, the aspect of lack of stability and its restoration by appropriate methods (*regularization*) can be treated in sufficient generality. The theory of regularization is well-developed for linear inverse problems and will be introduced, very briefly, in the next subsections.

2.1.1 Ill-posed linear operator equations

Formally, a mathematical problem is called *well-posed* if it fulfills Hadamard's conditions:

(i) For all admissible data, a solution exists.
(ii) For all admissible data, the solution is unique.
(iii) The solution depends continuously on the data.

A problem for which one or more of the above conditions are violated is called *ill-posed*. Note that the conditions mentioned above do not make a precise definition for well-posedness. To make a precise definition in a concrete situation, one has to specify the notion of a solution, which data are considered admissible, and which topology is used for measuring continuity.

Example 2.1. Let the linear operator equation

$$Ax = y \tag{2.1}$$

be defined by the continuous operator A that maps the elements x of a metric space $\mathcal{E}_1$ into elements y of another metric space $\mathcal{E}_2$. Consider that A is a Fredholm integral operator of the first kind so that

$$(Ax)(s) \triangleq \int_a^b K(s,t)x(t)\mathrm{d}t. \tag{2.2}$$

The kernel $K(s,t)$ is continuous on $[a\ b] \times [a\ b]$ and maps a function $x(t)$ continuous on $[a\ b]$ to a function $y(s)$ also continuous on $[a\ b]$.

It is well known that the Fourier series coefficients of a continuous function tend to zero at high frequencies (Jones, 1993, Chapter 14, Section I). This implies that the continuous function

$$g_\omega(s) \stackrel{\triangle}{=} \int_a^b K(s,t)\sin(\omega t)\mathrm{d}t, \tag{2.3}$$

which is formed by means of the continuous kernel $K(s,t)$ satisfies

$$\lim_{\omega\to\infty} g_\omega(s) = 0 \tag{2.4}$$

for all $s \in [a,b]$. Now, consider the perturbed equation

$$Ax = y + g_\omega, \tag{2.5}$$

where y is given and g_ω is defined in (2.3). Since the above equation is linear its solution $\hat{x}(t)$ has the form

$$\hat{x}(t) = x^*(t) + \sin(\omega t), \tag{2.6}$$

where $x^*(t)$ is a solution to the original integral equation $Ax = y$. For sufficiently large ω, the right hand side of (2.5) differs from the right hand side of (2.1) only by the "small" amount $g_\omega(s)$. However, its solution differs from that of (2.1) by the considerable amount $\sin(\omega t)$. This shows that the problem of solving (2.1) where A is a Fredholm integral operator of the first kind is ill-posed. ♢

In the early 1900s, Hadamard had already observed that under some (very general) circumstances the problem of solving the operator equation (2.1) is ill-posed. However, he believed that ill-posed problems are a pure mathematical phenomenon and that all real-life problems are "well-posed". In the second half of the 20th century, a number of very important real-life problems were found to be ill-posed.

The discovery of various *regularization methods* in the 60's made it possible to construct a sequence of well-posed solutions that converges to the desired one. Regularization theory was one of the first signs of existence of intelligent inference methods. It demonstrated that whereas the "self-evident" methods of solving an operator equation might not work, the "non-self-evident" methods of regularization theory do. See Vapnik (1999) for the influence of the regularization philosophy on statistical inference.

2.1.2 Regularization of ill-posed operator equations

One can easily verify that the problem of solving the operator equation (2.1) is equivalent to finding an element $x^* \in \mathcal{E}_1$ such that the functional

$$R(x) \stackrel{\triangle}{=} \|Ax - y\|_{\mathcal{E}_2} \tag{2.7}$$

is minimized[1]. If the right-hand side of (2.1) is not exact, that is, if we replace y by y_δ such that $\|y - y_\delta\|_{\mathcal{E}_2} < \delta$ where δ is a small value, a function $x_\delta \in \mathcal{E}_1$

[1] To save in notation, we write $\|a - b\|_{\mathcal{E}}$ to denote the *distance* between the two elements $a, b \in \mathcal{E}$ whether the metric space $\mathcal{E}$ is a normed space or not. If $\mathcal{E}$ is a normed space too, our notation is self-evident. Otherwise, it should be interpreted only as a *symbol* for the distance between a and b.

shall minimize the functional

$$R_\delta(x) \stackrel{\triangle}{=} \|Ax - y_\delta\|_{\mathcal{E}_2}. \tag{2.8}$$

However, if the operator equation $Ax = y$ is ill-posed, the new solution x_δ is not necessarily close to the desired solution x^* even if δ tends to zero. In other words, $\lim_{\delta \to 0} \|x^* - x_\delta\|_{\mathcal{E}_1} \neq 0$.

It was discovered by Tikhonov (1963) that, if instead of the functional $R_\delta(x)$ one minimizes

$$R_{reg}(x) \stackrel{\triangle}{=} \|Ax - y_\delta\|_{\mathcal{E}_2} + \xi(\delta) S(x), \tag{2.9}$$

where $S(x)$ is a *stabilizing functional* (that belongs to a certain class of functionals) and $\xi(\delta)$ is an appropriately chosen constant (whose value depends on the *noise* level δ), then one obtains a sequence of solutions x_δ that converges to the desired one as δ tends to zero. For the above result to be valid, it is required that

(i) the problem of minimizing $R_{reg}(x)$ is well-posed for fixed values of δ and $\xi(\delta)$, and
(ii) $\lim\limits_{\delta \to 0} \|x^* - x_\delta\|_{\mathcal{E}_1} \to 0$ when $\xi(\delta)$ is chosen appropriately.

Consider a real-valued lower semi-continuous[2] functional $S(x)$. The functional $S(x)$ can be a stabilizing functional if it possesses the following properties:

(i) The solution of the operator equation $Ax = y$ belongs to the domain of definition $\mathcal{D}(S)$ of the functional S.
(ii) $S(x) \geq 0, \quad \forall x \in \mathcal{D}(S)$.
(iii) The level sets $\{x : S(x) \leq c\}$, $c = const.$, are compact.

The above conditions essentially ensure that the problem of minimizing $R_{reg}(x)$ is well-posed (Tikhonov and Arsenin, 1977, Page 51). Now, the important remaining problem is to determine the functional relationship between δ and $\xi(\delta)$ such that the sequence of solutions obtained by minimizing (2.9) converges to the solution of (2.7) as δ tends to zero. The following theorem establishes sufficient conditions on such a relationship:

Theorem 2.1. *(Vapnik, 1999, Page 55) Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two metric spaces and let $A : \mathcal{E}_1 \to \mathcal{E}_2$ be a continuous and one-to-one operator. Suppose that for $y \in \mathcal{E}_2$ there exists a solution $x \in \mathcal{D}(S) \subset \mathcal{E}_1$ to the operator equation $Ax = y$. Let y_δ be an element in $\mathcal{E}_2$ such that $\|y - y_\delta\|_{\mathcal{E}_2} \leq \delta$. If the parameter $\xi(\delta)$ is chosen such that*

[2] A function $f : \mathbb{R}^N \to [-\infty, \infty]$ is called lower semi-continuous at $X \in \mathbb{R}^N$ if for any $t < f(X)$ there exists $\delta > 0$ such that for all $y \in \mathcal{B}(X, \delta)$, $t < \delta$. The notation $\mathcal{B}(X, \delta)$ represents a ball with centre at X and radius δ. This definition generalizes to functional spaces by using the appropriate metric in defining $\mathcal{B}(X, \delta)$.

(i) $\xi(\delta) \to 0$ *when* $\delta \to 0$,
(ii) $\lim_{\delta \to 0} \dfrac{\delta^2}{\xi(\delta)} < \infty$,

Then the elements $x_\delta \in \mathcal{D}(S)$ *minimizing the functional*

$$R_{reg}(x) = \|Ax - y_\delta\|_{\mathcal{E}_2} + \xi(\delta)S(x)$$

converge to the exact solution x *as* $\delta \to 0$.

If $\mathcal{E}_1$ is a Hilbert space, the stabilizing functional $S(x)$ may simply be chosen as $\|x\|^2$, which, indeed, is the original choice made by Tikhonov. In this case, the level sets of $S(x)$ will only be weakly compact. However, the convergence of the regularized solutions will be a strong one thanks to favorable properties of Hilbert spaces. The conditions imposed on the parameter $\xi(\delta)$ are, nevertheless, more stringent than those stated in the above theorem. (In this case, $\xi(\delta)$ should converge to zero *strictly slower* than δ^2. In more precise terms, $\lim_{\delta \to 0} \frac{\delta^2}{\xi(\delta)} = 0$ must hold.)

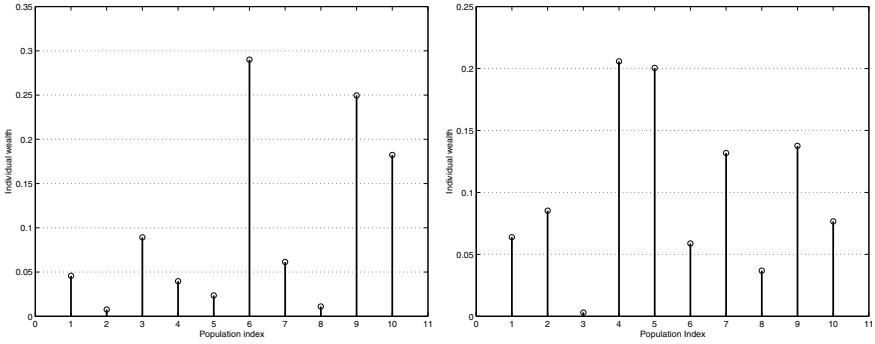
2.2 Measuring inequality

The book *History of the Peloponnesian War* by the ancient Greek historian Thucydides is widely considered the first work of scientific history, describing the human world as produced by men acting from ordinary motives, without the intervention of the gods.

In the Melian dialogue, Thucydides recounts the historical background of the invasion of the island of Melos by Athens in 416 BC. During the dialogue, the Athenians present Melos with an alternative: the island can pay tribute to Athens and thus survive, or fight Athens and be destroyed. Melos presents several counter-arguments: showing mercy towards Melos will win the Athenians more friends; the Spartans will come to their aid; the gods will protect them. The Athenians, however, refuse to discuss either the justice of their demand or any substantive argument by the Melians. Instead, the long-remembered Athenian commentary is one of hard realism:

> "We shall not trouble you with specious pretences — either of how we have a right to our empire because we overthrew the Mede, or are now attacking you because of wrong that you have done us— and make a long speech which would not be believed; and in return we hope that you ... will aim at what is feasible, holding in view the real sentiments of us both; since you know as well as we do that *right, as the world goes, is only in question between equals in power*, while *the strong do what they can and the weak suffer what they must.*"

The origin of inequality among mankind, and whether it is authorized by natural law is open to broad debate. However, as the above quote painfully



(a) (b)

Fig. 2.1. Wealth distribution data for two hypothetical communities.

reveals, it has always been a most influential factor in human affairs; leading to profound political and moral consequences.

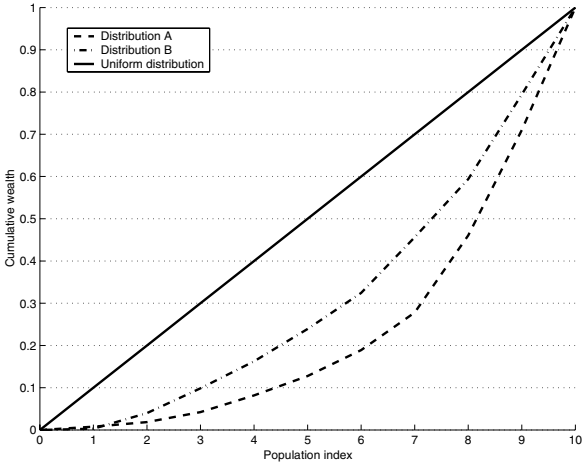
Over the years, scientists have tried to develop numerical indices to quantify various forms of inequality among human societies. In 1905, economist Max Lorenz suggested the idea of using *partial sums* to compare income distribution between two populations. He plotted cumulative partial sums of each distribution in a curve which is now called Lorenz curve. This curve is essentially a graph showing the proportion of the distribution assumed by the bottom x% of the values. It is often used to represent income distribution, where it shows for the bottom x% of households, what percentage y% of the total income they have.

Example 2.2. Consider two hypothetical communities each comprised of ten households. The annual household income for these two communities are depicted in Fig. 2.1. The income figures in each plot are normalized so that the total income for each community adds up to one. The question is to find out which community enjoys a more uniform or fairer income distribution?

To answer this question by using the method devised by Lorenz, we first sort the income figures reported in each plot from the lowest to the highest. Then, for each distribution, we add the sorted values together to form a series of successively increasing partial sums. These partial sums are plotted as two curves in Fig. 2.2. This figure also shows a straight line which represents a perfectly uniform distribution.

From Fig. 2.2 we can argue that distribution B whose curve is closer to the straight line is more uniform than distribution A. In other words, the households in community B enjoy a more even income distribution compared to community A. ♢

Lorenz's work in economics led to the abstract mathematical notion of *majorization* to quantify non-uniformity or inequality among the components of a vector or a set. In simple terms, a set of numbers majorizes another if the



(c)

Fig. 2.2. Lorenz curves associated with the wealth distributions shown in Fig. 2.1. The straight line here represents the uniform distribution.

Lorenz curve associated with the first set lies *below* the curve associated with the second. We will use the concept of majorization extensively in Chapter 6 where we develop a theory of optimal multirate signal decomposition. A brief introduction to the theory of majorization is given in the next subsections. The reader is referred to the excellent monograph by Marshall and Olkin (1979) for a detailed treatment of the history and the theory of majorization.

2.2.1 Definition of majorization

Let $\xi = \{a_0, \ldots, a_{M-1}\}$ be a set of M non-negative real numbers; thus, order is irrelevant. Let Ω denote the set of all such sets. One can always permute the indices such that elements of sets are indexed in a descending order. When such a permutation has been introduced, we use indices in brackets. Using this convention, for instance, ξ is equivalently represented as $\{a_{[0]}, a_{[1]}, \ldots, a_{[M-1]}\}$ where $a_{[0]} \geq a_{[1]} \geq a_{[2]} \geq \ldots \geq a_{[M-1]}$.

Definition 2.1. *The set* $\xi = \{a_0, \ldots, a_{M-1}\}$ *is said to be majorized by* $\eta = \{b_0, \ldots, b_{M-1}\}$*, in symbols* $\xi \preceq \eta$*, if*

(i) $\sum_{i=0}^{M-1} a_i = \sum_{i=0}^{M-1} b_i$ *and*
(ii) $\sum_{i=0}^{k} a_{[i]} \leq \sum_{i=0}^{k} b_{[i]}$ *for all* $0 \leq k \leq M-2$.

Example 2.3. Let $\xi = \{a_0, \ldots, a_{M-1}\}$ be any arbitrary set in Ω such that $\sum_{i=0}^{M-1} a_i = 1$. Then, $\{\frac{1}{M}, \ldots, \frac{1}{M}\} \preceq \xi \preceq \{1, 0, \ldots, 0\}$. $\diamondsuit$

The above example suggests that majorization can be used to quantify "inequality" or "nonuniformity" among the elements of a set. In fact, $\xi \preceq \eta$

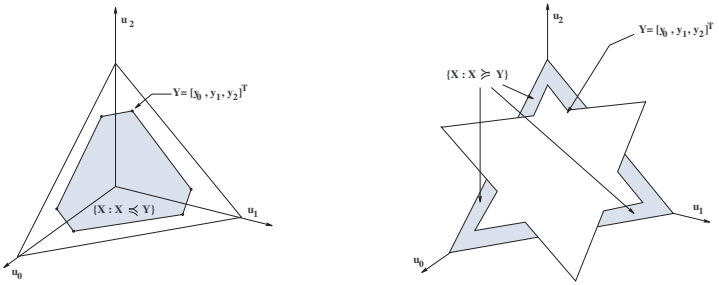


Fig. 2.3. Geometric representation of majorization. (a) The convex set $\{X \preceq Y\}$. (b) The non-convex set $\{Y \preceq X\}$.

means the elements of ξ are "more uniform" or "more equal" than those of η. We extend the definition of majorization to a vector in $\mathbb{R}_+^M$ simply by treating its components as elements of a set. In other words, for two vectors X and Y in $\mathbb{R}_+^M$ we write $X \preceq Y$ if $\{x_i\} \preceq \{y_i\}$.

2.2.2 Geometrical properties of majorization

Let Y be a point in $\mathbb{R}_+^M$. In general, we can form up to $M!$ different M-vectors by using the elements of $\{y_i\}$ as components. These vectors (points) simply constitute the orbit of Y under the group of permutation matrices. Connected together, these points form a plain polytope in $\mathbb{R}_+^M$. (See Fig. 2.3(a).)

Theorem 2.2. *A vector X is majorized by $Y \in \mathbb{R}_+^M$ if and only if X lies in the convex hull of the orbit of Y under the group of permutation matrices.*

An immediate result of the above theorem is that for a fixed Y the set $\{X : X \preceq Y\}$ is convex. It is, however, not true for $\{X : Y \preceq X\}$ as illustrated in the following example.

Example 2.4. Consider a point $Y \in \mathbb{R}_+^3$ as our reference point. Let $c = \sum_1^3 y_i$. Define $\Delta = \{X \in \mathbb{R}_+^3 : x_1 + x_2 + x_3 = c\}$. Obviously, Δ is a plane triangle that also includes Y. The sets $\{X : X \in \Delta, X \preceq Y\}$ and $\{X : X \in \Delta, Y \preceq X\}$ are denoted by gray color in Fig. 2.2. As seen in the figure, the first set is convex while the latter is not. Note also that the triangle Δ contains points that belong to neither set. These are the points which are not majorized by Y nor do they majorize it. This observation has important algebraic interpretations as discussed in the next subsection. ♢

2.2.3 Algebraic properties of majorization

In this subsection we review properties of the binary relation $\preceq$ as an ordering relation on sets. Our treatment is very brief and intended, mainly, to establish

terminology. See Rosenfeld (1968) or Mac Lane and Birkhoff (1967) for a well-written introduction to the theory of ordered sets.

Let ξ, η and ν be any three elements of Ω. You can yourself verify that the following statements hold:

$$\xi \preceq \xi \qquad \text{(Reflexivity)} \tag{2.10}$$

$$\xi \preceq \eta \ , \ \eta \preceq \xi \Longrightarrow \xi = \eta \qquad \text{(Antisymmetry)} \tag{2.11}$$

$$\xi \preceq \eta \ , \ \eta \preceq \nu \Longrightarrow \xi \preceq \nu \qquad \text{(Transitivity)} \tag{2.12}$$

The three properties above show that the binary relation $\preceq$ induces a *partial ordering* on Ω. A set endowed with a partial ordering is called a *partially ordered set*, or simply a *poset*. A poset Ω is called a *chain* or a *simply ordered* set when its elements satisfy the property that given any ξ and η in Ω, either $\xi \preceq \eta$ or $\eta \preceq \xi$.

Example 2.5. Let $\Omega = \{\xi : \xi = \{a_0, a_1, a_2\}, a_i \in \mathbb{R}_+, \sum_{i=0}^{2} a_i = 1\}$. One can easily find two elements in Ω (for instance $\xi_1 = \{.5, .4, .1\}$ and $\xi_2 = \{.6, .2, .2\}$) such that neither $\xi_1 \preceq \xi_2$ nor $\xi_2 \preceq \xi_1$. This shows that, in general, the ordering of majorization is not a simple ordering. $\diamondsuit$

For a poset, one can identify two types of extremal elements. Below, we define these elements following the terminology of Mac Lane and Birkhoff (1967).

Definition 2.2. *Let Ω be a poset supplied with a partial order $\preceq$. An element $\omega^* \in \Omega$ is called the "greatest element" if $\omega \preceq \omega^*$ for all $\omega \in \Omega$. An element $\hat{\omega} \in \Omega$ is called a "maximal element" if no $\omega \in \Omega$, $\omega \neq \hat{\omega}$, exists such that $\hat{\omega} \preceq \omega$.*

In words, the greatest element in a poset majorizes all other elements whereas a maximal element is not majorized by any other element.

Theorem 2.3. *The greatest element is unique. If it exists, it will be the unique maximal element as well. When the greatest element does not exist, a poset can have more than one maximal element.*

Definition 2.3. *Let $\preceq$ be an order relation on a set Δ and let Ω be a subset of Δ. We say that $\xi \in \Delta$ is an "upper bound" for Ω if $\omega \preceq \xi$ for all $\omega \in \Omega$.*

Note that if $\omega \in \Omega$ is an upper bound for Ω, it is the greatest element of Ω.

Closely related to the notion of partial ordering is the concept of order-preserving functions. A class of such functions related to partial ordering of majorization will be introduced in the next subsection.

2.2.4 Schur-convex functions

The first comprehensive study of functions preserving the order of majorization was made by I. Shur in 1923 (Brauer and Rohrbach, 1973). Such functions, called Schur-convex, are defined as follows.

Definition 2.4. *Let $\preceq$ denote the majorization ordering defined on some subset $\Omega \subset \mathbb{R}_+^M$. A function $\phi : \mathbb{R}_+^M \to \mathbb{R}$ is said to be Schur-convex if*

$$X \preceq Y \quad \Longrightarrow \quad \phi(X) \leq \phi(Y), \quad \forall X, Y \in \Omega.$$

If, in addition, $\phi(X) < \phi(Y)$ whenever $X \preceq Y$ but $Y \not\preceq X$, ϕ is called strictly Schur-convex.

Schur-concave and strictly Schur-concave functions are defined in an analogous manner. In words, $\phi(X)$ is Schur-convex if the functional value becomes smaller when the components of X are less diverse in the sense of majorization. Note that $\phi(X) \leq \phi(Y)$ does not necessarily imply $X \preceq Y$.

A Schur-convex function ϕ might operate on a vector X or a set $\xi = \{x_i\}$. This is simply because the components of vectors or elements of sets are reordered decreasingly in the definition of majorization. In fact, because of this very property, any Schur-convex function must necessarily be symmetric in its arguments. The following basic theorem, due to Schur, provides necessary and sufficient conditions for a function to be Schur-convex:

Theorem 2.4. *(Marshall and Olkin, 1979, Page 13) Let $\phi : \mathbb{R}_+^M \to \mathbb{R}$ have continuous partial derivatives $\phi_{(i)}(X) = \frac{\partial \phi(X)}{\partial x_i}$. Then, ϕ is Schur-convex if and only if*

(i) ϕ is permutation symmetric, that is, $\phi(X) = \phi(\Pi X)$ for all $X \in \mathbb{R}_+^M$ and all permutation matrices Π, and
(ii) $(x_1 - x_2)(\phi_{(1)}(X) - \phi_{(2)}(X)) \geq 0$ for all $X \in \mathbb{R}_+^M$.

Conditions for a Schur-concave function are the same except the $\geq$ sign in (*ii*) should be replaced by $\leq$. The following theorem, also due to Schur, provides sufficient conditions for a function to be strictly Schur-convex:

Theorem 2.5. *(Marshall and Olkin, 1979, pp. 56-58) Let $\phi : \mathbb{R}^M \to \mathbb{R}$ be twice differentiable with continuous partial derivatives $\phi_{(i)}(X) = \frac{\partial \phi(X)}{\partial x_i}$ and $\phi_{(i,j)}(X) = \frac{\partial^2 \phi(X)}{\partial x_i \partial x_j}$. Then, ϕ is strictly Schur-convex if*

(i) ϕ is permutation symmetric,
(ii) $(x_1 - x_2)(\phi_{(1)}(X) - \phi_{(2)}(X)) \geq 0$,
(iii) $\phi_{(1)}(X) = \phi_{(2)}(X) \implies \phi_{(1,1)}(X) + \phi_{(2,2)}(X) > \phi_{(1,2)}(X) + \phi_{(2,1)}(X)$.

There are a number of useful facts relating to compositions that involve Schur-convex or Schur-concave functions. Here, we mention a few such results. The reader is referred to the classic monographs by Hardy et al. (1934) and Marshall and Olkin (1979) for an exhaustive account of this topic.

Theorem 2.6. *If $\phi_i(X)$ is (strictly) Schur-convex, $i = 1, \ldots, k$, and $\phi_i(X) \geq 0$ for all i and X, then*

$$\psi(x) = \prod_{i=1}^{k} \phi_i(X)$$

is (strictly) Schur-convex.

Theorem 2.7. *If $I \subset \mathbb{R}$ is an interval and $g : I \to \mathbb{R}$ is (strictly) convex, then*

$$\phi(X) = \sum_{i=0}^{N-1} g(x_i)$$

is (strictly) Schur-convex on I^N.

Theorem 2.8. *If $\phi : \mathbb{R}^M \to \mathbb{R}$ is symmetric and strictly convex on sets of the form $\{X | \sum x_i = c\}$, then ϕ is strictly Schur-convex.*

Theorem 2.9. *Let $\phi : \mathbb{R}^N \to \mathbb{R}$ satisfy the following conditions:*

(i) $\phi(X) > 0$ when $X \neq 0$,
(ii) $\phi(\gamma X) = |\gamma|\phi(X)$ for all real γ,
(iii) $\phi(X + Y) \leq \phi(X) + \phi(Y)$,
(iv) $\phi(x_0, x_1, \ldots, x_{N-1}) = \phi(\epsilon_0 x_{i_0}, \epsilon_1 x_{i_1}, \ldots, \epsilon_{N-1} x_{i_{N-1}})$ where $(i_0, i_1, \ldots, i_{N-1})$ is a permutation of $(0, 1, \ldots, N-1)$ and $\epsilon_i = \pm 1$.

Then, ϕ is Schur-convex.

A function satisfying the conditions of the last theorem above is called a *symmetric gauge function*. As special cases, it follows that the following functions are Schur-convex:

$$\phi(X) = max|x_i| = \|X\|_\infty. \tag{2.13}$$

$$\phi(X) = (\sum_{i=0}^{N-1} |x_i|^p)^{1/p} = \|X\|_p, \ p > 1. \tag{2.14}$$

Finally, note that the regular L^p norms $\|X\|_p$, $p > 1$, are also strictly Schur convex but the infinity norm $\|X\|_\infty$ is not.

2.3 Measuring information

In communication theory, the term "information" refers to a measure of the freedom of choice with which a message is selected from the set of all possible messages. As such, information is a deterministic quantity and is measured in bits.

Entropy is a statistical concept which has its origins in thermodynamics. It quantifies how random a probabilistic event is. In his landmark paper,

Shannon (1948) defined a formula for entropy that, when applied to a probabilistic information source (i.e. discrete random variable) could determine the minimum number of bits required to reliably transmit the source's output. Shannon's work showed that the concepts of information and entropy have deep links with one another.

Kullback-Leibler divergence is a quantity that measures how separable or *distant* two probability distributions are from the point of view of a statistical observer. More precisely, it measures how difficult it is to distinguish between two candidate probability distributions given a series of sample data.

Entropy and Kullback-Leibler divergence are fundamental measures that quantify *statistical information*. In the following subsections, we introduce these two measures and discuss their interconnection.

2.3.1 Entropy

Let $X \in \mathbb{R}^N$ be a continuous random variable with PDF $p_X(X)$. The entropy[3] of X is denoted by $H(X)$ and defined by:

$$H(X) \triangleq -\int p_X(X) \ln p_X(X) \mathrm{d}X. \tag{2.15}$$

Note that $H(\cdot)$ is a functional of p_X so its OK if we use the notation $H(p_X)$ for entropy of X as well. It is easy to check that entropy is a convex functional of p_X and ranges from $-\infty$ to ∞.

Let X and Y be two random vectors that possess a joint probability density function $p_{XY}(X,Y)$, and marginal densities $p_X(X)$ and $p_Y(Y)$. We define the *mutual information* or *redundancy* $I(X;Y)$ between X and Y as

$$I(X;Y) \triangleq \int\int p_{XY}(X,Y) \ \ln \frac{p_{XY}(X,Y)}{p_X(X)p_Y(Y)} \mathrm{d}X\mathrm{d}Y. \tag{2.16}$$

Example 2.6. Let $X \in \mathbb{R}^N$ and $Y \in \mathbb{R}^M$ be two random variables with a jointly Gaussian probability density function. If $\mathbf{C}_{XY}$ represents the joint covariance matrix of the two random variables and $\mathbf{C}_X$ and $\mathbf{C}_Y$ represent the marginal covariance matrices associated with each variable, then (Kullback, 1954, Chapter 9)

$$I(X;Y) = \frac{1}{2} \ln \frac{|\mathbf{C}_X||\mathbf{C}_Y|}{|\mathbf{C}_{XY}|}. \tag{2.17}$$

$\diamondsuit$

The *joint information* in X and Y is denoted by $H(X,Y)$ and is defined as

$$H(X,Y) \triangleq -\int\int p_{XY}(X,Y) \ \ln p_{XY}(X,Y) \mathrm{d}X\mathrm{d}Y. \tag{2.18}$$

[3] For simplicity, in this book we use the term entropy for what is more precisely called "differential entropy" in the information theory literature.

Note that $H(X,Y)$ is the same as $H(p_{XY})$. Furthermore, it holds that

$$H(X,Y) = H(X) + H(Y) - I(X;Y). \tag{2.19}$$

Extension of the concepts of mutual information and joint information to more then two random variables is straightforward. Detailed discussions on entropy, mutual information and joint information can be found in any standard text on information theory, e.g. (Reza, 1961, Chapter 8) or (Csiszár and Körner, 1981, Chapter 1). For a more detailed exposition, including discussions on existence and interpretations, the reader is referred to Gray (1990) and MacKay (2003).

2.3.2 Kullback-Leibler divergence

Let $p_1(\cdot)$ and $p_2(\cdot)$ be two PDFs defined on $\mathcal{C} \subset \mathbb{R}^N$. If $\mathcal{H}_i$, $i = 1, 2$, is the hypothesis that a certain random variable is from the statistical population with PDF p_i, then the standard likelihood test would accept $\mathcal{H}_1$ if $Pr\{\mathcal{H}_1\} > Pr\{\mathcal{H}_2\}$. In other words, $\mathcal{H}_1$ is accepted if

$$\ln \frac{Pr\{\mathcal{H}_1\}}{Pr\{\mathcal{H}_2\}} > 0,$$

otherwise the hypothesis $\mathcal{H}_2$ would be accepted. Now, assume that we actually observe a sample value of that random variable and find out that its value is equal to X. It follows from Bayes' theorem, or the theorems on conditional probability, that

$$Pr\{\mathcal{H}_i|X\} = \frac{Pr\{\mathcal{H}_i\}p_i(X)}{Pr\{\mathcal{H}_1\}p_1(X) + Pr\{\mathcal{H}_2\}p_2(X)}, \quad i = 1, 2, \tag{2.20}$$

from which we obtain

$$\ln \frac{p_1(X)}{p_2(X)} = \ln \frac{Pr\{\mathcal{H}_1|X\}}{Pr\{\mathcal{H}_2|X\}} - \ln \frac{Pr\{\mathcal{H}_1\}}{Pr\{\mathcal{H}_2\}}, \tag{2.21}$$

where $Pr\{\mathcal{H}_i\}$, $i = 1, 2$, is the prior probability of $\mathcal{H}_i$ and $Pr\{\mathcal{H}_i|X\}$ is the posterior probability of $\mathcal{H}_i$, which is in fact the conditional probability of $\mathcal{H}_i$ given X. The right-hand side of the above equation is a measure of the difference between the logarithm of the odds in favor of $\mathcal{H}_1$ after the observation that the random variable under consideration assumed the value X and before the observation. Kullback and Leibler were the first to observe that this difference, which can be positive or negative, may be considered as the *information* resulting from the observation (Kullback and Liebler, 1951). They defined the logarithm of the likelihood ratio, $\ln \frac{p_1(X)}{p_2(X)}$, as the information in X for discrimination in favor of $\mathcal{H}_1$ against $\mathcal{H}_2$. The mean value of this information, that is, the mean information for discrimination in favor of $\mathcal{H}_1$ against

$\mathcal{H}_2$ per observation from the PDF p_1 is called the *Kullback-Leibler divergence* of p_1 from p_2:

$$D(p_1 \parallel p_2) \triangleq \int_{\mathcal{C}} p_1(X) \ln \frac{p_1(X)}{p_2(X)} \mathrm{d}X. \tag{2.22}$$

The base of the logarithm in the above definition is not important and only represents the choice of the unit of measurement. For discrete random variables, the Kullback-Leibler divergence is usually defined using base-2 logarithm in order to be consistent with the binary unit 'bit' as the unit of information:

$$D(p_1 \parallel p_2) \triangleq \sum_{\Omega} p_1(X) \log_2 \frac{p_1(X)}{p_2(X)}, \tag{2.23}$$

where $p_1(\cdot)$ and $p_2(\cdot)$ now represent two probability density functions defined on a countable set Ω.

Using (2.22) or (2.23) one can easily check that $D(p_1 \parallel p_2)$ has the following properties:

(i) $D(p_1 \parallel p_2) \geq 0$.
(ii) $D(p_1 \parallel p_2) = 0$ if and only if $p_1 = p_2$.
(iii) In general, $D(p_1 \parallel p_2) \neq D(p_2 \parallel p_1)$.

The property (*iii*) prevents $D(p_1 \parallel p_2)$ from satisfying the conditions of a metric on the space of probability density functions. However, as we will discuss shortly, $D(p_1 \parallel p_2)$ has a strong appeal as a measure of distance between probability density functions.

Example 2.7. Let p_1 and p_2 be N-dimensional Gaussian PDFs with zero mean and covariance matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ respectively. The Kullback-Leibler divergence of p_1 from p_2 is given by (Kullback, 1954, Chapter 9)

$$D(p_1 \parallel p_2) = \frac{1}{2} Tr(\mathbf{C}_1 \mathbf{C}_2^{-1}) - \frac{1}{2} \ln \frac{|\mathbf{C}_1|}{|\mathbf{C}_2|} - \frac{N}{2}. \tag{2.24}$$

$\diamondsuit$

Example 2.8. Let X and Y be two random variables with the joint probability density function $p_{XY}(X,Y)$ and marginal density functions $p_X(X)$ and $p_Y(Y)$. It can be easily verified that

$$I(X;Y) = D(p_{XY}(X,Y) \parallel p_X(X) \times p_Y(Y)). \tag{2.25}$$

$\diamondsuit$

Again, suppose that a probability density function of interest for the statistician is either $p_1(\cdot)$ or $p_2(\cdot)$. He has to decide between p_1 and p_2 on the basis of a *sample* of size k, i.e., the result of k independent drawings from the unknown distribution. Now, how separable or *distant* are p_1 and p_2 from the

point of view of a statistical observer? In other words, how difficult is it to distinguish between p_1 and p_2 having a series of sample data?

We can provide an answer to the above questions if we identify separability or *statistical distance* with *low probability of miss-classification using some sort of optimal decision procedure*. The following analysis shows that *statistical distance*, interpreted this way, is closely related to the Kullback-Leibler divergence of the two probability measures under consideration.

For simplicity, we consider a discrete and countable sample space. That is, p_1 and p_2 are two possible probability density functions associated with a discrete random variable X which takes values in a countable set Ω. A (non-randomized) test is characterized by a set $\Theta \subset \Omega^k$, in the sense that if the sample $\omega = \{X_1, \cdots X_k\}$ belongs to Θ, the statistician accepts p_1 and else he accepts p_2. It's important to note that the roles of the two hypotheses are not symmetric. It is, therefore, customary to prescribe a bound ϵ for the tolerated probability of wrong decision if p_1 is the true distribution. Then the task is to minimize the probability of wrong decision if hypothesis p_2 is true. The latter minimum is denoted by β and is given by

$$\beta(k, \epsilon) \stackrel{\triangle}{=} min\ p_2^k(\Theta),\ subject\ to\ \Theta \subset \Omega^k\ and\ p_1^k(\Theta) \geq 1 - \epsilon. \tag{2.26}$$

We are not interested in finding the optimal decision set Θ that achieves this minimum but in the minimum itself. The following theorem, known as Stein's Lemma, gives an asymptotic expression for $\beta(k, \epsilon)$:

Theorem 2.10. *(Csiszár and Körner, 1981, page 19) For any* $0 < \epsilon < 1$*,*

$$\lim_{k \to \infty} \frac{1}{k} \log_2 \beta(k, \epsilon) = - \sum_{X \in \Omega} p_1(X) log_2 \frac{p_1(X)}{p_2(X)}.$$

It follows from the above theorem that, for large k,

$$\beta(k, \epsilon) \simeq 2^{-kD(p_1 \| p_2)} \tag{2.27}$$

Therefore, one can say that the larger $D(p_1 \parallel p_2)$ is, the less is the risk of miss-classification between p_1 and p_2 or, in other words, the larger $D(p_1 \parallel p_2)$ is, the more distant (from the point of view of hypothesis testing) p_1 and p_2 are.

We can also interpret (2.27) in the sense that with a fixed probability of error, distant PDFs (for which $D(p_1 \parallel p_2)$ is large) can be classified with fewer samples available. In fact, the number of samples required is inversely proportional to $D(p_1 \parallel p_2)$.

2.4 Statistical inference

Probability theory is a mathematical discipline developed as an abstract model and it's calculations are *deductions* based on certain axioms. Statistics deals

with the application of probability theory to real problems and its conclusions are *inferences* based on observations, often without total justification.

Example 2.9 (A simple inference problem). Let x represent a random variable which is not observable to us. Let v denote an observable variable which is dependent on x through a known functional relation of the form $v = f(x)$. The problem is to determine a unique plausible value for x given the value of the dependent v. If $f(\cdot)$ invertible and the observed value of v is in the range of $f(\cdot)$, then the answer is simply $x = f^{-1}(v)$. However, if $f(\cdot)$ is not invertible, then knowing that $f(x) = v$ would, in general, specify a set of values for x. Which particular value of this set should we pick? ♢

In general, there is no answer to the question posed in the above example unless we make further assumptions about the nature of the variable x. We can go forward only if we submit to *inductive reasoning* which means that we are willing to accept *plausible* results generated by an *inference rule*.

2.4.1 The Maximum Likelihood principle

A popular statistical inference rule used widely by scientists and engineers is the *Maximum Likelihood principle*. According to this principle, to solve the inference problem mentioned in the above example we have to first assume that there exist a *known* probability distribution associated with the variable x. Then, we should calculate the *most probable value* for x given the data v according to the laws of probability theory. If this value exists and is unique, we shall use it as the inferred value of x.

A major issue with the maximum likelihood principle is that in many real-world inference problems, there is no straightforward way to assign an *a priori* probability distribution to the quantity to be inferred. For instance, in the above example where x is not observable, one might wonder how to assign a probability distribution to it?

One way to get out of this deadlock and move forward is to assume that we know the probability density function associated with the dependent variable v. Can this assumption help us infer a unique value for x subject to the condition that $f(x) = v$? The answer is, unfortunately, negative. This is because if $f(\cdot)$ is not invertible, it is not possible to specify a unique probability distribution for x given the probability distribution of v. To resolve this difficulty we need to look for an extra inference rule which enables us to *infer a probability distribution given certain conditions on that distribution*. This is the subject of the next subsection.

2.4.2 The Maximum Entropy principle

Mathematically speaking, the Maximum Entropy principle is a tool that enables one to infer a function $p(x)$ defined on a given set $\mathcal{X}$ when the available information specifies only a feasible set $\mathcal{P}$ of such functions.

Originally coming from the works of Boltzmann and Gibbs in statistical physics, the Maximum Entropy principle has been promoted to a general method of inference primarily by Edwin Jaynes (Jaynes, 1982), (Jaynes, 1983), (Jaynes, 2003). According to this principle, in an inference problem dealing with inferring probability distributions, one should choose a distribution which has maximum entropy among all those which are consistent with the problems' constraints.

In the following, we will outline two of the strongest arguments that have been put forth to support the principle of Maximum Entropy as *the* just rule of inference for choosing probability distributions.

Justification based on large deviation results: The idea of Maximum Entropy is closely connected to the subject of large deviations in probability theory, which, in turn, is closely related to information theory. Here, we provide some simple, yet fundamental, large deviation results which we may use to argue that Maximum Entropy is "right". The results stated below are regarded as key ingredients of information theory. However, they have been known in statistical physics much earlier, dating back to Boltzmann.

Theorem 2.11. *(Csiszár and Körner, 1981, Page 30) Given a finite set $\mathcal{X}$ of size $|\mathcal{X}|$, let $N_n(\hat{p})$ denote the number of n-tuples $(x_1, \cdots, x_n) \in \mathcal{X}^n$ with a given empirical density $\hat{p}$, where*

$$\hat{p}(x) \stackrel{\triangle}{=} \frac{1}{n}(\textit{number of indices } i \textit{ with } x_i = x). \tag{2.28}$$

Also, let $H(\hat{p})$ denote the entropy of $\hat{p}$. Then,

$$N_n(\hat{p}) = \exp[nH(\hat{p}) - r_n(\hat{p})], \tag{2.29}$$

where

$$0 \leq r_n(\hat{p}) \leq |\mathcal{X}| \log n. \tag{2.30}$$

Corollary 2.1. *If $x_1, x_2, \cdots, x_n$ are drawn independently from $\mathcal{X}$, governed by the probability density function q, then the empirical density will be $\hat{p}$ with probability*

$$\exp[-nD(\hat{p} \parallel q) - r_n(\hat{p})], \tag{2.31}$$

where $D(\hat{p} \parallel q)$ denotes the Kullback-Leibler divergence of $\hat{p}$ from q.

Suppose now that $\hat{p}$ is known to belong to a closed, convex feasible set $\mathcal{P}$, and let $p^* \stackrel{\triangle}{=} \arg\max_{p \in \mathcal{P}} H(p)$. Then, providing $\mathcal{P}$ contains empirical densities arbitrarily close to p^* if n is sufficiently large, the above theorem implies that all but an exponentially small fraction of the n-tuples $(x_1, x_2, \cdots, x_n)$ with empirical distribution $\hat{p} \in \mathcal{P}$ will be in an arbitrarily small neighborhood of p^*, if n is large. Similarly, the above corollary implies that if $x_1, x_2, \cdots, x_n$ are drawn independently from $\mathcal{X}$, governed by q, the conditional probability

(on the condition $\hat{p} \in \mathcal{P}$) that $\hat{p}$ will be in an arbitrarily small neighborhood of $p^* \stackrel{\triangle}{=} \arg\min_{p \in \mathcal{P}} D(p \parallel q)$ is exponentially close to 1, if n is large.

The above results represent very strong arguments for Maximum Entropy inference, providing the probability mass function to be inferred is an empirical distribution[4]. More general results of the above kind, for general rather than finite $\mathcal{X}$, are also available. See Van Campenhout and Cover (1981) and Csiszár (1984).

Justification based on axiomatization: A very attractive alternative is to take an axiomatic approach: Start from certain properties that a sound method of inference must have, and investigate whether the postulated properties uniquely characterize a distinguished method, or else what alternatives come into account. Such an approach was first put forward in the important paper by Shore and Johnson (1980) which inspired several later works including the excellent paper by Csiszár (1991). Below, we review certain axioms that, if one would require an inference method to satisfy them, one would arrive at "the inevitability of maximum entropy". Our account here is very brief and follows that of Csiszár (1991). The reader is referred to Shore and Johnson (1980), Skilling (1988), Paris and Vencovská (1990), Csiszár (1991), Csiszár (1996) and Jaynes (2003) for detailed discussions of this topic.

Assume, for simplicity, that $\mathcal{X}$ is a finite set of size $|\mathcal{X}| \geq 3$, and that $p(x)$ is a strictly positive function on $\mathcal{X}$. Our problem is to infer $p(x)$, knowing only a feasible set $\mathcal{P}$ and a "default" model $q(x)$. It is assumed that the feasible set is defined via linear constraints, that is,

$$\mathcal{P} \stackrel{\triangle}{=} \{p(x) : \sum_{\mathcal{X}} a_i(x)p(x) = b_i,\ i = 1, 2, \ldots, k\}$$

for some given functions $a_i(x)$ and constants b_i.

By an *inference method* we mean any rule that assigns, to every feasible set $\mathcal{P}$ defined be linear constraints and any default model q, an inferred function $p^* \in \mathcal{P}$, denoted by $p^*[\mathcal{P}; q]$. We emphasize that *it is not postulated that* p^* *is the minimizer of some measure of distance of* p *from* q. It is proved that this does follow from the regularity and locality axioms below and the proof constitutes a major step of the axiomatic approach.

(i) **(Regularity axiom)** If $\mathcal{P}_1 \subset \mathcal{P}$ and $p^*[\mathcal{P}; q] \in \mathcal{P}_1$ then $p^*[\mathcal{P}_1; q] = p^*[\mathcal{P}; q]$.

[4] This is *explicitly* the case in many applications of the Maximum Entropy principle in physics and *implicitly* the case for the problems we consider in this book. Note that if we accept Richard von Mises' "frequency definition" as our formal notion of probability, we have agreed that all probability measures, including those conceived for random processes, are limiting cases of their corresponding empirical distributions.

(ii) **(Locality axiom)** If the constraints defining $\mathcal{P}$ can be partitioned into two sets, the first involving functions $a_i(x)$ that vanish for $x \notin \mathcal{X}_1$ and the second involving functions $a_i(x)$ that vanish for $x \in \mathcal{X}_1$, where $\mathcal{X}_1$ is a subset of $\mathcal{X}$, then the values of $p^*[\mathcal{P}; q]$ for $x \in \mathcal{X}_1$ depend only on the first set of constraints and on the values of q for $x \in \mathcal{X}_1$.

(iii) **(Transitivity axiom)** If $\mathcal{P}_1 \subset \mathcal{P}$ then $p^*[\mathcal{P}_1; p^*[\mathcal{P}; q]] = p^*[\mathcal{P}_1; q]$.

(iv) **(Weak scaling axiom)** If $\mathcal{P} = \{p(x) : \ p(x_1) + p(x_2) = t\}$ for some $x_1, x_2 \in \mathcal{X}$ and $t > 0$, then $p^* = p^*[\mathcal{P}; q]$ satisfies $p^*(x_i) = \lambda q(x_i)$ for $i = 1, 2$ and $\lambda = t/(q(x_1) + q(x_2))$.

The regularity axiom formalizes the intuitive idea that if the inference based on some knowledge happens to be consistent also with some additional knowledge then the new knowledge provides no reason to change that inference. The locality axiom means that if the available knowledge contains pieces pertaining to disjoint subsets of $\mathcal{X}$, then, on each of these subsets, the inference must be based on the knowledge pertaining to that subset. The transitivity axiom is self-evident. The weak scaling axiom appears to be natural if we think of $p(x)$ as a probability mass function[5]. However, it is not self-evident and whether it is desirable or not-so-desirable would depend on the nature of the function $p(x)$ to be inferred.

Theorem 2.12. *(Csiszár, 1996, Theorem 7.1) The regularity, locality, transitivity and weak scaling axioms are satisfied if and only if*

$$p^*[\mathcal{P}; q] = \arg\min_{p \in \mathcal{P}} D(p \parallel q).$$

The inference rule suggested by the above theorem is called the principle of "Minimum Cross-Entropy." The Maximum Entropy principle follows immediately if we assume that $q(x)$ is the uniform prior.

Criticism: The Maximum Entropy principle has always remained controversial. This controversy, in part, derives from the fact that it relies on a *subjective* interpretation of probability as a measure of the degree of belief which a rational person ought to assign to the event. This contrasts with the *ensemble* or *frequency of occurrence* interpretations which are more common in many traditional applications of probability.

Another problem is that, the subjective interpretation of probability, even when equipped with the Principle of Maximum Entropy, does not resolve the issue of assigning prior probabilities in all cases. For instance, complications

[5] If a prior guess about a probability mass function has to be updated subject to a single constraint that specifies the probability of a given set, it is standard to assign probabilities to the elements of this set proportional to the prior ones. This is, in fact, what the weak scaling axiom requires.

arise in the infinite case, since there cannot be a flat distribution over denumerably many outcomes, on pain of violating Kolomogorov's probability calculus (with countable additivity).

The interested reader is referred to Uffink (1995), Uffink (1996) and Jaynes (2003) for further reading on this important topic. The fundamental texts by von Mises (1964) and Jeffreys (1967) represent the two major schools of thought in defining the *meaning* of probability. The authoritative book by Watanabe (1969) is highly recommended as well.

2.4.3 Probability density estimation

Let's go back to Example 2.9 at the beginning of this section. Under some mild mathematical conditions, the Maximum Entropy principle makes it possible to assign a unique probability distribution to the non-observable variable x given a probability distribution for the dependent variable v. The problem now remains as to how to choose a representative probability distribution for the variable v given a certain number of its observed values?

In probability theory, a probability distribution cannot be constructed based on finite-sample observations. This fundamental limitation causes major conceptual difficulties in applying both the axiomatic and the frequency theories of probability to modeling empirical data. J. Rissanen (Rissanen, 1989, Page 2) describes the problem with the axiomatic approach as follows:

> "We have grave difficulties in ensuring that the observed data behave as typical samples from the assumed distribution and checking the probabilistic consequences. Worse than that, while we know all about the behaviour of probabilities by Kolmogorov's axioms, no way has been found to give a constructive definition of probability which would allow us to recognize it in observed data in a clear-cut way. Similarly, there is no way for us to tell when a given data sequence is 'random' or 'randomly drawn' from a distribution. Indeed, even the 'mathematical' notion of randomness can only be defined in a relative sense ... As a consequence, all attempts to attach inherent probabilities to observed data have failed, as they indeed must."

The frequency definition of probability has nothing to say about finite data records either, as pointed out in the following words by Kolmogorov (Li and Vitányi, 1997, Page 55):

> "The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity does not contribute anything to substantiate the application of the results of probability theory to real practical problems where we always have to deal with a finite number of trials."

In spite of the fundamental difficulties mentioned above, statisticians have made up a wide variety of inference rules that, relying on various extra "assumptions", try to construct a probability distribution for a random variable

given some samples instances of that variable. These empirical methods collectively form the art of "statistical inference".

In statistical inference, the problem of estimating a probability distribution using sample observations is cast in several forms including *hypothesis testing* (deciding between two or more possible probability distributions), *parameter estimation* (estimating the value of some parameters that specify a probability distribution within a particular class) and *non-parametric estimation*, which considers the more general problem of estimating an arbitrary probability distribution based on observed data.

Hypothesis testing and parameter estimation are subjects that are standard fare in almost any book that deals with statistical inference. Such books range from the highly theoretical expositions written by statisticians, e.g. Kullback (1954) and Vajda (1989), to the more practical treatments contributed by the many users of applied statistics, e.g. Kay (1993), Kay (1998) and Duda et al. (2001). For an analysis of methods devised for non-parametric probability density estimation see Vapnik (1982), Devroye et al. (1996), Vapnik (1999), Duda et al. (2001) and Eggermont and LaRiccia (2001). Notable non-parametric probability density estimation methods include traditional histogram-based or Parzen-window methods (Duda et al., 2001, Chapter 4), kernel-based methods (Eggermont and LaRiccia, 2001) and methods based on structural risk minimization (Vapnik, 1982).

Any of the common probability density estimation methods may be used to complete the process of inferring x given the related variable v in Example 2.9.

2.4.4 Reliability of statistical inference principles

A very important question that comes to mind with regards to the empirical inference principles described above is how reliable they are?

For any density estimation method to be reliable, we need to have a sufficiently long series of independent data samples[6]. For the Maximum Entropy principle to be reliable, the density to be estimated should be among the so-called *typical densities* that satisfy the constraints (Csiszár, 1996). Finally, the maximum likelihood principle is reliable if the probability density function is highly concentrated around its maximum.

In real world, some of the above conditions might not hold for a particular application. In this case, using the inference methods described above might lead to nonsense. Ultimately, it is the user's responsibility to verify whether the assumptions based on which each inference step is formulated are indeed just, and, then, whether to associate any meaning to the results generated at the end. That's why statistical inference is an art not a science!

[6] Obviously, it is not clear what is meant by "sufficiently long" and this is yet another difficulty that exists in probability density estimation.

2.5 Stochastic processes

From a physical point of view, a stochastic process is a process whose behavior is governed at least in part by a random mechanism. Familiar examples of stochastic processes include stock market and exchange rate fluctuations, electronic signals such as speech, and audio; and medical signals such as a patient's EKG. A physical stochastic process is modelled by a random variable $x(t)$ with the continuous parameter $t \in \mathbb{R}$ representing time. A discrete-time stochastic process $x(n)$, $n \in \mathbb{Z}$, might be generated by sampling a physical continuous-time process at regular times $t = nT$.

2.5.1 Stationary stochastic processes

A discrete-time stochastic process $x(n)$ can be modelled by a set $\{x(n) : n \in \mathbb{Z}\}$ of real-valued random variables defined on the same probability space. In this book we work exclusively with discrete-time stochastic processes so we will drop mentioning the word "discrete-time" hereafter. A *wide-sense stationary* (WSS) stochastic process $x(n)$ is a process such that

(i) $E\{x^2(n)\} < \infty, \quad \forall n,$
(ii) $E\{x(n)\} = \mu \quad (constant), \quad \forall n,$
(iii) $E\{(x(n+k)-\mu)(x(n)-\mu)\} = E\{(x(k)-\mu)(x(0)-\mu)\} \quad \forall n, k;$

that is, the process has finite second moment at each instant and the mean values $E\{x(n)\}$ and covariance values $E\{(x(n+k)-\mu)(x(n)-\mu)\}$ are invariant with respect to shifts in the discrete time-parameter n.

We assume, without loss of generality, that $\mu = E\{x(n)\} = 0$. In this case the function $R_x(k) \stackrel{\triangle}{=} E\{x(n+k)x(n)\}$ will be called the *autocorrelation function* (ACF) associated with the process $x(n)$. It is sometimes more convenient to consider the set of ACF values $R_x(k)$, $k \in \mathbb{Z}$, as a sequence. Therefore, we sometimes call $R_x(k)$ the *autocorrelation sequence* (ACS) as well.

Due to the importance of the ACF, we review some of its main properties in this subsection. The reader is referred to any standard text on stochastic processes, e.g. Papoulis (1991), for proofs.

(i) $R_x(0) \geq 0$.
(ii) $R_x(-k) = R_x(k)$.
(iii) $|R_x(k)| \leq R_x(0)$.
(iv) $R_x(k)$ is a non-negative definite sequence. This is to say $\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j R_x(i-j) \geq 0$ for any N and any real numbers $a_1, a_2, \cdots, a_N$.
(v) Let $\mathbf{C}_{xx}$ be the N by N covariance matrix associated with the WSS process $x(n)$, that is, $[\mathbf{C}_{xx}]_{ij} = R_x(i-j)$, $1 \leq i, j \leq N$. Then, $\det \mathbf{C}_{xx} \geq 0$.

2.5.2 The power spectrum

One of the most important attributes of a stationary random process $x(n)$ is its *power spectrum*, also referred to as *power spectral density* or *PSD* for short. In this section, we provide a definition of power spectrum and clarify some issues regarding its interpretation. Our approach closely follows that of Mandel and Wolf (1995). The books by Middleton (1960), Caines (1988) and Papoulis (1991) contain well-written treatments of this topic as well.

Heuristically, one could try to introduce the spectrum of a random process as follows. Let's formally represent $x(n)$ as a Fourier integral,

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega t} \mathrm{d}\omega, \tag{2.32}$$

and let's assume that the integral exists and may be inverted, i.e., that

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}. \tag{2.33}$$

We might then attempt to define the spectrum $S(e^{j\omega})$ of $x(n)$ by the expectation value of $|X(e^{j\omega})|^2$, i.e.,

$$S(e^{j\omega}) = E\{|X(e^{j\omega})|^2\} \tag{2.34}$$

so that $S(e^{j\omega})$ would be a measure of the strength of the fluctuations associated with a particular Fourier component of $x(n)$. However, it turns out that the definition (2.34) is mathematically unsound. For if $x(n)$ is a stationary random process, it does not tend to zero as $n \to \infty$, simply because the underlying probability densities that characterize the fluctuations of $x(n)$ are invariant with respect to the translation of the origin of time. Thus, in the statistical sense, $x(n)$ can not behave any differently for large values of $|n|$ than it does for any other value of n. Consequently, $x(n)$ is neither square-integrable nor absolutely integrable and hence the Fourier integral (2.32) does not exist within the framework of the theory of ordinary functions. The difficulty just noted can be overcome only if we interpret our formulas within the framework of *generalized harmonic analysis*.

Generalized harmonic analysis originated by Wiener (1930). Wiener did not employ any statistical concepts in his analysis. Four years after publication of Wiener's classic paper, Khintchine (1934) derived results which were of the same nature as those of Wiener but were derived from a statistical point of view. It's worth noting that both Wiener and Khintchine employed the notion of *integrated spectrum* rather than *spectral density*, probably because the spectral density may become singular, though no more singular than the Dirac delta function. In the following, as in the rest of the book, we shall not hesitate to use the Dirac delta function and we will, therefore, work with the spectral density.

To have our historical perspective complete, we should also mention that it was not generally known till long after the publication of the papers by Wiener and Khintchine that the essential aspects of their results were discovered much earlier by Einstein[7] in 1914.

Now, let us return to the problem of defining the power spectrum. For this purpose, we will again use the Fourier transform relations (2.32) and (2.33), considering them now as symbolic formulas which, as just noted, can be given rigorous mathematical meaning if one goes beyond ordinary function theory.

For each realization of the real stationary random process $x(n)$, (2.33) will generate a complex function $X(e^{j\omega})$, and, hence, $X(e^{j\omega})$ is also a random function (process) in which frequency is the parameter rather than time. Let us now consider the expectation of the product $X(e^{j\omega})^* X(e^{j\omega'})$. From (2.33) we have, if we interchange the operations of expectation and summation, that

$$E\{X(e^{j\omega})^* X(e^{j\omega'})\} = \sum_{k=-\infty}^{+\infty} \sum_{k'=-\infty}^{+\infty} E\{x(k)x(k')\} e^{j(\omega k - \omega' k')}. \tag{2.35}$$

Since the process $x(n)$ is assumed to be stationary, $E\{x(k)x(k')\} = R_x(k-k')$, where $R_x(\cdot)$ is the ACF of $x(n)$. Using this fact, and setting $l = k - k'$, we find that

$$E\{X(e^{j\omega})^* X(e^{j\omega'})\} = \sum_{k'=-\infty}^{+\infty} e^{j(\omega-\omega')k'} \sum_{l=-\infty}^{+\infty} R_x(l) e^{-j\omega' l}, \tag{2.36}$$

which implies that

$$E\{X(e^{j\omega})^* X(e^{j\omega'})\} = P_x(e^{j\omega'}) \delta(\omega - \omega'), \tag{2.37}$$

where

$$P_x(e^{j\omega}) \triangleq \sum_{l=-\infty}^{+\infty} R_x(l) e^{-j\omega l}. \tag{2.38}$$

The two formulas (2.37) and (2.38) are both of fundamental importance. The first shows that the (generalized) Fourier components of a stationary random process belonging to different frequencies are uncorrelated, and that $P_x(e^{j\omega})$ is a measure of the strength of the fluctuations of the Fourier component at frequency ω, that is, $P_x(e^{j\omega})$ may be identified with the spectral density $S(e^{j\omega})$ of $x(n)$.

Formally, we regard the formula (2.38) as defining the power spectrum of the stationary random process $x(n)$ and use (2.37) as the justification of

[7] An English translation of Einstein's original 1914 paper has been published, along with interesting commentaries by A. M. Yaglom, in the IEEE ASSP magazine, Vol. 4, No. 4, November 1987.

our definition. Note that the singularity at $\omega' = \omega$ in (2.37) doesn't cause any problem since it can be removed if we integrate both sides over a small ω'-range around ω. Doing this, we might write our justifying formula (2.37) in the following form

$$P_x(e^{j\omega}) = \lim_{\Delta\omega \to 0} \int_{\omega-\Delta\omega}^{\omega+\Delta\omega} E\{X(e^{j\omega})^* X(e^{j\omega'})\} \mathrm{d}\omega'. \tag{2.39}$$

The similarity between the above formula and the naive definition (2.34) should be noted. Using the properties mentioned earlier for the ACF, one can prove the following properties for the power spectrum:

(i) $P_x(e^{j\omega})$ is a real function of ω.
(ii) $P_x(e^{-j\omega}) = P_x(e^{j\omega})$.
(iii) $P_x(e^{j\omega}) \geq 0$.

2.5.3 Processes with rational power spectra

In this book, we will consider WSS processes whose power spectral density $P_x(e^{j\omega})$ is a continuous function of ω and satisfies the following condition:

$$\int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) \mathrm{d}\omega > -\infty. \tag{2.40}$$

The above condition (known as the Paley-Wiener condition) ensures that the WSS process $x(n)$ is non-deterministic or linearly unpredictable (Caines, 1988, Section 1.3), (Papoulis, 1991, Section 14.2). A process that enjoys this property is called *regular*.

It is well known that the PSD of a regular WSS process may be factored into a product of the form

$$P_x(e^{j\omega}) = \sigma_0^2 |Q(e^{j\omega})|^2, \tag{2.41}$$

or, if we replace $e^{j\omega}$ by the complex variable z,

$$P_x(z) = \sigma_0^2 Q(z) Q(z^{-1}), \tag{2.42}$$

where

$$\sigma_0^2 = \exp\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) \mathrm{d}\omega\}, \tag{2.43}$$

and $Q(z)$ is the z-transform of a causal and stable sequence $q(k)$:

$$Q(z) = 1 + q(1)z^{-1} + q(2)z^{-2} + \cdots$$

It also turns out that $Q(z)$ has no poles or zeroes outside the unit circle, which, in turn, implies that $Q(z)$ has a stable and causal inverse $1/Q(z)$. Furthermore, one can show that

(i) Any regular WSS process $x(n)$ may be realized as the output of a causal and stable filter $Q(z)$ that is driven by white noise having a variance of σ_0^2. This is known as the innovations representation of the process. The filter $Q(z)$ is called the innovations filter or modeling filter of $x(n)$.
(ii) The inverse filter $1/Q(z)$ is a called a whitening filter. That is because, if $x(n)$ is filtered with $1/Q(z)$, then the output is white noise process with variance σ_0^2. This white noise process is called the innovations process associated with $x(n)$.

An important subset of regular WSS processes are those whose PSD is rational. In this case, the modelling filter $Q(z)$ might be written as the ratio of two monic polynomials $A(z)$ and $B(z)$ in z^{-1}:

$$Q(z) = \frac{A(z)}{B(z)} = \frac{1 + a(1)z^{-1} + a(2)z^{-2} + \cdots + a(q)z^{-q}}{1 + b(1)z^{-1} + b(2)z^{-2} + \cdots + b(p)z^{-p}}, \tag{2.44}$$

For completeness, let $a(0) = b(0) = 1$. A process whose modelling filter is of the above form is known as an *autoregressive moving average process* of order (p, q) and is referred to as an ARMA(p, q) process. Important special types of ARMA(p, q) processes result when either p or q is equal to zero. If $q = 0$, the process is generated by filtering white noise with an all-pole filter of the form

$$Q(z) = \frac{1}{1 + \sum_{k=1}^{p} b(k)z^{-k}}. \tag{2.45}$$

An ARMA$(p, 0)$ process is called an *autoregressive process* of order p and will be referred to as an AR(p) process. If, on the other hand, $p = 0$, the process is generated by filtering white noise with an FIR filter that has a system function of the form

$$Q(z) = 1 + \sum_{k=1}^{q} b(k)z^{-k}. \tag{2.46}$$

Such a process is known as a *moving average* process of order q and will be referred to as an MA(q) process. Note from (2.41) that the PSD of an MA(q) process is completely specified by a set of $q + 1$ numbers (parameters) $\{\sigma_0, b(1), \cdots, b(q)\}$. Similarly, to completely characterize the power spectrum of an AR(p) process, we need $p + 1$ parameters represented by the set $\{\sigma_0, a(1), \cdots, a(p)\}$.

The following properties of MA(q) and AR(p) processes are important and will be used frequently in our subsequent developments:

(i) The ACS $R_x(k)$ of an MA(q) process is zero for all values of k that lie outside the interval $-q \leq k \leq q$.
(ii) There is a one-to-one correspondence between the first $q + 1$ coefficients of the ACS of an MA(q) process and the $q + 1$ parameters that specify its PSD. This correspondence, however, is nonlinear and is given by the following convolution equation

$$R_x(k) = \sigma_0^2 b(k) \star b(-k). \tag{2.47}$$

(iii) The ACS $R_x(k)$ of an AR(p) process is non-zero for infinitely many values of k.

(iv) There is a one-to-one correspondence between the first $p+1$ coefficients of the ACS of an AR(p) process and the $p+1$ parameters that specify its PSD. This correspondence is also nonlinear and is given by the so-called Yule-Walker equations

$$R_x(k) + \sum_{l=1}^{p} b(l) R_x(k-l) = \sigma_0^2 \delta(k), \;\; k = 0, 1, \cdots, p. \tag{2.48}$$

2.5.4 Information rate of stochastic processes

The concept of entropy for random variables generalizes to the concept of *information rate* for WSS stochastic processes. Let $x(n)$ be a Gaussian WSS random process with PSD $P_x(e^{j\omega})$. The information rate of this process is defined by

$$H(x) \stackrel{\triangle}{=} \lim_{N \to \infty} \frac{1}{N} H(x(n), x(n+1), \cdots, x(n+N-1)).$$

It can be shown that the above limit exists and is given by (Papoulis, 1991, Page 568)

$$H(x) = \frac{1}{2} \ln 2\pi + \frac{1}{2} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) \mathrm{d}\omega, \tag{2.49}$$

where $P_x(e^{j\omega})$ is the power spectrum of $x(n)$. Since $H(\cdot)$ is a functional of P_x, we will use the notations $H(x)$ or $H(P_x)$ interchangeably for the entropy rate of a Gaussian WSS random process x.

For two or more correlated random processes, one can define *mutual information rate* and *joint information rate* as well but these concepts are not needed in this book. Information theory of WSS stochastic processes has been studied in detail in the classic book by Pinsker (1964). More recent books containing readable accounts on information rate are Caines (1988), Gray (1990), Cover and Thomas (1991), and Papoulis (1991).

The definition of Kullback-Leibler divergence given in (2.22) can be extended to zero-mean Gaussian WSS stochastic processes as well. As mentioned earlier in this chapter, the statistical properties of such processes are completely determined by their PSD function or, equivalently, by their ACS. So, let $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ represent two hypotheses for the PSD of a zero-mean Gaussian stochastic process $x(n)$. Now, define X_N as the vector containing N samples of the process $x(n)$, that is,

$$X_N \stackrel{\triangle}{=} [x(0) \; x(1) \; \cdots \; x(N-1)]^T. \tag{2.50}$$

Obviously, X_N is an N-dimensional Gaussian random vector with zero mean and an $N \times N$ covariance matrix which is uniquely determined by the PSD

of $x(n)$. We denote by D_N the Kullback-Leibler divergence between the two possible PDFs of X_N. From, (2.24), we can then write

$$D_N = \frac{1}{2} Tr(\mathbf{C}_1 \mathbf{C}_2^{-1}) - \frac{1}{2} \ln \frac{|\mathbf{C}_1|}{|\mathbf{C}_2|} - \frac{N}{2}, \tag{2.51}$$

where $\mathbf{C}_1$ represents the covariance matrix of X_N if $P_1(e^{j\omega})$ is the true PSD and $\mathbf{C}_2$ represents the covariance matrix of X_N for the case that $P_2(e^{j\omega})$ is true.

One can show that under mild conditions $\lim_{N\to\infty} D_N$ exists. This limit, when it exists, is defined as the Kullback-Leibler divergence $D(P_1 \parallel P_2)$ between the two power spectral densities $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$.

Theorem 2.13. *(Pinsker, 1964, Theorem 10.5.1) (Vajda, 1989, Proposition 8.29) Let $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ be two candidate PSDs for the zero-mean Gaussian WSS random process $x(n)$. If $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ are essentially bounded from below[8], then the Kullback-Leibler divergence of $P_1(e^{j\omega})$ from $P_2(e^{j\omega})$ exists and is given by*

$$D(P_1 \parallel P_2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{P_1(e^{j\omega})}{P_2(e^{j\omega})} - \ln \frac{P_1(e^{j\omega})}{P_2(e^{j\omega})} - 1 \right) \mathrm{d}\omega. \tag{2.52}$$

Note that if $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ are rational with no poles or zeroes on the unit circle, then the conditions of the above theorem are satisfied and $D(P_1 \parallel P_2)$ exists.

The Kullback-Leibler divergence between two candidate PSDs has important distance-like properties which can be used to introduce certain geometries on the space of PSD functions. The reader is referred to Amari and Nagaoka (2000) for further discussions on this very interesting topic.

[8] This is to say, there is a positive constant ϵ such that $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ are both greater than ϵ for almost all ω.

3

Multirate Spectrum Estimation

3.1 Introduction

A fundamental problem in statistical signal processing is how to estimate the power spectrum of a random process given a finite number of it's samples. Spectrum estimation has important applications in many fields of applied physics and engineering. Examples can be found, for instance, in oceanography, radar, geophysics, optics and oil exploration. See Kay (1988), Percival and Walden (1993), Hayes (1996) and Buttkus (2000).

In this chapter we consider the problem of estimating the power spectrum of a random process using a series of multirate statistical measurements. The sensor network setup shown in Fig. 3.1 represents a simple instance of this scenario. In this setup, a stationary sound source is to be monitored by a collection of wireless sensors (Motes) put at various known locations in a room. Assume that each Mote can process its observed signal and calculate it's first L correlation coefficients. The correlation coefficients are then transmitted to a central station for processing. Our goal is to consolidate the statistical measurements sent in by each Mote and make an accurate estimate of the power spectrum of the sound source.

The material in this chapter are based on the paper by Jahromi et al. (2004b).

3.2 Mathematical modelling of the problem

We discussed in Chapter 1 that in many cases multirate measurements can be modelled by a multirate filter bank. Consider the filter bank in Fig. 3.2. This filter bank can model an M-channel sensor system where each sensor produces a low-rate (low-resolution) measurement $v_i(n)$ of an original physical signal $x(n)$. The filters in Fig. 3.2 can model the bandwidth limitations of each sensor as well as reverberations, time-difference of arrival (TDOA), attenuation and other linear propagation artifacts. Each filter is followed by a down-sampler

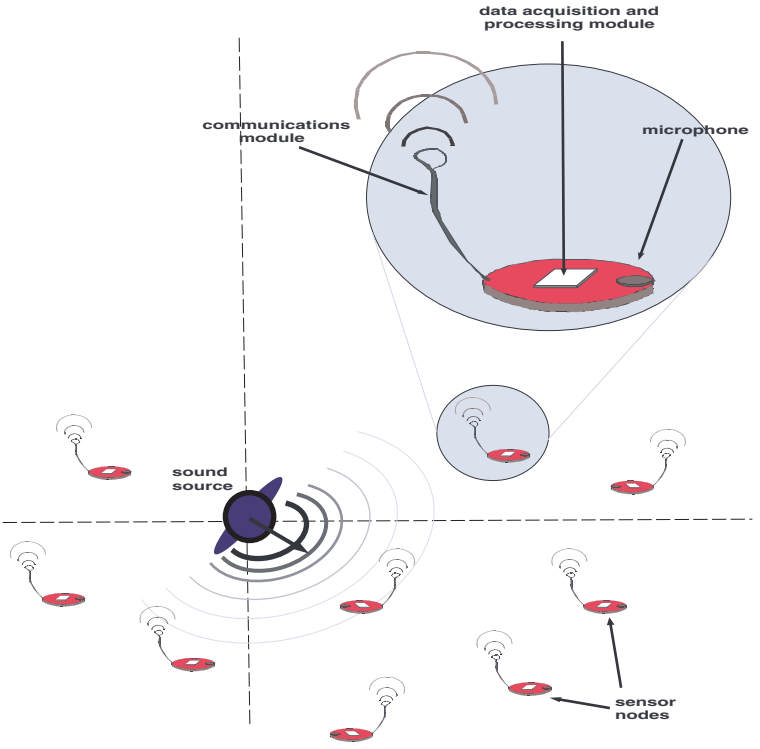


Fig. 3.1. A distributed network of sensors monitoring a stationary sound source in a room.

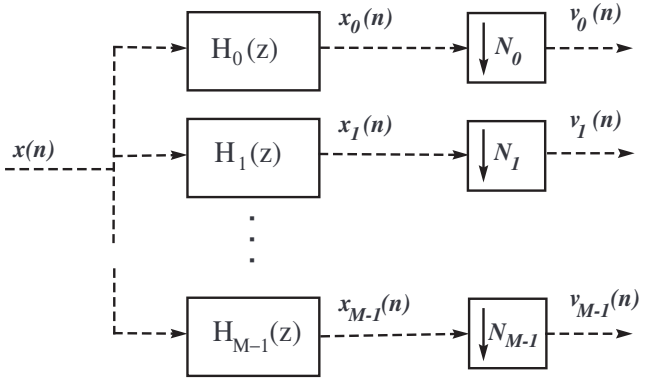


Fig. 3.2. An M-channel multirate filter bank can be used to model multirate sensors.

device which models the (possible) difference in sampling rate among the sensors. The decimators can also represent the limited processing speed the sensors' on-board processor.

Assume that the signal under measurement $x(n)$ is a realization of a zero-mean stationary random process. Recall from Chapter 2 that a complete statistical description of such a process is provided by its autocorrelation sequence (ACS)

$$R_x(k) \stackrel{\triangle}{=} E\{x(n)x(n+k)\}$$

or, equivalently, by its power spectrum which is also known as the power spectral density (PSD)

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_x(k)e^{-j\omega k}.$$

The autocorrelation sequence is a time-domain description of the second order statistics of a random process. The power spectrum provides a frequency domain description of the same statistics.

It is straightforward to show that the observations $v_0(n)$ to $v_{M-1}(n)$ in Fig. 3.2 are also WSS processes. The ACS of the low-rate measurements can be written as

$$R_{v_i}(k) = R_{x_i}(N_i k) \tag{3.1}$$

where

$$R_{x_i}(k) = (h_i(k) \star h_i(-k)) \star R_x(k), \tag{3.2}$$

and $h_i(k)$ denotes the impulse response of $H_i(z)$. We can express $R_{v_i}(k)$ as a function of the input PSD as well. Let's define $G_i(z) \stackrel{\triangle}{=} H_i(z)H_i(z^{-1})$. Now, (3.2) can be written as

$$R_{x_i}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jk\omega}d\omega. \tag{3.3}$$

From (3.1) and (3.3), we can then write

$$R_{v_i}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jN_ik\omega}d\omega. \tag{3.4}$$

The above relation shows that we can uniquely specify $R_{v_i}(k)$ for all values of i and k if we know $P_x(e^{j\omega})$. Here, we are interested in the following inverse problem:

Problem 3.1. Estimate $P_x(e^{j\omega})$ given $R_{v_i}(k)$ for $i = 0, \ldots, M-1$ and $k = 0, 1, \ldots, L-1$.

It is easy to check that, in general, knowing $R_{v_i}(k)$ for some limited values of i and k is not sufficient for characterizing $P_x(e^{j\omega})$ uniquely. That is, given a finite set of ACS values $R_{v_i}(k)$ there usually exist infinitely many $P_x(e^{j\omega})$ which can generate those values. Thus, Problem 3.1 is ill-posed.

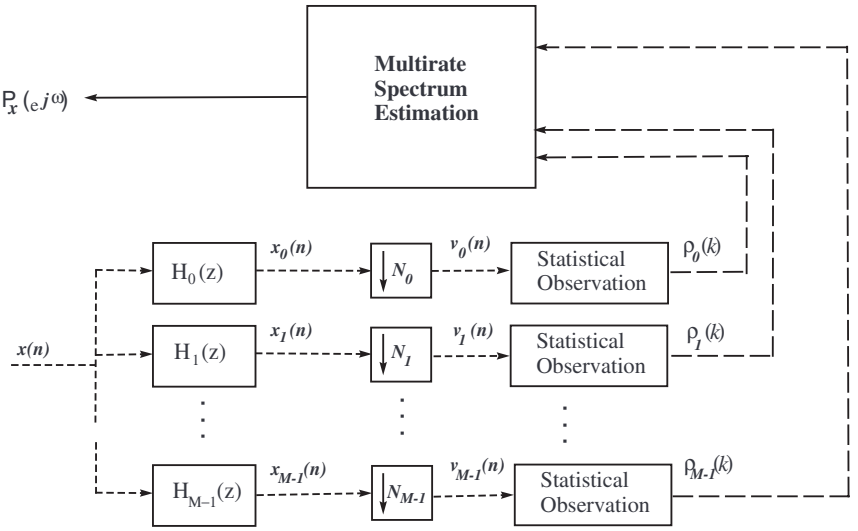


Fig. 3.3. The basic problem in multirate spectrum estimation is to estimate $P_x(e^{j\omega})$ given certain statistics of the low-rate observations $v_i(n)$.

3.3 The Maximum Entropy principle

We discussed in Chapter 2 that the Maximum Entropy principle can be used to resolve certain ill-posed statistical inference problems. According to this principle, one should choose *the most random* solution that satisfies the known constraints.

The Maximum Entropy principle was introduced in 1957 by E. T. Jaynes in the field of thermodynamics. Applying this principal to spectrum estimation is credited to Burg (1967) who used Jaynes' idea to estimate the PSD of a random signals given a limited number of its autocorrelation coefficients. In this chapter, we will generalize Burg's method to the multirate case. This is to say, we try to solve Problem 3.1 by finding a $P_x(e^{j\omega})$ which is

(i) consistent with the given ACS values $R_{v_i}(k)$,
(ii) associates with the most random input signal $x(n)$.

In the literature on the Maximum Entropy principle, e.g. Wu (1997), randomness is usually quantified using Shannon's entropy. Recall from Chapter 2 that the notion of entropy of a scalar random variable translates to the notion of *entropy rate* for WSS random signals. Recall also that when $x(n)$ is Gaussian, the entropy rate is given by

$$H(P_x) = \frac{1}{2}\ln 2\pi + \frac{1}{2} + \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln P_x(e^{j\omega})\mathrm{d}\omega. \tag{3.5}$$

Let $\Gamma \triangleq \{\rho_i(k) : i = 0, \ldots, M-1, k = 0, 1, \ldots, L-1\}$ denote the set of known autocorrelation coefficients obtained from the available low-rate signals $v_i(n)$. The set Ω of all input PSDs which are consistent with these data is given by

$$\Omega \triangleq \left\{ P_x : \begin{array}{l} \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jN_i k\omega}\mathrm{d}\omega = \rho_i(k); \forall \rho_i(k) \in \Gamma \\ P_x(e^{j\omega}) \in L^1(-\pi, \pi); \\ P_x(e^{j\omega}) \geq 0 \end{array} \right\} \tag{3.6}$$

Definition 3.1. *The data set Γ is called "admissible" if $\Omega \neq \emptyset$.*

Now, the Maximum Entropy principle asserts that we should pick a $P_x(e^{j\omega}) \in \Omega$ which has maximum entropy. This means we have to solve the following constrained optimization problem:

Problem 3.2. Find $P_x^*(e^{j\omega}) = \arg\max\ H(P_x)$ subject to $P_x \in \Omega$.

Theorem 3.1. *If Problem 3.2 has a solution, it is in the form*

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} G_i(e^{j\omega})F_i(e^{jN_i\omega})}, \tag{3.7}$$

where $F_i(z) \triangleq \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$. The coefficients λ_{ik} of the transfer functions $F_i(z)$ in (3.7) are specified such that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{G_i(e^{j\omega})e^{jN_i k\omega}}{\sum_{j=0}^{M-1} G_j(e^{j\omega})F_j(e^{jN_j\omega})} d\omega = \rho_i(k) \tag{3.8}$$

holds for all values of $i = 0, 1, \cdots M-1$ and $k = 0, 1, \cdots L-1$.

Proof. By using standard techniques from calculus of variations, we can convert Problem 3.2 to an unconstrained optimization problem. To do so, we form the functional

$$\begin{aligned} J \triangleq\ & \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln P_x^*(e^{j\omega}) d\omega + \\ & \sum_{i=0}^{M-1}\sum_{k=-(L-1)}^{L-1} \lambda_{ik}\left[\rho_i(k) - \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x^*(e^{j\omega})G_i(e^{j\omega})e^{jN_i k\omega} d\omega\right] \end{aligned} \tag{3.9}$$

where λ_{ik} are the Lagrange multipliers. Note that since both $\rho_i(k)$ and $R_{v_i}(k)$ are symmetric in k, we have $\lambda_{ik} = \lambda_{i(-k)}$. Taking the variation of J with respect to P_x^* we get

$$\begin{aligned} \delta J = & \frac{1}{4\pi}\int_{-\pi}^{\pi} \delta P_x^*(e^{j\omega}) \frac{1}{P_x^*(e^{j\omega})} d\omega - \\ & \frac{1}{2\pi}\int_{-\pi}^{\pi} \delta P_x^*(e^{j\omega}) \left[\sum_{i=0}^{M-1}\sum_{k=-(L-1)}^{L-1} \lambda_{ik} G_i(e^{j\omega})e^{jN_i k\omega}\right] d\omega. \end{aligned}$$

Putting δJ equal to zero[1] leads to

$$\int_{-\pi}^{\pi} \left[\frac{1}{P_x^*(e^{j\omega})} - 2 \sum_{i=0}^{M-1} G_i(e^{j\omega}) \left(\sum_{k=-(L-1)}^{L-1} \lambda_{ik} e^{jN_i k\omega} \right) \right] \delta P_x^*(e^{j\omega}) d\omega = 0.$$

Since $\delta P_x^*(e^{j\omega})$ is arbitrary, the quantity in the square brackets must be zero. Therefore, we obtain

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} G_i(e^{j\omega}) \left(\sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} e^{jN_i k\omega} \right)}. \tag{3.10}$$

It's convenient to define $F_i(z) \triangleq \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$ such that (3.10) can be written in the compact form

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} G_i(e^{j\omega}) F_i(e^{jN_i\omega})}. \tag{3.11}$$

To uniquely specify our solution, we have to calculate the Lagrange multipliers λ_{ik} which characterize the FIR transfer functions $F_i(z)$ in (3.11). As is standard in calculus of variations, λ_{ik} are obtained by the requirement that $P_x^*(e^{j\omega})$ must satisfy the constraints, i.e. $P_x^*(e^{j\omega}) \in \Omega$. Thus, $F_i(z)$ must be chosen such that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{G_i(e^{j\omega}) e^{jN_i k\omega}}{\sum_{j=0}^{M-1} G_j(e^{j\omega}) F_j(e^{jN_j\omega})} d\omega = \rho_i(k) \tag{3.12}$$

holds for all values of $i = 0, 1, \cdots M-1$ and $k = 0, 1, \cdots L-1$. □

Note that the transfer function $F_i(e^{j\omega})$ is a real function of ω. Thus, we might use the simpler expression $F_i(e^{j\omega}) = \sum_{k=0}^{L-1} \lambda_{ik} \cos(k\omega)$ in (3.7) and (3.8).

3.4 A geometric interpretation

Let $X = [x_0\ x_1\ x_2]^T$ be a vector in the Euclidean space $\mathbb{R}^3$ and define $\mathcal{N} = \{X \mid \langle X, X_0 \rangle = 3, X \geq 0\}$ where $X_0 = [1\ 1\ 1]^T$ is fixed reference vector. The equation $\langle X, X_0 \rangle = 3$ specifies a plane surface. The set $\mathcal{N}$ simply represents those vectors (points) on this surface which are confined to the first (positive) quadrant (Fig. 3.4).

The Burg entropy of each point in the set $\mathcal{N}$ is given by $\mathcal{H}(X) = \sum_{k=0}^{2} \ln x_k$. Fig. 3.5 shows the contours induced by the Burg entropy on the

[1] We'll show in the next section that the functional $H(P_x)$ is strictly concave on its domain of definition. Thus, putting the first variation δJ of the functional J is sufficient for obtaining a global maximizer.

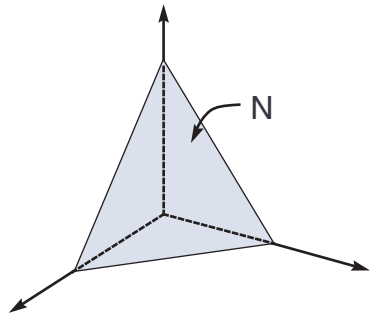


Fig. 3.4. Geometric visualization of $\mathcal{N}$ as a closed triangular region in the Euclidean space.

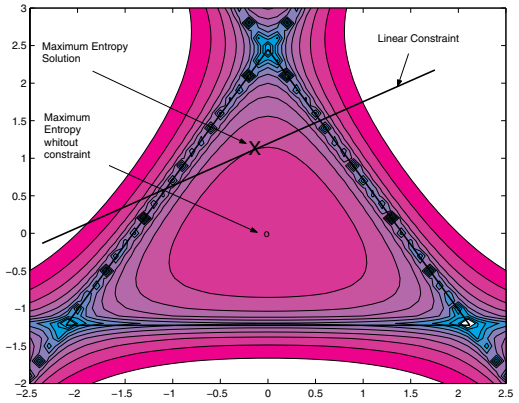


Fig. 3.5. A linear constraint will specify a line in $\mathbb{R}^3$. The set of feasible solutions is represented by the segment of this line which intersects with the triangular region that represents $\mathcal{N}$. The Maximum Entropy principle chooses a point on this segment which has the highest entropy. As entropy increase for points close the centre of the triangle $\mathcal{N}$, the selected point will be "close" to the centre.

triangular region specified by $\mathcal{N}$. Clearly, entropy increases as we approach the center of the triangle and decreases as we approach its borders. The Entropy is $-\infty$ on the side of this triangle. An extra linear constraint will specify a line in $\mathcal{N}$ (Fig. 3.5).

The above geometrical picture can be extended to the space of power spectra as follows: Many power spectra of practical interest are square integrable. Thus, we can imagine that $P(e^{j\omega})$ are "points" in the functional Hilbert space $L^2(-\pi, \pi)$. In this space, the constraint set Ω can be pictured as the intersection of the linear manifold defined by (3.14) and the positive quadrant defined by $P(e^{j\omega}) > 0$.

The Maximum Entropy principle chooses *the closest point* to the center of the triangle $\mathcal{N}$ among all the points in the feasible set, where "close" is interpreted as"high entropy". This is shown in Fig. 3.5.

3.5 Properties of the Maximum Entropy solution

The use of entropy-based stabilizing functions for regularizing various ill-posed inverse problems has been an active area of research in applied mathematics. See, e.g. Klaus and Smith (1988), Borwein and Lewis (1991), Engl and Landl (1993), Eggermont (1993), Engl et al. (1996) and Leonev (2000).

Entropy-based regularization methods usually lead to an optimization problem of the form

$$\begin{aligned} &\text{minimize} \int_a^b \phi(P(\omega))\mathrm{d}\omega,\ P(\omega) \in L^1(a,b) \\ &\text{subject to} \int_a^b G_i(\omega)P(\omega)\mathrm{d}\omega = \alpha_i, \end{aligned} \tag{3.13}$$

where $i = 0, 1, \cdots, N-1$ indexes the constraints, $G_i(\omega)$ are known functions and α_i are the measured data. The function $\phi(\cdot)$ reflects the choice of the entropy measure. Two popular choices for $\phi(\cdot)$ are (negative) Shannon entropy

$$\phi(P) \stackrel{\triangle}{=} P \log P$$

and (negative) entropy rate

$$\phi(P) \stackrel{\triangle}{=} -\log P$$

which is sometimes called Burg entropy.

Well-posedness of the minimization problem (3.13) when ϕ is the Shannon entropy has been established in several works including Klaus and Smith (1988), Amato and Hughes (1991), Borwein and Lewis (1991), Eggermont (1993) and Teboulle and Vajda (1993). Well-posedness results exist for certain classes of functions ϕ as well. See Teboulle and Vajda (1993) and Leonev (2000). Unfortunately, the function $\phi(P) = -\log P$ which leads to entropy rate (3.5) does not satisfy the conditions required in the above mentioned works[2].

While entropy rate (3.5) lacks theoretical support as a regularizing functional, it has been used extensively and successfully in many areas of application including spectral estimation and image restoration. What makes entropy rate attractive in the context of spectrum estimation is that its minimization subject to linear constraints implied by $P_x(e^{j\omega}) \in \Omega$ leads to a rational function. This is a very desirable property since the theory of stationary random processes with rational spectra is well known.

In the rest of this section we present some of the available (albeit partial) results on uniqueness, existence and stability of the Maximum entropy spectrum estimate parameterized in Theorem 3.1.

[2] The classes of stabilizing functionals studied by Teboulle and Vajda (1993) and Leonev (2000) require that $\lim_{P\to\infty} \frac{\phi(P)}{P} = \infty$. This property is not satisfied by $\phi(P) = -\log P$.

3.5.1 Uniqueness

Lemma 3.1. *When it exists, the solution to Problem 3.2 is unique.*

Proof. The proof is based on the fact that the entropy rate (3.5) is a concave functional. To see this, let $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$ be two distinct PSDs in the constraint set Ω. Define $P_\alpha(e^{j\omega}) = \alpha P_1(e^{j\omega}) + (1-\alpha)P_2(e^{j\omega})$. It is easy to check that Ω is a convex set. Thus $P_\alpha(e^{j\omega}) \in \Omega$ for $0 \leq \alpha \leq 1$ as well. Now, we define

$$\mathcal{H}(\alpha) \stackrel{\triangle}{=} \mathcal{H}(P_\alpha) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(\alpha P_1(e^{j\omega}) + (1-\alpha)P_2(e^{j\omega}))\mathrm{d}\omega.$$

It is easy to check that

$$\frac{\partial^2 \mathcal{H}(\alpha)}{\partial \alpha^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{-(P_1(e^{j\omega}) - P_2(e^{j\omega}))^2}{(P_\alpha(e^{j\omega}))^2}\mathrm{d}\omega < 0.$$

The above expression shows that (3.5) is a (strictly) concave functional which, in turn, proves that its maximum over Ω, when it exists, is unique. □

3.5.2 Existence

A comprehensive analysis of the problem of existence of a Maximum Entropy solution has been done by Borwein and Lewis (1993). Their results show that if one can specify the Lagrange multipliers λ_{ik} such that $P^*(e^{j\omega})$ is positive, then $P^*(e^{j\omega})$ *is* the solution to Problem 3.2. However, if $P^*(e^{j\omega})$ obtained after finding the Lagrange multipliers is not positive definite, then the solution to Problem 3.2 may contain a singular part as well.

In genera, it's not known if the solution given by Theorem 3.1 is positive for all feasible data $\rho_i(k)$. Because of this and other computational difficulties, we will use a modified parameterization of $P_x^*(e^{j\omega})$ in our computational algorithms introduced in the next section. This modified parameterization is only an approximation but has the practical advantage of being positive definite.

3.5.3 Stability

For stability analysis, it is more convenient if we represent the individual equality constraints imposed on the solution $P_x^*(e^{j\omega})$ by a single operator equation. To do this, we define

$$R \stackrel{\triangle}{=} [\rho_0(0), \cdots, \rho_0(L-1), \rho_1(0), \cdots \rho_1(L-1), \cdots \ \cdots, \rho_{M-1}(L-1)]^T$$

as the long vector containing all the known ACS values in Γ. Then, the condition that $P_x(e^{j\omega})$ must be consistent with all these given data can be simply expressed as

$$AP_x = R, \tag{3.14}$$

where A is a linear integral operator from $L^1(-\pi, \pi)$ to $\mathbb{R}^{ML}$. The operator A itself is given by

$$A \triangleq [A_{0,0}, \cdots, A_{0,L-1}, A_{1,0}, \cdots A_{1,L-1}, \cdots \cdots, A_{M-1,L-1}]^T, \tag{3.15}$$

where

$$A_{i,k}P_x \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jN_ik\omega}\mathrm{d}\omega. \tag{3.16}$$

Now, consider the functional

$$W(P_x) \triangleq \|AP_x - R\| + \alpha^2 D(P_x \parallel P_0), \tag{3.17}$$

where P_0 is a reference (flat) power spectrum and α is a small constant. Minimizing this functional can be regarded as an approximation to solving Problem 3.2.

Let R_δ denote a perturbed data vector (which might be obtained, for example, by a noisy measurement of the true autocorrelation values that constitute the data vector R) and assume that $\|R_\delta - R\| \leq \delta$. The following theorem, essentially due to Eggermont (1993), establishes a bound on the Kullback-Leibler divergence between the solutions of the original problem and the perturbed one:

Theorem 3.2. *Let R and R_δ be two admissible data vectors such that $\|R_\delta - R\| \leq \delta$. Also, let P, P_δ and P_0 denote three power spectra in $L^1(-\pi, \pi)$. If P minimizes the functional $W(P) = \|AP - R\| + \alpha^2 D(P \parallel P_0)$ over and P_δ minimizes the functional $W_\delta(P) = \|AP - R_\delta\| + \alpha^2 D(P \parallel P_0)$, then*

$$\|A(P - P_\delta)\| + \alpha^2 D(P \parallel P_\delta) \leq 4\delta^2. \tag{3.18}$$

The above theorem shows that the Kullback-Leibler divergence between the solution of the perturbed problem and that of the original problem is bounded by $\frac{4}{\alpha^2}\|R_\delta - R\|^2$. However, since the Kullback-Leibler divergence is not a metric, this does not imply continuous dependence of the solution of Problem 3.2 on the data. To obtain a formal continuity result, one has to further show that the distance (in the metric of some valid topology) between the solution of the perturbed problem and that of the original problem is bounded when $\|R_\delta - R\|$ is bounded. This is an open research problem.

3.6 Computing the Maximum Entropy solution

Theorem 3.1 indicated that the solution to Problem 3.2 has the functional form

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} |H_i(e^{j\omega})|^2 F_i(e^{jN_i\omega})}, \tag{3.19}$$

where $F_i(z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$, and λ_{ik} are Lagrange multipliers to be found. Since $P_x(e^{j\omega})$ is real, we must have $\lambda_{ik} = \lambda_{i(-k)}$. This means there are only L coefficients to be specified for each $F_i(z)$.

The above solution exists only if it is possible to choose the Lagrange multipliers embedded in $F_i(z)$ such that $P_x^*(e^{j\omega}) \in \Omega$. That is, only if it is possible to choose $F_i(z)$ such that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|H_i(e^{j\omega})|^2 e^{jN_i k\omega}}{\sum_{j=0}^{M-1} |H_j(e^{j\omega})|^2 F_j(e^{jN_j\omega})} \mathrm{d}\omega = \rho_i(k) \tag{3.20}$$

for all values of $i = 0, 1, \cdots M-1$ and $k = 0, 1, \cdots L-1$. In practice, it is often the case that the set Ω is empty due to inconsistent $\rho_i(k)$ found empirically. In this case, the system of equations (3.20) does not have a solution. To get around this difficulty, give up the strict equality constraints imposed by $P_x^*(e^{j\omega}) \in \Omega$ and instead seek a solution to the following least-squares problem:

Problem 3.3. Find λ_{ik} such that $F_i(z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$ minimize

$$J = \sum_{i=0}^{M-1}\sum_{k=0}^{L-1}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|H_i(e^{j\omega})|^2 e^{jN_i k\omega}}{\sum_{j=0}^{M-1} |H_j(e^{j\omega})|^2 F_j(e^{jN_j\omega})} \mathrm{d}\omega - \rho_i(k)\right)^2. \tag{3.21}$$

Note that when Problem 3.2 has a solution, it will be *the* solution to Problem 3.3 as well. However, if Problem 3.2 doesn't have a solution, then the solution to Problem 3.3 *which always exists* provides a viable approximation. We will use a semi-numerical algorithm for approximately solving Problem 3.3. We call this algorithm the *Maximum Entropy Inference Engine* or *MEIE* for short. For convenience, the main steps of this algorithm are listed in the box below.

Some of the steps outlined in Algorithm 3.1 still require special tricks and numerical methods. For instance, a key step is calculating the integrals necessary to compute the functions $g_{i,k}(\Lambda)$. Experiment shows that these integrals are very difficult to calculate numerically because the integrand tends to blow-up whenever any $H_j(e^{j\omega})$ gets close to zero. To avoid this difficulty, we have to work out an indirect method for calculating $g_{i,k}(\Lambda)$. A possible formulation which enables calculating $g_{i,k}(\Lambda)$ without having to perform ill-conditioned numerical integrations is presented below.

Recall from Chapter 2 that every PSD function $P_x(e^{j\omega})$ admits a spectral factorization of the form

$$P_x(e^{j\omega}) = \sigma_0^2 |Q(e^{j\omega})|^2, \tag{3.22}$$

or, if we replace $e^{j\omega}$ by the complex variable z,

$$P_x(z) = \sigma_0^2 Q(z) Q(z^{-1}), \tag{3.23}$$

where $Q(z)$ represents the transfer function of a causal and stable system.

Let $Q^*(z)$ represent the spectral factor of the rational PSD $P_x^*(e^{j\omega})$ given in (3.19). Algorithm 3.1 calculates $P_x^*(e^{j\omega})$ as an optimal element among a family $P_x(\Lambda; e^{j\omega})$ of PSDs parameterized by the parameter vector Λ. Let us denote by $Q(\Lambda; z)$ the spectral factor associated with this parameterized family of PSDs. Then, $Q^*(z)$ would be equal to $Q(\Lambda^*; z)$ where Λ^* denotes the optimal value of the parameters.

One easily observes that, for any Λ, $Q(\Lambda; z)$ can be identified with the *modeling filter* of an ARMA(p, q) process where p and q depend on the degree and nature of the analysis filters $H_i(z)$, the down-sampling ratios N_i and the number of ACS values available as data.

Any WSS process can be represented by its modeling filter driven by (unit-variance) white noise. Applying this fact to the input signal $x(n)$ in Fig. 3.2, it would be straightforward to calculate the ACS values $R_{v_i}(k)$ of the observable signals $v_i(n)$.

Define $q(\Lambda; k)$ as the impulse response of the modeling filters $Q(\Lambda; z)$ and $h_i(k)$ as the impulse response of the analysis filters $H_i(z)$. Then, we have

$$R_{x_i}(k) = (h_i(k) \star h_i(-k)) \star (q(\Lambda; k) \star q(\Lambda; -k)), \tag{3.24}$$

from which we can calculate

$$R_{v_i}(k) = R_{x_i}(N_i k). \tag{3.25}$$

Algorithm 3.1: Maximum Entropy Inference Engine

Input: Autocorrelation measurements $\rho_i(k)$, the transfer functions $H_i(z)$ and the down-sampling ratios N_i for $i = 0, \cdots, M-1$ and $k = 0, \cdots L-1$.
Output: An estimate $P_x^*(e^{j\omega})$ of the input power spectrum.
Procedure:

1. Let $\Lambda \triangleq [\lambda_{00}\ \lambda_{01}\ \ldots\ \lambda_{(M-1)(L-1)}]^T$.
2. Form the auxiliary functions $g_{i,k}(\Lambda) \triangleq$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|H_i(e^{j\omega})|^2 e^{jN_i k\omega} d\omega}{\sum_{j=0}^{M-1} |H_j(e^{j\omega})|^2 (\sum_{l=0}^{L-1} \lambda_{jl} \cos(N_j \omega))}.$$

3. Find $\Lambda^* \triangleq \arg\min J(\Lambda)$ where

$$J(\Lambda) = \sum_{i=0}^{M-1} \sum_{k=0}^{L-1} \left(g_{i,k}(\Lambda) - \rho_i(k)\right)^2.$$

4. Return

$$P_x^*(e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} |H_i(e^{j\omega})|^2 (\sum_{k=0}^{L-1} \lambda_{ik}^* \cos(N_i \omega))}.$$

The functions $g_{i,k}(\Lambda)$ in Algorithm 3.1 are, in fact, $R_{v_i}(k)$ calculated for a certain value of the parameter Λ. Thus, (3.24) and (3.25) enable one to calculate $g_{i,k}(\Lambda)$ *without integration*! There is a price, however, to be paid for this convenience: We have to know the functional form of the spectral factors $Q(\Lambda; z)$ in order to use (3.24).

From (3.19) we observe that $P_x(\Lambda; e^{j\omega})$ has the form

$$P_x(\Lambda; e^{j\omega}) = \frac{1}{\sum_{i=0}^{M-1} |H_i(e^{j\omega})|^2 F_i(\Lambda; e^{jN_i\omega})}, \tag{3.26}$$

where the parameter vector Λ is embedded in the (non-causal) FIR transfer functions $F_i(\Lambda; z) \stackrel{\triangle}{=} \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$. It is not, in general, true that $P_x(\Lambda; e^{j\omega}) \geq 0$ for all values of Λ. Thus, it is not possible to use standard packages for finding the spectral factors of $P_x(\Lambda; e^{j\omega})$ in its original form given in (3.26) while Λ is being changed for optimization purposes. To circumvent this difficulty, we enforce the extra condition that the transfer functions $F_i(\Lambda; e^{jN_i\omega})$ be *positive definite*. Thus, we force $F_i(\Lambda; z)$ to have the form

$$F_i(\Lambda; z) \stackrel{\triangle}{=} A_i(z) A_i(z^{-1}) \text{ where } A_i(z) \stackrel{\triangle}{=} \sum_{k=0}^{L-1} \lambda_{ik} z^{-k}. \tag{3.27}$$

The above parameterization ensures that $P_x(\Lambda; e^{j\omega}) \geq 0$ for all Λ[3]. Note that the number of parameters remains the same, but (3.27) is not equivalent to the original case where $F_i(\Lambda; z) = \sum_{k=-(L-1)}^{L-1} 2\lambda_{ik} z^{-k}$. Nonetheless, backed by the results of our numerical simulations, we believe that the new parameterization (3.27) has enough degrees of freedom to capture the essence of the original case for most practical applications. We leave it as an open problem to investigate, theoretically, how well this new parameterization approximates the original one and/or find some bounds on the error caused by (possible) lack of enough degrees of freedom.

The steps involved in calculating $g_{i,k}(\Lambda)$ using the alternative method are summarized in Algorithm 3.2 below. Note that several steps in Algorithm 3.2 require operations on infinite impulse-response sequences. Such operations, of course, can only be performed approximately after the sequence is truncated to a reasonable length. Also, the spectral factorization required in Step 6 may be performed, after truncating $e(n)$ to a reasonable length, by the MATLAB function `ac2poly`.

3.7 Simulated examples

Here, we demonstrate the performance of MEIE by providing two simulated examples.

[3] Note that $P_x(\Lambda; e^{j\omega})$ being positive definite is a weaker condition that $F_i(\Lambda; e^{jN_i\omega})$ being positive definite.

Algorithm 3.2: Computing $g_{i,k}(\Lambda)$ without integration

Input:

1. Two natural numbers i and k; $0 \leq i \leq M-1$, $0 \leq k \leq L-1$.
2. The parameter vector $\Lambda \triangleq [\lambda_{00}\ \lambda_{01}\ \ldots\ \lambda_{(M-1)(L-1)}]^T$.
3. Transfer functions $H_i(z)$ and down-sampling ratios N_i.

Output:

The functions $g_{i,k}(\Lambda)$ to be used by Algorithm 3.1.

Procedure:

1. For $0 \leq j \leq M-1$, form $A_j(z) = \sum_{l=0}^{L-1} \lambda_{jl} z^{-k}$.
2. For $0 \leq j \leq M-1$, form $B_j(z) = H_j(z) A_j(z^{N_j})$.
3. For $0 \leq j \leq M-1$, form $C_j(z) = B_j(z) B_j(z^{-1})$.
4. Form $D(z) = \sum_{j=0}^{M-1} C_j(z)$.
5. Find the impulse response $d(n)$ of $D(z)$.
6. Find $e(n)$ such that $d(n) = e(n) \star e(-n)$.
7. Form $E(z) = \dfrac{1}{\sum_{n=0}^{\infty} e(n) z^{-n}}$.
8. Form $U_i(z) = E(z) H_i(z)$.
9. Find the impulse response $u_i(n)$ of $U_i(z)$.
10. Return $g_{i,k}(\Lambda) = \displaystyle\sum_{n=-\infty}^{\infty} u_i(n) u_i(N_i k + n)$.

Example 3.1. Consider a 4-channel multirate observer system of the form shown in Fig. 3.2. Assume that the down-sampling ratio is equal to four for all channels. Thus, $N_0 = N_1 = N_2 = N_3 = 4$. Assume, further, that the linear filters $H_0(z)$ to $H_3(z)$ are given as follows[4].

$$H_0(z) = \frac{0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}}{1.0000 - 0.8877z^{-1} + 0.6738z^{-2} - 0.1206z^{-3} + 0.0225z^{-4}}$$

$$H_1(z) = \frac{0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}}{1.0000 - 0.6855z^{-1} + 0.3297z^{-2} - 0.0309z^{-3} + 0.0032z^{-4}}$$

$$H_2(z) = \frac{0.3732 - 0.8648z^{-1} + 0.7139z^{-2} - 0.1856z^{-3} - 0.0015z^{-4}}{1.0000 - 0.5800z^{-1} + 0.5292z^{-2} - 0.0163z^{-3} + 0.0107z^{-4}}$$

$$H_3(z) = \frac{0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}}{1.0000 + 0.2814z^{-1} + 0.3739z^{-2} + 0.0345z - 3 - 0.0196z^{-4}}$$

[4] The transfer functions used for this simulation example were chosen simply to show low-pass, band-pass and high-pass characteristics. They do not belong to any specific classes of filters (e.g., orthogonal or Perfect Reconstruction filters) used in classical filter bank theory.

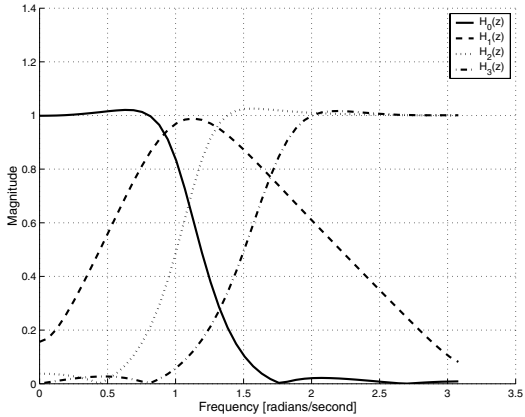


Fig. 3.6. Frequency response of the analysis filters used in Example 3.1.

The frequency response $|H_i(e^{j\omega})|$ associated with each filter is shown in Fig. 3.6.

For simplicity, we chose the input signal $x(n)$ to be a low-pass Gaussian WSS process with modelling filter of the form

$$Q_x(z) = \frac{\sum_{i=0}^{10} a(i) z^{-i}}{\sum_{j=0}^{10} b(j) z^{-j}}.$$

The coefficients $a(i)$ and $b(i)$ were calculated using the MATLAB command

```
[a,b]=YULEWALK( 10, [0 .5 .8 1], [1 1 0 0])
```

which implements the Yule-Walker filter design algorithm. We then calculated the ACS of the observable signals for the above input PSD using (3.4). The results are shown in Table 3.1. The numbers in Table 3.1 serve as the "observed" or "measured" values for the statistics of the low-rate observable signals $v_i(n)$.

We used the MEIE to estimate the input PSD $P_x(e^{j\omega}$ using some or all of the data shown in Table 3.1. The resulting estimated PSDs are shown in Fig. 3.7. As clearly seen in Fig. 3.7, the quality of estimation increases as we provide the MEIE with more data.

Table 3.1. Correlation coefficients used as data in Example 3.1.

$\rho_i(k)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
i=0	0.3536	-0.0649	0.0143	0.0004	-0.0010
i=1	0.3395	-0.0059	0.0052	0.0028	0.0012
i=2	0.2618	0.1180	0.0035	0.0003	0.0024
i=3	0.0947	0.0378	0.0002	0.0003	0.0002

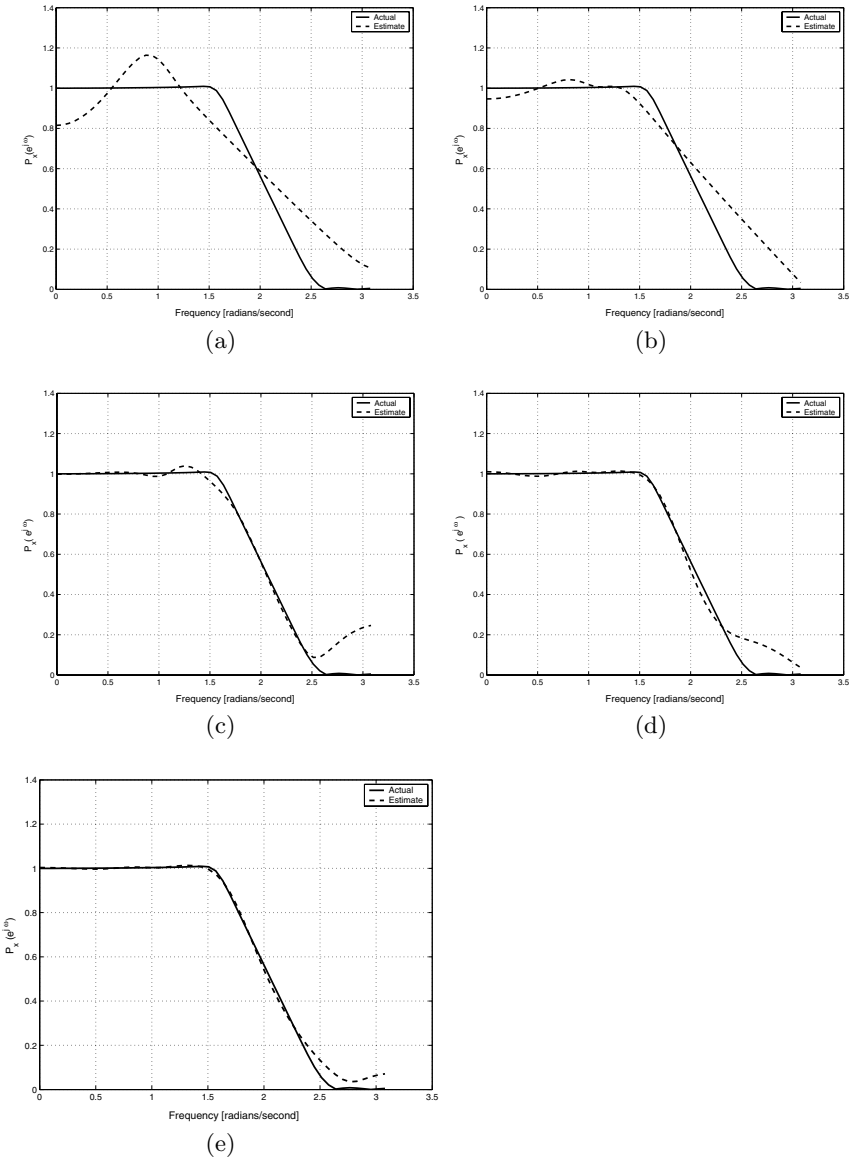


Fig. 3.7. Power spectral densities estimated using only part of the ACS data given in Table 3.1. Each plot shows the estimated PSD using $\rho_i(k)$ data up to and including the kth lag: (a) $k = 0$, (b) $k = 1$, (c) $k = 2$, (d) $k = 3$, (e) $k = 4$. Note how the matching quality increases as k increases.

We should emphasize, however, that increasing the number of data does not necessarily mean better matches because of i) the approximate nature of MEIE and ii) the nature of the optimization problem being solved. Please see Section 3.8 for further discussion on this topic. ♢

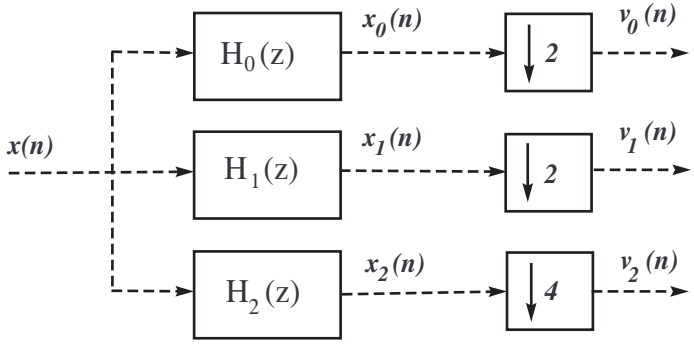


Fig. 3.8. A 3-channel nonuniform multirate observer system.

Example 3.2. Consider the 3-channel multi-rate observer system shown in Fig. 3.8. The down-sampling ratios for this system are $N_0 = 2$, $N_1 = 2$ and $N_2 = 4$. Therefore, it represents a non-uniform over-sampled analysis filter bank. The filters $H_0(z)$ to $H_2(z)$ are chosen to be FIR. The transfer function of each filter is as follows[5].

$$H_0(z) = 0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}$$
$$H_1(z) = 0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}$$
$$H_2(z) = 0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}$$

The frequency response $|H_i(e^{j\omega})|$ associated with each of the above filters is shown in Fig. 3.9. The non-observable input signal $x(n)$ was chosen to be the same low-pass WSS random process described in Example 3.1.

The calculated correlation coefficients associated with the low-rate observable signals $v_i(n)$ are shown in Table 3.2. The PSD estimates obtained by the MEIE using these correlation coefficients are shown in Fig. 3.10.

It's observed that the estimates quickly capture the shape of the actual PSD and approximate it more accurately as the number of input data (i.e., k)

[5] Again, we emphasize that the transfer functions used here were designed simply to show low-pass, band-pass and high-pass characteristics. They do not belong to any specific classes of filters used in classical filter bank theory.

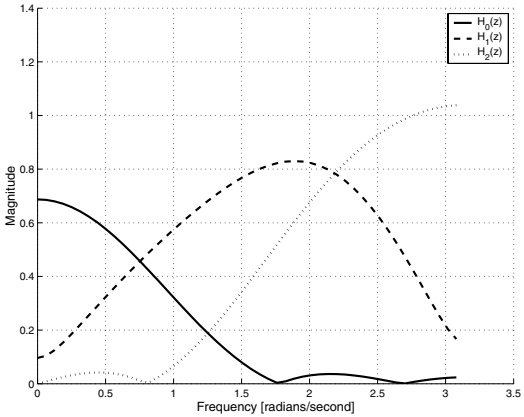


Fig. 3.9. Frequency response of the linear filters used in Example 3.2.

Table 3.2. Correlation coefficients used as data in Example 3.2.

$\rho_i(k)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$i=0$	0.1084	0.0583	0.0056	-0.0001	-0.0000	0.0000
$i=1$	0.1974	-0.1260	0.0423	-0.0169	0.0063	-0.0038
$i=2$	0.0438	0.0141	-0.0002	0.0008	0.0002	0.0001

increases. However, the approximation does not improve uniformly (compare, e.g., Fig. 3.10(d) with Fig. 3.10(e)). ♢

3.8 Complements

3.8.1 Does the estimate converge to the actual spectrum?

In classical Maximum Entropy spectrum estimation problems where the PSD of a given signal is being estimated using a few numbers of its own autocorrelation coefficients, it is possible, at least in principle, to achieve arbitrarily good estimates if a sufficiently large number of the ACS values are used. However, in the multirate case, the estimate $P_x^*(e^{j\omega})$ might or might not converge to the true PSD $P_x(e^{j\omega})$ depending on the nature of the filters $H_i(z)$ and the down-sampling rations N_i.

It is possible for the constraint set Ω to reduce to a single power spectrum as $k \to \infty$ in which case $P_x^*(e^{j\omega})$ will converge to the true PSD $P_x(e^{j\omega})$. However, in general, the set Ω may retain infinitely many elements even when $k \to \infty$. In this case, $P_x^*(e^{j\omega})$ will converge to the PSD in Ω which has the highest entropy and not necessarily to the actual PSD $P_x(e^{j\omega})$.

An investigation of the issues mentioned above leads to interesting topic of the *quantity of information* contained in individual low-rate measurements. This topic is covered in Chapter 7

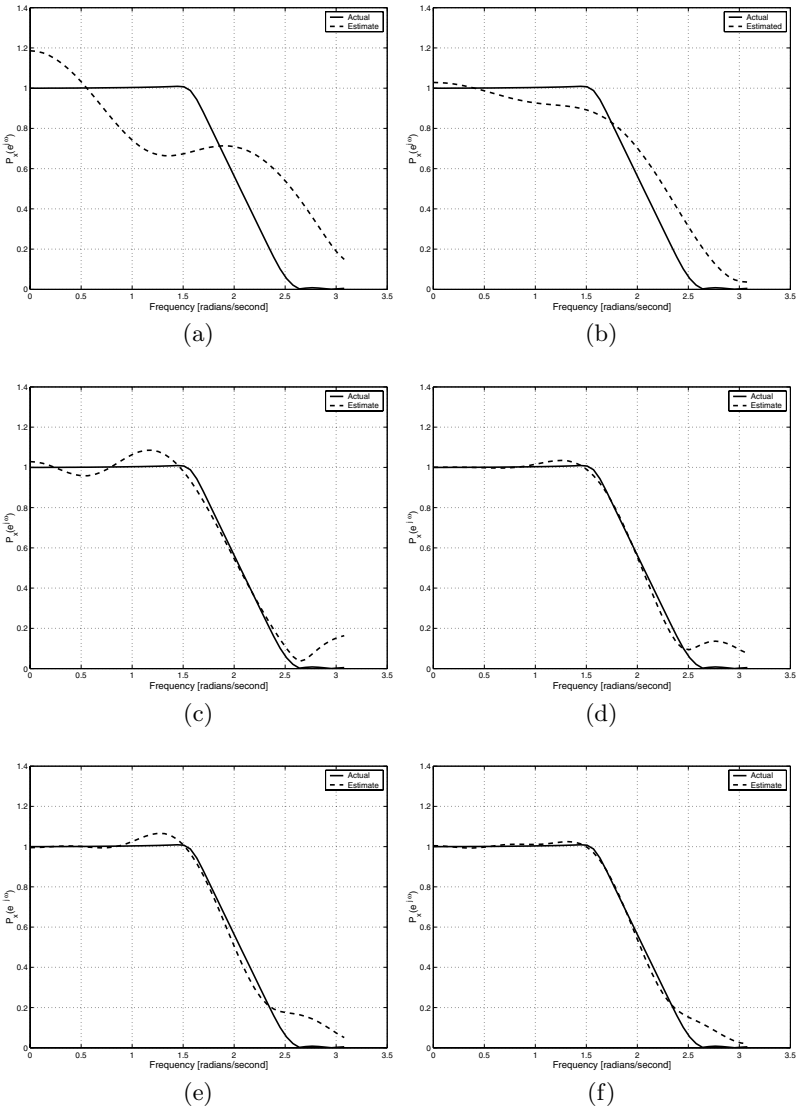


Fig. 3.10. Intermediate and final PSD estimates obtained from the ACS data given in Table 3.2. Each plot shows the estimated PSD using $\rho_i(k)$ data up to and including the kth lag: (a) $k = 0$, (b) $k = 1$, (c) $k = 2$, (d) $k = 3$, (e) $k = 4$, (f) $k = 5$,. Observe that the estimation quality improves (although not uniformly) from (a) to (f).

3.8.2 Why is the cross-correlation information not used?

It is possible to take the cross-correlation between the low-rate measurements into account while estimating the PSD of the high-rate signal. Technically, this would improve the accuracy of the estimates. However, we chose not to incorporate cross-correlation coefficients in our Maximum Entropy inference. There are several reasons for this choice. The most important one is that, in practice, cross-correlation estimates are far more erroneous than autocorrelation estimates[6]. To see why, imagine that we want to estimate the covariance matrix $\mathbf{C}_{XX}$ of a zero mean Gaussian random vector X of dimension N from L realizations $\{X_k\}_{0\leq k<L}$. If we try to do this using the sample-mean estimator

$$\bar{\mathbf{C}}_{XX} = \frac{1}{L}\sum_{k=0}^{L-1} X_k X_k^T, \tag{3.28}$$

it turns out (Mallat, 1999, Page 509) that the expected squared error for each element $[\bar{\mathbf{C}}_{XX}]_{lm}$ of $\bar{\mathbf{C}}_{XX}$ will be

$$E\{([\bar{\mathbf{C}}_{XX}]_{lm} - [\mathbf{C}_{XX}]_{lm})^2\} = \frac{1}{L}(([\mathbf{C}_{XX}]_{lm})^2 + [\mathbf{C}_{XX}]_{ll}[\mathbf{C}_{XX}]_{mm}) \tag{3.29}$$

and the total error variance is given by

$$E\{\|\bar{\mathbf{C}}_{XX} - \mathbf{C}_{XX}\|_{HS}^2\} = \frac{\|\mathbf{C}_{XX}\|_{HS}^2}{L} + \frac{E^2\{\|X\|^2\}}{L} \tag{3.30}$$

where $\|.\|_{HS}^2$ denotes the squared Hilbert-Schmidt norm[7]. The formula (3.29) shows that for off-diagonal elements, $E\{([\bar{\mathbf{C}}_{XX}]_{lm} - [\mathbf{C}_{XX}]_{lm})^2\}$ depends not only on $([\mathbf{C}_{XX}]_{lm})^2$ but also on the amplitude of the diagonal coefficients $[\mathbf{C}_{XX}]_{ll}$ and $[\mathbf{C}_{XX}]_{mm}$. Thus, even though $[\mathbf{C}_{XX}]_{lm}$ may be small, the estimation error is large if the diagonal coefficients are large:

$$E\{([\bar{\mathbf{C}}_{XX}]_{lm} - [\mathbf{C}_{XX}]_{lm})^2\} \geq \frac{[\mathbf{C}_{XX}]_{ll}[\mathbf{C}_{XX}]_{mm}}{L}. \tag{3.31}$$

Other reasons for ignoring the statistical information contained in cross-correlation coefficients include tremendous increase in the complexity of the MEIE and the necessity of synchronization between the measurement signals $v_i(n)$.

3.9 Open problems

In this chapter we used the Maximum Entropy principle to solve the ill-posed problem of multirate spectrum estimation. We formulated a parameterized

[6] Please see (Mallat, 1999, Section 10.6) for more details on this subject.

[7] For a matrix A we have $\|A\|_{HS}^2 \triangleq Tr(AA^T)$.

solution and introduced a semi-numerical method to calculate the parameters. We also analyzed the existence, uniqueness and stability of the solution.

While uniqueness of the Maximum Entropy solution is guaranteed, the issues of existence and stability are not resolved completely. A complete and rigorous analysis of the stability of the Maximum Entropy solution is an open research topic.

Another open problem is how to design an efficient and stable numerical procedure to compute the solution. We presented an ad-hoc algorithm for computing an approximate solution. However, this algorithm is neither the most accurate nor the most efficient algorithm that can be devised. For instance, our algorithm requires the use of some external optimization package which would consume considerable computational resources. This, and other computationally demanding steps make our algorithm unfit for large-scale or real-time applications.

4

Multirate Signal Estimation

4.1 Introduction

In this chapter we consider the problem of estimating samples of a random process $x(n)$ given a set of low-rate measurements $v_i(n)$ derived from it. We use the multirate filter bank model shown in Fig. 4.1 to formulate the problem.

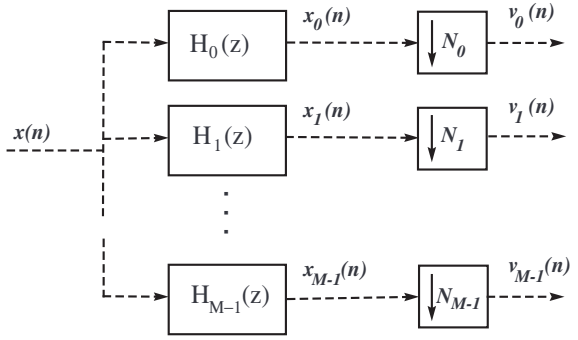


Fig. 4.1. An analysis filter bank can be used as a model for linear multirate measurement system.

In our model, $x(n)$ represents the non-observable signal under measurement. It is assumed that this signal is a zero-mean Gaussian wide-sense stationary (WSS) random process. The linear filters $H_i(z)$ model the bandwidth characteristics of each sensor. Each filter is followed by a down-sampling device which models the (possible) difference in sampling rate among the sensors. We assume that the filters $H_0(z)$ to $H_{M-1}(z)$ and the down-sampling ratios N_0 to N_{M-1} are known.

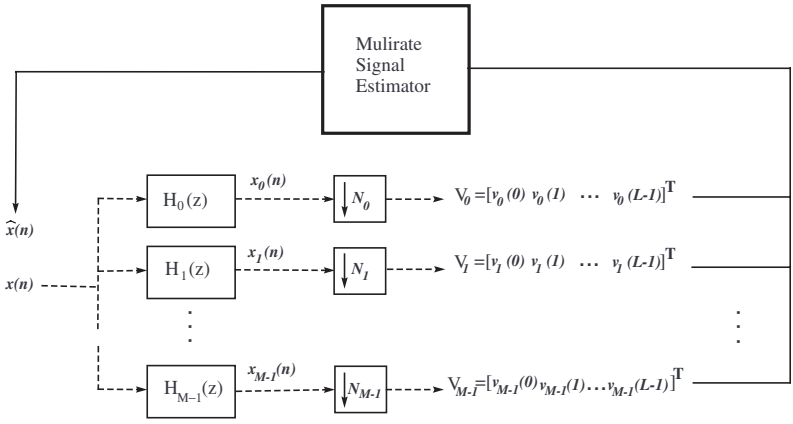


Fig. 4.2. A multirate signal estimator estimates a specified number of input signal's samples from a specified number of low-rate observations.

A multirate signal estimator is a function that calculates a unique estimate $\hat{x}(n)$ of $x(n)$ from the data $v_i(n)$ (Fig. 4.2). A popular way to specify an estimator function is to use the least-squares formulation. In subsequent sections, we show that the optimal least-squares estimator for samples of $x(n)$ is a linear operator (matrix) whose coefficients depend on the spectrum $P_x(e^{j\omega})$ of the non-observable signal $x(n)$. In practical applications, the exact functional form of $P_x(e^{j\omega})$ is often unknown. To get around this difficulty, we suggest using a Maximum Entropy estimate of $P_x(e^{j\omega})$ — which can be obtained from the low-rate signals $v_i(n)$ — in the least-squares estimation formulae. We show through simulated examples that the ad-hoc least-squares estimator obtained this way is still very viable. The chapter ends with a discussion of open problems and directions for future research.

4.2 Stochastic least-square estimation

4.2.1 Problem formulation

Let us assume that L samples for each signal $v_i(n)$ are known. We can pack these values in a column vector called V_i:

$$V_i \triangleq [v_i(0), v_i(1), \cdots, v_i(L-1)]^T.$$

Then, we can stack all the V_i, $i = 0, 1, \ldots, M-1$, together and form the long observation vector

$$V \triangleq [V_0^T \; V_1^T \; \cdots \; V_{M-1}^T]^T.$$

Similarly, we define

$$X \triangleq [x(Q), x(Q+1), \cdots, x(Q+P-1)]^T$$

to contain those samples $x(n)$ that we want to estimate. Here, the constant P represents the number of samples to be estimated and Q represents the time index where the input samples to be estimated start.

Problem 4.1 (Multirate Signal Estimation). Given the filter bank model in Fig. 4.1, the observation vector V and the input power spectrum $P_x(e^{j\omega})$, find an estimate $\hat{X}$ of X such that $E\{||X - \hat{X}||\}$ is minimum.

4.2.2 Solution

It is well-known that the optimal least squares estimator of a random vector X given the value of a correlated random vector V is the conditional expectation of X given V. That is, $\hat{X} = E\{X|V\}$. In general, the optimal least-square estimate $\hat{X}$ is a nonlinear function of the data V. But, luckily, it becomes a linear function of V in the important special case where the random variables X and V are jointly Gaussian. An introduction to the theory of stochastic least-squares estimation can be found in many texts on statistical signal processing. See, for example, Papoulis (1991) and Kailath et al. (2000).

Let X and V be a pair of random vectors with a jointly Gaussian distribution and zero mean. Let $\mathbf{C}_{XX}$ and $\mathbf{C}_{VV}$ represent the autocorrelation matrices associated with X and V, respectively, and let $\mathbf{C}_{XV}$ denote their cross-correlation matrix. Assume, further, that the augmented covariance matrix $\mathbf{R}$ constructed as

$$\mathbf{R} \triangleq \begin{bmatrix} \mathbf{C}_{XX} & \mathbf{C}_{XV} \\ \mathbf{C}_{XV}^T & \mathbf{C}_{VV} \end{bmatrix} \tag{4.1}$$

is non-singular. Under this assumption, it is straightforward to show that the optimal least-square estimate of X given V is

$$\hat{X} = \mathbf{F}V$$

where

$$\mathbf{F} \triangleq \mathbf{C}_{XV}\mathbf{C}_{VV}^{-1}. \tag{4.2}$$

Let $\mathbf{e} \triangleq X - \hat{X}$ represent the estimation error vector associated with the optimal least-squares estimator $\mathbf{F}$. The covariance matrix of $\mathbf{e}$ can be calculated as

$$\mathbf{C_{ee}} = \mathbf{C}_{XX} - \mathbf{C}_{XV}\mathbf{C}_{VV}^{-1}\mathbf{C}_{XV}^T. \tag{4.3}$$

The optimality of $\mathbf{F}$ means that $E\{||\mathbf{e}||^2\}$ —which is given by the trace of $\mathbf{C_{ee}}$ — is minimum.

4.3 More on linear least-squares estimation

A significant feature of the least-squares formulation is that it regularizes the *ill-posed* problem of calculating X from V. This is to say, the solution $\hat{X}$ provided by the least square formulation is unique and stable.

Another important property of the least-squares estimator matrix $\mathbf{F}$ is that it minimizes not only the overall expected estimation error $E\{||\mathbf{e}||\}$ but also the estimation error associated with each individual component of X! This means, if we define $e(n) \stackrel{\triangle}{=} x(n) - \hat{x}(n)$ for $Q \leq n \leq Q + P - 1$, then

$$E\{e(n)^2\} = [\mathbf{C_{ee}}]_{mm},\ m \stackrel{\triangle}{=} n - Q + 1, \tag{4.4}$$

are minimized individually. Note, however, that the expected estimation error $E\{e(n)^2\}$ depends on n and may vary for different samples of $x(n)$. This is not a very desirable property but, fortunately, $E\{e(n)^2\}$ can be specified a priori if the input signal's power spectrum is known.

In theory, one can try to estimate thousands of samples of $x(n)$ given a few low-rate data samples $v_i(n)$. However, the accuracy of the individual estimated samples $\hat{x}(n)$ — measured by $E\{e(n)^2\}$ — depends on the actual amount of correlation between each of these samples and the measured low-rate data. As the time lag between the available samples of $v_i(n)$ and those samples of $x(n)$ which are being estimated increases, the amount of correlation between these samples fades out. As a result, the expected estimation error $E\{e(n)^2\}$ increases to the same level as $E\{x(n)^2\}$, rendering the estimates useless at large time lags.

The interested reader is referred to (Kailath et al., 2000, Chapter 3) for a general discussion of the properties of linear least-squares estimators.

4.4 Computing the estimator matrix

In order to compute the least-square estimator $\mathbf{F}$ and the associated error variance estimates $E\{e(n)^2\}$, we need the covariance matrices $\mathbf{C}_{XX}$, $\mathbf{C}_{VV}$ and $\mathbf{C}_{XV}$. In this section we present a method to calculate these covariance matrices given $P_x(e^{j\omega})$ and the filter bank model shown in Fig. 4.1.

The computational method presented in this section is not the simplest way to calculate $\mathbf{F}$ nor is it the most computationally efficient. It is, nonetheless, simple enough to understand and can be readily implemented in MATLAB. Devising more elaborate ways for calculating $\mathbf{F}$ is open for research.

Calculating $\mathbf{C}_{XX}$:

Calculating $\mathbf{C}_{XX}$ is very simple. Let $P_x(e^{j\omega})$ denote the power spectrum of $x(n)$. The autocorrelation coefficients $R_x(k)$ associated with $x(n)$ are given by

$$R_x(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})e^{jk\omega}\mathrm{d}\omega. \tag{4.5}$$

The $P \times P$ matrix $\mathbf{C}_{XX}$ is then formed as $[\mathbf{C}_{XX}]_{mn} = R_x(m-n)$.

Calculating $\mathbf{C}_{VV}$:

The observation signals $v_i(n)$ are obtained from $x(n)$ by a filtering operation followed by down-sampling. It is straightforward to show that the autocorrelation coefficients of $v_i(n)$ are given by

$$R_{v_i}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})|H_i(e^{j\omega})|^2 e^{jkN_i\omega}\mathrm{d}\omega, \tag{4.6}$$

which enables us to construct $[\mathbf{C}_{V_iV_i}]_{ml} = R_{v_i}(m-l)$. Next, we calculate the intra-channel cross-correlation matrices $\mathbf{C}_{V_iV_j}$ associated with V_i and V_j from

$$[\mathbf{C}_{V_iV_j}]_{ml} = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})H_i(e^{j\omega})H_j^*(e^{j\omega})e^{j(mN_i-lN_j)\omega}D\omega. \tag{4.7}$$

The covariance matrix of the long observation vector V is formed by putting the autocorrelation matrices $\mathbf{C}_{V_iV_i}$ and the cross-correlation matrices $\mathbf{C}_{V_iV_j}$ together:

$$\mathbf{C}_{VV} \triangleq \begin{bmatrix} \mathbf{C}_{V_0V_0} & \cdots & \mathbf{C}_{V_0V_{M-1}} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{V_{M-1}V_0} & \cdots & \mathbf{C}_{V_{M-1}V_{M-1}} \end{bmatrix}. \tag{4.8}$$

Calculating $\mathbf{C}_{XV}$:

To calculate $\mathbf{C}_{XV}$, we start by calculating $\mathbf{C}_{XV_i} \triangleq E\{XV_i^T\}$ which represent the cross-correlation between X and V_i. Clearly,

$$[\mathbf{C}_{XV_i}]_{ml} = E\{x(Q+m-1)v_i(l-1)\}. \tag{4.9}$$

Let $h_i(k)$ denote the impulse response sequence of the filters $H_i(z)$ in the filter bank model shown in Fig. 4.1. This model implies that an expression of the form $E\{x(p)v_i(q)\}$ can be expanded as

$$\begin{aligned} &E\{x(p)v_i(q)\} \\ &= E\{x(p)x_i(N_iq)\} \\ &= E\left\{x(p)\left(\textstyle\sum_{k=-\infty}^{\infty} h_i(k)x(N_iq-k)\right)\right\} \\ &= \textstyle\sum_{k=-\infty}^{\infty} h_i(k)E\left\{x(p)x(N_iq-k)\right\} \\ &= \textstyle\sum_{k=-\infty}^{\infty} h_i(k)R_x(p-N_iq+k). \end{aligned} \tag{4.10}$$

Combining the above expansion and (4.9), we get

$$[\mathbf{C}_{XV_i}]_{ml} = \sum_{k=-\infty}^{\infty} h_i(k)R_x([Q+m-1-N_i(l-1)]-k). \tag{4.11}$$

Finally, the covariance matrix $\mathbf{C}_{XV}$ is constructed as

$$\mathbf{C}_{XV} \stackrel{\triangle}{=} [\mathbf{C}_{XV_0}, \mathbf{C}_{XV_1}, \cdots, \mathbf{C}_{XV_{M-1}}]. \tag{4.12}$$

Summary:

The necessary steps for calculating the least-square estimate $\hat{X}$ from the data vectors V_i are summarized in the algorithm below.

Estimates $x(n)$ given $v_i(n)$ and $P_x(e^{j\omega})$
Input:
1. $v_i(n)$ for $i = 0, \cdots, M-1$ and $n = 0, \cdots L-1$. 2. Transfer functions $H_i(z)$ and down-sampling ratios N_i for $i = 0, \cdots, M-1$. 3. $P_x(e^{j\omega})$ or, equivalently, $R_x(k)$. 4. The numbers Q (integer) and P (natural).
Output:
The vector $\hat{X}$ containing estimated values of $x(n)$ for $n = Q, Q+1, \cdots, Q+P-1$.
Procedure:
1. Calculate $\mathbf{C}_{V_iV_j}$ for $0 \le i, j \le M-1$ using (4.7) and then form $\mathbf{C}_{VV}$ using (4.8). 2. Calculate $\mathbf{C}_{XV_i}$ for $0 \le i \le M-1$ using (4.11) and then form $\mathbf{C}_{XV}$ using (4.12). 3. Calculate the estimator matrix $\mathbf{F} = \mathbf{C}_{XV}\mathbf{C}_{VV}^{-1}$. 4. Form the vectors $V_i = [v_i(0), \cdots, v_i(L-1)]$ and combine V_i together to form $V = [V_0^T \ \cdots \ V_{M-1}^T]^T$. 5. Calculate the estimate $\hat{X} = \mathbf{F}V$.

Once samples of $x(n)$ are estimated, the expected accuracy of estimation for each sample can be assessed by calculating the expected error variance $E\{e(n)^2$ associated with each sample. This is done by first computing

$$\mathbf{C_{ee}} = \mathbf{C}_{XX} - \mathbf{C}_{XV}\mathbf{C}_{VV}^{-1}\mathbf{C}_{XV}^T$$

and then calculating

$$E\{e(n)^2\} = [\mathbf{C_{ee}}]_{mm}$$

where $m = n - Q + 1$ and $Q \le n \le Q + P - 1$.

4.5 Simulated examples

In this section, we present two simulated examples to demonstrate the performance of the least-squares estimator derived in the above sections.

Remark 4.1. The computational procedures introduced in the previous section sometimes require operations on infinite length sequences. Such operations can only be performed approximately after the sequence is truncated to a reasonable length. Thus, in practice, the matrices $\mathbf{C}_{XV}$, $\mathbf{C}_{XX}$ and $\mathbf{C}_{VV}$ are not obtained exactly. This, in turn, might cause numerical difficulties in calculating $\mathbf{C_{ee}} = \mathbf{C}_{XX} - \mathbf{C}_{XV}\mathbf{C}_{VV}^{-1}\mathbf{C}_{XV}^{T}$ and $\mathbf{F} = \mathbf{C}_{XV}\mathbf{C}_{VV}^{-1}$. To circumvent such numerical we used generalized (Moore-Penrose) inverses when calculating $\mathbf{C_{ee}}$ and $\mathbf{F}$.

Example 4.1. Consider a 4-channel multirate measurement system where a discrete-time stochastic signal $x(n)$ is observed using linear sensors whose sampling rate is only 25% of the sampling rate associated with $x(n)$. This measurement setup can be modelled using the filter bank model shown in Fig. 4.1.

We use a Gaussian random process with a low-pass power spectrum as shown in Fig. 4.4 as input and assume that the sensors are modelled using linear filters whose transfer functions $H_0(z)$ to $H_3(z)$ are as follows:

$$H_0(z) = \frac{0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}}{1.0000 - 0.8877z^{-1} + 0.6738z^{-2} - 0.1206z^{-3} + 0.0225z^{-4}}$$

$$H_1(z) = \frac{0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}}{1.0000 - 0.6855z^{-1} + 0.3297z^{-2} - 0.0309z^{-3} + 0.0032z^{-4}}$$

$$H_2(z) = \frac{0.3732 - 0.8648z^{-1} + 0.7139z^{-2} - 0.1856z^{-3} - 0.0015z^{-4}}{1.0000 - 0.5800z^{-1} + 0.5292z^{-2} - 0.0163z^{-3} + 0.0107z^{-4}}$$

$$H_3(z) = \frac{0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}}{1.0000 + 0.2814z^{-1} + 0.3739z^{-2} + 0.0345z{-}3 - 0.0196z^{-4}}$$

The amplitude response $|H_i(e^{j\omega})|$ of these filters are shown in Fig. 4.3. Since the sensors have a sampling rate only 25% of that associated with $x(n)$, the down-sampling rate in the filter bank model of Fig. 4.1 is equal to four for all channels ($N_0 = N_1 = N_2 = N_3 = 4$). We assume that 6 samples of the observable signals $v_i(n)$ from $n = 1$ to $n = 6$ are available. These samples are shown in Fig. 4.5.

We designed a least-squares estimator to estimate 31 samples of the non-observable input signal $x(n)$ from $n = -10$ to $n = 30$. The estimation results are shown in Fig. 4.6(a). Fig. 4.6(b) shows the error associated with these estimated values. As seen in this figure, the estimation accuracy is not uniform: the estimates for $-3 \leq n \leq 20$ are very accurate while those obtained for $-10 \leq n \leq -4$ are not.

Fig. 4.6(c) shows $E\{e(n)^2\}$ i.e. the expected error of individual estimated samples. It is clear from this graph that $E\{e(n)^2\}$ is a very good predictor of the actual estimation error. ♢

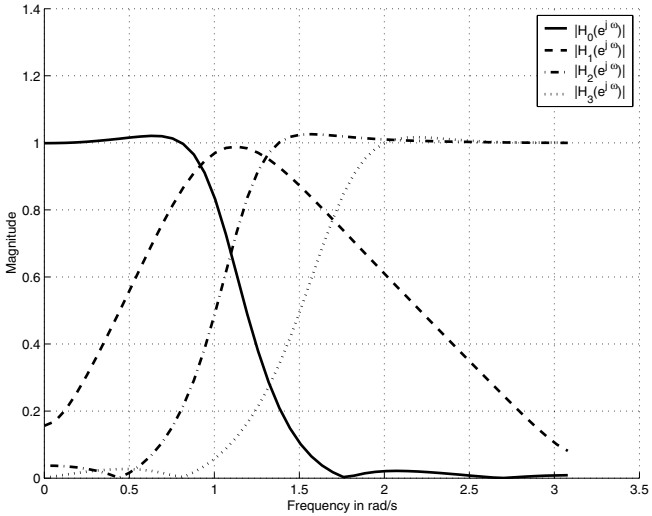


Fig. 4.3. Frequency response of the multirate observer filters used in Example 4.1.

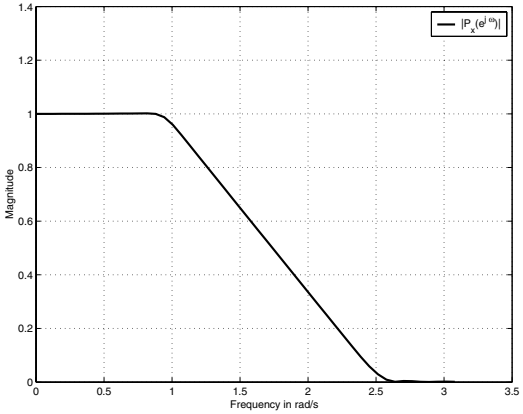


Fig. 4.4. Power spectrum of the ARMA$(10, 10)$ random signal used as input in Example 4.1.

Example 4.2. In this example we present an over-sampled multirate measurement scenario. Consider a measurement set up consisting of four linear sensors where the sensors can be modelled by linear filters $H_0(z)$ to $H_3(z)$ given below:

$$H_0(z) = \frac{0.1}{1.0000 - 0.9577z^{-1}}, \quad H_1(z) = \frac{0.1}{1.0000 - 0.8855z^{-1}},$$
$$H_2(z) = \frac{0.1}{1.0000 - 0.6800z^{-1}}, \quad H_3(z) = \frac{0.1}{1.0000 - 0.3140z^{-1}}.$$

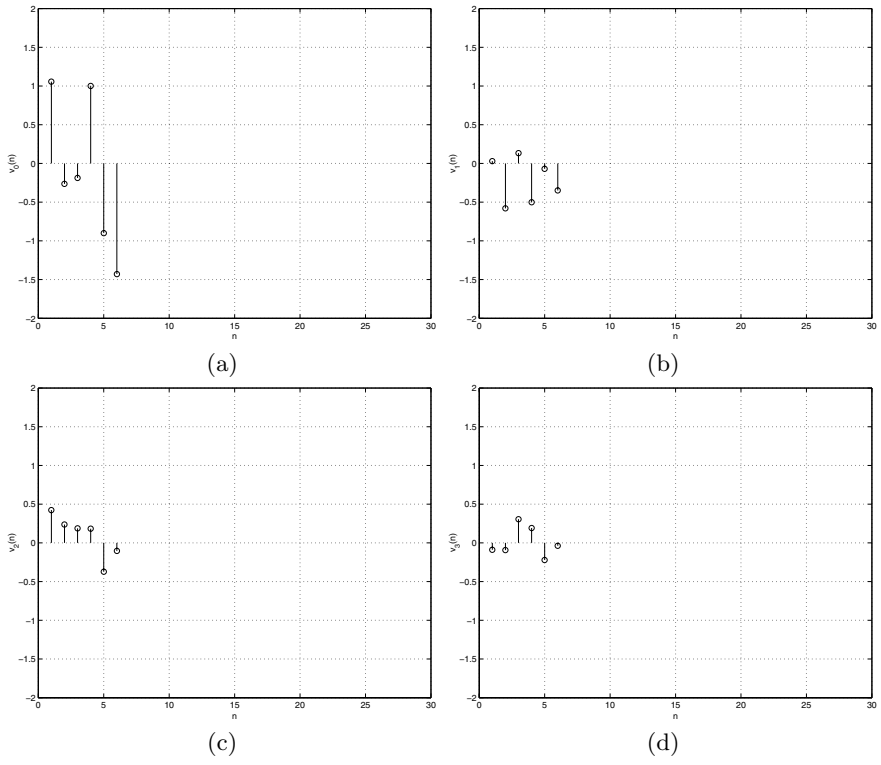
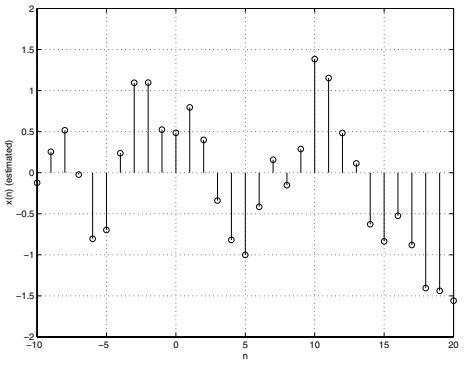


Fig. 4.5. Low-rate measurements $v_i(n)$, $1 \leq n \leq 6$ produced by the multirate observer system described in Example 4.1.

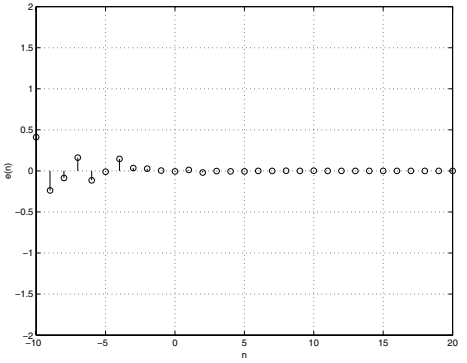
The frequency response amplitude $|H_i(e^{j\omega})|$ for these filters is shown in Fig. 4.7. Note that these filters are all low-pass! We Assume that each sensor (observer) outputs its data at a sampling rate which is 33% less than the sampling rate of $x(n)$. This means the down-sampling ratio N_i is equal to 3 for all channels. The non-observable full-rate signal $x(n)$ is a WSS random process with power spectrum shown in Fig. 4.8. The observed low-rate data (i.e. available samples of $v_i(n)$) are shown in Fig. 4.9.

We designed an optimal least-squares estimator to estimate 26 samples of $x(n)$ from the observed $v_i(n)$ samples shown in Fig. 4.9. The result is shown in Fig. 4.10(a). The estimation error is plotted in Fig. 4.10(b).

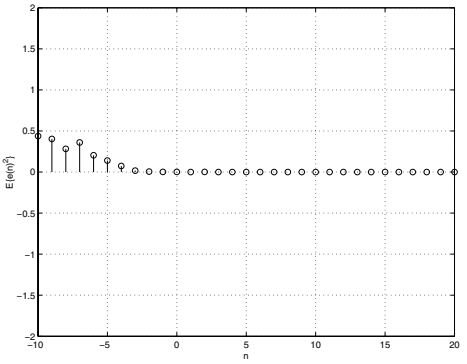
It is seen from the error plot that the estimated input values for $1 \leq n \leq 20$ are very accurate while those for $-5 \leq n \leq 0$ are not. As in the previous example, the accuracy of individual estimated samples can be reliably predicted from the expected error norm figures shown in Fig. 4.10(c). $\diamondsuit$



(a)



(b)



(c)

Fig. 4.6. Estimation results for the 4-channel multirate observer system described in Example 4.1. (a) Estimated samples of the full-rate input signal $x(n)$. (b) Estimation error. (c) Expected norm of estimation error calculated *a priori* using the trace of $\mathbf{C_{ee}}$. Note how well the expected error norm predicts the actual error magnitude.

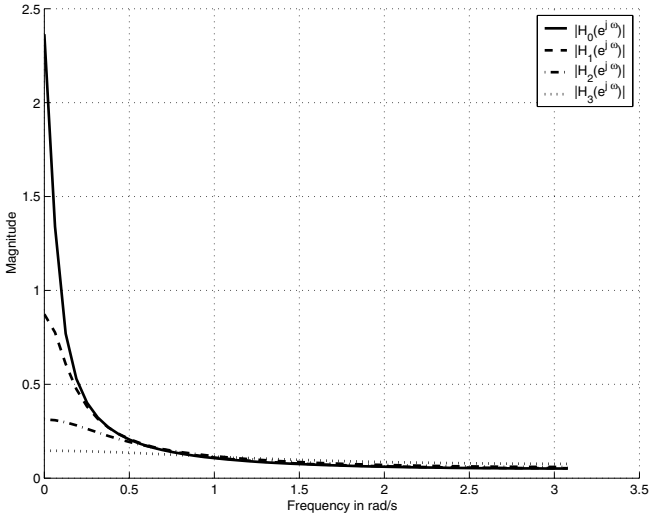


Fig. 4.7. Frequency response of the filters associated with the multirate observer system described in Example 4.2.

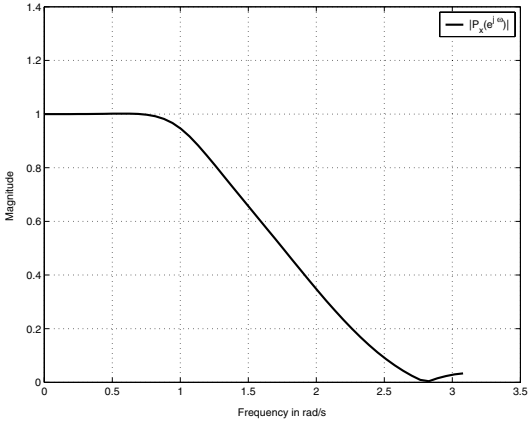


Fig. 4.8. Power spectrum of the random signal used as input in Example 4.2.

4.6 Multirate least-squares estimation in practice

In order to calculate the optimal least-squares estimator $\mathbf{F}$, we need to know the power spectrum of $x(n)$. In practice, the power spectrum of $x(n)$ is not known a priori so we have to somehow *estimate* it from the available observations $v_i(n)$.

In Chapter 3 we introduced an algorithm to estimate $P_x(e^{j\omega})$ from low-rate *auto-correlation coefficients*. Auto-correlation coefficients associated with the low-rate signals $v_i(n)$ are probabilistic quantities. As we discussed in

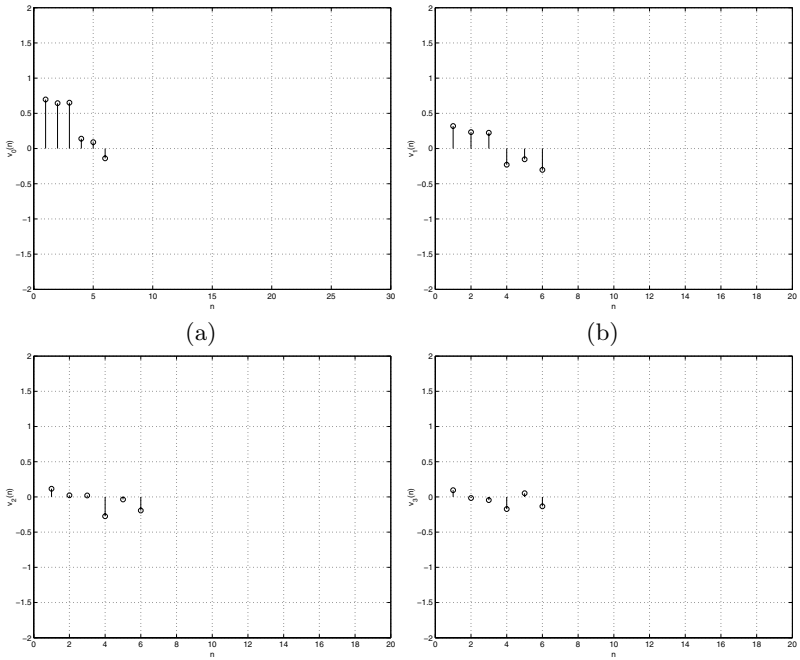


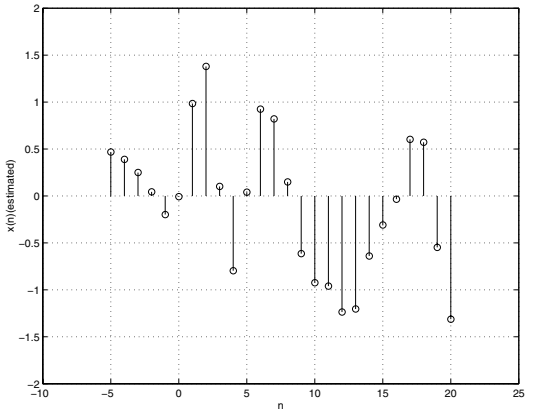
Fig. 4.9. Low-rate measurement signals $v_i(n)$ used in Example 4.2.

Chapter 2, the theory of probability does not provide a method to construct probabilistic measures such as power spectrum or autocorrelation coefficients from observed sample values of a random process. As a result, we have to resort to some practical trick. A widely used trick is to use appropriate time-averages as estimates for autocorrelation coefficients. This is to say, we may use

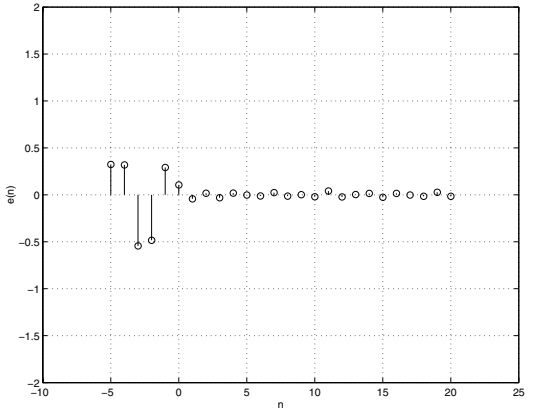
$$\rho_i(k) \stackrel{\triangle}{=} \frac{\sum_{n=0}^{L-k-1} v_i(n)v_i(n+k)}{L-k-1},\ 0 \le k \le L-1, \tag{4.13}$$

as estimates for the actual autocorrelation coefficients $R_i(k)$. Using these estimates, we can summarize our practical procedure for estimating *sample values* of $x(n)$ from *sample values* of $v_i(n)$ as follows.

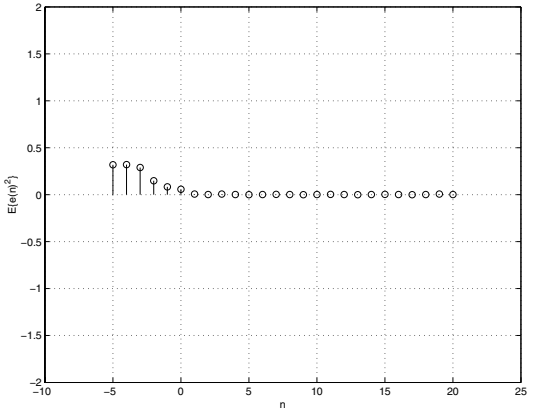
(i) Estimate a few autocorrelation coefficients $R_i(k)$ from available samples of $v_i(n)$ using the time average method mentioned above.
(ii) Use the Maximum Entropy algorithm described in Chapter 3 to estimate $P_x(e^{j\omega})$ from the autocorrelation coefficients estimated in the previous step.
(iii) Feed $P_x(e^{j\omega})$ obtained in the previous step into Algorithm 5.1 described in this chapter and estimate the desired samples of $x(n)$.



(a)



(b)



(c)

Fig. 4.10. (a) Estimated values of the full-rate signal $x(n)$ obtained using the low-rate measurement system described in Example 4.2. (b) Estimation error. (c) Expected norm of the estimation error.

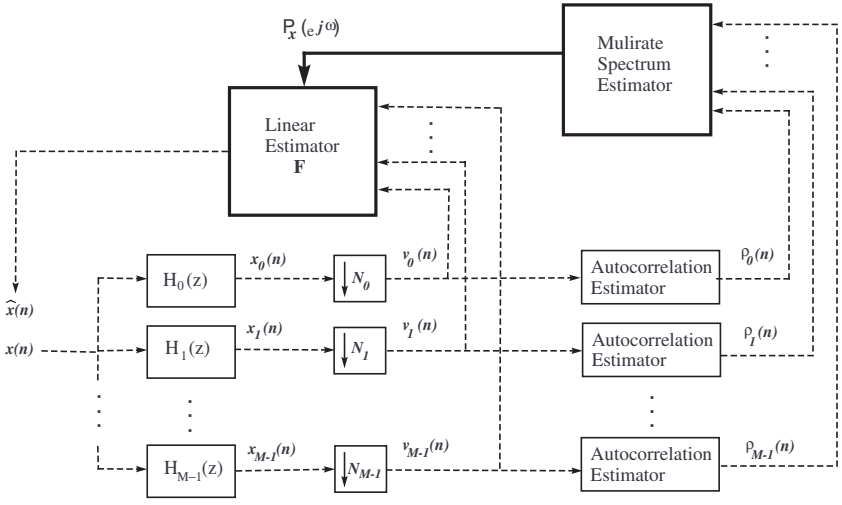


Fig. 4.11. A practical procedure for estimating the full-rate signal $x(n)$ from low-rate measurements $v_i(n)$.

A block-diagram representation of this procedure is shown in Fig. 4.11. Note that when the power spectrum of $x(n)$ is known, the linear estimator **F** is optimal in the sense of minimizing the mean-squared-error between the estimated values of $x(n)$ and the actual ones. However, it is not possible to make this optimality claim when $P_x(e^{j\omega})$ is itself estimated. Therefore, in practice, one must judge the viability of the three-step estimation procedure outlined above based on the actual results it produces for a particular application. The following example illustrates this point.

Example 4.3. Let's consider the 4-sensor multirate observer scenario described in Example 4.1 and try to estimate $P_x(e^{j\omega})$ from the observed data. To do this, we can take the available low-rate data (6 samples of $v_i(n)$ per channel, shown in Fig. 4.12) and put them in (4.13) to estimate the first two autocorrelation coefficients $R_i(0)$ and $R_i(1)$ for each channel. We can then pass these estimated autocorrelation coefficients on to the spectrum estimation algorithm described in Chapter 3 to estimate $P_x(e^{j\omega})$. The resulting power spectrum is shown in Fig. 4.13. The actual input power spectrum is shown in the same figure as well. Clearly the two spectra are different but, given the small number of available data, our estimation has captured the band-pass nature of the input spectrum fairly well.

Now, let's use this estimated spectrum to estimate samples of $x(n)$ for $-15 \leq n \leq 15$. The resulting estimates are shown in Fig. 4.14(b). The estimation error is depicted in Fig. 4.14(d). For comparison, the estimates which would have been obtained if we had used the correct power spectrum are shown in Fig. 4.14(a). The resulting error values for this case are shown in

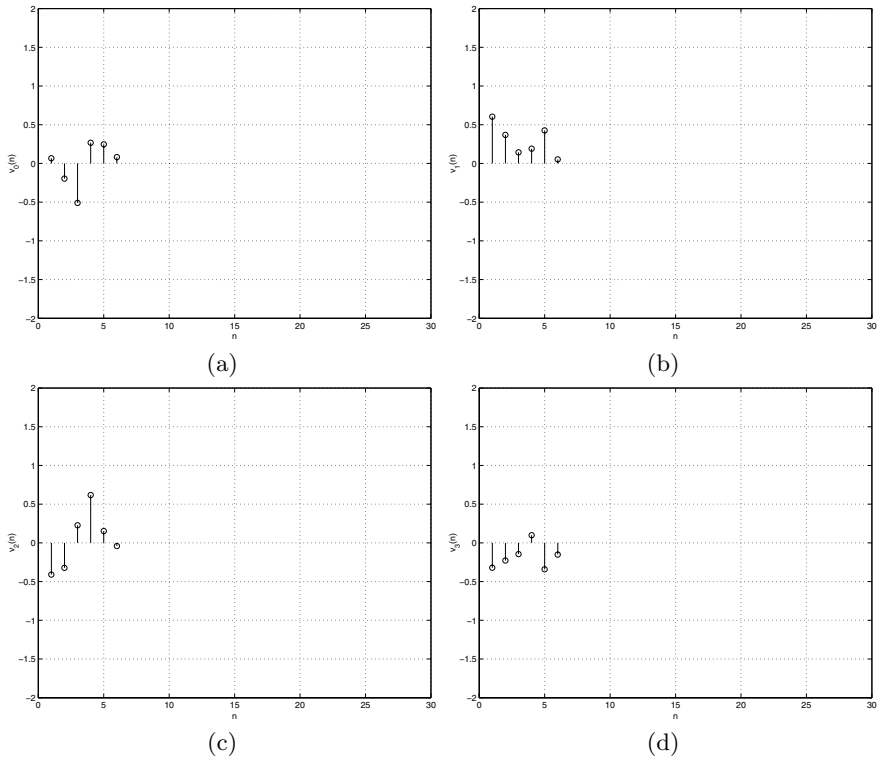


Fig. 4.12. Low-rate measurement data $v_0(n)$ to $v_3(n)$ used in Example 4.3.

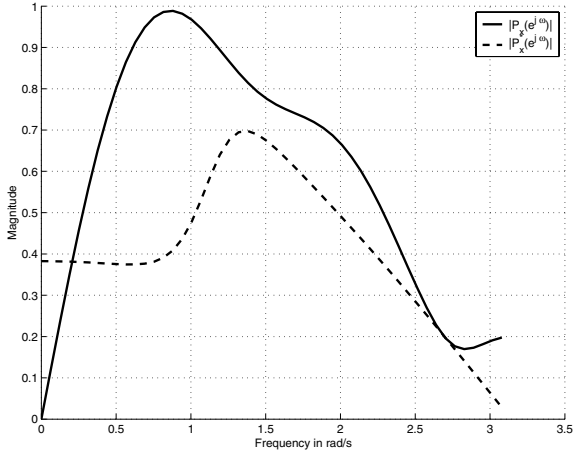


Fig. 4.13. Power spectral density of the full-rate signal $x(n)$ used in Example 4.3. The solid curve shows the actual PSD. The dashed curve shows the PSD estimated from available samples of the low-rate measurements $v_i(n)$.

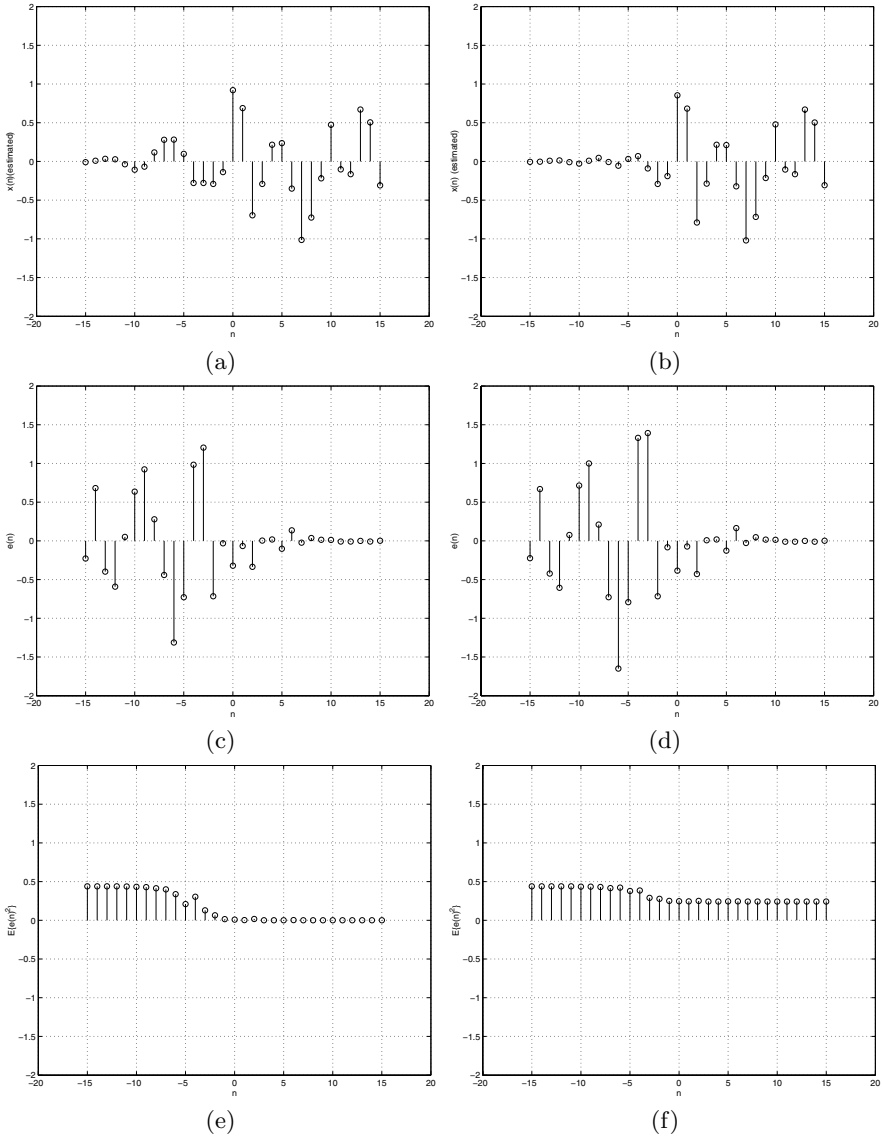


Fig. 4.14. (a),(b): Estimates of the full-rate signal $x(n)$ for $-15 \leq n \leq 15$ for the observer systems described in Example 4.3. (c), (d): Estimation error associated with the estimated values shown in (a) and (b). (e), (f): The expected norm of estimation error for each sample. In each row, the plot on the left-had side corresponds to the case where the power spectrum of $x(n)$ was known. The plots on the right had side correspond to the case where the full-rate power spectrum was estimated from the measured data.

Fig. 4.14(c). It is clear that the estimation error would have been slightly reduced if we have had access to the actual power spectrum. However, the extra error due to discrepancy between the actual input spectrum and the estimated one is not very large. This shows that the three-step estimation procedure suggested in this section is in fact viable.

It is instructive to compare the expected error norm figures calculated based on the true spectrum and the estimated one. These figures are plotted in Fig. 4.14(e) and Fig. 4.14(f). It is seen from the plots that the estimated error norm values are no longer an accurate indicator of the actual estimation error. This is justified because the expected error norms given in (4.3) and (4.4) depend on $P_x(e^{j\omega})$. The interesting observation is that actual errors are consistently less than what their expected value suggests! ♢

4.7 Open problems

Estimating the samples of a non-observable high-rate signal from low-rate measurements is an ill-posed inverse problem. In this chapter we opted for a statistical formulation to find a well-posed solution to this problem. The stochastic least-squares formulation is a very viable engineering solution in the sense that it can be applied, with appropriate modifications, to many real world engineering applications. However, the computational procedure introduced to calculate the least-squares solution was intended only as a "proof of concept". It is not suited for large-scale real-world signal estimation applications due to its very high computational complexity and lack of scalability. Finding practical numerical methods for signal estimation in multirate systems remain an open challenge.

In general, a high-rate signal can be reconstructed from low-rate observed data by finding a left-inverse for the linear operator that maps the high-rate samples to the low-rate data. The least-squares estimator is one possible left-inverse among a (usually infinite) number of valid solutions. It would be interesting to study other viable optimization criteria whereby alternative left-inverse operators might be selected.

5

Multirate Time-Delay Estimation

5.1 Introduction

Time-delay estimation is a key step in many sensor-array signal processing applications. Most notable among these applications are source localization, direction of arrival estimation and multi-sensor data fusion.

In this chapter, we will study time-delay estimation in multirate sensing systems. We will address questions such as "How to extend conventional time-delay estimation techniques to multirate signals"? and "How to design a multirate sensor system to allow signal fusion when time-delay is present?"

For simplicity, we will pose and solve these problems by reference to a basic model involving only two multirate sensors. The problem of multirate time-delay estimation when more than two sensors are present remains an open problem for research. The material in this chapter are based on the paper by Jahromi and Aarabi (2005).

5.2 Time-delay estimation techniques

Waves in nature propagate at a finite speed. This causes a time delay between the signals arriving at sensors located in different positions in a distributed sensor array. In signal processing literature, this delay is called Time Difference Of Arrival (TDOA). The basic discrete-time model for TDOA estimation can be stated as follows:

$$u_0(n) = x(n) + s_0(n) \tag{5.1}$$

$$u_1(n) = x(n - D) + s_1(n) \tag{5.2}$$

where $u_0(n)$ and $u_1(n)$ are the signals received at the observation points (i.e. sensors). Here, $x(n)$ is the signal of interest that is referenced (zero time-delay) according to the first sensor and will have a delay of D by the time it arrives at the second sensor. The signals $s_0(n)$ and $s_1(n)$ model the (possibly

dependent) noise/reverbration signals arriving at the first and second sensors, respectively. The goal of TDOA estimation is to estimate D given a segment of data obtained from each sensor, without any prior knowledge regarding the source signal $x(n)$ or the noise sources.

Time-delay estimation has been extensively explored in the past, and depending on the application at hand, different approaches have been proposed. See, e.g., Knapp and Carter (1976), Brandstein and Silverman (1997) and Aarabi (2001). The most commonly used TDOA estimation algorithm is the generalized cross-correlation based technique introduced by Knapp and Carter (1976). This approach suggests the use of an estimator of the form

$$\hat{D} = \arg\max_{D} \int_{\omega} Q(e^{j\omega}) U_0(e^{j\omega}) U_1^*(e^{j\omega}) e^{-j\omega D} d\omega \tag{5.3}$$

where $U_0(e^{j\omega})$ and $U_1(e^{j\omega})$ are the discrete-time Fourier transforms of the signals $u_0(n)$ and $u_0(n)$ respectively and $Q(e^{j\omega})$ is a cross-correlation weighting function.

While various weighting functions are possible, the so-called PHAse Transform (PHAT) whitening functions are commonly used due to the robustness of the resulting technique to reverberations. The PHAT weighting function is expressed as

$$Q(e^{j\omega}) = \frac{1}{|U_0(e^{j\omega}) U_1(e^{j\omega})|}. \tag{5.4}$$

Using this weighting function, the following form of the generalized cross-correlation formula is obtained:

$$\hat{D} = \arg\max_{D} \int_{\omega} \cos\left(\omega D - (\angle U_0(e^{j\omega}) - \angle U_1(e^{j\omega}))\right) d\omega \tag{5.5}$$

We will use the generalized cross-correlation technique with PHAT weighting function for TDOA estimation in this chapter. While there are many alternatives to this approach, PHAT was chosen because of its widespread use for microphone array based TDOA estimation as well as its proven robustness to reverberation artifacts. The reader is referred to Brandstein and Silverman (1997), Aarabi (2003) and Aarabi et al. (2005) for further practical discussions on this topic.

5.3 Time-delay estimation in multirate systems

Consider the multirate measurement model shown in Fig. 5.1. In this section we show that under some mild conditions, the unknown time delay D can be estimated by examining the phase of the cross spectral density (CSD) $P_{v_0 v_1}(e^{j\omega})$ of the low-rate signals $v_0(n)$ and $v_1(n)$. This will enable us to adapt conventional cross-correlation based techniques for estimating TDOA in multirate systems. Our key result is stated in the theorem below.

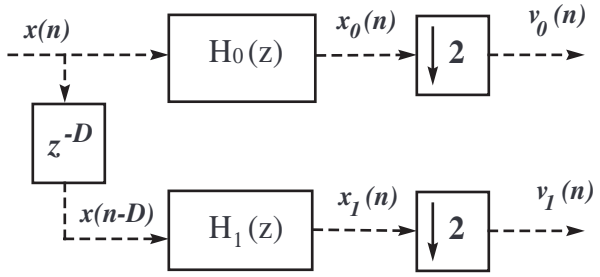


Fig. 5.1. Filter bank model for the multirate TDOA estimation problem.

Theorem 5.1. *Assume that the TDOA D is an integer and let $W(e^{j\omega}) \triangleq H_0(e^{j\omega})H_1^*(e^{j\omega})$ where $H_0(z)$ and $H_1(z)$ are the analysis filters shown in Fig. 5.1. If $\angle W(e^{j\frac{\omega}{2}}) = \angle W(e^{j(\pi-\frac{\omega}{2})})$, then*

$$\angle P_{v_0v_1}(e^{j\omega}) = \begin{cases} -D\frac{\omega}{2} + \angle W(e^{\frac{j\omega}{2}}) & D \text{ even} \\ -D\frac{\omega}{2} + \angle W(e^{\frac{j\omega}{2}}) + \lambda(\omega)\pi & D \text{ odd} \end{cases}$$

where $\lambda(\cdot)$ is a binary-valued function of ω assuming the values 0 and 1 only.

Proof. Consider the diagram shown in Fig. 5.1. It is straightforward to verify that the output signals $v_0(n)$ and $v_1(n)$ are jointly wide-sense stationary (Sathe and Vaidyanathan, 1993). Thus, the cross-correlation function $R_{v_0v_1}(k)$ defined by

$$R_{v_0v_1}(k) \triangleq E\{v_0(n)v_1(n+k)\} \tag{5.6}$$

exists. The signals $v_0(n)$ and $v_1(n)$ are down-sampled versions of $x_0(n)$ and $x_1(n-D)$. That is, $v_0(n) = x_0(2n)$ and $v_1(n) = x_1(2n-D)$. Thus we have

$$\begin{aligned} R_{v_0v_1}(k) &= E\{x_0(2n)x_1(2n+2k-D)\} \\ &= R_{x_0x_1}(2k-D). \end{aligned} \tag{5.7}$$

The above equation allows us to express the CSD $P_{v_0v_1}(e^{j\omega})$ of the low-rate signals in terms of the CSD $P_{x_0x_1}(e^{j\omega})$ associated with $x_0(n)$ and $x_1(n)$:

$$\begin{aligned} &P_{v_0v_1}(e^{j\omega}) \triangleq \textstyle\sum_{k=-\infty}^{\infty} R_{v_0v_1}(k)e^{-j\omega k} \\ &= \textstyle\sum_{k=-\infty}^{\infty} R_{x_0x_1}(2k-D)e^{-j\omega k} = \frac{1}{2}e^{-j\omega\frac{D}{2}} \times \\ &\begin{cases} P_{x_0x_1}(e^{j\frac{\omega}{2}}) + P_{x_0x_1}(e^{j\frac{\omega-2\pi}{2}}) & D \text{ even} \\ P_{x_0x_1}(e^{j\frac{\omega}{2}}) - P_{x_0x_1}(e^{j\frac{\omega-2\pi}{2}}) & D \text{ odd} \end{cases} \end{aligned} \tag{5.8}$$

In the last step of the above derivations we used the following properties of the discrete-time Fourier transform:

$$x(n) \overset{\mathcal{F}}{\rightleftharpoons} X(e^{j\omega}) \Rightarrow \begin{cases} x(2n) & \overset{\mathcal{F}}{\rightleftharpoons} \frac{X(e^{j\frac{\omega}{2}})+X(e^{j\frac{\omega-2\pi}{2}})}{2} \\ x(n-D) & \overset{\mathcal{F}}{\rightleftharpoons} e^{-j\omega D}X(e^{j\omega}) \end{cases}$$

It is straightforward to show that

$$P_{x_0x_1}(e^{j\omega}) = W(e^{j\omega})P_{xx}(e^{j\omega}), \tag{5.9}$$

where $P_{xx}(e^{j\omega})$ is the power spectral density (PSD) of the input signal. It follows from (5.8) and (5.9) that

$$\angle P_{v_0v_1}(e^{j\omega}) = \begin{cases} -D\frac{\omega}{2} + \\ \angle\left(P_{xx}(e^{j\frac{\omega}{2}})W(e^{j\frac{\omega}{2}}) + P_{xx}(e^{j\frac{\omega-2\pi}{2}})W(e^{j\frac{\omega-2\pi}{2}})\right) & D \text{ even} \\ -D\frac{\omega}{2} + \\ \angle\left(P_{xx}(e^{j\frac{\omega}{2}})W(e^{j\frac{\omega}{2}}) - P_{xx}(e^{j\frac{\omega-2\pi}{2}})W(e^{j\frac{\omega-2\pi}{2}})\right) & D \text{ odd} \end{cases}$$

The PSD of a real-valued WSS process is a real and positive function of frequency. Thus, $\angle P_{xx}(e^{j\omega}) = 0$. If the condition $\angle W(e^{j\frac{\omega}{2}}) = \angle W(e^{j\frac{\omega-2\pi}{2}})$ holds, we can simplify the above equation to get

$$\angle P_{v_0v_1}(e^{j\omega}) = \begin{cases} -D\frac{\omega}{2} + \angle W(e^{j\frac{\omega}{2}}) & D \text{ even} \\ -D\frac{\omega}{2} + \angle W(e^{j\frac{\omega}{2}}) + \\ \angle\left(P_{xx}(e^{j\frac{\omega}{2}})\left|W(e^{j\frac{\omega}{2}})\right| - P_{xx}(e^{j\frac{\omega-2\pi}{2}})\left|W(e^{j\frac{\omega-2\pi}{2}})\right|\right) & D \text{ odd} \end{cases}$$

In the above expression, the terms within the brackets are real. Thus, the phase contribution of the bracketed terms is either zero or π. □

The above theorem shows that, under suitable conditions on the phase of the analysis filters, the time delay D can be recovered by examining the phase of the cross spectral density of the low-rate measurements $v_0(n)$ and $v_1(n)$. In practice, one has to start with an estimate $\hat{P}_{v_0v_1}(e^{j\omega})$ of the cross spectral density of the low-rate measurements. The estimate $\hat{P}_{v_0v_1}(e^{j\omega})$ can be obtained using any of the standard spectral estimation methods discussed in texts such as Kay (1988) and Hayes (1996). Then, $\angle\hat{P}_{v_0v_1}(e^{j\omega})$ is used to calculate an estimate $\hat{D}$ of the actual time delay D by maximizing the PHAT integral

$$\hat{D} = \arg\max_D \int_\omega \cos\left(-D\frac{\omega}{2} - (\angle\hat{P}_{v_0v_1}(e^{j\omega}) - \angle W(e^{j\frac{\omega}{2}}))\right) d\omega. \tag{5.10}$$

In principle, PHAT is a nonlinear regression method which fits the linear model $-D\frac{\omega}{2}$ to the data represented by $\angle\hat{P}_{v_0v_1}(e^{j\omega}) - \angle W(e^{j\frac{\omega}{2}})$. If D is even and the available estimate $\hat{P}_{v_0v_1}(e^{j\omega})$ is accurate, then Theorem 5.1 shows

that $\angle\hat{P}_{v_0v_1}(e^{j\omega}) - \angle W(e^{j\frac{\omega}{2}})$ will be close to the linear function $-D\frac{\omega}{2}$. In this case, the PHAT integral (5.10) produces accurate TDOA estimates. When D is odd, Theorem 5.1 indicates that $\angle\hat{P}_{v_0v_1}(e^{j\omega}) - \angle W(e^{j\omega})$ will be close to $-D\frac{\omega}{2} + \lambda(\omega)\pi$. The binary-valued function

$$\lambda(\omega) \triangleq \begin{cases} 0 \text{ if } P_{xx}(e^{j\frac{\omega}{2}})\left|W(e^{j\frac{\omega}{2}})\right| > P_{xx}(e^{j\frac{\omega-2\pi}{2}})\left|W(e^{j\frac{\omega-2\pi}{2}})\right| \\ 1 \text{ if } P_{xx}(e^{j\frac{\omega}{2}})\left|W(e^{j\frac{\omega}{2}})\right| < P_{xx}(e^{j\frac{\omega-2\pi}{2}})\left|W(e^{j\frac{\omega-2\pi}{2}})\right| \end{cases} \tag{5.11}$$

is representative of the sign ambiguity which occurs in the determining the phase of $P_{v_0v_1}(e^{j\omega})$. As can be seen from the expression above, $\lambda(\omega)$ depends on the input signal statistics through $P_{xx}(e^{j\omega})$.

In principle, it is possible to estimate $P_{xx}(e^{j\omega})$ from $v_0(n)$ and $v_1(n)$ using the technique described in Jahromi et al. (2003) and then estimate $\lambda(\omega)$ from (5.11). However, we do not follow this possibility here due to its very high computational burden. Instead, we choose to ignore the term $\lambda(\omega)\pi$ while calculating the PHAT integral. Our rationale is that for most sensor array applications (e.g., microphone arrays), the low-frequency half of the spectrum hugely dominates the high frequency half in terms of energy. As a result, $\lambda(\omega)$ will be equal to zero much more frequently than 1. This makes the overall contribution of the term $\lambda(\omega)\pi$ to the PHAT integral (5.10) negligible. We will demonstrate the general validity of this assumption in Section 5.5 where we present actual TDOA estimation experiments. There, we will provide cases where this assumption fails to hold as well.

Note, also, that the results of Theorem 5.1 remain valid even when independent noise components $s_0(n)$ and $s_1(n)$ are added to the input signals $x(n)$ and $x(n-D)$, respectively. However, if the noise sources $s_0(n)$ and $s_1(n)$ are correlated, an extra term (which depends on the cross-correlation between the two noise signals) will be added to the right hand side of (5.7). This will introduce additional terms in the phase of $P_{v_0v_1}(e^{j\omega})$ and, hence, bias in the estimation of D. Our experiments with microphone arrays showed that nominal room noise (air conditioning systems, etc. resulting in 20dB SNR) had no noticeable effect on the accuracy of TDOA estimates (See the experiments in Section 5.5).

Practical experiments via microphone arrays have showen that the TDOA estimator $\hat{D}$ given by (5.10) is robust and does not collapse when the actual TDOA is not an integer multiple of the sampling interval. In fact, Theorem 5.1 is valid for the non-integer case too provided that the term $\lambda(\omega)\pi$ in its statement is replaced by a general ambiguous phase term. This ambiguous term can be neglected if the spectral domination condition discussed before is satisfied. In this case, the TDOA estimation procedure is the same as before except that, now, the search for the D which maximizes the PHAT integral (5.10) should include non-integer values as well.

5.4 Multirate sensors that allow time-delay estimation

In order to use the PHAT integral (5.10) as a valid estimator of D, one must choose sensor filters $H_0(z)$ and $H_1(z)$ whose phase response satisfy the symmetry condition

$$\angle\left(H_0(e^{j\frac{\omega}{2}})H_1^*(e^{j\frac{\omega}{2}})\right) = \angle\left(H_0(e^{j(\pi-\frac{\omega}{2})})H_1^*(e^{j(\pi-\frac{\omega}{2})})\right) \tag{5.12}$$

required by Theorem 5.1. Several classes of filters satisfy (5.12). For instance, the reader may observe that if $H_0(z)$ and $H_1(z)$ are linear-phase, FIR and with the same length N, then $\angle H_0(e^{j\omega})H_1^*(e^{j\omega})$ becomes a constant which, in turn, implies (5.12). It is possible to satisfy (5.12) by using certain types of IIR filters as well. In the examples that follow, we present three representative choices for $H_0(z)$ and $H_1(z)$.

5.4.1 Sensors based on linear-phase FIR filters

Linear-phase FIR filters with good frequency selectivity can be designed using a variety of methods, most notably the weighted-Chebyshev method of Parks and McClellan (1972a). In programming this method, an error function is formulated for the desired amplitude response in terms of a linear combination of cosine functions and is then minimized by using a very efficient multivariable optimization method known as the Remez exchange algorithm, see e.g. Parks and McClellan (1972b) and (Antoniou, 1993, Ch. 15).

Example 5.1. One can use the MATLAB function remez which implements the Parks-McClellan algorithm to obtain a low-pass filter $H_0(z)$ with symmetric impulse response of length $N = 9$. A high-pass filter $H_1(z)$ of the same length whose amplitude response is the mirror-image of $H_0(z)$ was obtained by simply replacing z in $H_0(z)$ with $-z$. The amplitude responses of $H_0(z)$ and $H_1(z)$ are shown in Fig. 5.3(a) where their phase responses are depicted in Fig. 5.3(b). The amplitude and phase of $H_0(z)H_1^*(z)$ for the filters designed in this example are shown in Fig. 5.4 (a) and (b). It is clear that the phase symmetry condition (5.12) is satisfied. ♢

5.4.2 Sensors based on Bessel IIR filters

In general, it is not possible to achieve linear phase response with IIR filters. However, it is possible to design IIR low-pass and high-pass filters $H_0(z)$ and $H_1(z)$ such that the product $H_0(z)H_1^*(z)$ has linear phase. This can be achieved, for instance, using second-order Butterworth filters. Another approach is to design almost-linear-phase $H_0(z)$ and $H_1(z)$ by discretizing analog Bessel filters via the impulse-invariant transformation. Here, we opt for the latter approach.

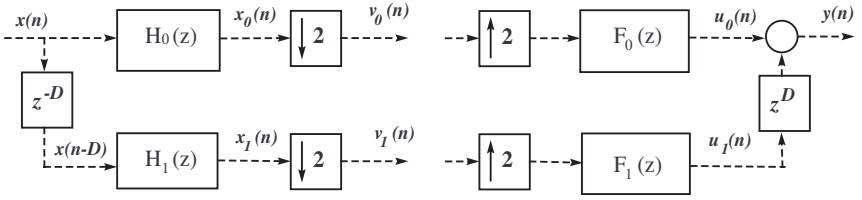


Fig. 5.2. A two-channel analysis/synthesis filter bank with TDOA.

Example 5.2. For the purpose of this example, we used the MATLAB function besself to design an 8-order analog Bessel filter with cutoff frequency of 40Hz. Then, we discretized it using the MATLAB function impinvar at a sampling frequency of 20 Hz to obtain the low-pass filter $H_0(z)$. A high-pass filter $H_1(z)$ of the same order was obtained by replacing z in $H_0(z)$ with $-z$. The amplitude responses of $H_0(z)$ and $H_1(z)$ are shown in Fig. 5.3(c) where their phase responses are depicted in Fig. 5.3(d). The amplitude and phase of $H_0(z)H_1^*(z)$ for the Bessel filters designed in this example are shown in Fig. 5.4 (c) and (d). It is clear from Fig. 5.4(d) that $H_0(z)$ and $H_1(z)$ obey (5.12) with a good approximation. ♢

5.4.3 Sensors based on perfect-reconstruction filter banks

Multirate filter banks for which it is possible to reconstruct the input signal from the low-rate components are called *perfect reconstruction filter banks*. It is possible to design perfect reconstruction filter banks for which the analysis and synthesis filters are both linear-phase and FIR (Vaidyanathan, 1993, Chap. 7). In this section we introduce a special class of such filter banks which can be used for TDOA estimation in multirate sensor systems.

Consider the two-channel analysis filter bank shown in Fig. 5.2 and assume that D is zero. It is well-known that the transfer functions $H_0(z)$ and $H_1(z)$ can be compactly represented by the transfer vector $\mathbf{h}(z) \triangleq [H_0(z)\ H_1(z)]^T$ and that $\mathbf{h}(z)$ can be factored as $\mathbf{h}(z) = \mathbf{E}(z^M)\mathbf{e}(z)$. In this factorization, $\mathbf{e}(z) \triangleq [1\ z^{-1}]^T$ and $\mathbf{E}(z)$ is called a *type-1 polyphase matrix* (Vaidyanathan, 1993).

We define the class $\mathcal{P}_N$ as the set of filter banks for which the following two conditions are satisfied:

(i) The filters $H_0(z)$ and $H_1(z)$ are of length $N \triangleq 2(K+1)$, where $K \in \mathbb{Z}^+$ is fixed. In other words, $\mathbf{E}(z)$ is FIR of order K.
(ii) The matrix $\mathbf{E}(z)$ has the factorization

$$\mathbf{E}(z) = \mathbf{A}_K\mathbf{D}(z)\mathbf{A}_{K-1}\mathbf{D}(z)\ldots\mathbf{D}(z)\mathbf{A}_0 \tag{5.13}$$

where

$$\mathbf{D}(z) \triangleq \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix} \tag{5.14}$$

and

$$\begin{aligned} \mathbf{A}_i &\triangleq \begin{bmatrix} 1 & \theta_i \\ \theta_i & 1 \end{bmatrix} \quad 0 \le i \le K-1, \\ \mathbf{A}_i &\triangleq \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad\qquad i = K. \end{aligned} \tag{5.15}$$

One can verify that the above conditions result in analysis filter banks for which the impulse response of one filter is symmetric while the impulse response of the other is anti-symmetric. Thus, $H_0(z)$ and $H_1(z)$ will have linear phase. Furthermore, we have

$$|\angle H_0(e^{j\omega}) - \angle H_1(e^{j\omega})| = \frac{\pi}{2}.$$

The synthesis filters are obtained by first constructing the adjoint polyphase matrix

$$\mathbf{R}(z) = \mathbf{A}_0^T \mathbf{C}(z) \mathbf{A}_1^T \mathbf{C}(z) \dots \mathbf{C}(z) \mathbf{A}_K^T \tag{5.16}$$

where

$$\mathbf{C}(z) = \begin{bmatrix} z^{-1} & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.17}$$

The synthesis filters $F_0(z)$ and $F_1(z)$ are then calculated from

$$\begin{bmatrix} f_0(z) & f_1(z) \end{bmatrix} = \mathbf{e}^T(z) \mathbf{R}(z^M). \tag{5.18}$$

The above filters lead to perfect reconstruction with an overall delay of $K+1$ samples.

Clearly, the filter banks in the class $\mathcal{P}_N$ are fully parameterized by the K free parameters θ_0 to θ_{K-1}. These parameters may be optimized such that certain frequency response requirements are satisfied or at least approximated.

Example 5.3. Let us choose a pair of analysis filters in $\mathcal{P}_1$. The filters in this class are parameterized by only one parameter θ_0! We can specify θ_0 by minimizing an objective function like

$$\gamma = \int_0^{\omega_{pass}} \left(1 - |H_0(e^{j\omega})|\right)^2 d\omega + \int_{\omega_{stop}}^{\pi} |H_0(e^{j\omega})|^2 \mathrm{d}\omega \tag{5.19}$$

$$+ \int_{\omega_{stop}}^{\pi} \left(1 - |H_1(e^{j\omega})|\right)^2 d\omega + \int_0^{\omega_{pass}} |H_1(e^{j\omega})|^2 \mathrm{d}\omega. \tag{5.20}$$

This objective function simply reflects the mean-square deviation of the frequency responses of the filters $H_0(z)$ and $H_1(z)$ from ideal low-pass and high-pass responses, respectively. With $\omega_{pass} = 0.45\pi$ and $\omega_{stop} = 0.55\pi$, the optimal value for θ_0 was found to be 78.04. This leads to

$$H_0(z) = 0.0091 + 0.7070z^{-1} + 0.7070z^{-2} + 0.0091z^{-3},$$

and

$$H_1(z) = 0.0091 + 0.7070z^{-1} - 0.7070z^{-2} - 0.0091z^{-3}.$$

The amplitude and phase responses of $H_0(z)$ and $H_1(z)$ designed above are shown in Fig. 5.3 (e) and (f). The amplitude and phase response of the product filter $H_0(z)H_1^*(z)$ are shown in Fig. 5.4 (e) and (f). It is clear from this figure that $H_0(z)$ and $H_1(z)$ obey (5.12). ♢

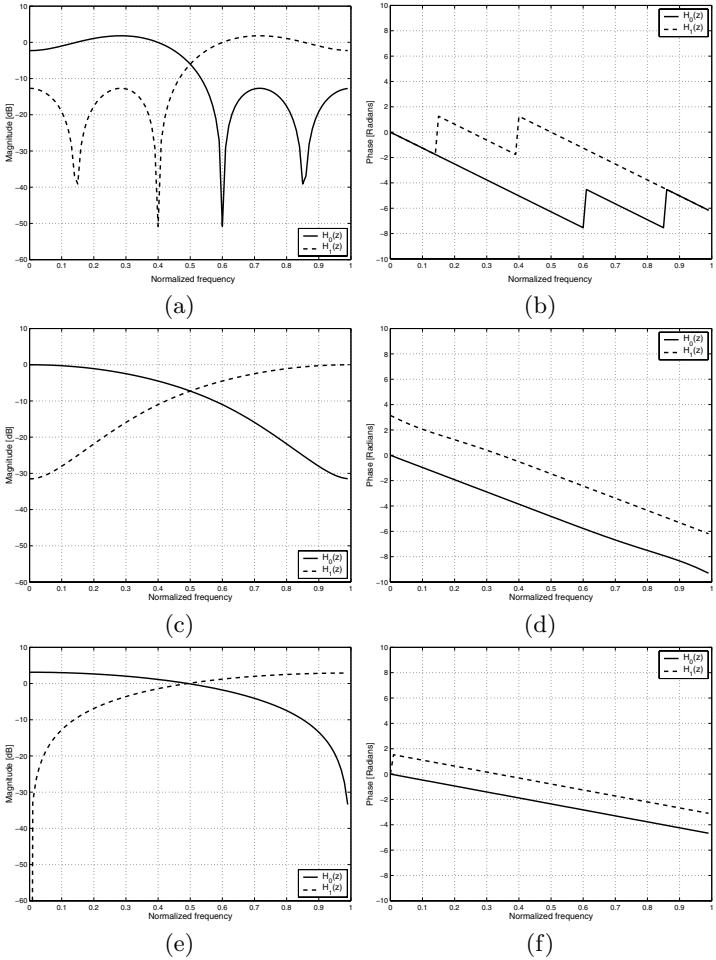


Fig. 5.3. Amplitude response and phase response of various sensor filters $H_0(z)$ and $H_1(z)$ introduced in Section 5.4. (a) and (b): ordinary linear-phase FIR filters. (c) and (d): Bessel IIR filters. (e) and (f) Perfect Reconstruction linear-phase FIR filters.

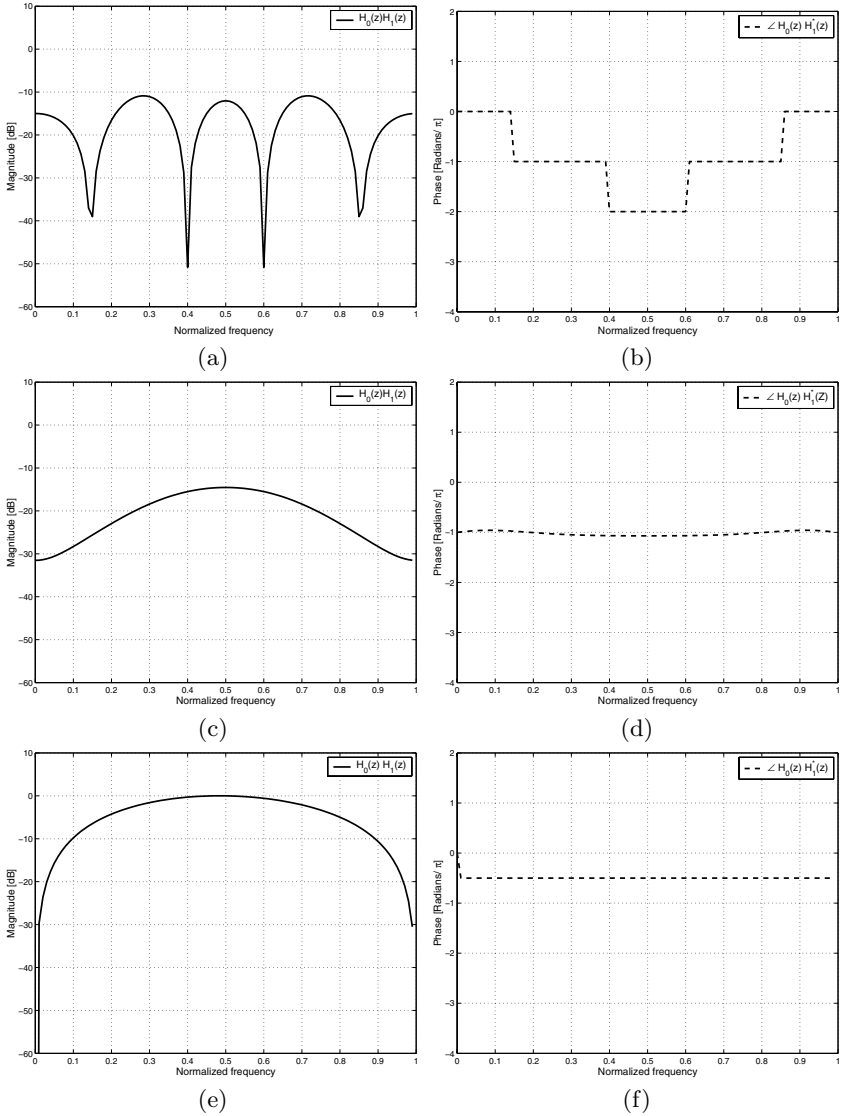


Fig. 5.4. Amplitude response and phase response of the product filter $H_0(z)H_1^*(z)$ for the sensor filters introduced in Section 5.4. (a) and (b): ordinary linear-phase FIR filters. (c) and (d): Bessel IIR filters. (e) and (f) perfect Reconstruction linear-phase FIR filters.

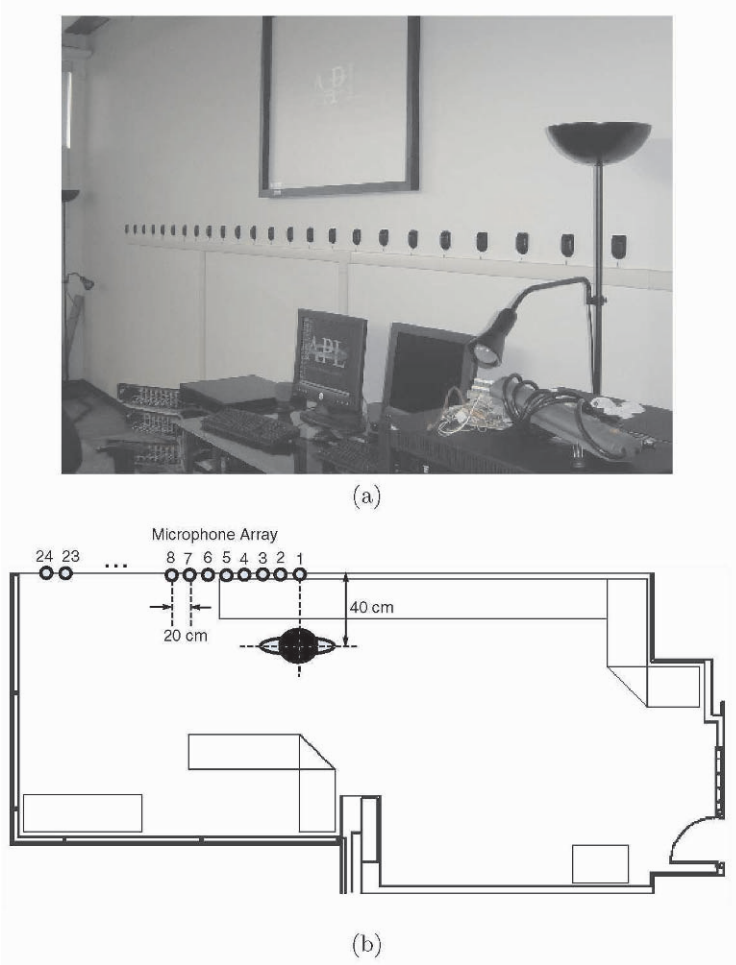


Fig. 5.5. The microphone array setup at the Artificial Perception Laboratory, University of Toronto.

5.5 Laboratory experiments

Here we present some actual multirate TDOA estimation experiments to back up the theoretical results presented in the previous sections. The experiments were performed at the Artificial Perception Laboratory, University of Toronto, using the linear microphone array shown in Fig. 5.5(a). A female student was asked to speak the sentence "Edward S. Rogers Sr. Department of Electrical and Computer Engineering" in front of a microphone array standing at the location specified in Fig. 5.5(b). The signal arriving at each microphone was sampled at 20 KHz and recorded for about 4 seconds.

We used two microphones in the microphone array (No. 3 and No. 5) to simulate a multirate sensor array measurement system similar to the one shown in Fig. 5.1. The output of microphone No. 3 was used as the reference signal $x(n)$. This signal and its spectrogram are shown in Fig. 5.6 (a) and (b), respectively. The signal recorded by microphone No. 5 was used as the delayed input. We used the example analysis filters described in Section 5.4 to filter these signals and then down-sampled the results to obtain $v_0(n)$ and $v_1(n)$.

An estimate $\hat{P}_{v_0v_1}(e^{j\omega})$ of the cross spectral density of the low-rate observations $v_0(n)$ and $v_1(n)$ was obtained by using the MATLAB function csd. This function estimates the cross spectral density of two signals using Welch's averaged periodogram method (see, e.g., Hayes (1996)). The parameters of the function csd were chosen such that it would operate as follows: First, the signals $v_0(n)$ and $v_1(n)$ were divided into overlapping sections of length 1024 and then each section was windowed by a von Hann window. The overlap length was set to 512. Several overlapping sections would form a "block". The products of the DFTs of the sections of $v_0(n)$ and $v_0(n)$ which were in the same block were averaged to form $\hat{P}_{v_0v_1}(e^{j\omega})$ for that block. The block length was set to 4096.

The above procedure provided us with a short-time cross spectrum estimate for each block (4096 samples or about 0.4 seconds) of the low-rate measurements. We used this estimate to calculate the PHAse Transform integral

$$\int_{\omega} \cos\left(-D\frac{\omega}{2} - (\angle\hat{P}_{v_0v_1}(e^{j\omega}) - \angle W(e^{j\frac{\omega}{2}}))\right) d\omega. \tag{5.21}$$

for each block and for all delay values D from -40 to 40. This process, which we call short-time PHAse Transform, was repeated until all the blocks in the signals $v_0(n)$ and $v_0(n)$ were covered. In Fig. 5.7, we have used shades of gray to depict the numerical value of the integral in (5.21) for all blocks in the signals $v_0(n)$ and $v_0(n)$ and for $-40 \leq D \leq 40$. For each block, the D value which maximizes the integral in (5.21) (i.e. the one which has produced the brightest color) represents the time delay estimate $\hat{D}$ for that block. Fig. 5.7 also shows the value of the integral (5.12) as a function of D averaged over the entire length of the signals $v_0(n)$ and $v_0(n)$. The D value which maximizes this quantity represents the TDOA estimate for the entire signal[1]. The plots in Fig. 5.7 show that $\hat{D} = 13$. This value was validated with estimates obtained from the original (full-rate) microphone signals.

Remark 5.1. Note that the in the above experiments, the TDOA value is odd which means the phase of $P_{v_0v_1}(e^{j\omega})$ contains the ambiguous component $\lambda(\omega)\pi$. Recall that in the odd TDOA case, the PHAT estimator is guaranteed to work only if the spectrum dominance condition mentioned in Theorem 5.1 is satisfied. This condition is not satisfied for the blocks centered at $t = 0.5$, $t = 1$ and $t = 2.75$ since these blocks contain strong components in both high

[1] Of course this assumes that D remains constant during the recording.

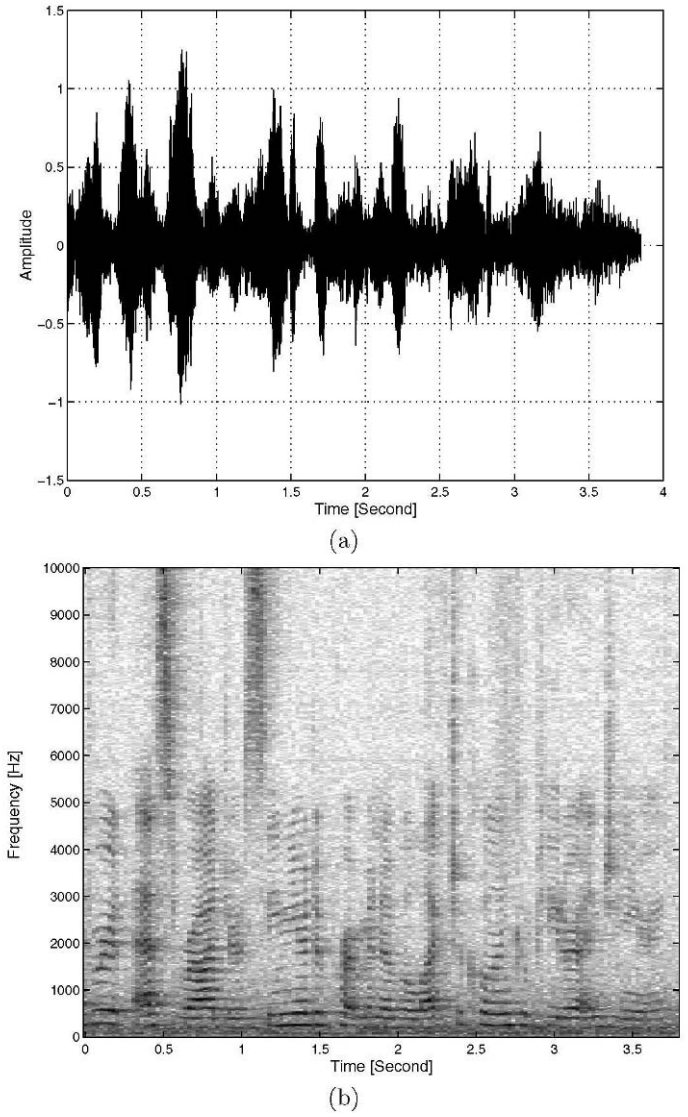


Fig. 5.6. (a) The reference speech signal $x(n)$ recorded by Microphone No. 3 in the array. The array was receiving the voice of a female speaker saying the sentence "Edward S. Rogers Sr. Department of Electrical and Computer Engineering". (b) The spectrogram of $x(n)$.

and low frequencies (see Fig. 5.5(b)). The short-time PHAT plots in Fig. 5.7 (a) and (c) do not show a prominent peak for these blocks. This indicates the failure of the PHAT-based TDOA estimation for these blocks in agreement with our theory.

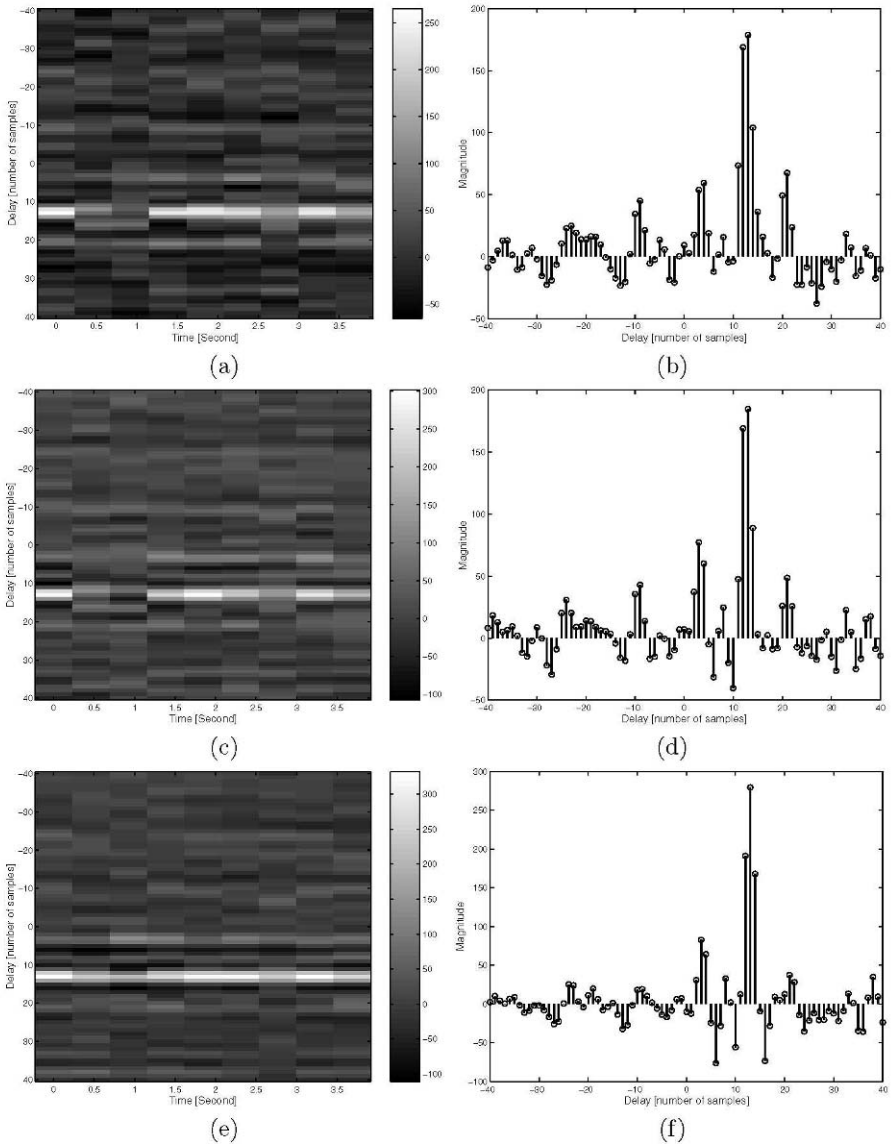


Fig. 5.7. TDOA estimation results using the filters $H_0(z)$ and $H_1(z)$ described in Section 5.4. Each row shows Short-time PHAse Transform results in the left and PHAse Transform averaged over the entire signal on the right. (a) and (b): ordinary linear-phase FIR filters. (c) and (d): Bessel IIR filters. (e) and (f) perfect reconstruction linear-phase FIR filters.

5.6 Multirate sensor fusion in the presence of time-delay

If we ignore the extra delay block, the multirate sensor model shown in Fig 5.1 resembles a standard two-channel analysis filter bank. It seems reasonable, therefore, to try to recover the high-rate signal $x(n)$ from the low-rate signals $v_i(n)$ using a standard synthesis filter bank equipped with a compensating *advance block* z^D as shown in Fig 5.2. However, theoretical analysis shows that it is impossible to achieve perfect reconstruction for all TDOA values in systems that use finite-order analysis and synthesis filters.

In this section, we first present a fundamental theorem which shows that it is impossible to design finite-order (practical) analysis and synthesis filters that achieve perfect reconstruction for all time-delay values. Then, we present an optimization method which enables us to design optimum synthesis filters for given analysis filters and a given (fixed) time delay. We show trough simulated design examples that this optimization technique is quite powerful in the sense that *almost perfect reconstruction* can be achieved in many practical cases.

5.6.1 Perfect reconstruction for arbitrary time-delays

The following theorem establishes necessary conditions under which the system shown in Fig 5.2 can achieve perfect reconstruction for arbitrary values of the time-delay D.

Theorem 5.2. *The multirate analysis/synthesis system shown in Fig 5.2 can achieve perfect reconstruction for all values of D only if*

$$H_0(e^{j\omega}) = F_0(e^{j\omega}) = 0, \forall\omega \in \Omega$$
$$H_1(e^{j\omega}) = F_1(e^{j\omega}) = 0, \forall\omega \in \widehat{\Omega}$$

where $\Omega = [0\ \frac{\pi}{2})$ and $\widehat{\Omega} = (\frac{\pi}{2}\ \pi]$ or the other way around.

Proof. Consider the block diagram shown in Fig. 5.2. When $x(n) \in \ell_2$, the Fourier transforms $V_0(e^{j\omega})$ and $V_1(e^{j\omega})$ of the low-rate signals $v_1(n)$ and $v_2(n)$ exist and can be expressed as

$$V_0(e^{j\omega}) = \frac{1}{2}[H_0(e^{j\frac{\omega}{2}})X(e^{j\frac{\omega}{2}}) + H_0(e^{j\frac{\omega-2\pi}{2}})X(e^{j\frac{\omega-2\pi}{2}})],$$
$$V_1(e^{j\omega}) = \frac{1}{2}[e^{-j\frac{\omega}{2}D}H_1(e^{j\frac{\omega}{2}})X(e^{j\frac{\omega}{2}}) + e^{-j\frac{\omega-2\pi}{2}D}H_1(e^{j\frac{\omega-2\pi}{2}})X(e^{j\frac{\omega-2\pi}{2}})].$$

The Fourier transforms $U_0(e^{j\omega})$ and $U_1(e^{j\omega})$ of the synthesized signals $u_0(n)$ and $u_1(n)$ can be written as

$$U_0(e^{j\omega}) = F_0(e^{j\omega})V_0(e^{j2\omega}),$$
$$U_1(e^{j\omega}) = e^{j\omega D}F_1(e^{j\omega})V_1(e^{j2\omega}).$$

Finally, $Y(e^{j\omega}) = U_0(e^{j\omega}) + U_1(e^{j\omega})$. It is straightforward to combine the previous four equations and express $Y(e^{j\omega})$ in terms of the *true spectrum* $X(e^{j\omega})$ and the *image spectrum* $X(e^{j(\pi-\omega)})$ as follows

$$Y(e^{j\omega}) = \frac{1}{2}A(e^{j\omega})X(e^{j\omega}) + \frac{1}{2}B(e^{j\omega})X(e^{j(\pi-\omega)});$$

where

$$A(e^{j\omega}) \triangleq [F_0(e^{j\omega})H_0(e^{j\omega}) + F_1(e^{j\omega})H_1(e^{j\omega})],$$
$$B(e^{j\omega}) \triangleq [F_0(e^{j\omega})H_0(e^{j(\pi-\omega)}) + e^{j\pi D}F_1(e^{j\omega})H_1(e^{j(\pi-\omega)})].$$

A necessary condition for perfect reconstruction is that terms pertaining to the image spectrum $X(e^{j(\pi-\omega)})$ are completely eliminated in the output. That is $B(e^{j\omega}) = 0$. Another necessary condition is that $A(e^{j\omega})$ becomes a non-zero constant. The condition that $B(e^{j\omega}) = 0$ is possible for all values of $D \in \mathbb{Z}$ only if both $F_0(e^{j\omega})H_0(e^{j(\pi-\omega)}) = 0$ and $F_1(e^{j\omega})H_1(e^{j(\pi-\omega)}) = 0$ for all ω. To satisfy the latter condition, the products $F_0(e^{j\omega})H_0(e^{j\omega})$ and $F_1(e^{j\omega})H_1(e^{j\omega})$ should not vanish simultaneously. Using logical notation we can write these necessary conditions in the compact form:

$$F_0(e^{j\omega})H_0(e^{j\omega}) = 0 \text{ XOR } F_1(e^{j\omega})H_1(e^{j\omega}) = 0 \tag{5.22}$$
$$F_0(e^{j\omega})H_0(e^{j(\pi-\omega)}) = 0 \text{ AND } F_1(e^{j\omega})H_1(e^{j(\pi-\omega)}) = 0 \tag{5.23}$$

Let $\Gamma \triangleq [0\ \frac{\pi}{2}) \bigcup (\frac{\pi}{2}\ \pi]$. Denote by Ω the set of all frequencies in Γ for which $F_0(e^{j\omega})$ is identically zero and define $\overline{\Omega} \triangleq \Gamma - \Omega$. In other words, assume

$$F_0(e^{j\omega}) \triangleq 0,\ \omega \in \Omega, \tag{5.24}$$
$$F_0(e^{j\omega}) \neq 0,\ \omega \in \overline{\Omega}. \tag{5.25}$$

It follows from (5.22) that the synthesis filters $F_0(e^{j\omega})$ and $F_1(e^{j\omega})$ cannot vanish at the same frequencies. They cannot be both nonzero at the same frequency either. The reason is that if they become nonzero at the some frequencies, (5.23) will require that $H_0(e^{j(\pi-\omega)})$ and $H_1(e^{j(\pi-\omega)})$ both be zero at those frequencies. This implies that $H_0(e^{j\omega})$ and $H_1(e^{j\omega})$ will vanish simultaneously and therefor contradicts (5.22). Thus, in summary, $F_1(e^{j\omega})$ should vanish wherever $F_0(e^{j\omega})$ is nonzero and vice versa:

$$F_1(e^{j\omega}) \neq 0,\ \omega \in \Omega, \tag{5.26}$$
$$F_1(e^{j\omega}) = 0,\ \omega \in \overline{\Omega}. \tag{5.27}$$

To satisfy (5.24)-(5.27) along with the original conditions in (5.22) and (5.23), it is further required that

$$H_1(e^{j\omega}) \neq 0 \ \text{ AND } \ H_1(e^{j(\pi-\omega)}) = 0,\ \omega \in \Omega, \tag{5.28}$$
$$H_0(e^{j\omega}) \neq 0 \ \text{ AND } \ H_0(e^{j(\pi-\omega)}) = 0,\ \omega \in \overline{\Omega}. \tag{5.29}$$

The above conditions can be satisfied only if

$$\omega \in \Omega \Longrightarrow \pi - \omega \in \overline{\Omega} \text{ AND } \omega \in \overline{\Omega} \Longrightarrow \pi - \omega \in \Omega \tag{5.30}$$

which is possible only if $\Omega = [0\ \frac{\pi}{2})$ or $\Omega = (\frac{\pi}{2}\ \pi]$. Thus, to satisfy (5.28) and (5.29) simultaneously it is required that $\Omega = [0\ \frac{\pi}{2})$, $\overline{\Omega} = (\frac{\pi}{2}\ \pi]$ or $\overline{\Omega} = [0\ \frac{\pi}{2})$, $\Omega = (\frac{\pi}{2}\ \pi]$ and that

$$H_0(e^{j\omega}) = 0,\ \omega \in \Omega, \tag{5.31}$$

$$H_1(e^{j\omega}) = 0,\ \omega \in \overline{\Omega}. \tag{5.32}$$

□

The ideal brick-wall frequency responses that satisfy the conditions of the above theorem cannot be realized using FIR or IIR structures. Therefore, it is not possible to perfectly recover the high-rate signal $x(n)$ from the low-rate measurements $v_i(n)$ using fixed finite-order filters.

One way to get around this difficulty is to use different pairs of synthesis filters for different TDOA values. Let us write the TDOA D as $D = 2K + \tilde{D}$ where $K \in \mathbb{Z}$ and $\tilde{D} \in [0, 2)$. In this case, the delay block in the sensors' filter bank model (Fig. 5.2) can be decomposed into an integer, even, delay z^{-2K} and a residual delay $z^{-\tilde{D}}$. z^{-2K} commutes with down-sampling and up-sampling operation. Thus, it can be readily compensated for in the receiving end by adding an equivalent delay z^{-2K} to the other channel. The residual part $z^{-\tilde{D}}$, however, does not commute with the down-sampling or up-sampling blocks. Thus, when designing the fusion system, we have to consider it along with the filter $H_1(z)$ in Fig. 5.2. In other words, $H_1(z)$ subsumes $z^{-\tilde{D}}$ as shown in Fig. 5.8.

When $\tilde{D}$ is different from 0 or 1, the transfer function $z^{-\tilde{D}}$ cannot be realized using a finite-order physical system. In this case, one must approximate $z^{-\tilde{D}}$ by a finite-order rational transfer function. This can be done by using the classic Padé approximation (Lam, 1993) or other more recent techniques (Yoon and Lee, 1997), (Philipp et al., 1999).

Once $z^{-\tilde{D}}$ is approximated and combined with $H_1(z)$, a synthesis filter bank must be designed to reconstruct $x(n)$ as faithfully as possible (Fig. 5.8). The synthesis filters $F_0(z)$ and $F_1(z)$ depend on $\tilde{D}$. Thus, they must be designed (in real time if needed) after the TDOA D is estimated. An efficient way to do this is the model-matching technique first proposed by Shenoy in the context of multirate systems (Shenoy, 1994) and (Shenoy et al., 1994).

In the next subsection, we will use an elegant variation of this approach due to Chen and Francis (1995a) to design optimal synthesis filters for a given time-delay. These authors used $\mathcal{H}_\infty$ optimization theory to minimize the ℓ_2-induced norm between a pure delay system and the multirate system to be designed. A version of the $\mathcal{H}_\infty$ optimization method which uses linear matrix inequalities (LMIs) and achieves reduced-order solutions has been recently proposed by Li and Kok (2003).

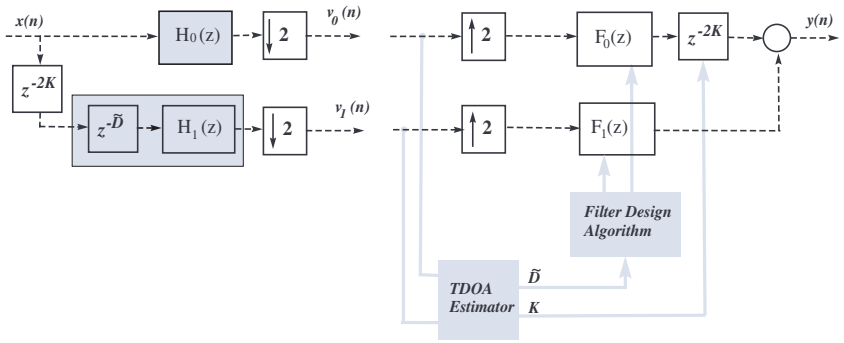


Fig. 5.8. The structure of the central fusion system along with the analysis filter bank model used for designing $F_0(z)$ and $F_1(z)$.

5.6.2 A practical design method using $\mathcal{H}_\infty$ optimization

$\mathcal{H}_\infty$ optimization is central in modern control theory. See e.g. Francis (1987), Green and Limebeer (1995) and Chen and Francis (1995b). The Hardy space $\mathcal{H}_\infty$ consists of all complex-valued functions $H(z)$ that are analytic and bounded outside the unit disc, that is for all $|z| > 1$. Therefore, $\mathcal{H}_\infty$ is the space of transfer functions of causal and LTI systems which are stable in the bounded-input, bounded-output (BIBO) sense. The norm of a multi-input multi-output transfer function $\mathbf{H}(z) \in \mathcal{H}_\infty$ is defined as the peak magnitude of its maximum singular value on the unit circle:

$$||\mathbf{H}||_\infty \triangleq \sup_\omega \sigma_{max}[\mathbf{H}(e^{j\omega})]. \tag{5.33}$$

If $\mathbf{H}(z)$ is the transfer function of a stable, causal LTI system with input $X(n)$ of dimension m and output $Y(n)$ of dimension p, so that $\mathbf{H}(z)$ is $p \times m$, then the induced norm from the input space ℓ_2^m to the output space ℓ_2^p equals the $\mathcal{H}_\infty$-norm of $\mathbf{H}(z)$. That is,

$$\sup_{||X||_2=1} ||Y||_2 = ||\mathbf{H}||_\infty \tag{5.34}$$

where the norm of a signal $X(n)$ in ℓ_2^m is defined to be

$$||X||_2 \triangleq \left(\sum_n X^T(n) X(n) \right)^{\frac{1}{2}}. \tag{5.35}$$

Now, consider the analysis/synthesis filter bank shown in Fig. 5.8. Because of the down-sampling and up-sampling operations, the system which relates the output signal $y(n)$ to the input signal $x(n)$ is, in general, a linear periodically time-varying (LPTV) system. Thus, it does not admit a transfer function. However, we can "block" the input and output signals to obtain an LTI

input-output equivalent system. This latter system has the two-dimensional input and output

$$X(n) \triangleq \left[x(2n)\ x(2n+1)\right]^T,\ Y(n) \triangleq \left[y(2n)\ y(2n+1)\right]^T \tag{5.36}$$

and a 2×2 transfer matrix which we denote by $\mathbf{P}(z)$.

To find an expression for $\mathbf{P}(z)$, we have to use the *polyphase representation* of the analysis and synthesis filters (Vaidyanathan, 1993). Let us represent the analysis filters $H_0(z)$ and $H_1(z)$ compactly by defining the transfer vector $\mathbf{h}(z) \triangleq [H_0(z)\ H_1(z)]^T$. It is possible to factor $\mathbf{h}(z)$ as the product of a 2×2 transfer matrix $\mathbf{E}(z)$ and a delay vector $\mathbf{e}(z)$. That is,

$$\mathbf{h}(z) = \mathbf{E}(z^2)\mathbf{e}(z) \tag{5.37}$$

where $\mathbf{e}(z) \triangleq [1\ z^{-1}]^T$. The matrix $\mathbf{E}(z)$ is called the *type-1 polyphase matrix* associated with the analysis filter bank $\mathbf{h}(z)$. Similarly, the synthesis filters can be represented in the compact form $\mathbf{f}(z) \triangleq [F_0(z)\ F_1(z)]$ which, in turn, may be factored as

$$\mathbf{f}(z) \triangleq \mathbf{e}^T(z)\mathbf{R}(z^2). \tag{5.38}$$

The matrix $\mathbf{R}(z)$ is called a *type-2 polyphase matrix*. Using the polyphase notation, it is straightforward to show that

$$\mathbf{P}(z) \triangleq \mathbf{R}(z)\mathbf{E}(z). \tag{5.39}$$

Our objective is to design the synthesis filters $F_0(z)$ and $F_1(z)$ given the analysis filters $H_0(z)$ and $H_1(z)$, and a tolerable delay $T_d \in \mathbb{Z}_+$ such that $y(n)$ is "as close as possible" to $x(n - T_d)$. This objective can be made precise by defining the *error signal* $e(n) \triangleq y(n) - x(n - T_d)$ and then minimizing the *performance measure*

$$J \triangleq \sup_{||x||=1} ||e||_2 \tag{5.40}$$

which measures the worst-case ℓ_2-induced norm from the input signal $x(n)$ to $e(n)$. Blocking preserves ℓ_2 norm, that is, the norm of a signal in ℓ_2 is equal to the norm of its blocked version in ℓ_2^m. Using this fact and (5.34), it can be shown (Chen and Francis, 1995a, Theorem 2.1) that

$$J = ||\mathbf{C} - \mathbf{RE}||_\infty \tag{5.41}$$

where

$$\mathbf{C}(z) \triangleq \begin{cases} z^{-k}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ if } T_d = 2k+1, \\ z^{-k}\begin{bmatrix} 0 & z \\ 1 & 0 \end{bmatrix}, \text{ if } \quad T_d = 2k. \end{cases} \tag{5.42}$$

Based on the above result, our design problem can be precisely stated as follows: Given causal and stable (FIR or IIR) analysis filters $H_0(z)$ and $H_1(z)$

and given a tolerable overall delay T_d, find causal, stable IIR synthesis filters $F_0(z)$ and $F_1(z)$ such that J is minimized. The optimum performance measure J_{opt} is therefore

$$J_{opt} = \inf_{F_0(z),F_1(z)} \sup_{||x||=1} ||e||_2 = \inf_{\mathbf{R}(z)\in\mathcal{H}_\infty} ||\mathbf{C} - \mathbf{RE}||_\infty. \tag{5.43}$$

The latter optimization is a standard $\mathcal{H}_\infty$ model matching problem and can be solved using existing software tools, e.g. the μ-Analysis and Synthesis Toolbox of MATLAB. For reader's convenience, the $\mathcal{H}_\infty$-optimal synthesis filter design procedure is outlined in the text box below.

$\mathcal{H}_\infty$-optimal synthesis filter design

Input: The analysis filters $H_0(z)$ and $H_1(z)$, and the tolerable system delay T_d.
Output: The synthesis filters $F_0(z)$ and $F_1(z)$, and the worst-case reconstruction error J_{opt}.
Procedure:

1. Construct the polyphase matrix $\mathbf{E}(z)$ associated with the analysis filters $H_0(z)$ and $H_1(z)$.
2. Construct the delay matrix

$$\mathbf{C}(z) \triangleq \begin{cases} z^{-k}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, & \text{if } T_d = 2k+1, \\ z^{-k}\begin{bmatrix}0 & z\\ 1 & 0\end{bmatrix}, & \text{if } \quad T_d = 2k. \end{cases}$$

3. Find $\mathbf{R}_{opt}(z) \in \mathcal{H}_\infty$ which minimizes $J = ||\mathbf{C} - \mathbf{RE}||_\infty$.
4. Return $[F_0(z)\ F_1(z)] = \mathbf{e}^T(z)\mathbf{R}_{opt}(z^2)$.
5. Return $J_{opt} = ||\mathbf{C} - \mathbf{R}_{opt}\mathbf{E}||_\infty$.

5.6.3 Example designs

Assume that D takes integer values only. In this very simple case, two pairs of synthesis filters are sufficient for all TDOA circumstances since $\tilde{D}$ is either 0 or 1. Here, we provide three design examples for this simple case assuming that the sensor filters are those designed in Section 5.4.

The $\mathcal{H}_\infty$-optimal synthesis filters $F_0(z)$ and $F_1(z)$ (designed for even- and odd-TDOA values separately) are shown in Fig. 5.9. The total system delay T_d which was chosen for each case is also quoted in this figure. The worst-case reconstruction error norm J_{opt} for the synthesis filter bank pairs depicted in Fig. 5.9 are shown in Table 5.1.

The figures reported in Table 5.1 are quite impressive once we note that the reconstruction error norm J_{opt} represents *the worst-case scenario* and the

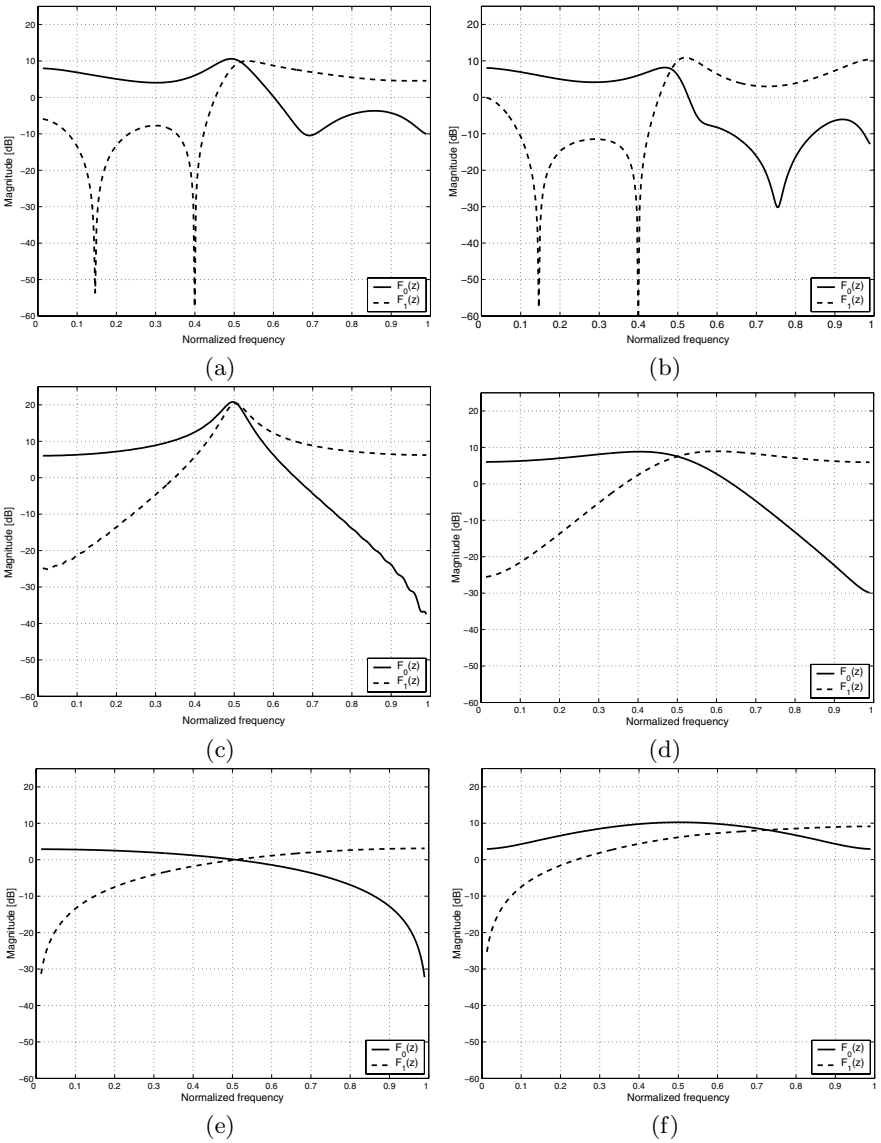


Fig. 5.9. $\mathcal{H}_\infty$ optimal synthesis filters designed for each pair of analysis filters introduced in the design examples of Section 5.4. The right hand plot in each row shows the synthesis filters $F_i(z)$ used for even TDOA values while those on the left are $G_i(z)$ used for odd TDOA values. (a) and (b): ordinary linear-phase FIR analysis filters, $T_d = 27$ samples. (c) and (d): Bessel IIR analysis filters, $T_d = 35$ samples. (e) and (f) Perfect Reconstruction linear-phase FIR analysis filters, $T_d = 7$ samples.

Table 5.1. Worst-case reconstruction error norm J_{opt} for the optimal synthesis filters shown in Fig. 5.9.

Type of analysis filters	J_{opt} (even TDOA values)	J_{opt} (odd TDOA values)
ordinary linear-phase FIR	-70.4 dB	-65.3 dB
Bessel IIR	- 54.6 dB	less than -80 dB
Linear-phase FIR filter banks	less than -80 dB	less than -80 dB

actual reconstruction error for a concrete case can be much less. Moreover, the reader is reminded that the peak reconstruction error J_{opt} depends on both the analysis filters and the value chosen for T_d. Chen and Francis (Chen and Francis, 1995a, Theorem 4.1) show that under a mild condition[2] on the analysis filters $\lim_{T_d \to \infty} J_{opt} = 0$. This means arbitrary good reconstruction is possible if a sufficiently large time delay is tolerated. We were able to achieve $J_{opt} \leq -80$dB for all the example analysis filter banks introduced in Section 5.4 by choosing a large enough T_d.

5.7 Open problems

In this chapter we studied the effects of time-delay in multirate sensor systems. First, we extended a well-known method of time delay estimation to multirate signals and showed that it can lead to reliable TDOA estimations even where reverberations and noise are present. Then, we turned our focus to the design of practical analysis and synthesis filters to allow signal fusion when TDOA is present.

The algorithms and techniques presented in this chapter where developed with reference to a simple model involving only two sensors. We did not address signal fusion in M-channel systems with $M > 2$ neither did we discuss important issues such as the accuracy of the low-rate TDOA estimator and optimal selection of sensor filters to facilitate both TDOA estimation and signal fusion. These remain as challenging topics for future research.

[2] The condition is that the polyphase matrix $\mathbf{E}(e^{j\omega})$ associated with the analysis filters must be nonsingular for all ω.

6

Optimal Multirate Decomposition of Signals

6.1 Introduction

In many multimedia applications, it's desirable to store or broadcast a signal in *multiple resolutions*. This is to say, it is desirable to have a signal representation where various "coarse" or "fine" approximations to the signal are readily retrieved.

In this chapter, we examine the problem of decomposing an information bearing signal $x(n)$ into a set of lower-rate *subband* components $v_0(n)$, $v_1(n)$, ..., $v_{M-1}(n)$ for efficient storage or broadcasting purposes.

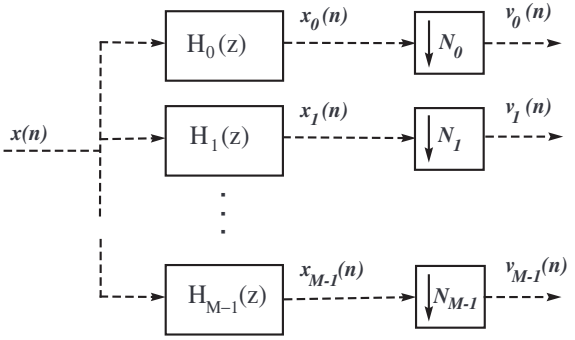


Fig. 6.1. An M-channel multirate analysis filter bank.

An analysis filter bank as shown in Fig. 6.1 is used to perform the decomposition. We assume that at the receiving end or during retrieval, some or all of the low-rate components $v_0(n)$, $v_1(n)$, etc. will be used to reconstruct an approximation $y(n)$ to the original signal $x(n)$. The approximation $y(n)$ is obtained via a synthesis filter as shown in Fig. 6.2.

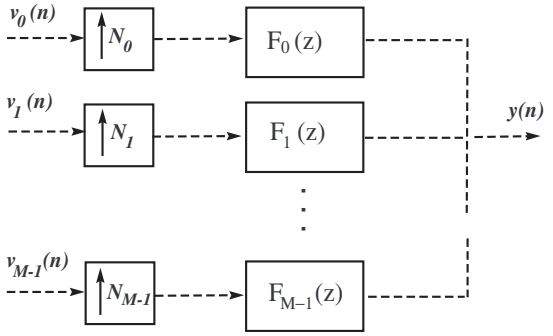


Fig. 6.2. An M-channel synthesis filter bank.

Let $y_0(n)$ denote the output of the synthesis filet bank when only the subband component $v_0(n)$ is present. Then, define $y_1(n)$ as the approximation obtained by feeding *both* subband signals $v_0(n)$ and $v_1(n)$ to the synthesis filter bank. Similarly, define $y_2(n)$, $y_3(n)$, ..., $y_{M-1}(n)$ to denote the approximations obtained by including successively more subbands.

A very interesting question is whether it is possible to design the filters $H_i(z)$ and $F_i(z)$ such that *all of* the approximations $y_0(n), y_1(n), \cdots, y_{M-1}(n)$ are as close to the original measurement signal $x(n)$ as possible. When the answer is yes, the resulting decomposition is called *scalable* or *optimal in the sense of scalability*[1].

When $x(n)$ is a stationary random process, Tsatsanis and Giannakis (1995) demonstrated that it is possible to achieve scalable decomposition using what they called a *Principal Component Filter Bank*. A Principal Component Filter Bank, or PCFB for short, is an M-channel uniform filter bank (i.e. $N_0 = N_1 = \cdots = N_{M-1} = M$) that decomposes the input signal into M low-rate components. It has the property that all the approximations $y_0(n)$, $y_1(n)$, etc. obtained by feeding successively more components $v_0(n)$, $v_1(n)$, etc. to its synthesis filter bank are optimal in the mean-squared-error sense[2]. However, the analysis and synthesis filter of the PCFB have ideal straight-wall frequency response characteristics. As a result, it cannot be realized using practical FIR or IIR filters.

[1] In the source compression and information theory literature, the broad concept of scalability is also described by such terms as *progressive transmission*, *multi-resolution coding*, *successively refinable coding*, *multi-scale representation* and so on, c.f. Bull et al. (1999) and Goyal (2001).

[2] In general, the definition of the best approximation depends on the specific application and the type of signals to be dealt with. In a deterministic setting, where $x(n)$ is in the space $l^2(\mathbb{Z})$, different techniques for achieving a scalable decomposition can be envisioned; c.f. Mallat (1999) and references therein.

The concept of scalability is inherently associated with a multi-objective optimization problem. Naturally, the objective functions to be minimized might not admit a common solution. The existence of the PCFB proves that a common solution always exist if we do not impose realizability constraints on the filters'frequency response. However, if we require the filters in our system to be realizable, then it turns out that in many cases a scalable decomposition is not possible[3].

Since in general it is not possible to achieve a scalable decomposition using practical filter banks, it is very desirable to develop a *quantitative* notion of scalability. That is, it is desirable to be able to measure *how scalable* a decomposition produced by a certain filter bank is. The main goal of this chapter is to develop such a theory.

To construct a quantitative theory of scalability, we follow a rigorous algebraic approach: We consider a specific set $\mathcal{L}$ of analysis filter banks. For simplicity, we use the adjoint adjoint of the analysis filter banks for synthesis. Next, we endow the set $\mathcal{L}$ with an algebraic ordering that represents scalability. Then, we establish an equivalence relation between the order of scalability and another order called majorization. Using either ordering, specifying the best (optimal) analysis system in $\mathcal{L}$ translates to finding the so-called *greatest element* associated with the order.

It turns out that the ordering of scalability is partial. While we prove that this partial ordering has maximal elements in $\mathcal{L}$, the existence of a greatest element is not guaranteed. We move forward by embedding the partial ordering of scalability in a simple (total) ordering. This is done by using a special class of functions known as Schur-convex functions. Using the simple ordering generated on $\mathcal{L}$ by these functions, we are able to introduce a formal way for measuring the scalability of the filter banks in $\mathcal{L}$. As is always the case in real life, there is catch to this achievement: There are infinitely many possible choices for Schur-convex functions and our measure of scalability depends on this (subjective) choice!

Remark 6.1. For reasons of rigor and clarity, our usage of some terminology is more precise than that in the existing literature. For instance, we shall reserve the term PCFB for the (non-realizable) system defined in section 6.5.1 only. If a filter bank exists in $\mathcal{L}$ better than all other elements in the sense of scalability, it will be called an SC-optimal filter bank and not a PCFB.

The sections in the rest of this chapter are organized as follows: In the next section, we introduce some basic concepts pertaining to FIR analysis and synthesis filter banks. We also define the admissible class $\mathcal{L}$ of the filter banks that we shall deal with in the rest of the chapter. The main body of the algebraic theory of scalability starts with Section 6.3 where we establish a partial ordering representing scalability on the set $\mathcal{L}$. In Section 6.4, we embed this partial

[3] See Example 6.2 in this chapter and Kirac and Vaidyanathan (1998a) for counter examples.

ordering in a total order by appealing to a set of order-preserving functions. We explore the connection between the quantitative theory of scalability developed in Sections 6.3 and 6.4 and the theory of optimum subband coding in Section 6.6. In Section 6.5 we review in more detail the connection between the concepts of optimum scalability in finite-order systems and the PCFB. This section shows that the quantitative notion of scalability developed in Sections 6.3.2 and 6.4 is consistent in the sense that it has an interpretation as approximating PCFB using filter banks in $\mathcal{L}$. Section 6.7 contains complementary material including possible generalizations and the connection between our work and the existing literature on the design of optimal filter banks.

The material in this chapter are based on the paper by Jahromi et al. (2003).

6.2 Review of FIR filter banks

6.2.1 Some basic notions

An M-channel analysis filter bank is shown in Fig. 6.1. The filters $H_0(z)$ to $H_{M-1}(z)$ together with the decimators following them generate the low-rate subband signals $v_0(n), v_1(n), \ldots, v_{M-1}(n)$. In this chapter, we assume the same down-sampling rate $N_i = 1/M$ for all the channels. Thus, sampling rate of each subband signal is $1/M$ of the sampling rate of the input signal $x(n)$.

In the synthesis filter bank (Fig. 6.2) the subband signals are up-sampled to the original rate using M expanders and then combined, after passing through the synthesis filters $F_0(z)$ to $F_{M-1}(z)$, to generate the output signal $y(n)$. To learn more about the basic theory of multirate filter banks, see the excellent books by Vaidyanathan (1993), Fliege (1994) or Mertinz (1999).

A filter bank for which the output signal $y(n)$ is an exact reproduction of the input $x(n)$ (within, perhaps, a multiplicative constant and a delay) is called a *perfect reconstruction* (PR) filter bank. An interesting class of PR filter banks are those whose analysis and synthesis filters satisfy an orthogonality condition. Such filter banks, called *orthogonal* or *paraunitary* filter banks, admit several nice properties which make their analysis, design and implementation much easier. In this chapter, we consider this class of filter banks only.

6.2.2 Orthogonal FIR filter banks (the class $\mathcal{L}$)

We denote by $\mathcal{L}$ the class of M-channel orthogonal FIR analysis filters of order less than or equal to $N = (K+1)M$, where K is a fixed integer. To specify this class more precisely, it is convenient to represent the filters $H_i(z)$ in the *polyphase form*. In this representation, the analysis filters are packed in a transfer vector

$$\mathbf{h}(z) \triangleq [H_0(z)\ H_1(z)\ \cdots\ H_{M-1}(z)]^T.$$

Then, $\mathbf{h}(z)$ is factorized as the product of a matrix transfer function and a vector transfer function consisting only of delay elements:

$$\mathbf{h}(z) = \mathbf{E}(z^M)\mathbf{e}(z) \tag{6.1}$$

In the above formula, $\mathbf{e}(z) \triangleq [1\ z^{-1}\ \cdots\ z^{-(M-1)}]^T$ and $\mathbf{E}(z)$ is called the *polyphase matrix* associated with the analysis filter bank $\mathbf{h}(z)$.

Definition 6.1. $\mathbf{h}(z) \in \mathcal{L}$ *if its polyphase matrix* $\mathbf{E}(z)$ *satisfies the following conditions*

(i) $\mathbf{E}(z)$ *is paraunitary, i.e.,* $\mathbf{E}(z)^{-1} = \mathbf{E}(z^{-1})^T$.
(ii) $\mathbf{E}(z)$ *is FIR of order at most* K.

The filter banks in the class $\mathcal{L}$ possess many desirable properties (Vaidyanathan, 1993) . The following properties will be used later in this chapter:

1. Implementation based on rotation matrices: The matrix $\mathbf{E}(z)$ can be factored as

$$\mathbf{E}(z) = \mathbf{A}_K\mathbf{D}(z)\mathbf{A}_{K-1}\mathbf{D}(z)\ldots\mathbf{D}(z)\mathbf{A}_0 \tag{6.2}$$

where

$$\mathbf{D}(z) = \begin{bmatrix} I_{M-1} & 0 \\ 0 & z^{-1} \end{bmatrix}$$

and $\mathbf{A}_i$ are constant orthogonal matrices. Not all of these matrices have to be general unitary matrices in order for the factorization in (6.2) to cover all possible orthogonal filter banks in $\mathcal{L}$. In fact, the matrices $\mathbf{A}_0, \cdots, \mathbf{A}_{K-1}$ have only to belong to the subset of all possible $M \times M$ unitary matrices that can be written as a sequence of $M-1$ Givens rotations (Vaidyanathan, 1993, Section 14.6). The last matrix $\mathbf{A}_K$ has to be a general $M \times M$ unitary matrix which, in turn, can be decomposed into $M(M-1)/2$ Givens rotations. This leads to the following theorem:

Theorem 6.1. *A filter bank* $\mathbf{h}$ *in* $\mathcal{L}$ *is uniquely parameterized by* P *rotation angles, where* $P = K(M-1) + M(M-1)/2$. *Thus, elements of* $\mathcal{L}$ *are indexed by the points in the closed and bounded cube* $\Gamma = [-\pi, \pi]^P \subset \mathbb{R}^P$.

Note that the factorization presented here is not canonical. This means different parameter vectors in Γ can lead to the same transfer vector $\mathbf{h}(z) \in \mathcal{L}$. For a discussion on canonical realization of lossless transfer functions see (Dewilde and van der Veen, 1998, Chapter 14).

2. The adjoint system is FIR and achieves PR: The factorization introduced above suggests that the corresponding synthesis polyphase matrix can be designed as

$$\mathbf{R}(z) = \mathbf{A}_0^T \mathbf{C}(z) \mathbf{A}_1^T \mathbf{C}(z) \dots \mathbf{C}(z) \mathbf{A}_K^T \tag{6.3}$$

where

$$\mathbf{C}(z) = \begin{bmatrix} z^{-1} I_{M-1} & 0 \\ 0 & 1 \end{bmatrix}.$$

Observe that $\mathbf{R}(z)$ is the adjoint of $\mathbf{E}(z)$ and FIR too! The synthesis filters $F_0(z)$ to $F_{M-1}(z)$ are then obtained from

$$\mathbf{f}(z) = \mathbf{e}^T(z) \mathbf{R}(z^M) \tag{6.4}$$

where $\mathbf{f}(z) \triangleq [F_0(z) \; F_1(z) \; \cdots \; F_{M-1}(z)]$. It's easy to verify that, by the above choice of the synthesis filters, the overall analysis/synthesis system is PR within a delay of $D = MK + (M-1)$ samples.
When analysis filters are specified, the synthesis filters are given by (6.4). For filter banks in the class $\mathcal{L}$, therefore, design can be carried out by defining an objective function and optimizing the P rotation angles that parameterize the analysis bank.

3. Making a perfect transmultiplexer: Switching the role of analysis and synthesis banks, one can use a multirate filter bank as a *transmultiplexer* (Vaidyanathan, 1993, Sec. 5.9). It is easy to verify that filter banks in $\mathcal{L}$ generate perfect reconstruction transmultiplexers. This property of the class $\mathcal{L}$ is used in Section 6.3.2 to show a direct connection between scalability and power distribution.

4. Power preservation: When driven by a stationary zero-mean stochastic signal, the filter banks in $\mathcal{L}$ preserve the input signals' power (variance) within a multiplicative constant. In other words, the input power is distributed among the subband signals such that

$$\sum_{i=0}^{M-1} E\{x_i^2(n)\} = M \times E\{x^2(n)\}. \tag{6.5}$$

6.3 Scalability in the class $\mathcal{L}$

We mentioned in the introduction section that a multirate decomposition is scalable if subband signals generated by it, and proper combinations of several of them, provide best possible approximations to the original full-rate signal. In this section we use the mathematical concept of majorization (introduced in Section 2.2) to give a precise definition for scalability.

6.3.1 Ordering the filter banks in $\mathcal{L}$ based on their scalability

Consider the setup shown in Fig. 6.3 where the analysis filter bank belongs to class $\mathcal{L}$. Assume that the input signal $x(n)$ is wide-sense stationary (WSS)

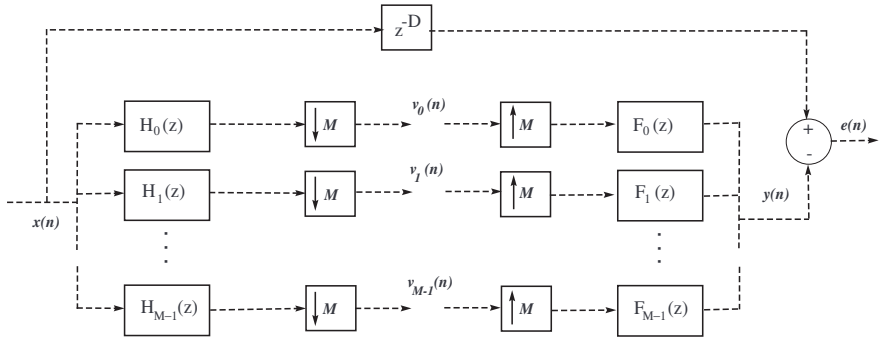


Fig. 6.3. A setup which shows how the approximation error is measured for scalability analysis.

with zero mean and known power spectral density $P_x(e^{j\omega})$. The delay D is introduced to synchronize the original signal with the filter bank output.

Let $e_i(n) = x(n-D) - y_i(n),\ 0 \le i \le M-1$, represent the reconstruction error when only subband signals v_0 to v_i are used to synthesize the output. Obviously, when the filter banks are Perfect Reconstruction, $e_{M-1}(n) = 0$. In general the error signals $e_i(n)$ are cyclostationary random processes with period M. To measure their expected power, therefore, one needs to consider both time and ensemble averaging. In view of this fact, we define

$$\epsilon_i \stackrel{\triangle}{=} \frac{1}{M} \sum_{k=0}^{M-1} E\{e_i^2(n-k)\}$$

to represent the *expected error power* when only i out of M subband signals, namely, v_0 to v_{i-1}, are used for reconstruction. It's convenient to pack the set of all expected error powers ϵ_i associated with $\mathbf{h}$ in the *error vector*:

$$\mathbf{e_h} \stackrel{\triangle}{=} [\epsilon_{M-1}\ \epsilon_{M-2}\ \cdots\ \epsilon_0]^T. \tag{6.6}$$

The following definition allows us to formally rank the filter banks in $\mathcal{L}$ based on their error vectors:

Definition 6.2. (Scalability) *A filter bank* $\mathbf{h}_1(z) \in \mathcal{L}$ *is more scalable*[4] *than* $\mathbf{h}_2(z) \in \mathcal{L}$, *in symbols* $\mathbf{h}_1(z) \stackrel{SC}{\ge} \mathbf{h}_2(z)$, *if* $\mathbf{e}_{\mathbf{h}_1} \le \mathbf{e}_{\mathbf{h}_2}$.

The binary relation $\stackrel{SC}{\ge}$ introduced in the above definition induces an ordering on the set $\mathcal{L}$. One might easily verify that it satisfies the conditions of a partial ordering as discussed in Section 2.2. Now, we are ready to search in $\mathcal{L}$ for a filter bank more scalable than all others. If such a filter bank exists, we would call optimal in the sense of scalability or *SC-optimal* for short.

[4] The modifier *more* used here is meant to imply equality as well.

Definition 6.3. (SC-optimality) *A filter bank* $\mathbf{h}^*(z) \in \mathcal{L}$ *is optimal in terms of scalability (SC-optimal) if*

$$\mathbf{h}^*(z) \overset{SC}{\geq} \mathbf{h}(z), \ \forall \ \mathbf{h}(z) \in \mathcal{L}.$$

In the next subsection we consider the issue of existence of $\mathbf{h}^*(z)$. For reasons that will become clear later, we do this in an indirect way through a new, yet equivalent, partial ordering on $\mathcal{L}$.

6.3.2 Scalability in terms of power distribution

In this subsection, we compare the analysis filter banks in class $\mathcal{L}$ in terms of the way they distribute the input signal's power among the M subband signals. We show that distribution of power among the subbands is intimately connected to the scalability of a filter bank.

Again, consider the setup in Fig. 6.3, where the analysis filter bank belongs to class $\mathcal{L}$ and the input $x(n)$ is WSS with power spectral density $P_x(e^{j\omega})$. It is straightforward to calculate the variances of the subband signals $x_i(n)$. Denote by $\mathbf{p_h}$ the vector whose components are these variances; that is,

$$\mathbf{p_h} = [E\{x_0^2(n)\} \ E\{x_1^2(n)\} \ \cdots \ E\{x_{M-1}^2(n)\}]^T.$$

The set of all variance vectors $\mathbf{p_h}$ generated by the whole class $\mathcal{L}$ will be denoted by Ω: $\Omega = \{\mathbf{p_h} | \mathbf{h} \in \mathcal{L}\}$. Each $\mathbf{h}(z) \in \mathcal{L}$, therefore, produces some[5] $\mathbf{p_h} \in \Omega$. We can now establish the following theorem:

Theorem 6.2. *For any two filter banks* $\mathbf{h}_1(z)$ *and* $\mathbf{h}_2(z)$ *in* $\mathcal{L}$, $\mathbf{h}_1(z) \overset{SC}{\geq} \mathbf{h}_2(z)$ *if and only if* $\mathbf{p}_{\mathbf{h}_1} \succeq \mathbf{p}_{\mathbf{h}_2}$.

Proof. Consider the setup shown in Fig. 6.4. In this figure, an extra analysis filter bank is used to decompose the error signal $e_i(n)$ into M low-rate error signals. This is taken to be the same analysis bank used to decompose $x(n)$. Using the power preservation property of class $\mathcal{L}$, it is not difficult to show that

$$\epsilon_i = \sum_{j=0}^{M-1} E\{u_j^2(n)\}, \forall i. \tag{6.7}$$

The error signal $e_i(n)$ is made up of two components, a delayed version of the input signal $x(n)$ and the reconstructed signal $y(n)$. We recall that filter banks in the class $\mathcal{L}$ make perfect transmultiplexers. Thus, the contribution of $y(n)$ to the error subbands will be, within a delay D and a sign change, those subband signals that *were not dropped* while synthesizing $y(n)$. Contribution of $x(n-D)$, however, will produce error subbands which are, within a delay D,

[5] Note that the mapping between $\mathcal{L}$ and Ω is not one-to-one. Different $\mathbf{h} \in \mathcal{L}$ can lead to the same $\mathbf{p_h} \in \Omega$.

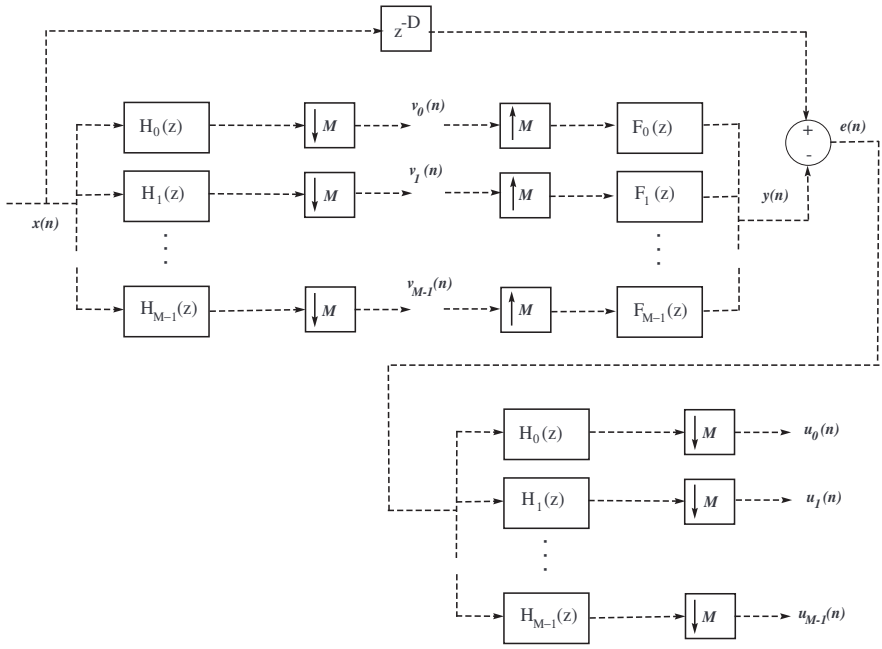


Fig. 6.4. Setup to show the relation between scalability and power distribution.

exactly the same as those available at the output of the original analysis bank. In other words, when subband signals $v_0(n)$ to $v_{i-1}(n)$ are used to produce $y(n)$, we have

$$u_j(n) = \begin{cases} 0 & \text{if } 0 \le j \le i-1 \\ v_j(n-D) & \text{if } i \le j \le M-1 \end{cases}. \tag{6.8}$$

The signals $v_j(n)$ and $u_j(n)$ are both stationary. From (6.7) and (6.8), it then follows that

$$\epsilon_i = \sum_{j=i}^{M-1} E\{v_j^2(n)\}. \tag{6.9}$$

The above formula, and the fact that $E\{v_j^2(n)\} = E\{x_j^2(n)\}$, allow us to expand the relation $\mathbf{e}_{\mathbf{h}_1} \le \mathbf{e}_{\mathbf{h}_2}$ as

$$\sum_{j=i}^{M-1} E\{x_j^2(n)\} \; for \; \mathbf{h}_1 \le \sum_{j=i}^{M-1} E\{x_j^2(n)\} \; for \; \mathbf{h}_2, \quad \forall i.$$

For filter banks in $\mathcal{L}$ the total sum $\sum_{j=0}^{M-1} E\{x_j^2(n)\}$ is a constant. Hence, the above result can be modified as

$$\mathbf{e}_{\mathbf{h}_1} \le \mathbf{e}_{\mathbf{h}_2} \Longleftrightarrow \sum_{j=0}^{i-1} E\{x_j^2(n)\} \; for \; \mathbf{h}_1 \ge \sum_{j=0}^{i-1} E\{x_j^2(n)\} \; for \; \mathbf{h}_2, \quad \forall i; \tag{6.10}$$

which, in the language of majorization, means

$$\mathbf{e}_{\mathbf{h}_1} \leq \mathbf{e}_{\mathbf{h}_2} \Longleftrightarrow \mathbf{p}_{\mathbf{h}_1} \succeq \mathbf{p}_{\mathbf{h}_2}. \tag{6.11}$$

□

The above theorem has important consequences. Recall from Section 2.2 that majorization is a measure of nonuniformity. Hence, $\mathbf{p}_{\mathbf{h}_1} \succeq \mathbf{p}_{\mathbf{h}_2}$ means that $\mathbf{h}_1(z)$ distributes the input signal's power among the subbands more nonuniformly than $\mathbf{h}_2(z)$ does. Thus, Theorem 6.2 states that a filter bank is good in the sense of scalability when it distributes the input signal's power among the subbands in a very nonuniform way.

One might use the notion of majorization introduced above to induce an ordering on the set Ω. It is easily verified that this ordering satisfies all the requirements of a partial ordering as well. From Definition 6.3 and Theorem 6.2 we then conclude that an optimal filter bank in $\mathcal{L}$ is associated with the greatest element in Ω. This is to say,

Lemma 6.1. *An element $\mathbf{h}^*(z) \in \mathcal{L}$ is SC-optimal if and only if $\mathbf{p}_{\mathbf{h}^*}$ is the greatest element of Ω.*

The greatest element $\mathbf{p}_{\mathbf{h}^*}$ is necessarily a maximal element of Ω (see Theorem 2.3). Hence, to find the greatest element, one has to search among the maximal elements only. The following theorem asserts that maximal elements always exist in Ω.

Theorem 6.3. *The set Ω has a maximal element.*

Proof. Our proof is by construction and consists of three steps: First, we observe that components of $\mathbf{p}_{\mathbf{h}}$ are continuous functions of a parameter vector $V \in \Gamma \subset \mathbb{R}^P$. Then, we show that the level sets of these functions are closed and bounded subsets of the domain set Γ themselves. Finally, we use this result to construct a maximal element in Ω.

Step I: Recall from Proposition 6.1 that a filter bank $\mathbf{h}_1(z) \in \mathcal{L}$ is uniquely parameterized by P rotation angles where $P = K(M-1) + M(M-1)/2$. In other words, elements of $\mathcal{L}$ are indexed by the points in the closed and bounded cube $\Gamma = [-\pi, \pi]^P \in \mathbb{R}^P$.

From the discussion in section 6.2.2 we also observe that:

(i) Each element of the polyphase matrix $\mathbf{E}(z)$ is itself a polynomial in z^{-1} whose coefficients depend on the vector $V = [\theta_0, \theta_1, \ldots, \theta_{P-1}]^T$ of rotation angles.

(ii) The dependence in Item 1 is through a combination of sine, cosine and polynomial functions only.

(iii) Let the vector $L_i = [l_0 \; l_1 \; \cdots \; l_{N-1}]^T$ denote the impulse response of the filter $H_i(z)$. Components of L_i are functions of V through (the time-domain version of) (6.1). These components are, therefore, polynomials in the sine and cosine of θ_j as well.

(iv) The observation in Item 3 implies that each component of L_i is a *continuous* function of V.
(v) Variance of the *ith* subband signal is a quadratic function of the impulse response vector L_i. That is, $E\{x_i^2(n)\} = L_i^T \mathbf{R}_{xx} L_i$, where $\mathbf{R}_{xx}$ is the $N \times N$ autocorrelation matrix of $x(n)$.

The last observation together with the one in Item 4 justifies that the components of $\mathbf{p_h}$, that is, $E\{x_i^2(n)\}$, are all continuous functions of $V \in \Gamma$. For simplicity, let's denote components of $\mathbf{p_h}$ by p_i and their functional relation with V by $f_i(.)$. That is, $\mathbf{p_h} = [p_0\ p_1\ \cdots\ p_{M-1}]^T$ where $p_i = f_i(V)$.
Step II: Recall the following two results from the theory of continuous functions (Taylor and Mann, 1972):

Theorem 6.4. *Let Γ be a non-empty bounded and closed set, and suppose f is a function defined on Γ and continuous at each point of Γ. Let m and M be the greatest lower bound and least upper bound, respectively, of the values of f on Γ. Then there is some point of Γ at which f has the value M, and there is also a point at which f has the value m.*

Theorem 6.5. *Level curves of a continuous function are closed. In other words, $S = \{X : f(X) = c\}$ is a closed set where f is continuous on its domain.*

Step III: We use the results of the previous two steps to construct a maximal element in Ω. Consider the first component of $\mathbf{p_h}$, that is, $p_0 = f_0(V)$. Let $\Gamma_0 \subset \Gamma$ denote the set of all points in Γ for which $f_0(V)$ attains its maximum. Call this maximum p_0^{Max}. Theorem 6.4 guarantees that Γ_0 is not empty. It then follows from continuity of $\mathbf{f}(V)$ that Γ_0 is closed and bounded as well (Theorem 6.5 above).
Now, consider the second component of $\mathbf{p_h}$, namely, $p_1 = f_1(V)$. The function $f_1(V)$ and the set Γ_0 satisfy the conditions of Theorem 6.4. Therefore, for the points in a (nonempty) closed and bounded set $\Gamma_1 \subset \Gamma_0$, the function $f_1(V)$ attains its maximum value p_1^{Max}. Proceeding in this way, we can find nonempty, closed and bounded sets $\Gamma_i \subset \Gamma_{i-1}$ that maximize the functions $f_i(V)$ over Γ_{i-1} and, accordingly, specify the maximum values p_i^{Max} for $2 \leq i \leq M-1$. The maximum values found this way constitute an element $\mathbf{p}_{Max} = [p_0^{Max}\ p_1^{Max} \cdots p_{M-1}^{Max}]^T \in \Omega$.
By its very construction, no point in Ω can majorize $\mathbf{p}_{Max}$. Therefore, it is a maximal element of Ω. □

Remark 6.2. Note that more than one $\mathbf{h}(z) \in \mathcal{L}$ might have $\mathbf{p_h} = \mathbf{p}_{Max}$. We regard all such $\mathbf{h}(z)$ as maximal elements of $\mathcal{L}$.

Although Theorem 6.3 confirms the existence of a maximal element in Ω, it can be shown that this element does not always satisfy the requirements of a greatest element. In fact, one can find examples[6] for which several maximal

[6] See Example 6.2 in the next section.

elements exist in Ω and, by Theorem 2.3, this can only happen when a greatest element doesn't exist.

In summary, the best one can do, in general, is to find a maximal element in $\mathcal{L}$. Theoretically, this can be done using the procedure outlined in the proof of Theorem 6.3. However, this is not an easy thing to do numerically. The difficulty arises because there does not exist any systematic numerical method to specify the level sets Γ_i as required by this procedure. Luckily, there exist alternative methods for characterizing a maximal element. One such method is provided later in Theorem 6.6.

In the next section, we consider converting the partial ordering of scalability into (non-unique) total orderings. This relaxes our notion of SC-optimality, originally stated in Definition 6.3, and solves the problem of non-existence of an optimal solution. The price we have to pay, however, is the subjectivity introduced in the definition of an optimal solution. Now, one has to choose among an infinite number of possibilities for defining an optimal solution!

6.4 Embedding the ordering of scalability in a total ordering

We showed in the previous section that the ordering introduced on $\mathcal{L}$ by scalability is associated with the ordering induced by majorization on Ω. It is sufficient, therefore, to study the implications of the latter ordering. Recall that majorization is a partial ordering. Therefore, it does not always specify an SC-optimal solution in $\mathcal{L}$. In this section, we embed the partial ordering of majorization in a total ordering by introducing appropriate order-preserving functions.

By choosing a function ϕ which preserves the order of majorization, we can come up with a more relaxed definition of scalability. That is to say, we can rank the elements in Ω based on the value of an order-preserving function rather than comparing its elements directly via the relation $\succeq$. This way, we can always specify an optimal point in Ω. The problem now is that the optimal solution might change if we use a different order preserving function!

6.4.1 SC_ϕ-optimality

Definition 6.4. (Scalability with respect to ϕ) *Let ϕ be strictly Schur-convex. A filter bank $\mathbf{h}_1(z) \in \mathcal{L}$ is more scalable than $\mathbf{h}_2(z) \in \mathcal{L}$ with respect to ϕ, in symbols $\mathbf{h}_1(z) \overset{\phi}{\geq} \mathbf{h}_2(z)$, if $\phi(\mathbf{p}_{\mathbf{h}_1}) \geq \phi(\mathbf{p}_{\mathbf{h}_2})$.*

This new definition has the important advantage that it induces a simple (total) ordering on $\mathcal{L}$. This is because, now, any two elements in $\mathcal{L}$ are comparable using this ordering:

$$\mathbf{h}_1, \mathbf{h}_2 \in \mathcal{L} \quad \Longrightarrow \quad \phi(\mathbf{p}_{\mathbf{h}_1}) \geq \phi(\mathbf{p}_{\mathbf{h}_2}) \text{ or } \phi(\mathbf{p}_{\mathbf{h}_2}) \geq \phi(\mathbf{p}_{\mathbf{h}_1}).$$

It is clear that the set $\mathcal{L}$ with the ordering induced by the binary relation $\overset{\phi}{\geq}$ constitutes a chain. The greatest element of this chain is called *SC-optimal with respect to function* ϕ, or SC_ϕ*-optimal* for short.

In the following example we explore the geometry of some typical Schur-convex functions.

Example 6.1. **(Geometry of Schur-convex functions)** Consider a 3-channel filter bank in the class $\mathcal{L}$. In this case, $M = 3$ so the variance vector $\mathbf{p}$ will have three components, namely $\mathbf{p} = [p_0\ p_1\ p_2]$ where $p_i = E\{x_i^2(n)\},\ i = 0, 1, 2$.

Recall that $\mathbf{p} \in \mathbb{R}_+^3$. Since $p_0 + p_1 + p_2$ is a constant, assumed to be 1 here, the set Ω of all error vectors $\mathbf{p}$ will be located on the intersection of the plane $p_0 + p_1 + p_2 = 1$ and $\mathbb{R}_+^3$. This intersection forms a triangle which we call Δ.

The level sets of Schur-convex functions are contours on the two-dimensional plane that includes Δ. This is shown in Fig. 6.5 for three strictly Schur-convex functions. The horizontal and vertical scales in Fig. 6.5 represent an orthogonal coordinate system on the plane of Δ. The coordinates are

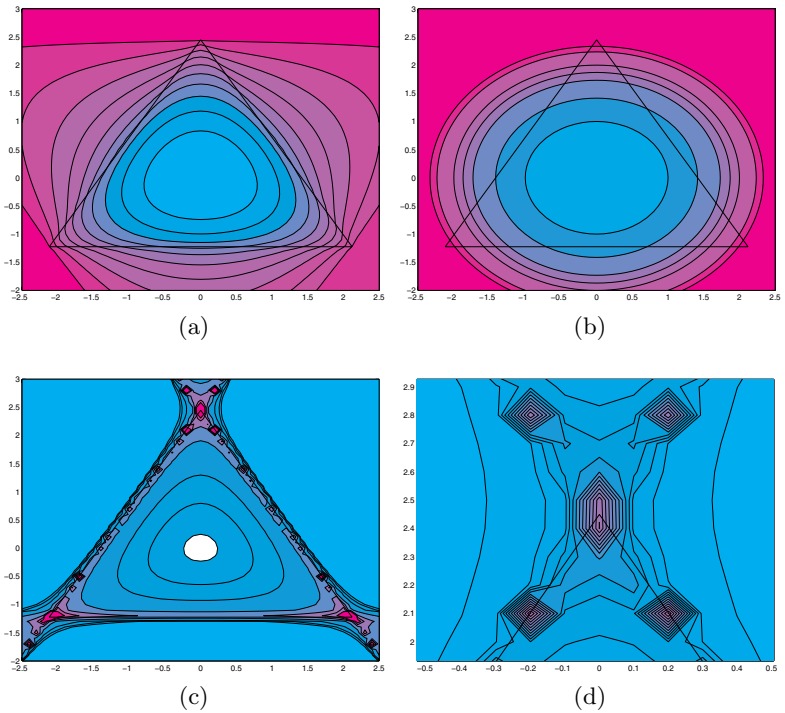


Fig. 6.5. Contours generated on the surface of the triangle Δ by typical Schur-convex functions: (a) Negative entropy, (b) Euclidean norm, (c) Coding Gain, (d) Top-central part of (c) enlarged to show singularity of Coding Gain on the sides of Δ. Contours that lay outside Δ are irrelevant.

chosen such that their center coincides with the center of Δ but are otherwise arbitrary.

The contours in Fig. 6.5(a) represent the (negative) entropy

$$\phi_1(\mathbf{p}) = \sum_{i=0}^{2} p_i \log(p_i),$$

while those in Fig. 6.5(b) represent the familiar Euclidean norm

$$\phi_2(\mathbf{p}) = \sum_{i=0}^{2} p_i^2.$$

Fig. 6.5(c) shows contours associated with the *Coding Gain*[7] function

$$\phi_3(\mathbf{p}) = \frac{p_1 + p_2 + p_3}{3\sqrt[3]{p_1 \times p_2 \times p_3}}.$$

It is easily observed that, for the points on Δ, all these functions achieve their minimum at the central point and their maximum at any of the three corners[8]. The corners of Δ represent the vectors $[1\ 0\ 0]^T, [0\ 1\ 0]^T$ and $[0\ 0\ 1]^T$. These vectors (points) majorize all other points in Δ. The central point $1/\sqrt{3}[1\ 1\ 1]^T$ is majorized by all other points in Δ. ♢

Note from Fig. 6.5 that the sample Schur-convex functions we presented generated fairly similar contours on the triangle S (at least as long as we are not very close to the corners). This means that the SC_ϕ-optimal solutions obtained via different Schur-convex functions ϕ tend to be close to each other although they are not necessarily the same.

As we mentioned earlier, using the relaxed definition of scalability, the optimal filter bank picked in $\mathcal{L}$ depends on the choice of the specific Schur-convex function used. In fact, each SC_ϕ-optimal solution is a maximal element of Ω:

Theorem 6.6. *Let $\mathbf{h}^* \in \mathcal{L}$ be an SC_ϕ-optimal filter bank. Then, $\mathbf{h}^*$ is a maximal element of $\mathcal{L}$.*

Proof. Assume that there exist $\mathbf{h}_1 \in \mathcal{L}$ deferent from $\mathbf{h}^*$ such that $\mathbf{p}_{\mathbf{h}_1} \succeq \mathbf{p}_{\mathbf{h}^*}$. This implies that $\phi(\mathbf{p}_{\mathbf{h}_1}) > \phi(\mathbf{p}_{\mathbf{h}^*})$ since ϕ preserves the order of majorization. However, this will contradict our original assumption that $\mathbf{h}^*$ is SC_ϕ-optimal. Therefore, no other element exists in Ω that majorizes $\mathbf{p}_{\mathbf{h}^*}$. Thus, $\mathbf{h}^*$ is a maximal element of $\mathcal{L}$. □

[7] This function arises in the study of optimal subband encoders and is proved to be strictly Schur-convex. See section 6.6.

[8] Note that ϕ_3 is not defined on the sides of Δ and is particularly singular at its corners.

6.4.2 An illustrative design example

In the following, we provide a simple design example to illustrate the concepts and ideas discussed so far. We choose two Schur-convex functions and design two SC_ϕ-optimal filter banks accordingly. It turns out that neither is better than the other in the sense of scalability ordering. This demonstrates that maximal elements can be found in $\mathcal{L}$ which do not satisfy the requirement of being a greatest element. Since a greatest element is necessarily a maximal element, this proves that, for the case considered, $\mathcal{L}$ doesn't have a greatest element, i.e. an SC-optimal filter bank does not exist.

Example 6.2. **(SC_ϕ-optimal Filter Bank Design)** Let $\mathcal{L}$ be the class of filter banks which have 3 channels. Assume, further, that the filters are of length 6. This means $M = 3$, $N = 6$ and $K = N/M - 1 = 1$. It follows from the results in Section 6.2.2 that the class $\mathcal{L}$ is parameterized by $P = K(M-1) + M(M-1)/2 = 5$ rotation angles. Let us pack these angles in the parameter vector $V = [\theta_0, \theta_1, \cdots, \theta_4]^T \in [-\pi\ \pi]^5$.

Table 6.1. Correlation coefficients used in Example 6.2.

$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(4)$	$\rho(5)$
0.2624	-0.2746	0.1781	0.4023	0.1780

For our design, we assume that $x(n)$ is a WSS random signal with autocorrelation coefficients $\rho(i) \stackrel{\triangle}{=} E\{x(n)x(n-i)\}$ as shown in Table 6.1. Given these autocorrelation coefficients[9], and the analysis filters, it is possible to calculate the power vector $\mathbf{p}$. For the sake of brevity, we omit the details of this process here. It suffices to know that, once the input signal statistics are fixed, the output power vector $\mathbf{p}$ is only a function of the rotation parameters that specify the analysis filter bank. This is to say, $\mathbf{p} = \mathbf{f}(V)$ where $\mathbf{f}$ is a fixed function $\mathbf{f} : [-\pi\ \pi]^5 \longrightarrow \Omega$.

Now, we may use the strictly Schur-convex function

$$\psi_1(\mathbf{p}) = \|\mathbf{p}\|_2$$

to find our first SC_ϕ-optimal element in Ω. We can then use the "Coding Gain"

$$\psi_2(\mathbf{p}) = \frac{p_1 + p_2 + p_3}{3\sqrt[3]{p_1 \times p_2 \times p_3}}$$

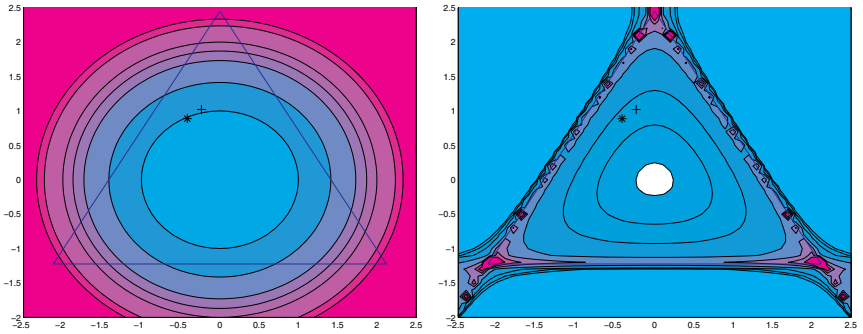
which is also strictly Schur-convex to find a second solution.

The resulting maximizing vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ obtained by maximizing $\psi_1(\mathbf{f}(V))$ and $\psi_2(\mathbf{f}(V))$, respectively, over the parameter space $[-\pi\ \pi]^5$, are

[9] In general, the variance of $x(n)$, i.e. $\rho(0)$, is irrelevant in scalability analysis. Also, due to the finite order of the analysis filters used here, higher-order autocorrelation coefficients not shown in the table are irrelevant as well.

Table 6.2. Optimization results for Example 6.2.

$\mathbf{p}$	$\psi_1(\mathbf{p})$	$\psi_2(\mathbf{p})$
$\mathbf{p}_1 = [0.4170\ 1.8333\ 0.7497]^T$	4.0969	1.2038
$\mathbf{p}_2 = [0.3441\ 1.7244\ 0.9315]^T$	3.9597	1.2185



(a) (b)

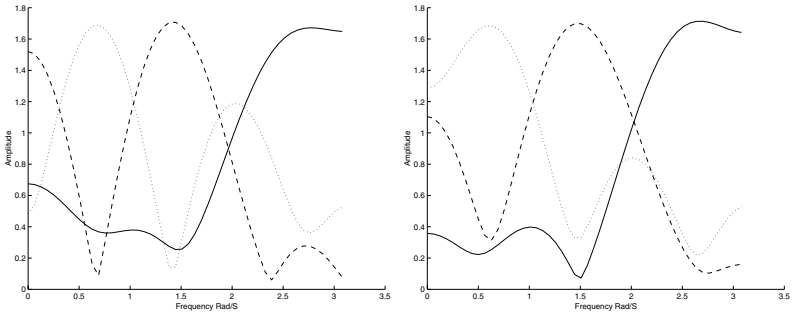
Fig. 6.6. Positions of $\mathbf{p}_1$ and $\mathbf{p}_2$ obtained in Example 6.2 on the triangle Δ. The two positions are denoted by '+' and '*' respectively. The plot in (a) shows the contours of $\psi_1(\mathbf{p})$ while the one in (b) shows the contours of $\psi_2(\mathbf{p})$. The two plots are otherwise the same.

shown in Table 6.2. These vectors along with contours of ψ_1 and ψ_2 are shown in Fig. 6.6 as well. Note that we are mainly interested in the power vectors $\mathbf{p}$ associated with each solution. The parameter vectors that produce the solutions are not important to us. The frequency response of the SC_ϕ-optimal filter banks associated with $\mathbf{p}_1$ and $\mathbf{p}_2$ are shown in Fig. 6.7.

The important observation here is that $\mathbf{p}_1$ is better than $\mathbf{p}_2$ with respect to ψ_1, while the reverse is true with respect to ψ_2. So neither is better than the other with respect to the ordering of scalability. Since SC_ϕ-optimal solutions are maximal elements, we conclude that, for the class $\mathcal{L}$ defined in this example, a greatest element (i.e. an SC-optimal solution) does not exist.

Finally, we would like to mention in passing that, for this example, we calculated a third SC_ϕ-optimal solutions using entropy as the order-preserving function. This solution was virtually the same as the one obtained using Coding Gain. ♢

In general, all Schur-convex functions are equivalent as far as the mathematical formulation of SC_ϕ-optimality is concerned. In other words, maximal elements do not have any preference over each other. One might, nevertheless, consider other factors such computational cost and prefer one Schur-convex function over another.



(a) (b)

Fig. 6.7. Frequency response of optimal 3-channel filter banks designed in Example 6.2. The filters associated with $\mathbf{p}_1$ are shown in (a) while those associated with $\mathbf{p}_2$ are depicted in (b).

In subsequent sections, we will show that several well-known notions of optimality for filter banks are intimately related to scalability via particular choices of Schur-convex functions. which itself is not in $\mathcal{L}$.

6.5 SC-Optimality vs PCFB

Let $\bar{\mathcal{L}}$ denote the class of all orthogonal M channel filter banks without any restriction on realizability or order. That is,

$$\bar{\mathcal{L}} = \{\mathbf{h}(z) : \mathbf{h}(z) = \mathbf{E}(z^M)\mathbf{e}(z),\ \mathbf{E}(z)^{-1} = \mathbf{E}(z^{-1})^T\}.$$

It was shown by Tsatsanis and Giannakis (1995) that a filter bank $\mathbf{h}^P(z)$ can be constructed in $\bar{\mathcal{L}}$ such that $\mathbf{e}_{\mathbf{h}^P} \leq \mathbf{e}_{\mathbf{h}}$ for all $\mathbf{h} \in \bar{\mathcal{L}}$. They called $\mathbf{h}^P(z)$ the Principal Component Filter Bank (PCFB).

Recall that we defined $\mathcal{L}$ as the class of orthogonal FIR filter banks of fixed order. Clearly, $\mathcal{L} \subset \bar{\mathcal{L}}$. Recall also that PCFB is not in $\mathcal{L}$. In this section, we show that an SC_ϕ-optimal element (which is in $\mathcal{L}$) can be interpreted as the best approximation to PCFB from among the elements in $\mathcal{L}$.

6.5.1 A constructive definition of the PCFB

To characterize PCFB, we need to make a couple of definitions.

Definition 6.5. *Let S denote a (measurable) subset of the interval $\Omega_0 = [0\ 2\pi]$. It is called an invariant subset if the following two properties hold:*

1. $\int_S d\omega = \frac{2\pi}{M}$. *This is to say, S has measure (bandwidth)* $\frac{2\pi}{M}$.
2. $\forall \omega \in S \Rightarrow (\omega + k\frac{2\pi}{M})_{mod\ 2\pi} \notin S$ *for* $k = 1, 2, \ldots, M-1$. *This is to say, S is an* alias free *frequency range.*

Band-pass signals whose PSD is an invariant subset, have the important property that they will remain invariant under combined multirate operations of decimation, expansion and indeal interpolation (Sathe and Vaidyanathan, 1993). There are an infinite number of invariant subsets of Ω_0. We are only interested in certain ones that can gather a random signal's energy as much as possible. Such subsets are defined below.

Definition 6.6. *An invariant subset $S_0 \subset \Omega_0$ is called first principal subset of Ω_0 if it maximizes the integral*

$$\int_{S_0} P_x(e^{j\omega})d\omega.$$

In words, the frequency band represented by S_0 contains the maximum energy of the random process $x(n)$ that one can collect in a frequency band subject to the condition that the portion of the signal in that frequency band remains invariant under multirate operations of decimation by M and upsampling by M. The second principal subset is obtained by first defining $\Omega_1 = \Omega_0 - S_0$ and then finding an invariant set $S_1 \subset \Omega_1$ such that it maximizes the integral

$$\int_{S_1} P_x(e^{j\omega})d\omega.$$

Following a similar procedure, we can find M principal subsets S_0, S_1, ..., S_{M-1} in Ω_0. These subsets partition Ω_0 into M disjoint frequency bands: $\Omega_0 = \cup_{i=0}^{M-1} S_i$. Using the notion of principal subsets, the PCFB is defined as follows:

Definition 6.7. *A bank of M linear filters $H_0(z), H_1(z), \ldots, H_{M-1}(z)$ is called a Principal Component Filter Bank (PCFB) if*

$$|H_i(e^{j\omega})| = \begin{cases} 1 & \omega \in S_i \\ 0 & \omega \notin S_i \end{cases} \text{ for } i = 1, 2, \ldots, M-1. \tag{6.12}$$

Characterizing the synthesis bank associated with PCFB is straightforward. One might simply choose the synthesis filters such that

$$F_i(e^{j\omega}) = \begin{cases} M/H_i(e^{j\omega}) & \omega \in S_i \\ 0 & \omega \notin S_i \end{cases} \text{ for } i = 1, 2, \ldots M-1. \tag{6.13}$$

Note that the above definitions do not impose any constraint over the filters' phase response. Therefore, a PCFB is an equivalence class of filter banks whose magnitude response satisfy the above definition.

Here, we are not going to describe how one can actually specify the PCFB for a given PSD. We just mention that the construction procedure for PCFB is similar to the *water pouring* technique well-known in information theory, c.f. (Gallager, 1968, Section 9.7) and (Berger and Gibson, 2000, Section III). The only difference is that one now has to construct invariant sets rather

than simple bandlimited ones. The invariance restriction makes construction of PCFB more involved than traditional water pouring. The reader is referred to Tsatsanis and Giannakis (1995) and Vaidyanathan (1998) for details on how to construct the PCFB.

6.5.2 PCFB is an upper bound for $\mathcal{L}$

As mentioned at the beginning of this section, our interest in PCFB stems from the fact that it provides the most scalable decomposition of a random signal possible by uniform orthogonal filter banks.

Theorem 6.7. *(Tsatsanis and Giannakis (1995), Unser (1993)) Let* $\mathbf{h}^P \in \bar{\mathcal{L}}$ *be the PCFB. Then* $\mathbf{h}^P \overset{SC}{\geq} \mathbf{h}$ *for all* $\mathbf{h} \in \mathcal{L}$.

It is not difficult to verify that PCFB satisfies the requirements of Theorem 6.2. Thus, $\mathbf{p}_{\mathbf{h}^P} \succeq \mathbf{p}_{\mathbf{h}}$, $\forall \mathbf{h} \in \mathcal{L}$. Since $\mathbf{h}^P \notin \mathcal{L}$, $\mathbf{h}^P$ specifies an upper bound for the poset $\mathcal{L}$ of FIR filter banks. An equally valid statement is that $\mathbf{p}_{\mathbf{h}^P}$ is the greatest element of $\bar{\mathcal{L}}$ with respect to the ordering induced by $\succeq$.

6.5.3 Historical notes

Since $\mathbf{p}_{\mathbf{h}^P}$ is the greatest element of $\bar{\mathcal{L}}$ it will maximize any Schur-convex function of the variance vector $\mathbf{p}_{\mathbf{h}}$. Recall that, this statement is a purely mathematical result due to Schur. Yet Schur's results seem to have been unnoticed in the early literature on PCFB. An initial comment on the relation between majorization and certain convex functions seems to be the one made by Unser (1993). Unser remarks in passing that $\mathbf{p}_{\mathbf{h}^P}$ maximizes convex functions of the form $\phi_1(\mathbf{p}) = \sum_{i=0}^{M-1} g(p_i)$. Later, Akkarakaran and Vaidyanathan (2001a) proved a result slightly more general than Unser's: They showed that $\phi_2(\mathbf{p}) = \sum_{i=0}^{M-1} g_i(p_i)$, where $g_i(\cdot)$ are possibly different convex functions, is also maximized by $\mathbf{p}_{\mathbf{h}^P}$. One easily notices that ϕ_2 is nothing but a special case in the broad class of Schur-convex functions.

The authors of (Akkarakaran and Vaidyanathan, 2001a) make a very brief note of Schur-convex functions in the last paragraph of Section III of this paper. As noted above, it appears that a major part of the paper (e.g. the whole of Section III) is devoted to deriving a result which can be interpreted as a special case of Schur's theory. These authors have published two other papers on this topic which again do not contain references to Schur's work (Akkarakaran and Vaidyanathan, 2001b), (Vaidyanathan and Akkarakaran, 2001).

The systematic theory of scalability presented in this chapter and the clear connection between the PCFB and scalable FIR filter banks were developed by Jahromi et al. (2003).

6.5.4 Approximating PCFB using the filter banks in $\mathcal{L}$

The PCFB is a bank of ideal band-pass filters. Obviously, such filters are not realizable using any finite-order FIR or IIR discrete-time system. The PCFB, therefore, can only be approximated using realizable structures of finite order. In this subsection, we define the notion of approximating PCFB in a mathematically precise manner.

From the viewpoint of ordering induced by scalability, it is reasonable to choose the greatest element of $\mathcal{L}$ as the best approximation of $\mathbf{h}^P$. This is simply because should a greatest element $\mathbf{h}^*$ exist in $\mathcal{L}$ its associated power vector $\mathbf{p}_{\mathbf{h}^*}$ would always sit immediately below $\mathbf{p}_{\mathbf{h}^P}$ when the elements in the augmented set $\Omega \cup \{\mathbf{p}_{\mathbf{h}^P}\}$ are ordered[10]:

$$\cdots \preceq \mathbf{p}_{i-1} \preceq \mathbf{p}_i \preceq \mathbf{p}_{i+1} \ \cdots \preceq \mathbf{p}_{\mathbf{h}^*} \preceq \mathbf{p}_{\mathbf{h}^P}. \tag{6.14}$$

Recall from the previous section that a greatest element usually doesn't exist in $\mathcal{L}$. Hence, the problem of finding the closest element in $\mathcal{L}$ to PCFB is indeterminate. To remove this indeterminacy, we need to specify a well-behaved measure of closeness to PCFB which preserves the ordering of scalability as well (see Fig. 6.8). Such a measure of closeness, or *divergence function* $D(\mathbf{h} : \mathbf{h}^P)$, should satisfy the following condition

$$\mathbf{p}_{\mathbf{h}_1} \succeq \mathbf{p}_{\mathbf{h}_2} \Rightarrow D(\mathbf{h}_1 : \mathbf{h}^P) \leq D(\mathbf{h}_2 : \mathbf{h}^P). \tag{6.15}$$

Note that we are not interested in inducing any topology on the set $\mathcal{L}$, simply because a notion of distance between two filter banks in $\mathcal{L}$ doesn't seem to have any application! Therefore, we regard $D(\mathbf{h} : \mathbf{h}^P)$ as a function of $\mathbf{h}$ only. We suggest a simple divergence function in the following definition.

Definition 6.8. (Approximate PCFB) *A filter bank $\mathbf{h}^* \in \mathcal{L}$ is called the best approximation to PCFB if it minimizes the divergence function*

$$D(\mathbf{h} : \mathbf{h}^P) = \phi(\mathbf{p}_{\mathbf{h}^P}) - \phi(\mathbf{p}_{\mathbf{h}});$$

where ϕ is strictly Schur-convex and $\mathbf{h}^P$ is PCFB.

The main property of $D(:)$ introduced in Definition 6.8 is that it preserves the ordering of scalability as required by (6.15). In addition, it is non-negative and becomes zero if and only if $\mathbf{h} = \mathbf{h}^P$. Having defined the notion of approximate PCFB, we have, in fact, established a direct connection between SC-optimality and approximating PCFB: Let $\mathbf{h}^* \in \mathcal{L}$ be an SC_ϕ-optimal filter bank; it is the best approximation to PCFB as well. In other words, $D(\mathbf{h}^* : \mathbf{h}^P)$ is minimum.

In summary, PCFB represents an orthogonal multirate system from which one can obtain the best low-rate approximations of the original (full-rate)

[10] The ordering presented in (6.14) is for the purpose of illustration only. The set Ω is not countable.

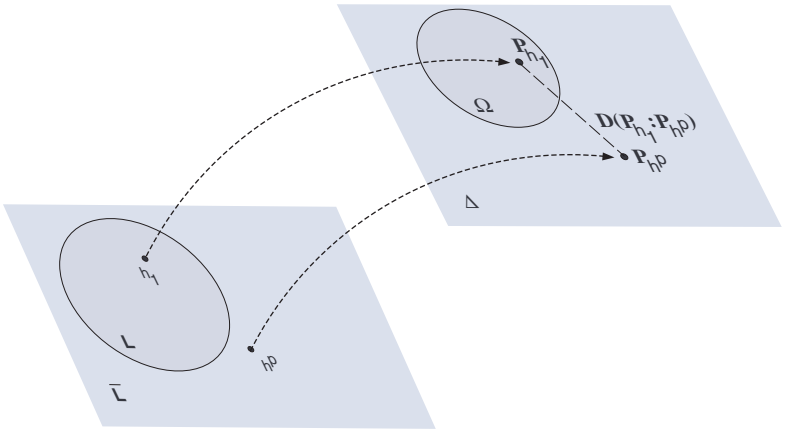


Fig. 6.8. A symbolic set diagram showing how the "distance" between an element in $\mathcal{L}$ and PCFB is defined.

signal. This system, however, is an ideal concept and one can only approximate it using FIR filter banks (Fig. 6.8). In our discussion above, we provided a precise (but not objective) definition of approximating PCFB using filter banks in $\mathcal{L}$. Objectivity is lost because the definition of the best approximation depends on the choice of a Schur-convex function. Our definition, however, has the interesting consequence that finding an SC-optimal filter bank in $\mathcal{L}$ and approximating the PCFB become equivalent.

6.6 SC-Optimality vs Subband Coding optimality

In this section, we consider the relation between optimality in the sense of scalability and optimality in the sense of minimum quantization noise in a subband coding (SBC) system. We show that the optimality of a filter bank for subband coding falls within the scope of SC_ϕ-optimality. In other words, an optimal filter bank for subband coding is an SC_ϕ-optimal filter bank where ϕ is a particular Schur-convex function called Coding Gain.

The basic block diagram of a subband coding system is shown in Fig. 6.9. The input signal $x(n)$ is decomposed into M subband signals which are then decimated and quantized by a set of M memoryless quantizers Q_0 to Q_{M-1}. An approximation to the original input is then synthesized by up-sampling the quantized subbands and combining them using a set of synthesis filters. One can express the basic subband coding problem as follows:

Problem 6.1. (Optimal Subband Coding) Given that the input signal is zero-mean wide-sense stationary, find

(i) a way to distribute b_{total} bits among the quantizers

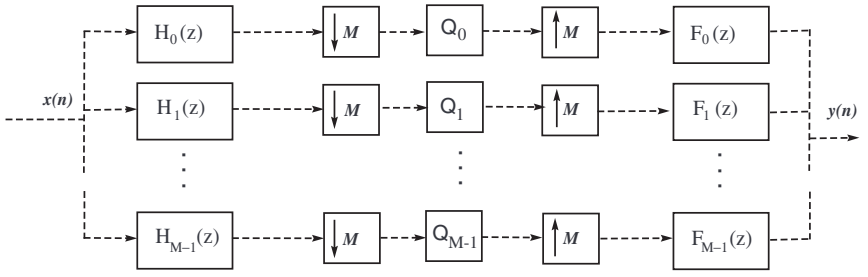


Fig. 6.9. An M-channel uniform filter bank with quantizers in the subbands.

(ii) a filter bank in $\mathcal{L}$

such that the expected mean square error between the input $x(n)$ and output $y(n)$ is minimized.

Designing an optimal system with regards to the above requirements has been a subject of research since the early 90s, e.g. Soman and Vaidyanathan (1993), Ase and Mullis (1996), Vaidyanathan (1998), Moulin et al. (2000). Assuming that the bit rate is high enough[11] to replace the nonlinear effect of quantizers with uncorrelated noise sources, it turns out that the optimal bit allocation depends on the variance of the subband signals only. The ultimate performance of the system, therefore, is determined by the way the analysis filter bank distributes the input signal's variance among the subbands.

Detailed analysis (e.g., Soman and Vaidyanathan (1993)) shows that for a filter bank with orthogonal filters, the mean-squared error between $x(n)$ and $y(n)$ is minimized when

(i) The following function of the subband signals' power is maximized

$$G = \frac{\frac{1}{M}\sum_{k=0}^{M-1} E\{x_k^2(n)\}}{(\prod_{k=0}^{M-1} E\{x_k^2(n)\})^{\frac{1}{M}}}. \tag{6.16}$$

(ii) The available bits are distributed among the quantizers according to the following formula[12]:

$$b_k = b_{total} + 0.5\log_2 \frac{E\{x_k^2(n)\}}{\prod_{k=0}^{M-1} E\{x_k^2(n)\}}. \tag{6.17}$$

From the above discussion, we can define an optimal filter bank for subband coding as follows:

[11] Designing an optimal subband coding system for low bit rates is an open problem. The reader is referred to Mallat and Falzon (1998) for an analysis of this situation.

[12] In general, (6.17) doesn't lead to integer values for b_k. One should round up the results to the closest integer which will, then, lead to suboptimal quantization. It may also result in negative values which should be replaced by zero in practice.

Definition 6.9. (*CG*-optimality) *A filter bank* $\mathbf{h} \in \mathcal{L}$ *is optimal for subband coding (CG-optimal for short) if* $\mathbf{p_h}$ *maximizes, among all elements of* $\mathcal{L}$*, the coding gain function given in (6.16).*

Note that the performance of a subband coder is a *signal dependent* quantity whose value is determined by the input signal's PSD, number of bits available and frequency response of analysis filters. The dependency on the filters, however, is solely through the way they distribute the input signal's power among the subbands as expressed quantitatively by (6.16). This is reflected in the fact that G is a function only of the subband power vector $\mathbf{p}$ introduced earlier in Section 6.3.2. Therefore, as in the case for scalability, choosing an optimal filter bank $\mathbf{h} \in \mathcal{L}$ is determined by the relative position of $\mathbf{p_h}$ in Ω. The following theorem establishes the fundamental connection between the two notions of optimality:

Theorem 6.8. *If an SC-optimal filter bank exists in* $\mathcal{L}$*, then it is CG-optimal too. More generally,* $\mathbf{h}_1 \overset{SC}{\geq} \mathbf{h}_2 \Longrightarrow G(\mathbf{p_{h_1}}) \geq G(\mathbf{p_{h_2}}), \forall \mathbf{h}_1, \mathbf{h}_2 \in \mathcal{L}$.

Proof. It suffices to show that $G(.)$ is Schur-convex. To show this, we first consider the function $S(\mathbf{p}) = p_0 \times p_1 \ldots \times p_{M-1}$ where $p_i > 0$ are the components of $\mathbf{p}$. This function is symmetric in its arguments and satisfies the property

$$(p_0 - p_1)(\frac{\partial S}{\partial p_0} - \frac{\partial S}{\partial p_1}) \leq 0.$$

Therefore, by Theorem 2.4, $S(\mathbf{p})$ is Schur-concave. Next, we note that for the filter banks in $\mathcal{L}$, the numerator of G is a (positive) constant. Thus, $G(\mathbf{p}) = \frac{const.}{\sqrt[M]{S(\mathbf{p})}}$. Since the function $\frac{1}{\sqrt[M]{S}}$ is strictly decreasing for $S \geq 0$, we conclude that $G(\mathbf{p})$ is Schur-convex[13]. □

The reader is referred to Fig. 6.5 (c) and (d) for an illustration of the contours generated by the Coding Gain function G. Note that this function follows the limits introduced by the sides of the triangle Δ more notably than other Schur-convex functions shown in Example 6.1. In fact, it grows very quickly as one approaches the sides of Δ and eventually becomes singular on them (Fig. 6.5 (d)).

6.7 Complements

In this chapter, we showed that for the filter banks in class $\mathcal{L}$ several notions of optimality, each arising from a different viewpoint, can be unified under one algebraic framework, namely the set-theoretic concept of partial ordering.

[13] It is further verified that $G(\mathbf{p})$ is strictly Schur-convex. The proof is straightforward.

The presented algebraic approach, however, seems to be more powerful than covering (and unifying) the optimality concepts of coding gain, scalability and PCFB! In this section, we shall probe some generalizations and discuss, in some detail, connections between concepts developed here and those that already exist in the literature.

6.7.1 Algorithmic aspects of finding an SC_ϕ-optimal element in $\mathcal{L}$ and previous works

Optimal filter banks are replacements for fixed ones where statistical properties of the input signal can be incorporated in their design for improving the performance. Many practical applications require this optimized solution to be estimated in an on-line fashion. In other words, the input signal's statistics are not known a priori so the optimization should eventually be formulated as an adaptive algorithm. This introduces many concerns on both the Schur-convex function to be used as an objective function and the way it is to be estimated and maximized on-line.

Some results have appeared with regards to designing CG-optimal filter banks[14]. An early work in this area is the adaptive algorithm devised by Delsarte and co-workers Delsarte et al. (1992). It is designed around a two-channel orthogonal filter bank and searches for an optimal solution in an on-line (adaptive) fashion. Other results in this direction can be found in Regalia and Huang (1995), (Jahromi, 1997, Chapters 6,7), Jahromi and Francis (1999) and Douglas et al. (1999). None of these algorithms, however, is guaranteed to converge to a global optimum nor are there any rigorous results available on their convergence performance. A major challenge here is that, even for the simple case of two-channel filter banks, the optimization involved is highly ill-conditioned (Jahromi and Francis, 1999, Section 6). Another issue, as severe as the previous one, is the existence of local maxima!

In addition to on-line algorithms, several off-line optimization techniques (analytic and semi-numerical) have been proposed for designing what the authors have considered optimal FIR filter banks. Notable among these are the linear programming techniques developed by Moulin et al. Moulin and Mihçak (1998) and a constrained optimization method proposed by Xuan and Bamberger Xuan and Bamberger (1998)[15]. The method proposed in Moulin and Mihçak (1998) leads to a maximal element which, in a loose sense, can be interpreted as an SC_ϕ-optimal solution where ϕ is selected as in (2.13). The optimization technique in Xuan and Bamberger (1998) achieves a CG-optimal filter bank which the authors wrongly call PCFB.

---

[14] Recall that from the viewpoint of our algebraic theory, this corresponds to finding an SC_ϕ-optimal solution where ϕ is the coding gain function (6.16).

[15] The reader is referred to (Kirac and Vaidyanathan, 1998b, section II.A) for pointers to other related works.

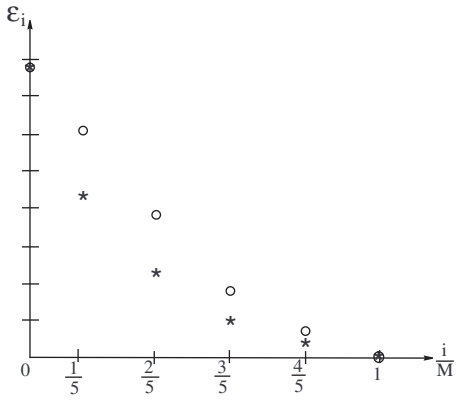


Fig. 6.10. Components of the error vector **e** plotted vs the fraction of the original sampling rate used in synthesis. '*' represents PCFB and 'o' represents an arbitrary element in $\mathcal{L}$.

6.7.2 Similarity with rate-distortion theory

Let $x(n)$ be a stationary random signal with given PDF and let $\mathbf{e}_{\mathbf{h}^P} = [\epsilon_{M-1}\ \epsilon_{M-2} \cdots \epsilon_0]^T$ denote the error vector when PCFB is used to approximate $x(n)$ using its low-resolution components. Recall that ϵ_i represents the mean-squared reconstruction error where the synthesis filter uses i out of M subband signals to generate the output. In other words, reconstruction is performed using a fraction (i/M) of the original sampling rate. We can plot components of $\mathbf{e}_{\mathbf{h}^P}$ vs this fraction. This is done in Fig. 6.10 for $M = 5$. In this figure, we have also plotted an extra point to represent the reconstruction error (zero) when all the subbands are present (i.e. $i/M = 1$).

If we connect the '*'s in Fig. 6.10 using a smooth curve, the result will look very similar to a rate-distortion curve Berger and Gibson (2000), Gallager (1968). Like the rate-distortion curve, the error vs sampling rate curve generated by the PCFB is a characteristic of the signal. The significance of the error vs sampling rate curve is that *its shape shows how scalable a signal is.*

No matter how high is the order of the filter bank used, it is not possible to achieve an error-sampling rate curve which lies below that of PCFB. The performance (i.e. scalability) of a finite-order filter bank is determined by the closeness of its error-sampling rate curve to that of PCFB. The divergence function introduced in Definition 6.8 is, simply, a quantitative measure of this closeness!

6.7.3 On partial ordering and subjectivity

The most notable instance of the occurrence of a partial ordering is, perhaps, in the philosophy of science where one wants to choose among empirical theories based on their degree of falsifiability (Popper, 1959, Chapter VI). While the notion of partial ordering is not a well-known one in signal processing, one can find its traces in related areas like information theory Shannon (1958),(Csiszár and Körner, 1981, pp. 115-116), statistical inference (Vajda, 1989, Chapter 6) and system identification (Caines, 1988, Chapter 5).

Basically, a partial ordering is likely to occur when one tries to rank entities based on more than one criteria. In system identification, for example, one selects a model for the observations at hand based on several factors including model complexity and its accuracy in predicting the observed signal. Having chosen suitable *measures* of model complexity and prediction error, one can then induce a partial ordering on the set of feasible models. The partial ordering is embedded in a total ordering by selecting a scalar-valued *criterion function* that penalizes model complexity and its prediction error at the same time. A great deal of subjectivity has been introduced into the field with regards to the choice of this criterion function Caines (1988), Ljung (1999).

6.8 Summary

We considered the problem of characterizing an optimal filter bank from the scalability point of view. We presented a quantitative notion of scalability and, based on that, derived an optimality theory applicable to the class $\mathcal{L}$ of FIR filter banks. Our approach to the problem was abstract and algebraic: We endowed the set $\mathcal{L}$ with an ordering that represented scalability. Then, we established an equivalence relation between the order of scalability and another order called majorization.

We showed that, using either ordering, specifying the *optimal* filter bank in $\mathcal{L}$ translates to finding the *greatest element* associated with the order. The ordering of scalability is partial. Thus, while we are able to prove that it has maximal elements, the existence of a greatest element is not guaranteed. In view of this fact, we relaxed our original notion of optimality (i.e. SC-optimality) to that of SC_ϕ-optimality. An SC_ϕ-optimal solution is viable since

(i) it is equal to an SC-optimal solution if the latter exists,
(ii) it always exists in $\mathcal{L}$.

We showed, moreover, that an SC_ϕ-optimal solution has the interpretation as an approximation to PCFB when the approximation is from the scalability point of view. Table 3 summarizes various notions of optimality we considered in this chapter and their connection to the poset $\mathcal{L}$.

Table 6.3. Various types of optimal solutions considered.

SC-optimal	The greatest element of $\mathcal{L}$
SC_ϕ-optimal	A maximal element of $\mathcal{L}$
CG-optimal	A maximal element of $\mathcal{L}$
PCFB	An upper bound for the poset $\mathcal{L}$

6.9 Open problems

The algebraic theory of scalability developed in this chapter is not restricted to uniform or orthogonal filter banks. In this section we provide some hints with regards to some possible directions in which this theory might be extended. Graduate students and other interested researchers might use this directions as starting points for new research on scalable multirate systems.

6.9.1 Extension to non-perfect-reconstruction filter banks

It should be possible to apply the quantitative concept of scalability introduced in Section 6.3 to filter bank classes which are not PR. In this case, the error vector $\mathbf{e_h}$ should be augmented to include one more component representing the reconstruction error when all the channels are used by the synthesis bank[16]. Then, one might use the same definitions of scalability and SC-optimality as in Section 6.1.

Again, the ordering introduced by $\overset{SC}{\geq}$ will be a partial ordering. But,in general, Theorem 6.2 is not valid for non-PR systems. A way forward might be apply an order-preserving function directly on the ordering $\overset{SC}{\geq}$. Let $\phi(.)$ be such a function. An SC_ϕ-optimal filter bank can then be defined as the one that maximizes $\phi(\mathbf{e_h})$ among all filter banks $\mathbf{h}$ in the considered non-PR class.

Note that any l_p norm $\|.\|_p$, $p \geq 1$, has the property that $\mathbf{e_{h_1}} \leq \mathbf{e_{h_2}} \Rightarrow \|\mathbf{e_{h_1}}\|_p \leq \|\mathbf{e_{h_2}}\|_p$. Hence, $\phi = -||\mathbf{e}||_p$ for some fixed p can be used as an order preserving function.

6.9.2 Extension to tree-structured filter banks

The optimality theory presented in this chapter can be extended to orthogonal tree-structured filter banks as well. Some initial work in this direction has been reported in Jahromi and Francis (1999) but a more thorough construction is needed.

[16] This means, in a plot similar to Fig. 6.10, the 'o' associated with full-rate reconstruction will not necessarily be located on the horizontal axis.

6.9.3 Scalability with respect to other error measures

In Section 6.3.1, we defined the error vector $\mathbf{e}$ such that its components represent expected mean square errors when the input signal is reconstructed using i out of M subbands. One, nevertheless, can define $\mathbf{e}$ such that its components represent any other meaningful measure of the difference between the original signal and its low-resolution reconstructions. For instance, one can take into account properties of human perceptual system when defining ϵ_i if applications in digital audio are concerned (Gold and Morgan, 2000, Part IV).

6.9.4 Scalability when an optimal synthesis system is used

We mentioned at the beginning of this chapter that, in general, the adjoint of an analysis system is *not* the best system (from a minimum-mean-square-error point of view) to use for approximating the input signal using a selected number $i < M$ of subbands. We already showed in Chapter 4 how to formulate an *optimal* synthesis procedure for a given analysis filter bank. The synthesis technique introduced in Chapter 4 is quite general and works for analysis systems which are not in class $\mathcal{L}$ or even PR. One might use this optimal procedure (or any other feasible synthesis system) and calculate an error vector $\mathbf{e}$ such that its components represent the expected mean-square-errors when the input signal is reconstructed using i out of M subbands. Once $\mathbf{e}$ is defined, the definitions of scalability, SC-optimality and SC_ϕ-optimality introduced earlier in this chapter would become applicable again.

7

Information in Multirate Systems

7.1 Introduction

Recall from Chapter 2 that all the statistical information about a Gaussian WSS random signal $x(n)$ is contained in its power spectrum $P_x(e^{j\omega})$ or, equivalently, in its autocorrelation sequence $R_x(k)$. This is because if we know $P_x(e^{j\omega})$, we can calculate the probability density functions that govern the statistical dependence among any number of samples of $x(n)$.

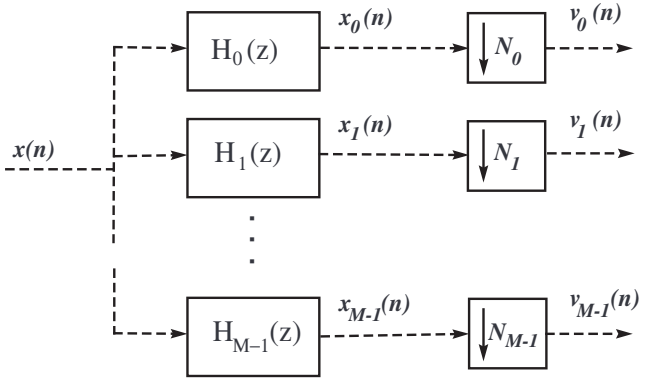


Fig. 7.1. An analysis filter bank can model a multirate measurement system.

Assume that a WSS signal $x(n)$ is not available for statistical experiment directly but we have access some multi-rate measurement signals $v_i(n)$ obtained indirectly via a multirate measurement system (Fig. 7.1). One may pose questions like these: How much statistical information about $x(n)$ can be gained if we calculate statistical properties of $v_0(n)$? Which signal, $v_2(n)$ or $v_3(n)$, gives *more* statistical information about $x(n)$? If we know statistical

properties of $v_0(n)$, how much *more* information about $x(n)$ will be gained by doing statistical experiments on $v_1(n)$?

To answer questions of the type mentioned above, we need to establish a quantitative measure of statistical information gained about $x(n)$ by statistical experiments performed on the multirate measurements $v_i(n)$. The purpose of this chapter is to introduce one such measure.

The material in this chapter are based on the paper by Jahromi et al. (2004a).

7.2 Information as distance from uniform spectrum

Consider the multirate filter bank in Fig. (7.2) and assume that we know a priori that $x(n)$ is a regular WSS random signal with zero mean and unit variance[1].

We say that the state of our knowledge regarding the statistics of $x(n)$ is "complete ignorance" if all we know about the statistics of $x(n)$ is the assumptions stated in the above paragraph. By definition, we assign the value zero to the quantity of information associated with this state of knowledge.

When in the state of complete ignorance, we assume based on the principle of Maximum Entropy that the power spectrum of $x(n)$ is a white (constant) spectrum with unit variance. In other words, in the state of complete ignorance we assume that $P_x = \bar{P}_x$ where

$$\bar{P}_x(e^{j\omega}) \stackrel{\triangle}{=} 1, \ \omega \in [-\pi \ \pi]. \tag{7.1}$$

We say that the state of our knowledge regarding the statistics of $x(n)$ is "complete information" if we know $P_x(e^{j\omega})$ exactly. By definition, the amount of information associated with this state of knowledge is given by

$$I(x) \stackrel{\triangle}{=} D(P_x \| \bar{P}_x), \tag{7.2}$$

where $D(P_x \| \bar{P}_x)$ is the Kullback-Leibler divergence of P_x from $\bar{P}_x$.

Obviously, "complete information" and "complete ignorance" are two extreme states of knowledge with regards to the statistics of $x(n)$. These states, therefore, mark the upper and lower limits on the information scale (Fig. 7.2) which represents *the quantity of information* that can be gained via multirate measurements.

Recall from Chapter 2 that the entropy rate $H(x)$ of a Gaussian random process $x(n)$ is given by

[1] The assumption that the input signal's variance is normalized to one is necessary to avoid the ambiguity caused by the scale-dependence of our information measure. This is a regrettable limitation but is shared by all entropy-based information measures

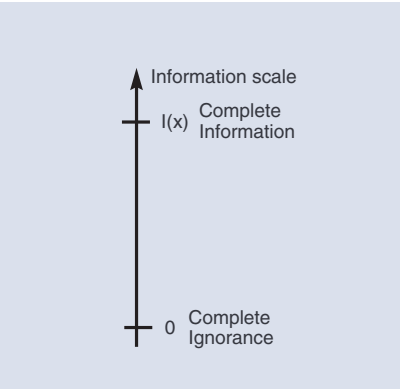


Fig. 7.2. "Complete information" and "complete ignorance" mark the upper and lower limits of the quantity of information that can be gained about $x(n)$.

$$H(x) = \frac{1}{2}\ln 2\pi + \frac{1}{2} + \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln P_x(e^{j\omega})d\omega, \tag{7.3}$$

and that $D(P_2\|P_1)$ is given by

$$D(P_2\|P_1) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{P_2(e^{j\omega})}{P_1(e^{j\omega})} - \ln\frac{P_2(e^{j\omega})}{P_1(e^{j\omega})} - 1\right)d\omega. \tag{7.4}$$

It follows from (7.1) - (7.4) that

$$H(x) + I(x) = c \tag{7.5}$$

where $c \stackrel{\triangle}{=} \frac{1}{2}\ln 2\pi + \frac{1}{2}$ is a constant. The statistical information $I(x)$ is always positive or zero while $H(x)$ can become negative or even, for a non-regular random process, $-\infty$. This is shown graphically in Fig. 7.3.

Let $\mathcal{Q}$ denote the set of all power spectra associated with regular WSS processes with unit variance. Let $\mathcal{S}^{(i;N)} \subset \mathcal{Q}$ denote the set of those spectra in $\mathcal{Q}$ which are consistent with first N autocorrelation coefficients $R_{v_i}(k)$, $k \in \{0, 1, \cdots, N-1\}$ obtained from the low-rate signals $v_0(n)$ to $v_i(n)$. In Chapter 3, we developed a Maximum Entropy inference algorithm which would pick a unique power spectrum from the infinitely many spectra in $\mathcal{S}^{(i;N)}$. Let $P_x^{(i;N)}(e^{j\omega})$ denote this power spectrum. Now, we are ready to make the following fundamental definition:

Definition 7.1 (Information content of low-rate measurements). *The quantity of statistical information contained in $v_0(n)$ to $v_i(n)$ about $x(n)$ is given by*

$$I(v_0, \cdots, v_i) \stackrel{\triangle}{=} \lim_{N\to\infty} D(P_x^{(i;N)}\|\bar{P}_x). \tag{7.6}$$

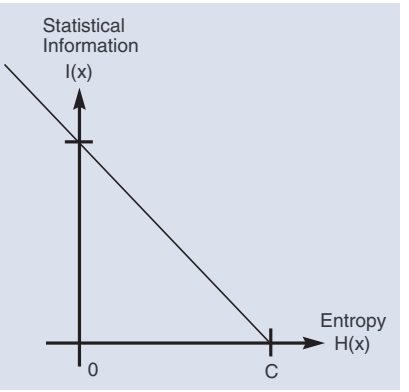


Fig. 7.3. Statistical information $I(x)$ and entropy rate $H(x)$ of a Gaussian WSS random process $x(n)$ add up to a constant if the variance of $x(n)$ is held fixed.

The above definition allows us to measure quantitatively the statistical information provided by one or a multitude of low-rate measurements. Note, however, that this definition ignores any (potential) information that could have been obtained by considering the cross-correlation among the low-rate measurements $v_0(n)$ to $v_i(n)$.

Theorem 7.1. *The limit used in Definition 7.1 exists.*

Proof. It follows from the definition of $\mathcal{S}^{(i;N)}$ that

$$(\mathcal{S}^{(i;N+1)} \bigcap \mathcal{Q}) \subset (\mathcal{S}^{(i;N)} \bigcap \mathcal{Q}). \tag{7.7}$$

Also, it is obvious that $P_x(e^{j\omega}) \in (\mathcal{S}^{(i;N)} \bigcap \mathcal{Q})$ for all values of i and N. Now, let $P_x^{(i;N)}(e^{j\omega})$ and $P_x^{(i;N+1)}(e^{j\omega})$ represent the Maximum Entropy estimates corresponding to the constraint sets $(\mathcal{S}^{(i;N)} \bigcap \mathcal{Q})$ and $(\mathcal{S}^{(i;N+1)} \bigcap \mathcal{Q})$ respectively. It follows from this fact and (7.7) that

$$H(P_x^{(i;N)}(e^{j\omega})) \geq H(P_x^{(i;N+1)}(e^{j\omega})) \geq H(P_x(e^{j\omega})). \tag{7.8}$$

Using (7.1), (7.3) and (7.4) we can then write

$$D(P_x^{(i;N)} \| \bar{P}_x) \leq D(P_x^{(i;N+1)} \| \bar{P}_x) \leq D(P_x \| \bar{P}_x). \tag{7.9}$$

The above relation shows that the sequence $D(P_x^{(i;N)} \| \bar{P}_x)$ is bounded from above and also non-decreasing in N. Thus, it has a limit. □

The following theorem shows that the information content of low-rate measurements increases (or in the worst case remains constant) as we include more channels. It also indicates that the information provided by multirate measurements cannot exceed that of the original signal:

Theorem 7.2. *If $i \geq j$ then*

$$I(v_0, \ldots, v_j) \leq I(v_0, \ldots, v_i) \leq I(x). \tag{7.10}$$

Proof. Using a procedure similar to the one used in the proof of the above lemma, one can show that

$$H(P_x^{(i;N)}(e^{j\omega})) \geq H(P_x^{(i+1;N)}(e^{j\omega})) \geq H(P_x(e^{j\omega})), \tag{7.11}$$

and then proceed to

$$D(P_x^{(i;N)} \| \bar{P}_x) \leq D(P_x^{(i+1;N)} \| \bar{P}_x) \leq D(P_x \| \bar{P}_x) \tag{7.12}$$

from which the asserted inequality follows in the limiting case when $N \to \infty$ □

It follows from the above theorem that, as we include more channels in our multirate measurement system, we can expect to gain more information. However, the total amount of information will eventually saturate to a limiting value below $I(x)$ which represents the "complete information" needed to specify statistics of $x(n)$. This is shown graphically in Fig. 7.4

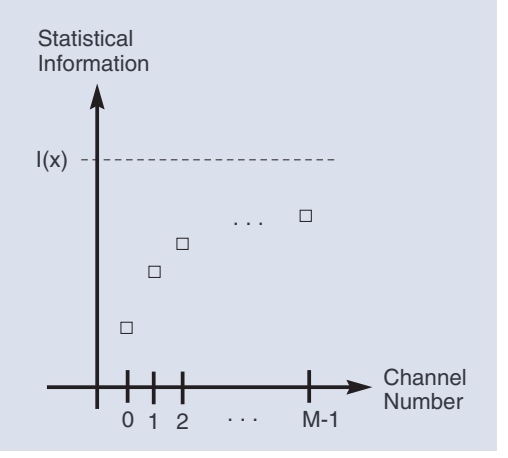


Fig. 7.4. A fictional plot depicting the amount of statistical information $I(v_0, \cdots, v_i)$ as a function of the number of channels included in the measurement process.

7.3 An illustrative example

It is time to illustrate the ideas developed in the previous sections using a concrete example.

In practice, it is not possible to c $P_x^{(i;N)}(e^{j\omega})$ when $N \to \infty$. For this reason we have to limit ourselves to a fixed number of autocorrelation coefficients per channel. In this section we use $P_x^{(i;4)}(e^{j\omega})$ as an approximation to $P_x^{(i;\infty)}(e^{j\omega})$ calculations.

Another practical difficulty is that the inference algorithm introduced in Chapter 3 does not satisfy the constraints exactly. This means that the solution provided by it may not be a unit variance spectrum. We got around this issue by modifying the algorithm in Chapter 3 such that it considered the input variance as a specific constraint. Then, to make the variance of the solution as close to one as possible, we made sure that the algorithm would emphasize the variance constraint 10 times more than the constraints imposed by the measured autocorrelation coefficients.

Example 7.1 (A 3-channel multirate measurement system). Consider a 3-channel analysis filter bank in a structure similar to the one shown in Fig. 7.1. Assume that the down-sampling rate is equal to four for all channels ($N_0 = N_1 = N_2 = 4$) and that the linear filters $H_0(z)$, $H_1(z)$ and $H_2(z)$ are specified as follows:

$$H_0(z) = \frac{0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}}{1.0000 - 0.8877z^{-1} + 0.6738z^{-2} - 0.1206z^{-3} + 0.0225z^{-4}}$$

$$H_1(z) = \frac{0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}}{1.0000 - 0.6855z^{-1} + 0.3297z^{-2} - 0.0309z^{-3} + 0.0032z^{-4}}$$

$$H_2(z) = \frac{0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}}{1.0000 + 0.2814z^{-1} + 0.3739z^{-2} + 0.0345z{-3} - 0.0196z^{-4}}$$

The frequency response curves $|H_i(e^{j\omega})|$ for these filters are shown in Fig. 7.5. For this example, the input signal $x(n)$ is chosen to be a Gaussian WSS process[2] whose power spectrum $P_x(e^{j\omega})$ is a simple low-pass function as shown in Fig. 7.6.

In this example, we use $P_x^{(i;4)}(e^{j\omega})$ calculated by our Maximum Entropy inference algorithm as an approximation to $P_x^{(i;\infty)}(e^{j\omega})$ in our information calculations. The spectra estimated using the data provided by each channel of the analysis system described above are shown in Fig. 7.7. Fig. 7.8 shows

[2] The process used in this example is an ARMA$(4, 4)$ process whose modeling filter was calculated using the MATLAB command

```
[a,b]=YULEWALK( 4, [0 .35 .85 1],[1 1 0.1 0])
```

that implements the Yule-Walker filter design algorithm. The process' amplitude was then scaled such that $E\{x^2(n)\} = 1$.

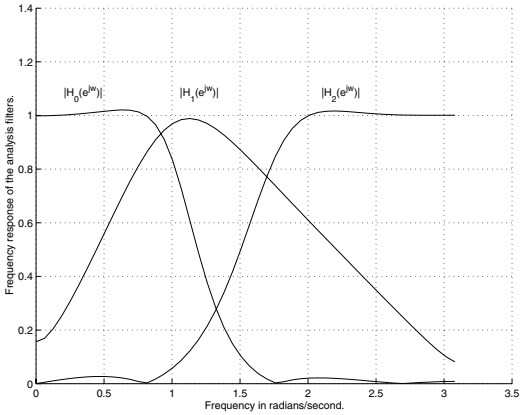


Fig. 7.5. Frequency response of the analysis filters used in Example 7.1.

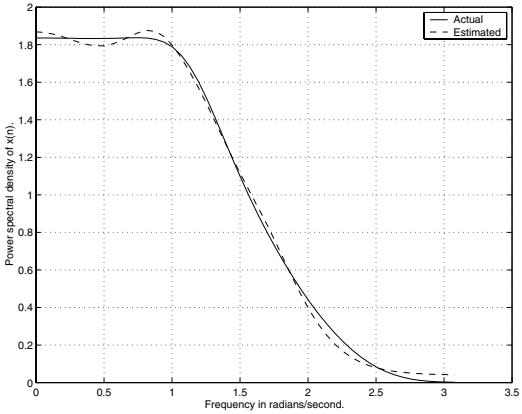


Fig. 7.6. The input power spectrum estimated using 4 autocorrelation coefficients per channel and all the channels of the filter bank described in Example 7.1. The doted curve shows the estimate while the solid one shows the actual spectrum.

the PSDs estimated using data produced by three different pairs of the channels of this analysis system. All PSD estimates shown have a variance in the range 1 ± 0.01. Given these estimates, one can easily calculate the quantity of information provided by each channel and the quantity of information provided by various combinations of the channels. These values, along with the value $I(x)$ corresponding to complete statistical information about $x(n)$ are plotted in Fig. 7.9. As can be seen in these plots, the collective quantity of information provided the channels increase as we include new ones but there exists a considerable amount of redundancy as well.

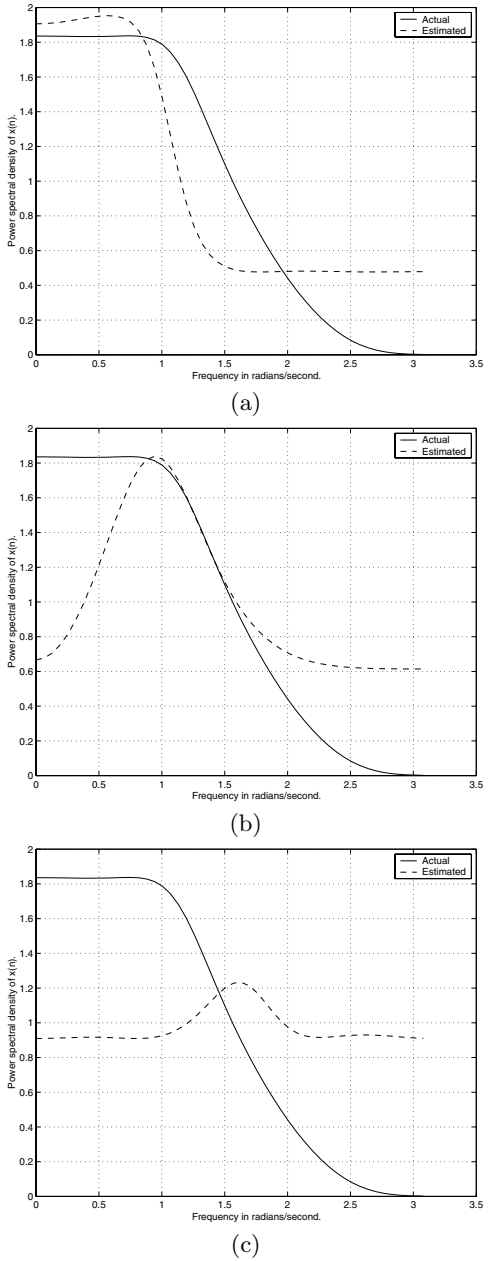


Fig. 7.7. The input power spectra estimated using 4 autocorrelation coefficients calculated for each individual channels of the filter bank described in Example 7.1. (a) The estimate obtained using Channel 1 data, (b) the estimate obtained using Channel 2 data, (c) the estimate obtained using Channel 3 data. In each plot, the doted curve shows the estimate while the solid one shows the actual spectrum for comparison.

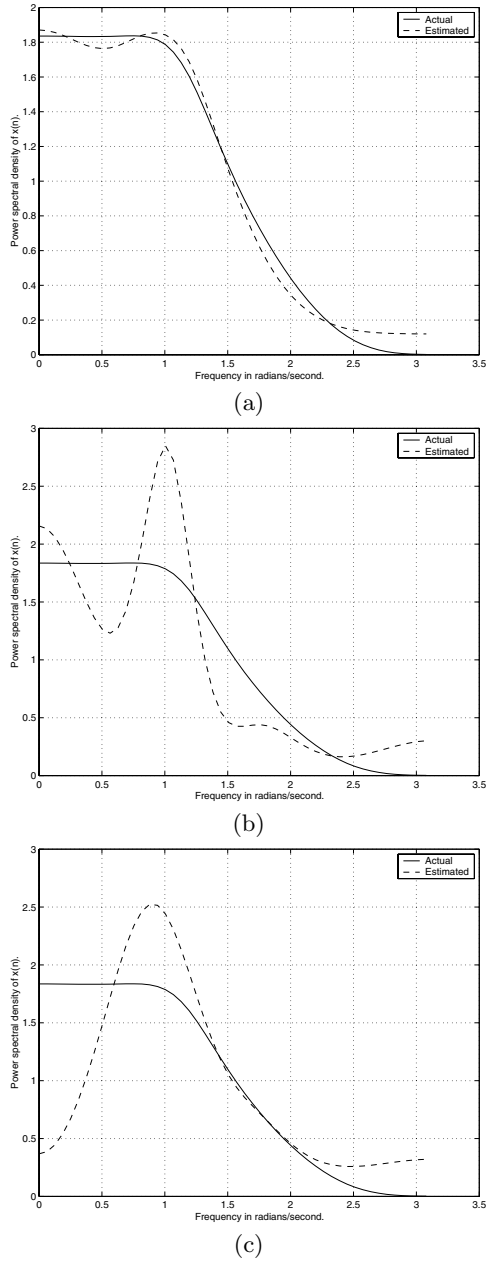


Fig. 7.8. The input power spectra estimated using 4 autocorrelation coefficients per channel for various 2-channel combinations of the filter bank described in Example 7.1. (a) Channels 1 and 2, (b) Channels 1 and 3, (c) Channels 2 and 3. Note that the scale in plots (b) and (c) has changed to accommodate the overshoots that occurs in these estimated spectra. In each plot, the doted curve shows the estimate while the solid one shows the actual spectrum for comparison.

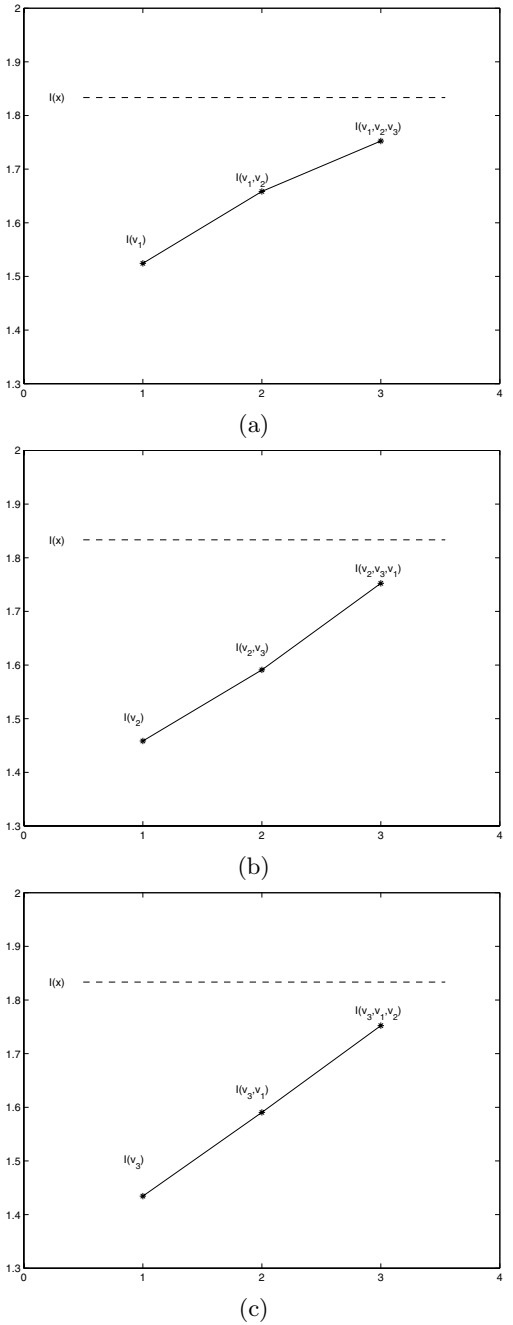


Fig. 7.9. The quantity of statistical information gained by using different combinations of channels of the filter bank described in Example 7.1: (a) Channels 1, 1,2 and 1,2,3, (b) Channels 2, 2,3 and 2,3,1, (c) Channels 3, 3,1 and 3,1,2.

7.4 Redundancy

The saturating nature of $I(v_0, \cdots, v_i)$ as i increases does not mean that the new channels contribute a small amount of information. It means that the new channels contribute a small amount of *additional* information. In other words, the information provided by the new channels becomes more and more redundant as the number of channels increases. This observation motivates defining a quantitative measure for the *redundancy* of the statistical information provided by a multitude of low-rate signals.

To define redundancy, we first need to define the information content of a single low-rate signal $v_i(n)$. This can be done similar to what we did for multiple measurements before. Namely, we let $\mathcal{Q}$, as before, denote the set of PSDs with unit variance. We use the notation $\tilde{\mathcal{S}}^{(i;N)}$ to denote the set of all PSDs which are consistent with N ACS values measured from the signal $v_i(n)$. We then define $\tilde{P}_x^{(i;N)}(e^{j\omega})$ as the estimate of $P_x(e^{j\omega})$ provided by the MEIE from the constraint set $\tilde{\mathcal{S}}^{(i;N)} \bigcap \mathcal{Q}$. The definition of the information content of the ith channel —independent of any other channels— is then made as follows:

Definition 7.2 (Information content of a single measurement signal). *The statistical information contained in the signal $v_i(n)$ about $x(n)$ is given by*

$$I(v_i) \stackrel{\triangle}{=} \lim_{N\to\infty} D(\tilde{P}_x^{(i;N)} \| \bar{P}_x). \tag{7.13}$$

By using the above definition, the definition of redundancy follows naturally[3]:

Definition 7.3 (Redundancy). *The redundancy $R(v_0, \cdots, v_i)$ of the statistical information contained in $v_0(n)$ to $v_i(n)$ about $x(n)$ is given by*

$$R(v_0, \cdots, v_i) \stackrel{\triangle}{=} \sum_{k=0}^{i} I(v_k) - I(v_0, \cdots, v_i). \tag{7.14}$$

Finally, it would be illuminating to recall the notion of "mutual information" introduced in classical information theory (Section 2.3.1) and see how it compares to our definition of information above. Let $X \in \mathbb{R}^N$ be a continuous random variable with PDF $p_X(X)$ and let $Y \in \mathbb{R}^M$, $M < N$, be a non-invertible function of X so that the conditional density function $p_{X|Y}(X|Y)$ exists. Then, the (average) information contained in Y about X is defined as the mutual information between X and Y, which, in turn, can be written as the difference between the entropy of X and the conditional entropy of X given Y (Gallager, 1968, Equation 2.2.17). In other words, in classical

[3] The reader should easily notice the similarity between the definition of redundancy here and the definitions of "mutual information" and "mutual information rate" made in information theory (see Sections 2.3.1 and 2.5.4).

information theory the "information contained in Y about X" is quantified by $R(X;Y) = H(X) - H(X|Y)$.

It is easy, using equations (7.1) - (7.5), to show that $\lim_{N\to\infty} D(P_x^{(i;N)} \| \bar{P}_x)$ used in Definition 7.1 can be written as $H(\bar{P}_x) - \lim_{N\to\infty} H(P_x^{(i;N)})$. Thus, $I(v_0, \cdots, v_i)$ is similar to mutual information in the sense that it too measures the reduction in entropy of a random variable after a related variable is observed. There is, however, an important conceptual difference between mutual information and the quantity measured by $I(v_0, \cdots, v_i)$: The PDFs used in the classical formula $H(X) - H(X|Y)$ are "actual" while the PSDs used in Definition 7.1 are "inferred". (Recall that there is no deductive way to specify these PSDs even after doing statistical experiments on the low-rate signals.)

7.5 Scalability in terms of information

In the previous sections we discussed in detail how a multirate observation is done and how much information is gained in each observation. In this section, we are going to compare *multirate analysis systems* in terms of the amount of information provided by their output about their input.

A very desirable property of a multirate analysis system is that it can supply most of the information about the input signal in a few number of its low-rate outputs. If this is the case, we might save in our computational or communication resources by picking those low-rate signals as representatives of the original full-rate signal. The property just described is often referred to as "scalability". Here, we consider scalability of multirate analysis systems with regard to the "statistical information" contained in their outputs. The reader is referred to Chapter 6 for an indepth description of the concept of scalability and its quantification in general.

To make things formal, let's consider a class $\mathcal{K}$ of M-channel analysis systems.

Definition 7.4 (Information vector). *Let $v_i(n)$, $0 \le i \le M-1$, denote the low-rate outputs of a multirate analysis system $\mathbf{h} \in \mathcal{K}$ and let $\mu_i = I(v_0, \cdots, v_i)$ for $0 \le i \le M-1$. The vector $\mathbf{J_h} = [\mu_0, \mu_1, \cdots, \mu_{M-1}]^T$ is called the information vector associated with $\mathbf{h}$.*

In simple terms, the components of the information vector $\mathbf{J_h}$ show the cumulative amount of information gained by the channels of the analysis system $\mathbf{h}$. Using the same approach used in Chapter 6, we consider an analysis system $\mathbf{h}_1 \in \mathcal{K}$ to be *more scalable from the information point of view* or, simply, *more informative* than $\mathbf{h}_2 \in \mathcal{K}$ if all the low-rate observations on the outputs of $\mathbf{h}_1$ are more informative than similar observations made using outputs of $\mathbf{h}_2$.

Definition 7.5 (Information scalability). *A multirate analysis system* $\mathbf{h}_1 \in \mathcal{K}$ *is more informative than* $\mathbf{h}_2 \in \mathcal{K}$*, in symbols* $\mathbf{h}_1 \overset{INF}{\geq} \mathbf{h}_2$*, if* $\mathbf{J}_{\mathbf{h}_1} \geq \mathbf{J}_{\mathbf{h}_2}$.

Let $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$ be any three elements in $\mathcal{K}$. It is easily verified that the following statements are true:

$$\mathbf{h}_1 \overset{INF}{\geq} \mathbf{h}_1 \qquad \text{(Reflexivity)}$$

$$\mathbf{h}_1 \overset{INF}{\geq} \mathbf{h}_2 \ , \ \mathbf{h}_2 \overset{INF}{\geq} \mathbf{h}_1 \quad \Rightarrow \mathbf{h}_1 = \mathbf{h}_2 \qquad \text{(Antisymmetry)}$$

$$\mathbf{h}_1 \overset{INF}{\geq} \mathbf{h}_2 \ , \ \mathbf{h}_2 \overset{INF}{\geq} \mathbf{h}_3 \Rightarrow \mathbf{h}_1 \overset{INF}{\geq} \mathbf{h}_3 \qquad \text{(Transitivity)}$$

The three properties above show that the binary relation $\overset{INF}{\geq}$ induces a *partial ordering* on $\mathcal{K}$. The set $\mathcal{K}$ is therefore a poset (see Section 2.2.3). We call the greatest element of the poset $\mathcal{K}$ the optimal multirate analysis system from the information viewpoint:

Definition 7.6 (INF-optimality). *A multirate analysis system* $\mathbf{h}^* \in \mathcal{K}$ *is called optimal in the sense of informativity (INF-optimal for short) if* $\mathbf{h}^* \overset{INF}{\geq} \mathbf{h}, \forall \mathbf{h} \in \mathcal{K}$.

When it exists, $\mathbf{h}^*$ is the most informative multirate analysis system to observe $x(n)$ with. In other words, *all* the low-rate Maximum Entropy estimates of $P_x(e^{j\omega})$ obtained by observing the outputs of $\mathbf{h}^*$ are the most accurate possible. For a poset the existence of a greatest element is not guaranteed. Therefore, an INF-optimal analysis system $\mathbf{h}^*$ might not exist in $\mathcal{K}$. It is possible, nonetheless, to embed the partial ordering introduced by $\overset{INF}{\geq}$ into a total (simple) ordering using appropriate order-preserving functions. This helps to re-define the notion of INF-optimality such that an optimal solution exists. Such issues might be worked out similar to those in Chapter 6 and will not be considered here.

7.6 Open problems

In this chapter we introduced a quantitative measure for the *statistical information* gained in multirate observations. We also introduced the notion of INF-optimality for comparing different analysis systems. Here, we point out some aspects of our definition of statistical information that deem further attention and may be investigated in future research.

7.6.1 Cross-correlation data are ignored

Recall from Chapter 3 that the PSD estimates provided by the Maximum Entropy Inference Engine are based on the autocorrelation coefficients of the

low-rate signals. This means that any (potential) information that could have been obtained by considering the cross-correlation between the low-rate measurements is ignored.

7.6.2 Information rate of the low-rate signals

The notion of "statistical information" defined in this chapter is conceptually different from "information rate" defined in Chapter 2. A major distinction is that the statistical information provided by a low-rate signal depends both on that signal's statistical properties and on the characteristics of the analysis system that produced it. Nonetheless, there are certainly connections between, say, the "mutual information rate" of the low-rate signals and their redundancy. Investigating relations of this type is an open topic.

7.6.3 INF-optimality, SC-optimality and PCFB

Investigating the connection between the concept of SC-optimality introduced in Chapter 6 and the concept of INF-optimality introduced in this chapter is another open problem. Specifically, it would be interesting to apply the notion of INF-optimality to the class $\mathcal{L}$ of analysis systems introduced in Chapter 6 and investigate whether PCFB is optimal from the information point of view as well.

8

Distributed Algorithms

8.1 The need for distributed algorithms

In recent years, a new information collection paradigm which advocates connecting a large number of inexpensive and small sensors in a *wireless sensor network* has emerged. The building blocks of a sensor network, often called "Motes", are self-contained, battery-powered computers that measure light, sound, temperature, humidity, and other environmental variables (See Fig 8.1).

As we mentioned in Chapter 1, a wireless sensor network can be very flexible, cost effective, and robust with respect to individual Mote's failure. However, there are many theoretical and engineering obstacles that must be overcome for sensor networks to become viable. For example, new networking protocols must be devised to allow the sensor nodes to spontaneously create an

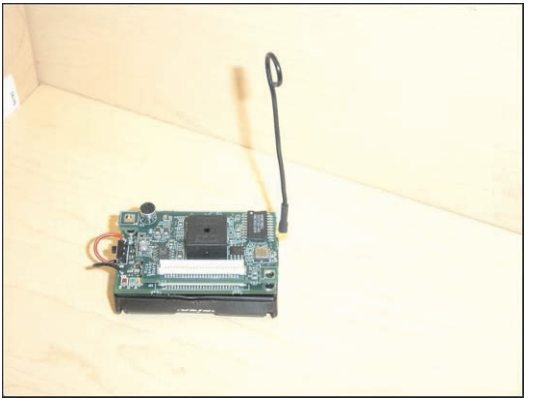


Fig. 8.1. A wireless sensor node or "Mote" made by Crossbow Technology, Inc. in San Jose, California.

impromptu network, dynamically adapt to device failure, manage movement of sensor nodes, and react to changes in task and network requirements.

From the signal processing point of view, the main challenge is the distributed fusion of sensor data across the network. In order to be viable, the distributed data fusion algorithm must meet the following requirements:

(i) It must be very efficient in utilizing the limited range and data rate of each sensor's communication module.
(ii) It must work even if some sensors in the network fail.

In this chapter, we introduce a class of multirate information fusion algorithms which are highly suited for distributed applications such as wireless sensor networks. In our algorithms, the data fusion problem is formulated as a problem of finding a "point" in the intersection of some "convex sets". The key advantage of this viewpoint is that the solution can be found using a series of projections onto the individual sets. The projections can be computed locally at each sensor node allowing the fusion process to be done in a parallel and distributed fashion.

For clarity and simplicity, we will introduce our method by applying it to the problem of multirate spectrum estimation — which we are already familiar with. However, the algorithms discussed in this chapter are separate from the specific problem that is used to illustrate them. They are much more general, much bigger, than multirate spectrum estimation. Somewhere over the horizon, in the direction in which these algorithms point, a whole new field of "distributed signal processing" is waiting for us!

The material in this chapter are based on the author's research originally published in (Jahromi and Aarabi, 2006).

8.2 Spectrum estimation as a convex feasibility problem

Consider a hypothetical scenario where a sound source (say a speaker) is monitored by a collection of Motes put at various known locations in a room (Fig. 8.2). Because of reverberation, noise and other artifacts, the signal arriving at each Mote location is different. The Motes (which constitute the sensor nodes in our network) are equipped with microphones, sampling devices, sufficient signal processing hardware and some communication means. Each Mote can process its observed data, come up with some statistical inference about it and share the result with other nodes in the network. However, to save energy and communication bandwidth, *the Motes are not allowed to share their raw observed data with each other.*

Our goal is to design a fusion algorithm so that an estimate of the power spectrum of the sound source *consistent with the observations made by all Motes* is obtained after a number of *local calculations* and *local communications* between the Motes. One way to achieve this goal is to formulate the

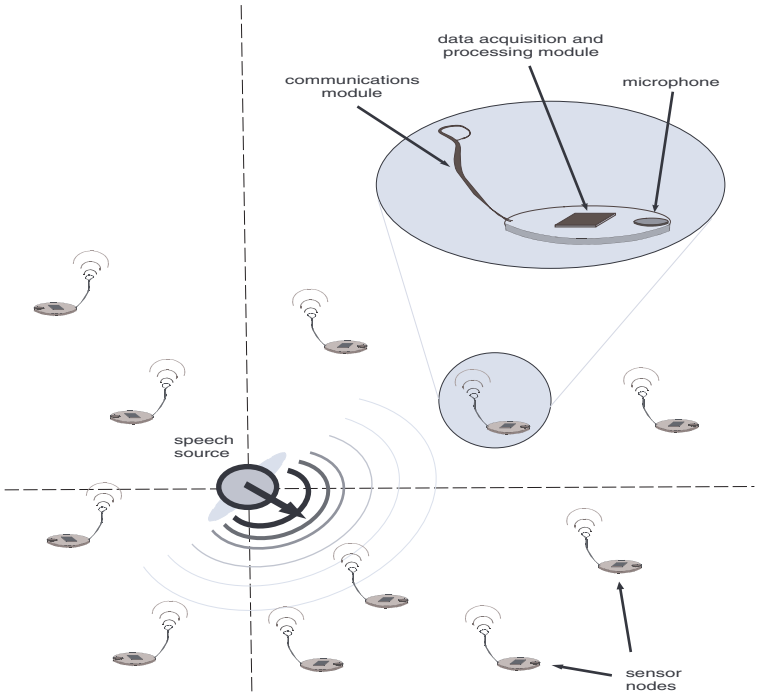


Fig. 8.2. A sensor network monitoring a stationary sound source in room.

spectrum estimation problem as a *convex feasibility problem*. This is done below.

Let $x(n)$ denote a discrete version of the signal produced by the source and assume that it is a zero-mean Gaussian wide-sense stationary (WSS) random process. The sampling frequency f_s associated with $x(n)$ is arbitrary and depends on the frequency resolution desired in the spectrum estimation process.

We denote by $v_i(n)$ the signal produced at the front end of the ith sensor node. We assume that $v_i(n)$ are related to the original source signal $x(n)$ by the model shown in Fig. 8.3. The linear filter $H_i(z)$ in this figure models the combined effect of room reverberations, microphone's frequency response and an additional filter which the system designer might want to include. The decimator block which follows the filer represents the (potential) difference between the sampling frequency f_s associated with $x(n)$ and the actual sampling frequency of the Mote's sampling device. Here, it is assumed that the sampling frequency associated with $v_i(n)$ is f_s/N_i where N_i is a fixed natural number.

It is straightforward to show that the signal $v_i(n)$ in Fig. 8.3 is also a WSS processes. The autocorrelation coefficients $R_{v_i}(k)$ associated with $v_i(n)$ are

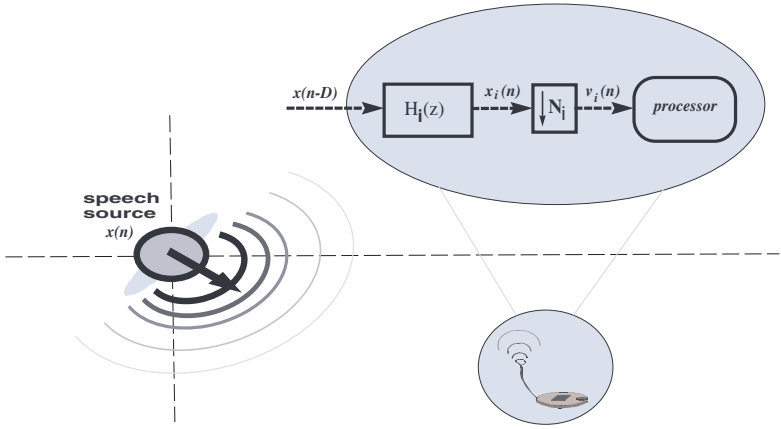


Fig. 8.3. The relation between the signal $v_i(n)$ produced by the front end of the ith sensor and the original source signal $x(n)$.

given by

$$R_{v_i}(k) = R_{x_i}(N_i k) \tag{8.1}$$

where

$$R_{x_i}(k) = (h_i(k) \star h_i(-k)) \star R_x(k), \tag{8.2}$$

and $h_i(k)$ denotes the impulse response of $H_i(z)$. We can express $R_{v_i}(k)$ as a function of the source signal's power spectrum as well. To do this, we define $G_i(z) \stackrel{\triangle}{=} H_i(z)H_i(z^{-1})$ and then use it to write (8.2) in the frequency domain:

$$R_{x_i}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) G_i(e^{j\omega}) e^{jk\omega} d\omega. \tag{8.3}$$

Combining (8.1) and (8.3), we then get

$$R_{v_i}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) G_i(e^{j\omega}) e^{jN_i k\omega} d\omega. \tag{8.4}$$

The above formula shows that $P_x(e^{j\omega})$ uniquely specifies $R_{v_i}(k)$ for all values of k. However, the reverse is not true. That is, in general, knowing $R_{v_i}(k)$ for some or all values of k is not sufficient for characterizing $P_x(e^{j\omega})$ uniquely.

Recall that $v_i(n)$ is a WSS signal so all the statistical information that can be gained about it is confined in its autocorrelation coefficients. One might use the signal processing hardware available at each sensor node and estimate the autocorrelation coefficients $R_{v_i}(k)$ for some k, say $0 \leq k \leq L-1$. Now, we may pose the sensor network spectrum estimation problem as follows:

Problem 8.1. Let $\mathcal{Q}_{i,k}$ denote the set of all power spectra which are consistent with the kth autocorrelation coefficient $R_{v_i}(k)$ estimated at the ith sensor node. That is, $P_x(e^{j\omega}) \in \mathcal{Q}_{i,k}$ if

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} P_x(e^{j\omega})G_i(e^{j\omega})e^{jMk\omega}d\omega = R_{v_i}(k),$$
$$P_x(e^{j\omega}) \geq 0,$$
$$P_x(e^{j\omega}) = P_x(e^{-j\omega}),$$
$$P_x(e^{j\omega}) \in L^1(-\pi,\pi).$$

Define $\mathcal{Q} \triangleq \bigcap_{i=1}^{N}\bigcap_{k=0}^{L-1}\mathcal{Q}_{i,k}$ where N is the number of nodes in the network and L is the number of autocorrelation coefficients estimated at each node. Find a $P_x(e^{j\omega})$ in $\mathcal{Q}$.

The above problem essentially means finding a $P(e^{j\omega})$ in the intersection of the feasible sets $\mathcal{Q}_{i,k}$. It is easy to verify that the sets $\mathcal{Q}_{i,k}$ are closed and convex. The problem of finding a point in the intersection of finitely many closed convex sets is known as the *convex feasibility problem* and is an active area of research in applied mathematics.

In the ideal case where the observed autocorrelation coefficients $R_{v_i}(k)$ are exact, the sets $\mathcal{Q}_{i,k}$ are non-empty and admit a non-empty intersection $\mathcal{Q}$ as well. In this case $\mathcal{Q}$ contains infinitely many $P_x(e^{j\omega})$. In the case that the measurements $v_i(n)$ are contaminated by noise or $R_{v_i}(k)$ are estimated based on finite-length data records, the intersection set $\mathcal{Q}$ might be empty due to the potential inconsistency of the autocorrelation coefficients used. So Problem 8.1 has either infinitely many solutions or no solution at all! In any case, it is ill-posed.

8.3 Solution using generalized projections

An elegant way to solve a convex feasibility problem is to employ a series of *generalized projections* (Censor and Zenios, 1997). A generalized projection is essentially a regularization method with a *generalized distance* serving as the stabilizing functional. A great advantage of using the generalized projections formulation is that the solution $P^* \in \mathcal{Q}$ can be found using a series of projections onto the intermediate sets $\mathcal{Q}_{i,k}$. These intermediate projections can be computed locally at each sensor node thus allowing the computations to be done simultaneously and in a highly distributed fashion.

A generalized distance is a real-valued non-negative function of two vector variable $D(X,Y)$ defined in a specific way such that it's value may represent the distance between X and Y in some generalized sense. When defining generalized distances, it is customary not to require the symmetry condition. Thus, $D(X,Y)$ may not be the same as $D(Y,X)$. Moreover, we do not insist on the triangle inequality that a traditional metric must obey either.

Example 8.1. Let $P_1(e^{j\omega}) > 0$ and $P_2(e^{j\omega}) > 0$ be two power spectra in $\mathbf{L}^1(-\pi, \pi)$. The functions

$$D_1(P_1, P_2) = \int_{-\pi}^{\pi} (P_1 - P_2)^2 \, d\omega,$$
$$D_2(P_1, P_2) = \int_{-\pi}^{\pi} \left(P_1 \ln \frac{P_1}{P_2} + P_2 - P_1 \right) d\omega,$$
$$D_3(P_1, P_2) = \int_{-\pi}^{\pi} \left(\frac{P_1}{P_2} - \ln \frac{P_1}{P_2} - 1 \right) d\omega,$$

can be used to measure the generalized distance between $P_1(e^{j\omega})$ and $P_2(e^{j\omega})$. These functions are non-negative and become zero if and only if $P_1 = P_2$. Note that D_1 is simply the Euclidean distance between P_1 and P_2. The functions D_2 and D_3 have roots in information theory and statistics. They are known as the Kullback-Leibler divergence and Burg cross entropy, respectively. $\diamondsuit$

By using a suitable generalized distance, we can convert our original sensor network spectrum estimation problem (Problem 8.1) into the following minimization problem:

Problem 8.2. Let $\mathcal{Q}$ be defined as in Problem 8.1. Find $P_x^*(e^{j\omega})$ in $\mathcal{Q}$ such that

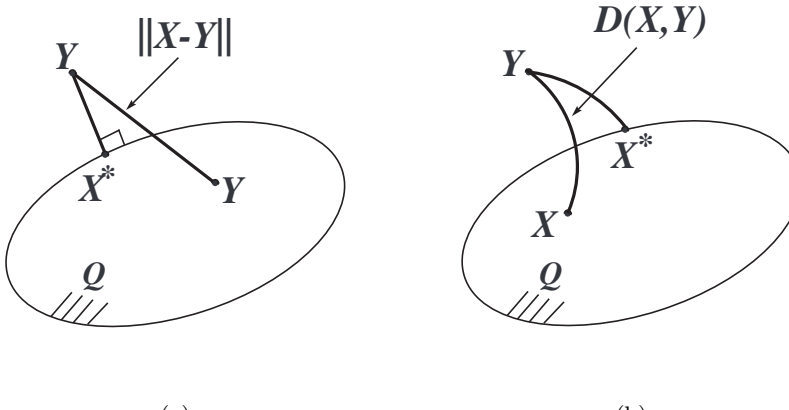
$$P^* = \arg \min_{P \in \mathcal{Q}} D(P, P_0), \tag{8.5}$$

where $P_0(e^{j\omega})$ is an arbitrary power spectrum, say $P_0(e^{j\omega}) = 1, -\pi \leq \omega < \pi$.

When a unique P^* exists, it is called the *generalized projection* of P_0 onto $\mathcal{Q}$ (Bauschke and Borwein, 1996). In general, a projection of a given point onto a convex set is defined as another point which has two properties: First, it belongs to the set onto which the projection operation is performed and, second, it renders a minimal value to the distance between the given point and any point of the set (Fig. 8.4).

If the Euclidean distance $||X - Y||$ is used in this context then the projection is called a metric projection. In some cases, such as the spectrum estimation problem considered here, it turns out to be very useful to introduce more general means to measure the "distance" between two vectors. The main reason is that the functional form of the solution will depend on the choice of the distance measure used in the projection. Often, a functional form which is easy to manipulate or interpret (for instance, a rational function) can not be obtained using the conventional Euclidean metric.

It can be shown that the distances D_1 and D_2 in Example 8.1 lead to well-posed solutions for P^*. The choice D_3 will lead to a unique solution given that ceratin singular power spectra are excluded from the space of valid solutions (Borwein and Lewis, 1993). It is not known whether D_3 will lead to a stable solution. As a result, the well-posedness of Problem 8.2 when D_3 is used is not yet established.



(a) (b)

Fig. 8.4. Symbolic depiction of metric projection (a) and generalized projection (b) of a vector Y into a closed convex set $\mathcal{Q}$. In (a) the projection X^* is selected by minimizing the metric $||X - Y||$ over all $X \in \mathcal{Q}$ while in (b) X^* is found by minimizing the generalized distance $D(X, Y)$ over the same set.

Remark 8.1. Well-posedness of the minimization problem (8.5) when D is the Kullback-Leibler divergence D_2 has been established in several works including Klaus and Smith (1988), Amato and Hughes (1991), Borwein and Lewis (1991), Eggermont (1993) and Teboulle and Vajda (1993). Well-posedness results exist for certain classes of generalized distance functions as well Teboulle and Vajda (1993), Leonev (2000). Unfortunately, the Burg cross entropy D_3 does not belong to any of these classes. While Burg cross entropy lacks theoretical support as a regularizing functional, it has been used successfully to resolve ill-posed problems in several applications including spectral estimation and image restoration. See, for example, Wu (1997) and references therein. The desirable feature of Burg cross entropy in the context of spectrum estimation is that its minimization subject to linear constraints $P_x^*(e^{j\omega}) \in \mathcal{Q}$ leads to rational power spectra.

8.4 Distributed algorithms based on local generalized projections

As we mentioned before, a very interesting aspect of the generalized projections formulation is that the solution $P^* \in \mathcal{Q}$ can be found using a series of projections onto the intermediate sets $\mathcal{Q}_{i,k}$. In this section, we first calculate the generalized projection of a given power spectrum onto the sets $\mathcal{Q}_{i,k}$ for the sample distance functions introduced in Example 8.1. Then, we propose a distributed algorithm for calculating the final solution P^* from these intermediate projections.

Let $P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_j]}$ denote the power spectrum resulting from projecting a given power spectrum P_1 onto the set $\mathcal{Q}_{i,k}$ using a given distance functions D_j. That is,

$$P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_j]} \stackrel{\triangle}{=} \arg \min_{P \in \mathcal{Q}_{i,k}} D_j(P, P_1). \tag{8.6}$$

Using standard techniques from calculus of variations we can show that the generalized distances D_1, D_2 and D_3 introduced in Example 8.1 result in projections of the form

$$\begin{aligned} P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_1]} &= P_1(e^{j\omega}) - \alpha G_i(e^{j\omega}) \cos(Mk\omega), \\ P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_2]} &= P_1(e^{j\omega}) \exp\left(-\beta G_i(e^{j\omega}) \cos(Mk\omega)\right), \\ P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_3]} &= \left(P_1(e^{j\omega})^{-1} + \gamma G_i(e^{j\omega}) \cos(Mk\omega)\right)^{-1}, \end{aligned}$$

where α, β and γ are parameters (Lagrange multipliers). These parameter should be chosen such that in each case $P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_j]} \in \mathcal{Q}_{i,k}$. That is,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} P_{[P_1 \mapsto \mathcal{Q}_{i,k}; D_j]} G_i(e^{j\omega}) e^{jMk\omega} d\omega = R_{v_i}(k). \tag{8.7}$$

The reader may observe that the above equation leads to a closed-form formula for α but in general finding β and γ requires numerical methods. The projection formulae developed above can be employed in a variety of iterative algorithms to fined a solution in the intersection of $\mathcal{Q}_{i,k}$. We discuss two example algorithms below.

8.4.1 The Ring Algorithm

Ring Algorithm is a very simple algorithm: it starts with an initial guess $P^{(0)}$ for $P_x(e^{j\omega})$ and then calculates a series of successive projections onto the constraint sets $\mathcal{Q}_{i,k}$. Then, it takes the last projection, now called $P^{(1)}$, and projects it back onto the first constraint set. Continuing this process will generates a sequence of solutions $P^{(0)}, P^{(1)}, P^{(1)}, \cdots$ which will eventually converge to a solution $P^* \in \bigcap_{i,k} \mathcal{Q}_{i,k}$ (Censor and Zenios, 1997). Steps of the Ring Algorithm are summarized in the text box below. A graphical representation of this algorithm is shown in Fig. 8.5.

Example 8.2. Consider a simple 4-sensor network similar to the one shown in Fig. 8.4. Assume that the down-sampling ratio in each Mote is equal to 4. Thus, $N_0 = N_1 = N_2 = N_3 = 4$. Assume, further, that the transfer functions $H_0(z)$ to $H_3(z)$ which relate the Motes' front-end output $v_i(n)$ to the original source signal $x(n)$ are given as follows:

$$H_0(z) = \frac{0.0753 + 0.1656z^{-1} + 0.2053z^{-2} + 0.1659z^{-3} + 0.0751z^{-4}}{1.0000 - 0.8877z^{-1} + 0.6738z^{-2} - 0.1206z^{-3} + 0.0225z^{-4}}$$

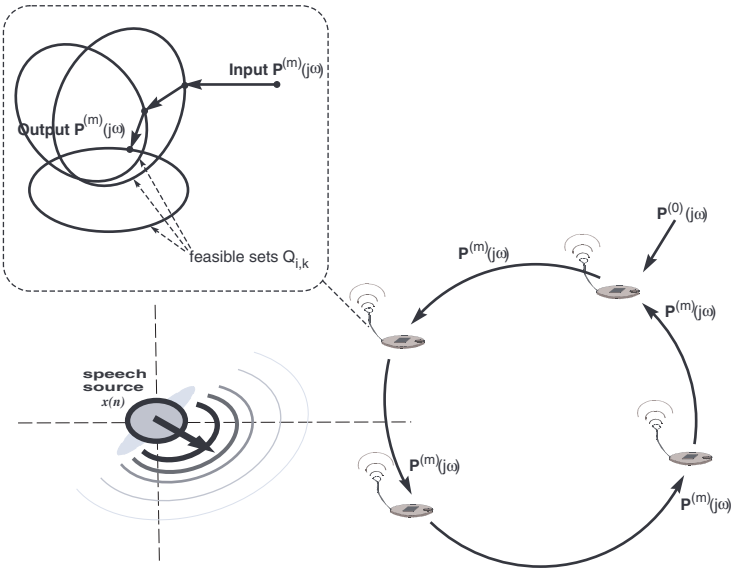


Fig. 8.5. Graphical depiction of the Ring Algorithm. For illustrative reasons, only three feasible sets $Q_{i,k}$ are shown in the inside picture. Also, it's shown that the output spectrum $P^{(m)}(e^{j\omega})$ is obtained from the input $P^{(m)}(e^{j\omega})$ only after three projections. In practice, each sensor node has L feasible sets and has to repeat the sequence of projections many times before it can successfully project the input $P^{(m)}(e^{j\omega})$ into the intersection of its feasible sets.

$$H_1(z) = \frac{0.4652 - 0.1254z^{-1} - 0.3151z^{-2} + 0.0975z^{-3} - 0.0259z^{-4}}{1.0000 - 0.6855z^{-1} + 0.3297z^{-2} - 0.0309z^{-3} + 0.0032z^{-4}}$$
$$H_2(z) = \frac{0.3732 - 0.8648z^{-1} + 0.7139z^{-2} - 0.1856z^{-3} - 0.0015z^{-4}}{1.0000 - 0.5800z^{-1} + 0.5292z^{-2} - 0.0163z^{-3} + 0.0107z^{-4}}$$
$$H_3(z) = \frac{0.1931 - 0.4226z^{-1} + 0.3668z^{-2} - 0.0974z^{-3} - 0.0405z^{-4}}{1.0000 + 0.2814z^{-1} + 0.3739z^{-2} + 0.0345z{-}3 - 0.0196z^{-4}}$$

The above transfer functions were chosen to show typical low-pass, band-pass and high-pass characteristics (Fig. 8.6) They were obtained using standard filter design techniques. The input signal whose power spectrum is to be estimated was chosen to have a smooth low-pass spectrum . We used the Ring Algorithm with $L = 4$ and the Euclidean metric D_1 as the distance function to estimate the input signal's spectrum. The results are shown in (Fig. 8.7). As seen in this figure, the algorithm converges to a solution which is in this case almost identical to the actual input spectrum in less than 100 rounds. $\diamondsuit$

8.4.2 The Star Algorithm

The Ring Algorithm is completely decentralized. However, it will not converge to a solution if the feasible sets $\mathcal{Q}_{i,k}$ do not have an intersection (which can

The Ring Algorithm

Input: A distance function $D_j(P_1, P_2)$, an initial power spectrum $P_0(e^{j\omega})$, the squared sensor frequency responses $G_i(e^{j\omega})$, and the autocorrelation estimates $R_{v_i}(k)$ for $k = 0, 1, \cdots, L-1$ and $i = 1, 2, \cdots, N$.
Output: A power spectrum $P_*(e^{j\omega})$.
Procedure:

1. Let $m = 0$, $i = 1$ and $P^{(m)} = P_0$.
2. Send $P^{(m)}$ to the ith sensor node.
   At the ith sensor:
   (i) Let $k = 0$ and define $\tilde{P}_k = P^{(m)}$.
   (ii) Calculate $\tilde{P}_k = P_{[\tilde{P}_{k-1} \mapsto \mathcal{Q}_{i,k}; D_j]}$ for $k = 1, 2, \cdots, L-1$.
   (iii) If $D(\tilde{P}_{L-1}, \tilde{P}_0) > \epsilon$ then let $\tilde{P}_0 = \tilde{P}_{L-1}$ and go back to item (ii). Otherwise, let $i = i + 1$ and go to Step 3.
3. If $(i \bmod N) = 1$ then set $m = m+1$ and reset i to 1. Otherwise, set $P^{(m)} = \tilde{P}_{L-1}$ and go back to Step 2.
4. Define $P^{(m)} = \tilde{P}_{L-1}$. If $D(P^{(m)}, P^{(m-1)}) > \epsilon$, go back to Step 2. Otherwise output $P^* = P^{(m)}$ and stop.

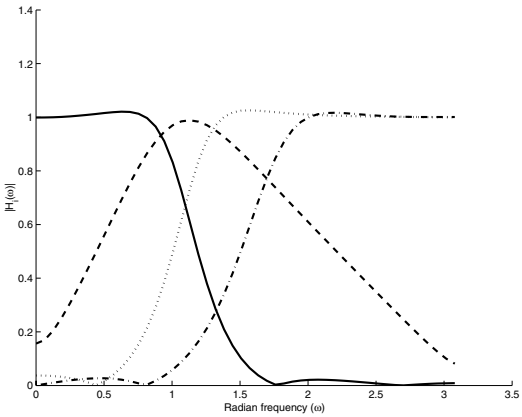


Fig. 8.6. Frequency response amplitude of the transfer functions used in Example 8.2. The curves show, from left to right, $|H_0(e^{j\omega})|$, $|H_1(e^{j\omega})|$, $|H_2(e^{j\omega})|$ and $|H_3(e^{j\omega})|$.

happen due to measurement noise) or one or more sensors in the network are faulty. The Star Algorithm is an alternative distributed algorithm for fusing individual sensors' data. It combines successive projections onto $\mathcal{Q}_{i,k}$ with a kind of averaging operation to generate a sequence of solutions $P^{(m)}$. This sequence will eventually converge to a solution $P^* \in \bigcap_{i,k} \mathcal{Q}_{i,k}$ if one exists. The Star Algorithm is fully parallel and hence much faster than the Ring Algorithm. It provides some degree of robustness to individual node's failure as well. However, it includes a centralized step which needs to be accommodated

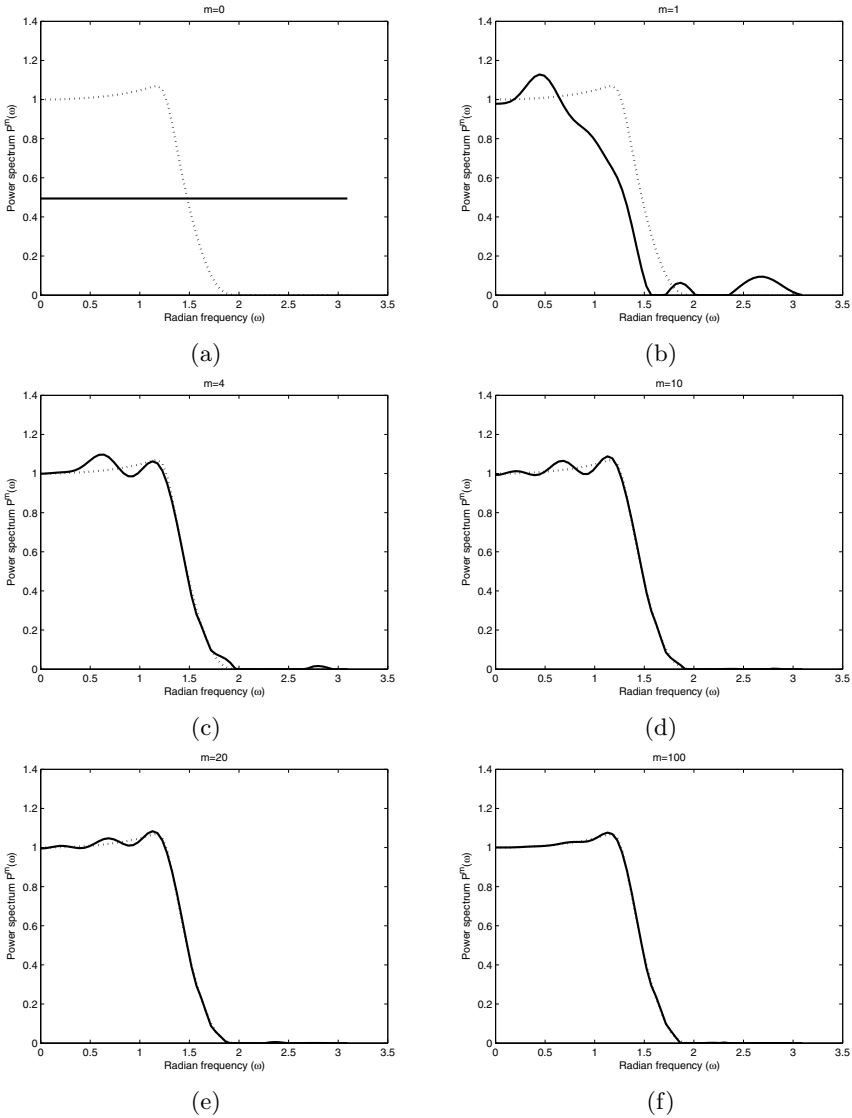


Fig. 8.7. Ring algorithm convergence results. In each figure, the dashed curve shows the source signal's actual power spectrum while the solid curve is the estimate obtained by the Ring Algorithm after m rounds. A "round" means projections have been passed through all the nodes in the network.

for when the system's network protocol is being designed. Steps of the Star Algorithm are summarized in the text box below. A graphical representation of this algorithm is shown in Fig. 8.8.

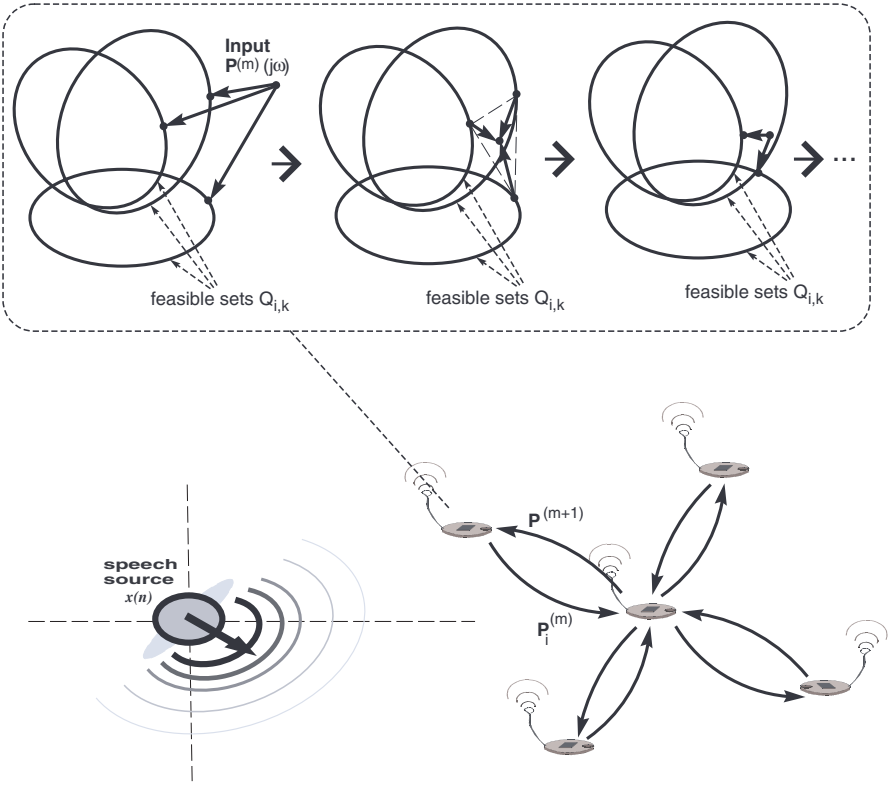


Fig. 8.8. The Star Algorithm. Again, only three feasible sets $Q_{i,k}$ are shown in the inside picture. In practice, each sensor node has to repeat the sequence of projections and averaging many times before it can successfully project the input $P^{(m)}(e^{j\omega})$ supplied by the central node into the intersection of its feasible sets. The projection result, which is called $P_i^{(m)}(e^{j\omega})$ is sent back to the central node. The central node then averages all the $P_i^{(m)}(e^{j\omega})$ it has received and averages them to produce $P^{(m+1)}(e^{j\omega})$. This is sent back to the individual nodes and the process repeats.

Example 8.3. Consider a simple 5-sensor network similar to the one shown in Fig. 8.8. Assume that the down-sampling ratio in each Mote is equal to 4. Thus, again, $N_0 = N_1 = N_2 = N_3 = 4$. Assume, further, that the transfer functions $H_0(z)$ to $H_3(z)$ which relate the Motes' front-end output $v_i(n)$ to the original source signal $x(n)$ are the same as those introduced in Example 8.2. We simulated the Star Algorithm with $L = 4$ and the Euclidean metric D_1 as the distance function to estimate the input signal's spectrum. The results are shown in (Fig. 8.9). Like the Ring Algorithm, the Star Algorithm also converges to a solution which is almost identical to the actual input spectrum in less than 100 rounds. ♢

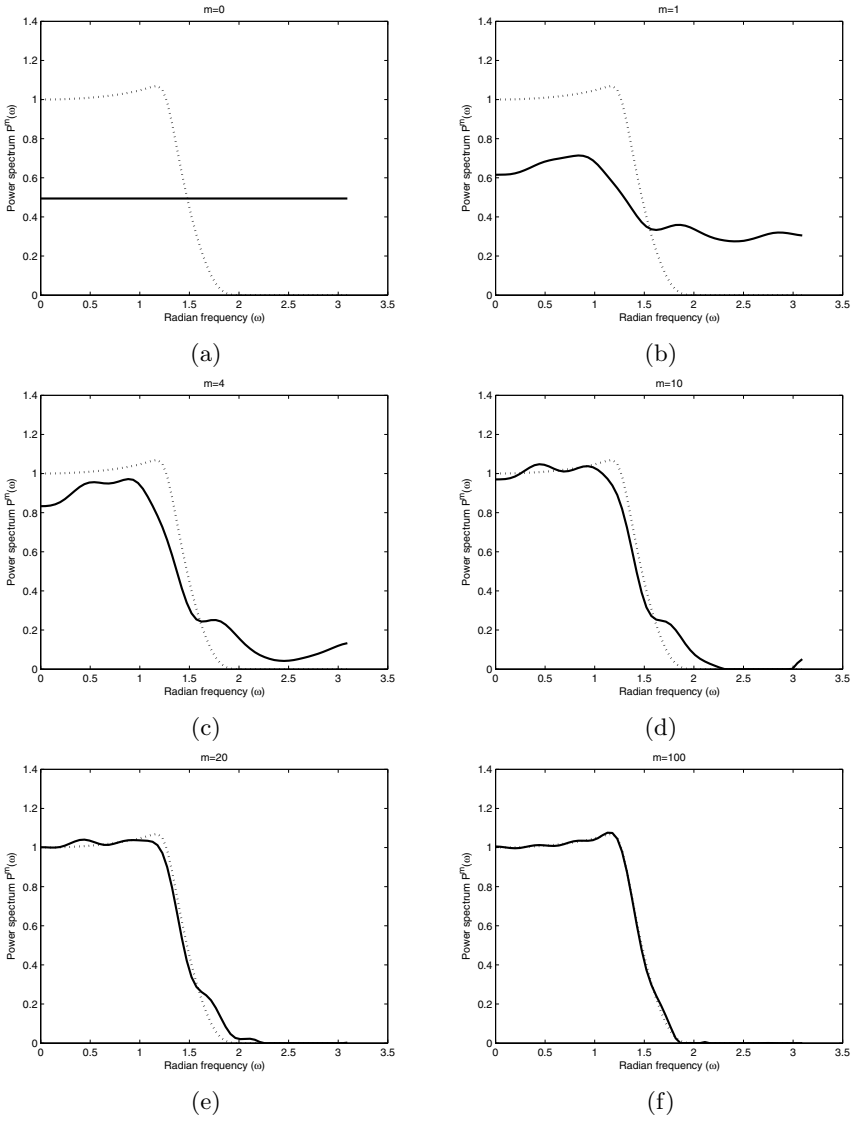


Fig. 8.9. Star algorithm convergence results.

8.5 Open problems

In this chapter we considered the problem of fusing the statistical information gained by a distributed network of sensors. We constructed a mathematical model for this problem where the solution is obtained by finding a point in the intersection of finitely many closed convex sets. Some key advantages of the convex feasibility formulation are summarized below.

The Star Algorithm

Input: A distance function $D_j(P_1, P_2)$, an initial power spectrum $P_0(e^{j\omega})$, the squared sensor frequency responses $G_i(e^{j\omega})$, and the autocorrelation estimates $R_{v_i}(k)$.
Output: A power spectrum $P_*(e^{j\omega})$.
Procedure:

1. Let $m = 0$ and $P^{(0)} = P_0$.
2. Send $P^{(m)}$ to all sensor nodes.
   At the ith sensor:
   (i) Let $n = 0$ and define $\tilde{P}^{(n)} = P^{(m)}$.
   (ii) Calculate $\tilde{P}_k = P_{[\tilde{P}^{(n)} \mapsto \mathcal{Q}_{i,k}; D_j]}$ for all k.
   (iii) Calculate $\tilde{P}^{(n+1)} = \arg\min_P \sum_k D(P, \tilde{P}_k)$.
   (iv) If $D(\tilde{P}^{(n+1)}, \tilde{P}^{(n)}) > \epsilon$ go to item (ii) and repeat. Otherwise, define $P_i^{(m)} = \tilde{P}^{(n+1)}$ and send it to the central unit.
2. Receive $P_i^{(m)}$ from all sensor and calculate $P^{(m+1)} = \arg\min_P \sum_i D(P, P_i^{(m)})$.
3. If $D(P^{(m+1)}, P^{(m)}) > \epsilon$, go to step 2 and repeat. Otherwise stop and output $P^* = P^{(m+1)}$.

(i) The global solution is unique and stable in the sense that small perturbations in the observed data will cause a small change in the solution.
(ii) The functional form of the solution will depend on the choice of the generalized distance used in the projections. Therefore, a functional form which is easy to manipulate or interpret for a specific application (for instance, a rational function) can be obtained using a proper generalized distance.
(iii) The formulation can be extended to a variety of network topologies. Some topologies might allow for the most efficient computation, some might generate a more robust setup, and others lead to varies degrees of compromise between these desirable properties.
(iv) The formulation has a very rich mathematical structure relying on recent results in several fields of applied mathematics including convex analysis, parallel optimization and regularization theory.

The information fusion theory presented in this chapter is by no means complete. Many issues regarding theoretical performance and practical implementation of the two algorithms we introduced need to be investigated. Also, various other distributed algorithms for solving the problem of finding a solution in the intersection of the feasible sets can be devised. We encourage the reader to use the ideas introduced in this chapter as a starting point and develop a more complete theory for "distributed signal processing" in networked environments!

9

Epilogue

In this book we studied a variety of signal processing problems that mainly involved processing low-rate components of a full-rate information bearing signal.

In developing mathematical models for these problems, we opted for a statistical viewpoint where the original full-rate signal is assumed to be a stationary random process. This allowed us to find unique and stable solutions to the ill-posed inverse problem of reconstructing the samples of the original signal from the available low-rate observations. It also allowed us to pose and solve other interesting problems such as estimating the spectrum of a full-rate signal from low-rate observations, quantifying the information content of individual low-rate measurements and specifying multirate filter banks that lead to scalable decompositions.

The statistical theory presented in this book is very general in the sense that it can be used to pose and solve multirate signal processing problems involving both FIR and IIR filter banks. Furthermore, there are no restrictions on the number of channels and/or the down-sampling ratio in each channel. These are important advantages that distinguish the statistical approach from conventional deterministic formulations considered in the past.

While I have tried my best to provide a systematic introduction to the theory of statistical multirate signal processing, it is a fact that this theory is still far from complete. I invite the reader to examine the initial results presented in this book and extend them to overcome their numerous gaps and shortcomings. Somewhere over the horizon, in the direction in which this book points, a beautiful unified theory of "multirate statistical signal processing" is waiting for us!

References

Aarabi, P. (2001). *The Integration and Localization of Distributed Sensor Arrays*. PhD thesis, Stanford University.

Aarabi, P. (2003). The fusion of distributed microphone arrays for sound localization. *EURASIP Journal of Applied Signal Processing (Special Issue on Sensor Networks)*, 2003(4):338–347.

Aarabi, P., Shi, G., Shanechi, M. M., and Rabi, A. (2005). *Phase-Based Speech Processing*. World Scientific Press.

Akkarakaran, S. and Vaidyanathan, P. P. (2001a). Filter bank optimization with convex objective functions, and the optimality of principal component forms. *IEEE Transactions on Signal Processing*, 49(1):100–114.

Akkarakaran, S. and Vaidyanathan, P. P. (2001b). Results on principal component filter banks: color noise suppression and existence issues. *IEEE Transactions on Information Theory*, 47(3):1003–1020.

Amari, S. I. and Nagaoka, H. (2000). *Methods of Information Geometry*. American Mathematical Society and Oxford University Press.

Amato, U. and Hughes, W. (1991). Maximum entropy regularization of Fredholm integral equations of the first kind. *Inverse Problems*, 7:793–808.

Antoniou, A. (1993). *Digital Filters: Analysis, Design and Applications*. McGraw-Hill, 2nd edition.

Ase, K. C. and Mullis, C. T. (1996). Minimum mean-squared error transform coding and subband coding. *IEEE Transactions on Information Theory*, 42(4):1179–1192.

Basseville, M., Benveniste, A., Chou, K. C., Golden, S. A., Nikoukhah, R., and Willsky, A. S. (1992a). Modeling and estimation of multiresolution stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784.

Basseville, M., Benveniste, A., and Willsky, A. (1992b). Multiscale autoregressive processes, Part I: Schur-Levinson parametrization. *IEEE Transactions on Signal Processing*, 40(8):1915–1934.

Basseville, M., Benveniste, A., and Willsky, A. (1992c). Multiscale autoregressive processes, Part II: Lattice structures for whitening and modeling. *IEEE Transactions on Signal Processing*, 40(8):1935–1953.

Bauschke, H. H. and Borwein, J. M. (1996). On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38:367–426.

Berger, T. and Gibson, J. D. (2000). Lossy source coding. In Verdú, S. and McLaughlin, S. W., editors, *Information Theorey, 50 Years of Discovery*, pages 649–679, Piscataway, NJ. IEEE Press.

Borwein, J. M. and Lewis, A. S. (1991). Convergence of best maximum entropy estimates. *SIAM Journal of Optimization*, 1:191–205.

Borwein, J. M. and Lewis, A. S. (1993). Partially-finite programming in $\mathbf{L}_1$ and the existence of maximum entropy estimates. *SIAM Journal of Optimization*, 3:248–267.

Brandstein, M. and Silverman, H. (1997). A robust method for speech signal time-delay estimation in reverberant rooms. In *Proceedings of Intl. Conf. Acoustic, Speech and Signal Process. (ICASSP)*, pages 375–378, Munich, Germany.

Brauer, A. and Rohrbach, H., editors (1973). *Issai Schur Collected Works*, volume 2, pages 416–427. Springer-Verlag, Berlin. (in German).

Bull, D., Canagarajah, N., and Nix, A., editors (1999). *Insights into Mobile Multimedia Communications*. Academic Press, London.

Burg, J. P. (1967). Maximum entropy spectral analysis. In *37th Annual International Meeting of Society of Exploration Geophysicists*, Oklahoma City, Oklahoma.

Buttkus, B. (2000). *Spectral Analysis and Filter Theory in Applied Geophysics*. Springer-Verlag, Berlin.

Caines, P. E. (1988). *Linear Stochastic Systems*. John Wiley, New York.

Censor, Y. and Zenios, S. A. (1997). *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press.

Chen, T. and Francis, B. A. (1995a). Design of multirate filter banks by $\mathcal{H}_\infty$ optimization. *IEEE Transactions on Signal Processing*, 43(12):2822–2830.

Chen, T. and Francis, B. A. (1995b). *Optimal Sampled Data Control Systems*. Springer-Verlag, London.

Chou, K. C., Willsky, A. S., and Benveniste, A. (1994a). Multiscale recursive estimation, data fusion and regularization. *IEEE Transactions on Automatic Control*, 39(3):464–478.

Chou, K. C., Willsky, A. S., and Nikoukhah, R. (1994b). Multiscale systems, Kalman filters and Riccati equations. *IEEE Transactions on Automatic Control*, 39(3):479–492.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley Interscience, New York.

Crochiere, R. E. and Rabiner, L. R. (1983). *Multirate Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ.

Csiszár, I. (1984). Sanov property, generalized I-projection and a conditional limit theorem. *The Annals of Probability*, 12:768–793.

Csiszár, I. (1991). Why least squares and maximum entropy? An axiomatic approch to inference for linear inverse problems. *The Annals of Statistics*, 19:2032–2066.

Csiszár, I. (1996). Maxent, mathematics and information theory. In Hanson, K. M. and Silver, R. N., editors, *Maximum entropy and Bayesian methods*, pages 35–50, The Netherlands. Kluwer Academic Publishers.

Csiszár, I. and Körner, J. (1981). *Information Theory : coding theorems for discrete memoryless systems*. Akadémiai Kiadó, Budapest.

Delsarte, P., Macq, B., and Slock, D. T. M. (1992). Signal-adapted multiresolution transform for image coding. *IEEE Transactions on Information Theory*, 38:897–904.

Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Regognition*. Speringer-Verlag.

Dewilde, P. and van der Veen, A. J. (1998). *Time-Varying Systems and Computations*. Kluwer Academic Press, Boston.

Douglas, S. C., Amari, S. I., and Kung, S. Y. (1999). Adaptive paraunitary filter banks for spatio-temporal principal and minor subspace analysis. In *Proceedings of Intl. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, volume 2, pages 1089–1092, Phoenix, AZ.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley Interscience, 2nd edition.

Eggermont, P. and LaRiccia, V. (2001). *Maximum penalized likelihood estimation*. Springer-Verlag, New York.

Eggermont, P. P. B. (1993). Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.*, 24(6):1557–1576.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Engl, H. W. and Landl, G. (1993). Convergence rates for maximum entropy regularization. *SIAM Journal on Numerical Analysis*, 30(5):1509–1536.

Fliege, N. J. (1994). *Multirate Digital Signal Processing*. John Wiley, Chichester.

Francis, B. A. (1987). *A Course in $\mathcal{H}_\infty$ control theory*, volume 88 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin.

Gallager, R. G. (1968). *Information Theory and Reliable Communication*. John Wiley, New York.

Gold, B. and Morgan, N. (2000). *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiely, New York, NY.

Goyal, V. K. (2001). Multiple description coding: Compression meets the network. *IEEE Signal Processing Magazine*, 18(5):74–93.

Gray, R. M. (1990). *Entropy and Information Theory*. Springer-Verlag, New York.

Green, M. and Limebeer, D. J. N. (1995). *Linear Robust Control*. Printice-Hall, Englewood Clifes, NJ.

Hardy, G. H., Littlewood, J. E., and Pòlya, G. (1934). *Inequalities*. Cambridge University Press, London.

Hayes, M. H. (1996). *Statistical Signal Processing and Modeling*. Wiely, New York, NY.

Jahromi, O. S. (1997). An adaptive approach to digital filter bank design. Master's thesis, Department of Electrical Engineering, Shiraz University, Shiraz, Iran.

Jahromi, O. S. and Aarabi, P. (2005). Theory and design of multirate sensor arrays. *IEEE Transactions on Signal Processing*, 53(5):1739– 1753.

Jahromi, O. S. and Aarabi, P. (2006). Distributed signal processing in sensor networks. In Zurawski, R., editor, *Embedded Systems Handbook*, chapter 38. Taylor and Francis, Boca Raton, FL.

Jahromi, O. S. and Francis, B. A. (1999). An information theoretic method for multiresolution principal component analysis. In *Proceedings of Intl. Joint Conf. Neural Networks (IJCNN)*, Washington, DC.

Jahromi, O. S., Francis, B. A., and Kwong, R. H. (2003). Algebraic theory of optimal filter banks. *IEEE Transactions on Signal Processing*, 51(2):442– 457.

Jahromi, O. S., Francis, B. A., and Kwong, R. H. (2004a). Relative information of multirate sensors. *Information Fusion*, 5(2):119– 129.

Jahromi, O. S., Francis, B. A., and Kwong, R. H. (2004b). Spectrum estimation using multirate observations. *IEEE Transactions on Signal Processing*, 52(7):1878–1890.

Jayant, N. S. and Noll, P. (1984). *Digital Coding of Waveforms*. Prentice-Hall, Upper Saddle River, NJ.

Jaynes, E. T. (1982). On the rationale of maximum entropy methods. *Proceedings of the IEEE*, 70:939–952.

Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*. Reidel, Dordrecht. (Editted by R. D. Rosenkrantz).

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffreys, H. (1967). *Theory of Probability*. Oxford University Press, London, 3rd edition.

Jones, F. (1993). *Lebesgue Integration on Euclidean Space*. Jones and Bartlett Publishers, Boston, MA.

Kailath, T., Sayed, A. H., and Hassibi, B. (2000). *Linear Estimation*. Prentice Hall, Upper Saddle River, NJ.

Kay, S. M. (1988). *Modern Spectrum Estimation: Theory and Applications*. Prentice Hall, Upper Saddle River, NJ.

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall PTR, Upper Saddle River, NJ.

Kay, S. M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall PTR, Upper Saddle River, NJ.

Khintchine, A. (1934). Korrelationstheorie der stasionären stochastischen prozesse. *Math. Ann.*, 109(604). (in German).

Kirac, A. and Vaidyanathan, P. P. (1998a). On existence of principal component filter banks. In *Proceedings of Intl. Conf. Acoustic, Speech and Signal Process. (ICASSP)*, pages 1329–32, Seattle, WA.

Kirac, A. and Vaidyanathan, P. P. (1998b). Theory and design of optimum FIR compaction filters. *IEEE Transactions on Signal Processing*, 46(4):903–919.

Klaus, M. and Smith, R. T. (1988). A Hilbert space approach to maximum entropy requlatization. *Mathematical Methods in Applied Sciences*, 10:397–406.

Knapp, C. H. and Carter, G. (1976). The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-24:320–327.

Kullback, S. (1954). *Information Theory and Statistics*. John Wiley and Sons, New York, NY.

Kullback, S. and Liebler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22:79–86.

Lam, J. (1993). Model reduction of delay systems using Padé approximation. *International Journal of Control*, 57(2):377–391.

Leonev, A. S. (2000). A generalization of the maximal entropy method for solving ill-posed problems. *Siberian Mathematical Journal*, 41(4):716–724.

Li, M. and Kok, C. W. (2003). Linear phase filter bank design using LMI-based $\mathcal{H}_\infty$ optimization. *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, 50(3):143–150.

Li, M. and Vitányi, P. (1997). *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, New York, 2nd edition.

Ljung, L. (1999). *System Identification*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 2nd edition.

Mallat, S. and Falzon, F. (1998). Analysis of low bit rate image transform coding. *IEEE Transactions on Signal Processing*, 46(4):1027–42.

Mandel, L. and Wolf, E. (1995). *Optical Coherence and Quantum Optics*. Cambridge University Press, New York, NY.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press.

Mac Lane, S. and Birkhoff, C. (1967). *Algebra*. MacMillan, New York.

Mertinz, A. (1999). *Signal Analysis, Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. John Wiley, Chichester.

Meyr, H., Moeneclaey, M., and Fechtel, S. A. (1998). *Digital Communication Receivers, Synchronization, Channel estimation, and Signal Processing*. Wiley Interscience, New York, NY.

Middleton, D. (1960). *An Introduction to Statistical Communication Theory*. International Series in Pure and Applied Physics. McGraw-Hill, New York. Reprinted by IEEE Press, 1996.

Moulin, P., Anitescu, M., and Ramachandran, K. (2000). Theory of rate-distortion-optimal constrained filter banks-application to IIR and FIR biorthogonal dessigns. *IEEE Transactions on Signal Processing*, 48(4):1120–1132.

Moulin, P. and Mihçak, M. K. (1998). Theory and design of signal-adapted FIR paraunitary filter banks. *IEEE Transactions on Signal Processing*, 46(4):920–929.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing*. Prentice Hall, New Jersey.

Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, 3rd edition.

Paris, J. B. and Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Inexact Reasoning*, 4:183–223.

Parks, T. W. and McClellan, J. H. (1972a). Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Transactions on Circuit Theory*, 19:189–194.

Parks, T. W. and McClellan, J. H. (1972b). A program for the design of linear phase finite impulse response digital filters. *IEEE Transactions on Audio and Electroacoustics*, 20:195–199.

Percival, D. B. and Walden, A. T. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

Philipp, L. D., Mahmood, A., and Philipp, B. L. (1999). An improved refinable rational approximation to the ideal time delay. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 46(5):637–640.

Pinsker, M. S. (1964). *Information and information stability of random variables and processes*. Holden-Day, San Francisco.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. University of Toronto Press, Toronto.

Regalia, P. A. and Huang, D. Y. (1995). Attainable error bonds in multirate adaptive lossless FIR filters. In *Proceedings of Intl. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, pages 1460–63, Detroit, MI.

Reza, F. M. (1961). *An Introduction to Information Theory.* McGraw-Hill, New York.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry.* World Scientific, Singapore.

Rosenfeld, A. (1968). *An Introduction to Algebraic Structures.* Holden-Day, San Fransisco.

Sathe, V. and Vaidyanathan, P. P. (1993). Effect of multirate systems on the statistical properties of random signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 41(1):131–146.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656.

Shannon, C. E. (1958). A note on partial ordering of communication channels. *Information and Control*, 1:390–397.

Shenoy, R. G. (1994). Formulation of multirate filter design as an approximation problem. *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '94)*, 2:173–176.

Shenoy, R. G., Burnside, D., and Parks, T. W. (1994). Linear periodic systems and multirate filter design. *IEEE Transactions on Signal Processing*, 42(9):2242–2256.

Shore, J. E. and Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, 26(1):26–37.

Skilling, J. (1988). The axioms of maximum entropy. In Erickson, G. J. and Smith, C. R., editors, *Maximum entropy and Bayesian methods in science and engineering*, volume 1, pages 173–187, The Netherlands. Kluwer Academic Publishers.

Soman, A. K. and Vaidyanathan, P. P. (1993). Coding gain in paraunitary analysis/synthesis systems. *IEEE Transactions on Signal Processing*, 41(5):1824–35.

Taylor, A. E. and Mann, W. R. (1972). *Advanced Calculus.* Xerox College Publishing, Lexington, MA, 2nd edition.

Teboulle, M. and Vajda, I. (1993). Convergence of best ϕ-entropy estimates. *IEEE Transactions on Information Theory*, 39(1).

Tikhonov, A. N. (1963). On solving ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, 151(3):501–504. (in Russian), English translation in Soviet Math. Dokl.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems.* V. H. Winston & Sons, Washington, D.C.

Tikhonov, A. N., Leonov, A. S., and Yagola, A. G. (1998). *Nonlinear ill-posed problems.* Chapman and Hall, London. (2 volumes).

Tsatsanis, M. K. and Giannakis, G. B. (1995). Principal component filter banks for optimal multiresolution analysis. *IEEE Transactions on Signal Processing*, 43(8).

Uffink, J. (1995). Can the Maximum Entropy principle be explained as a consistency requirement? *Studies on the History and Philosophy of Modern Physics*, 26(3):223–261.

Uffink, J. (1996). The constraint rule of the Maximum Entropy principle. *Studies on the History and Philosophy of Modern Physics*, 27(1):47–79.

Unser, M. (1993). An extension of the Karhunen-Loève transform for wavelets and perfect-reconstruction filter banks. In *Proceedings of SPIE*, volume 2034, pages 45–56.

Vaidyanathan, P. P. (1993). *Multirate Sytems and Filter Banks*. Prentice-Hall, Upper Saddle River, NJ.

Vaidyanathan, P. P. (1998). Theory of optimal orthonormal subband coders. *IEEE Transsactions on Signal Processing*, 46(6):1528–1543.

Vaidyanathan, P. P. and Akkarakaran, S. (2001). A review of the theory and application of optimal subband and transform coders. *Applied and Computational Harmnic Analysis*, 10:254–289.

Vajda, I. (1989). *Theory of Statistical Inference and Information*. Kluwer Academic Publishers, Boston.

Van Campenhout, J. M. and Cover, T. (1981). Maximum entropy and conditional probability. *IEEE Transactions on Information Theory*, 27:483–489.

Vapnik, V. (1982). *Estimation of Dependencies based on empirical data*. Springer-Verlag, New York.

Vapnik, V. (1999). *Statistical Learning Theory*. Wiely, New York.

Vasin, V. V. and Ageev, A. L. (1995). *Ill-posed problems with a priori information*. VSP, Utrecht, The Netherlands.

von Mises, R. (1964). *Mathematical Theory of Probability and Statistics*. Academic Press, New York.

Watanabe, S. (1969). *Knowing and Guessing, A Quantitative Study of Inference and Information*. Wiley, New York.

Wiener, N. (1930). Generalized harmonic analysis. *Acta Math.*, 55(117).

Wu, N. (1997). *The Maximum Entropy Method*. Springer, Berlin.

Xuan, B. and Bamberger, R. H. (1998). FIR principal component filter banks. *IEEE Transactions on Signal Processing*, 46(4):930–938.

Yoon, M. G. and Lee, B. H. (1997). A new approximation method for time-delay systems. *IEEE Transactions on Automatic Control*, 42(7):1008–1012.