SPRINGER BRIEFS IN
ELECTRICAL AND COMPUTER ENGINEERING

Jacob Benesty · Jingdong Chen

Optimal Time-Domain Noise Reduction Filters

A Theoretical Study

Springer

SpringerBriefs in Electrical and Computer Engineering

For further volumes:
http://www.springer.com/series/10059

Jacob Benesty · Jingdong Chen

Optimal Time-Domain Noise Reduction Filters

A Theoretical Study

Springer

Prof. Dr. Jacob Benesty
INRS-EMT
University of Quebec
de la Gauchetiere Ouest 800
Montreal, H5A 1K6, QC
Canada
e-mail: benesty@emt.inrs.ca

Jingdong Chen
Northwestern Polytechnical University
127 Youyi West Road
Xi’an, Shanxi 710072
China
e-mail: jingdongchen@ieee.org

ISSN 2191-8112   e-ISSN 2191-8120

ISBN 978-3-642-19600-3   e-ISBN 978-3-642-19601-0

DOI 10.1007/978-3-642-19601-0

Springer Heidelberg Dordrecht London New York

Cover design: eStudio Calamar, Berlin/Figueres

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Introduction

1.1 Noise Reduction

Signal enhancement is a fundamental topic of signal processing in general and of speech processing in particular [1]. In audio and speech applications such as cell phones, teleconferencing systems, hearing aids, human–machine interfaces, and many others, the microphones installed in these systems always pick up some interferences that contaminate the desired speech signal. Depending on the mechanism that generates them, these interferences can be broadly classified into four basic categories: additive noise originating from various ambient sound sources, interference from concurrent competing speakers, filtering effects caused by room surface reflections and spectral shaping of recording devices, and echo from coupling between loudspeakers and microphones. These four categories of distortions interfere with the measurement, processing, recording, and communication of the desired speech signal in very distinct ways and combating them has led to four important research areas: noise reduction (also called speech enhancement), source separation, speech dereverberation, and echo cancellation and suppression. A broad coverage of these research areas can be found in [2, 3]. This work is devoted to the theoretical study of the problem of speech enhancement in the time domain.

Noise reduction consists of recovering a speech signal of interest from the microphone signals, which are corrupted by unwanted additive noise. By additive noise we mean that the signals picked up by the microphones are a superposition of the convolved clean speech and noise. Schroeder at Bell Laboratories in 1960 was the first to propose a single-channel algorithm for that purpose [4]. It was basically a spectral subtraction method implemented with analog circuits.

Frequency-domain approaches are usually preferred in real-time applications as they can be implemented efficiently thanks to the fast Fourier transform. However, they come with some well-known problems such as the so-called "musical noise," which is very unpleasant to hear and difficult to get rid off. In the time domain, this problem does not exist and, contrary to what some readers might believe, time-domain algorithms are at least as flexible as their counterparts in the frequency domain as it

J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters*,
SpringerBriefs in Electrical and Computer Engineering, 1,
DOI: 10.1007/978-3-642-19601-0_1,

will be shown throughout this work; but they can be computationally more complex in terms of multiplications. However, with little effort, it is not hard to make them more efficient by exploiting the Toeplitz or close-to-Toeplitz structure of the matrices involved in these algorithms.

In this work, we propose a general framework for the time-domain noise reduction problem. Thanks to this formulation, it is easy to derive, study, and analyze all kinds of algorithms.

1.2 Organization of the Work

The material in this work is organized into five chapters, including this one. The focus is on the time-domain algorithms for both the single and multiple microphone cases. The work discussed in these chapters is as follows.

In Chap. 2, we study the noise reduction problem with a single microphone by using a filtering vector for the estimation of the desired signal sample.

Chapter 3 generalizes the ideas of Chap. 2 with a rectangular filtering matrix for the estimation of the desired signal vector.

In Chap. 4, we study the speech enhancement problem with a microphone array by using a long filtering vector.

Finally, Chap. 5 extends the results of Chap. 4 with a rectangular filtering matrix.

References

1. J. Benesty, J. Chen, Y. Huang, I. Cohen, *Noise Reduction in Speech Processing* (Springer, Berlin, 2009)
2. J. Benesty, J. Chen, Y. Huang, *Microphone Array Signal Processing* (Springer, Berlin, 2008)
3. Y. Huang, J. Benesty, J. Chen, *Acoustic MIMO Signal Processing* (Springer, Berlin, 2006)
4. M.R. Schroeder, Apparatus for suppressing noise and distortion in communication signals, U.S. Patent No. 3,180,936, filed 1 Dec 1960, issued 27 Apr 1965

Chapter 2
Single-Channel Noise Reduction with a Filtering Vector

There are different ways to perform noise reduction in the time domain. The simplest way, perhaps, is to estimate a sample of the desired signal at a time by applying a filtering vector to the observation signal vector. This approach is investigated in this chapter and many well-known optimal filtering vectors are derived. We start by explaining the single-channel signal model for noise reduction in the time domain.

2.1 Signal Model

The noise reduction problem considered in this chapter and Chap. 3 is one of recovering the desired signal (or clean speech) $x(k)$, k being the discrete-time index, of zero mean from the noisy observation (microphone signal) [1–3]

$$y(k) = x(k) + v(k), \tag{2.1}$$

where $v(k)$, assumed to be a zero-mean random process, is the unwanted additive noise that can be either white or colored but is uncorrelated with $x(k)$. All signals are considered to be real and broadband. To simplify the derivation of the optimal filters, we further assume that the signals are Gaussian and stationary.

The signal model given in (2.1) can be put into a vector form by considering the L most recent successive samples, i.e.,

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k), \tag{2.2}$$

where

$$\mathbf{y}(k) = \left[y(k) \ y(k-1) \ \cdots \ y(k-L+1) \right]^T \tag{2.3}$$

is a vector of length L, superscript ^{T} denotes transpose of a vector or a matrix, and $\mathbf{x}(k)$ and $\mathbf{v}(k)$ are defined in a similar way to $\mathbf{y}(k)$. Since $x(k)$ and $v(k)$ are

J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters*, SpringerBriefs in Electrical and Computer Engineering, 1,
DOI: 10.1007/978-3-642-19601-0_2,

uncorrelated by assumption, the correlation matrix (of size $L \times L$) of the noisy signal can be written as

$$\begin{aligned}\mathbf{R}_{\mathbf{y}} &= E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right] \\ &= \mathbf{R}_{\mathbf{x}} + \mathbf{R}_{\mathbf{v}},\end{aligned} \tag{2.4}$$

where $E[\cdot]$ denotes mathematical expectation, and $\mathbf{R}_{\mathbf{x}} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ and $\mathbf{R}_{\mathbf{v}} = E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively. The objective of noise reduction in this chapter is then to find a "good" estimate of the sample $x(k)$ in the sense that the additive noise is significantly reduced while the desired signal is not much distorted.

Since $x(k)$ is the signal of interest, it is important to write the vector $\mathbf{y}(k)$ as an explicit function of $x(k)$. For that, we need first to decompose $\mathbf{x}(k)$ into two orthogonal components: one proportional to the desired signal, $x(k)$, and the other one corresponding to the interference. Indeed, it is easy to see that this decomposition is

$$\mathbf{x}(k) = \boldsymbol{\rho}_{\mathbf{x}x} \cdot x(k) + \mathbf{x}_{\mathrm{i}}(k), \tag{2.5}$$

where

$$\begin{aligned}\boldsymbol{\rho}_{\mathbf{x}x} &= [1 \ \rho_x(1) \ \cdots \ \rho_x(L-1)]^T \\ &= \frac{E\left[\mathbf{x}(k)x(k)\right]}{E\left[x^2(k)\right]}\end{aligned} \tag{2.6}$$

is the normalized [with respect to $x(k)$] correlation vector (of length L) between $\mathbf{x}(k)$ and $x(k)$,

$$\rho_x(l) = \frac{E\left[x(k-l)x(k)\right]}{E\left[x^2(k)\right]}, \quad l = 0, 1, \ldots, L-1 \tag{2.7}$$

is the correlation coefficient between $x(k-l)$ and $x(k)$,

$$\mathbf{x}_{\mathrm{i}}(k) = \mathbf{x}(k) - \boldsymbol{\rho}_{\mathbf{x}x} \cdot x(k) \tag{2.8}$$

is the interference signal vector, and

$$E\left[\mathbf{x}_{\mathrm{i}}(k)x(k)\right] = \mathbf{0}_{L\times 1}, \tag{2.9}$$

where $\mathbf{0}_{L\times 1}$ is a vector of length L containing only zeroes.

Substituting (2.5) into (2.2), the signal model for noise reduction can be expressed as

$$\mathbf{y}(k) = \boldsymbol{\rho}_{\mathbf{x}x} \cdot x(k) + \mathbf{x}_{\mathrm{i}}(k) + \mathbf{v}(k). \tag{2.10}$$

This formulation will be extensively used in the following sections.

2.2 Linear Filtering with a Vector

In this chapter, we try to estimate the desired signal sample, $x(k)$, by applying a finite-impulse-response (FIR) filter to the observation signal vector $\mathbf{y}(k)$, i.e.,

$$\begin{aligned} z(k) &= \sum_{l=0}^{L-1} h_l y(k-l) \\ &= \mathbf{h}^T \mathbf{y}(k), \end{aligned} \tag{2.11}$$

where $z(k)$ is supposed to be the estimate of $x(k)$ and

$$\mathbf{h} = \begin{bmatrix} h_0 \; h_1 \; \cdots \; h_{L-1} \end{bmatrix}^T \tag{2.12}$$

is an FIR filter of length L. This procedure is called the single-channel noise reduction in the time domain with a filtering vector.

Using (2.10), we can express (2.11) as

$$\begin{aligned} z(k) &= \mathbf{h}^T \left[\boldsymbol{\rho}_{\mathbf{x}x} \cdot x(k) + \mathbf{x}_{\mathrm{i}}(k) + \mathbf{v}(k) \right] \\ &= x_{\mathrm{fd}}(k) + x_{\mathrm{ri}}(k) + v_{\mathrm{rn}}(k), \end{aligned} \tag{2.13}$$

where

$$x_{\mathrm{fd}}(k) = x(k)\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} \tag{2.14}$$

is the filtered desired signal,

$$x_{\mathrm{ri}}(k) = \mathbf{h}^T \mathbf{x}_{\mathrm{i}}(k) \tag{2.15}$$

is the residual interference, and

$$v_{\mathrm{rn}}(k) = \mathbf{h}^T \mathbf{v}(k) \tag{2.16}$$

is the residual noise.

Since the estimate of the desired signal at time k is the sum of three terms that are mutually uncorrelated, the variance of $z(k)$ is

$$\begin{aligned} \sigma_z^2 &= \mathbf{h}^T \mathbf{R}_{\mathbf{y}} \mathbf{h} \\ &= \sigma_{x_{\mathrm{fd}}}^2 + \sigma_{x_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2, \end{aligned} \tag{2.17}$$

where

$$\begin{aligned} \sigma_{x_{\mathrm{fd}}}^2 &= \sigma_x^2 \left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} \right)^2 \\ &= \mathbf{h}^T \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} \mathbf{h}, \end{aligned} \tag{2.18}$$

$$\begin{aligned}\sigma_{x_{\mathrm{ri}}}^2 &= \mathbf{h}^T \mathbf{R}_{\mathbf{x}_{\mathrm{i}}} \mathbf{h} \\ &= \mathbf{h}^T \mathbf{R}_{\mathbf{x}} \mathbf{h} - \mathbf{h}^T \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} \mathbf{h},\end{aligned} \tag{2.19}$$

$$\sigma_{v_{\mathrm{rn}}}^2 = \mathbf{h}^T \mathbf{R}_{\mathbf{v}} \mathbf{h}, \tag{2.20}$$

$\sigma_x^2 = E\left[x^2(k)\right]$ is the variance of the desired signal, $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}} = \sigma_x^2 \boldsymbol{\rho}_{\mathbf{x}x} \boldsymbol{\rho}_{\mathbf{x}x}^T$ is the correlation matrix (whose rank is equal to 1) of $\mathbf{x}_{\mathrm{d}}(k) = \boldsymbol{\rho}_{\mathbf{x}x} \cdot x(k)$, and $\mathbf{R}_{\mathbf{x}_{\mathrm{i}}} = E\left[\mathbf{x}_{\mathrm{i}}(k)\mathbf{x}_{\mathrm{i}}^T(k)\right]$ is the correlation matrix of $\mathbf{x}_{\mathrm{i}}(k)$. The variance of $z(k)$is useful in the definitions of the performance measures.

2.3 Performance Measures

The first attempts to derive relevant and rigorous measures in the context of speech enhancement can be found in [1, 4, 5]. These references are the main inspiration for the derivation of measures in the studied context throughout this work.

In this section, we are going to define the most useful performance measures for speech enhancement in the single-channel case with a filtering vector. We can divide these measures into two categories. The first category evaluates the noise reduction performance while the second one evaluates speech distortion. We are also going to discuss the very convenient mean-square error (MSE) criterion and show how it is related to the performance measures.

2.3.1 Noise Reduction

One of the most fundamental measures in all aspects of speech enhancement is the signal-to-noise ratio (SNR). The input SNR is a second-order measure which quantifies the level of noise present relative to the level of the desired signal. It is defined as

$$\mathrm{iSNR} = \frac{\sigma_x^2}{\sigma_v^2}, \tag{2.21}$$

where $\sigma_v^2 = E\left[v^2(k)\right]$ is the variance of the noise.

The output SNR[1] helps quantify the level of noise remaining at the filter output signal. The output SNR is obtained from (2.17):

[1] In this work, we consider the uncorrelated interference as part of the noise in the definitions of the performance measures.

$$\text{oSNR}(\mathbf{h}) = \frac{\sigma_{x_{\text{fd}}}^2}{\sigma_{x_{\text{ri}}}^2 + \sigma_{v_{\text{rn}}}^2} = \frac{\sigma_x^2 \left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x}\right)^2}{\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}}, \tag{2.22}$$

where

$$\mathbf{R}_{\text{in}} = \mathbf{R}_{\mathbf{x}_{\text{i}}} + \mathbf{R}_{\mathbf{v}} \tag{2.23}$$

is the interference-plus-noise correlation matrix. Basically, (2.22) is the variance of the first signal (filtered desired) from the right-hand side of (2.17) over the variance of the two other signals (filtered interference-plus-noise). The objective of the speech enhancement filter is to make the output SNR greater than the input SNR. Consequently, the quality of the noisy signal will be enhanced.

For the particular filtering vector

$$\mathbf{h} = \mathbf{i}_{\text{i}} = [1 \; 0 \; \cdots \; 0]^T \tag{2.24}$$

of length L, we have

$$\text{oSNR}(\mathbf{i}_{\text{i}}) = \text{iSNR}. \tag{2.25}$$

With the identity filtering vector $\mathbf{i}_{\text{i}}$, the SNR cannot be improved.

For any two vectors $\mathbf{h}$ and $\boldsymbol{\rho}_{\mathbf{x}x}$ and a positive definite matrix $\mathbf{R}_{\text{in}}$, we have

$$\left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x}\right)^2 \leq \left(\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}\right)\left(\boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}\right), \tag{2.26}$$

with equality if and only if $\mathbf{h} = \varsigma \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}$, where $\varsigma (\neq 0)$ is a real number. Using the previous inequality in (2.22), we deduce an upper bound for the output SNR:

$$\text{oSNR}(\mathbf{h}) \leq \sigma_x^2 \cdot \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}, \quad \forall \mathbf{h} \tag{2.27}$$

and clearly

$$\text{oSNR}(\mathbf{i}_{\text{i}}) \leq \sigma_x^2 \cdot \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}, \tag{2.28}$$

which implies that

$$\sigma_v^2 \cdot \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x} \geq 1. \tag{2.29}$$

The maximum output SNR is then

$$\text{oSNR}_{\max} = \sigma_x^2 \cdot \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x} \tag{2.30}$$

and

$$\text{oSNR}_{\max} \geq \text{iSNR}. \tag{2.31}$$

The noise reduction factor quantifies the amount of noise being rejected by the filter. This quantity is defined as the ratio of the power of the noise at the microphone over the power of the interference-plus-noise remaining at the filter output, i.e.,

$$\xi_{\text{nr}}(\mathbf{h}) = \frac{\sigma_v^2}{\mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h}}. \tag{2.32}$$

The noise reduction factor is expected to be lower bounded by 1; otherwise, the filter amplifies the noise received at the microphone. The higher the value of the noise reduction factor, the more the noise is rejected. While the output SNR is upper bounded, the noise reduction factor is not.

2.3.2 Speech Distortion

Since the noise is reduced by the filtering operation, so is, in general, the desired speech. This speech reduction (or cancellation) implies, in general, speech distortion. The speech reduction factor, which is somewhat similar to the noise reduction factor, is defined as the ratio of the variance of the desired signal at the microphone over the variance of the filtered desired signal, i.e.,

$$\begin{aligned} \xi_{\text{sr}}(\mathbf{h}) &= \frac{\sigma_x^2}{\sigma_{x_{\text{fd}}}^2} \\ &= \frac{1}{\left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x}\right)^2}. \end{aligned} \tag{2.33}$$

A key observation is that the design of filters that do not cancel the desired signal requires the constraint

$$\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} = 1. \tag{2.34}$$

Thus, the speech reduction factor is equal to 1 if there is no distortion and expected to be greater than 1 when distortion happens.

Another way to measure the distortion of the desired speech signal due to the filtering operation is the speech distortion index,[2] which is defined as the mean-square error between the desired signal and the filtered desired signal, normalized by the variance of the desired signal, i.e.,

---

[2] Very often in the literature, authors use $1/\upsilon_{\text{sd}}(\mathbf{h})$ as a measure of the SNR improvement. This is wrong! Obviously, we can define whatever we want, but in this is case we need to be careful to compare "apples with apples." For example, it is not appropriate to compare $1/\upsilon_{\text{sd}}(\mathbf{h})$ to iSNR and only $\text{oSNR}(\mathbf{h})$ makes sense to compare to iSNR.

$$\begin{aligned}\upsilon_{\text{sd}}(\mathbf{h}) &= \frac{E\left\{[x_{\text{fd}}(k) - x(k)]^2\right\}}{E\left[x^2(k)\right]} \\ &= \left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} - 1\right)^2 \\ &= \left[\xi_{\text{sr}}^{-1/2}(\mathbf{h}) - 1\right]^2. \end{aligned} \tag{2.35}$$

We also see from this measure that the design of filters that do not distort the desired signal requires the constraint

$$\upsilon_{\text{sd}}(\mathbf{h}) = 0. \tag{2.36}$$

Therefore, the speech distortion index is equal to 0 if there is no distortion and expected to be greater than 0 when distortion occurs.

It is easy to verify that we have the following fundamental relation:

$$\frac{\text{oSNR}(\mathbf{h})}{\text{iSNR}} = \frac{\xi_{\text{nr}}(\mathbf{h})}{\xi_{\text{sr}}(\mathbf{h})}. \tag{2.37}$$

This expression indicates the equivalence between gain/loss in SNR and distortion.

2.3.3 Mean-Square Error (MSE) Criterion

Error criteria play a critical role in deriving optimal filters. The mean-square error (MSE) [6] is, by far, the most practical one.

We define the error signal between the estimated and desired signals as

$$\begin{aligned} e(k) &= z(k) - x(k) \\ &= x_{\text{fd}}(k) + x_{\text{ri}}(k) + v_{\text{rn}}(k) - x(k), \end{aligned} \tag{2.38}$$

which can be written as the sum of two uncorrelated error signals:

$$e(k) = e_{\text{d}}(k) + e_{\text{r}}(k), \tag{2.39}$$

where

$$\begin{aligned} e_{\text{d}}(k) &= x_{\text{fd}}(k) - x(k) \\ &= \left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} - 1\right) x(k) \end{aligned} \tag{2.40}$$

is the signal distortion due to the filtering vector and

$$\begin{aligned} e_{\text{r}}(k) &= x_{\text{ri}}(k) + v_{\text{rn}}(k) \\ &= \mathbf{h}^T \mathbf{x}_{\text{i}}(k) + \mathbf{h}^T \mathbf{v}(k) \end{aligned} \tag{2.41}$$

represents the residual interference-plus-noise.

The mean-square error (MSE) criterion is then

$$\begin{aligned} J(\mathbf{h}) &= E\left[e^2(k)\right] \\ &= \sigma_x^2 + \mathbf{h}^T \mathbf{R_y} \mathbf{h} - 2\mathbf{h}^T E\left[\mathbf{x}(k)x(k)\right] \\ &= \sigma_x^2 + \mathbf{h}^T \mathbf{R_y} \mathbf{h} - 2\sigma_x^2 \mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} \\ &= J_{\mathrm{d}}(\mathbf{h}) + J_{\mathrm{r}}(\mathbf{h}), \end{aligned} \tag{2.42}$$

where

$$\begin{aligned} J_{\mathrm{d}}(\mathbf{h}) &= E\left[e_{\mathrm{d}}^2(k)\right] \\ &= \sigma_x^2 \left(\mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} - 1\right)^2 \end{aligned} \tag{2.43}$$

and

$$\begin{aligned} J_{\mathrm{r}}(\mathbf{h}) &= E\left[e_{\mathrm{r}}^2(k)\right] \\ &= \mathbf{h}^T \mathbf{R}_{\mathrm{in}} \mathbf{h}. \end{aligned} \tag{2.44}$$

Two particular filtering vectors are of great interest: $\mathbf{h} = \mathbf{i}_{\mathrm{i}}$ and $\mathbf{h} = \mathbf{0}_{L\times 1}$. With the first one (identity filtering vector), we have neither noise reduction nor speech distortion and with the second one (zero filtering vector), we have maximum noise reduction and maximum speech distortion (i.e., the desired speech signal is completely nulled out). For both filters, however, it can be verified that the output SNR is equal to the input SNR. For these two particular filters, the MSEs are

$$J(\mathbf{i}_{\mathrm{i}}) = J_{\mathrm{r}}(\mathbf{i}_{\mathrm{i}}) = \sigma_v^2, \tag{2.45}$$

$$J(\mathbf{0}_{L\times 1}) = J_{\mathrm{d}}(\mathbf{0}_{L\times 1}) = \sigma_x^2. \tag{2.46}$$

As a result,

$$\mathrm{iSNR} = \frac{J(\mathbf{0}_{L\times 1})}{J(\mathbf{i}_{\mathrm{i}})}. \tag{2.47}$$

We define the normalized MSE (NMSE) with respect to $J(\mathbf{i}_{\mathrm{i}})$ as

$$\begin{aligned} \widetilde{J}(\mathbf{h}) &= \frac{J(\mathbf{h})}{J(\mathbf{i}_{\mathrm{i}})} \\ &= \mathrm{iSNR} \cdot \upsilon_{\mathrm{sd}}(\mathbf{h}) + \frac{1}{\xi_{\mathrm{nr}}(\mathbf{h})} \\ &= \mathrm{iSNR}\left[\upsilon_{\mathrm{sd}}(\mathbf{h}) + \frac{1}{\mathrm{oSNR}(\mathbf{h}) \cdot \xi_{\mathrm{sr}}(\mathbf{h})}\right], \end{aligned} \tag{2.48}$$

where

$$\upsilon_{\text{sd}}(\mathbf{h}) = \frac{J_{\text{d}}(\mathbf{h})}{J_{\text{d}}(\mathbf{0}_{L\times 1})}, \tag{2.49}$$

$$\text{iSNR} \cdot \upsilon_{\text{sd}}(\mathbf{h}) = \frac{J_{\text{d}}(\mathbf{h})}{J_{\text{r}}(\mathbf{i}_{\text{i}})}, \tag{2.50}$$

$$\xi_{\text{nr}}(\mathbf{h}) = \frac{J_{\text{r}}(\mathbf{i}_{\text{i}})}{J_{\text{r}}(\mathbf{h})}, \tag{2.51}$$

$$\text{oSNR}(\mathbf{h}) \cdot \xi_{\text{sr}}(\mathbf{h}) = \frac{J_{\text{d}}(\mathbf{0}_{L\times 1})}{J_{\text{r}}(\mathbf{h})}. \tag{2.52}$$

This shows how this NMSE and the different MSEs are related to the performance measures.

We define the NMSE with respect to $J(\mathbf{0}_{L\times 1})$ as

$$\begin{aligned} \overline{J}(\mathbf{h}) &= \frac{J(\mathbf{h})}{J(\mathbf{0}_{L\times 1})} \\ &= \upsilon_{\text{sd}}(\mathbf{h}) + \frac{1}{\text{oSNR}(\mathbf{h}) \cdot \xi_{\text{sr}}(\mathbf{h})} \end{aligned} \tag{2.53}$$

and, obviously,

$$\widetilde{J}(\mathbf{h}) = \text{iSNR} \cdot \overline{J}(\mathbf{h}). \tag{2.54}$$

We are only interested in filters for which

$$J_{\text{d}}(\mathbf{i}_{\text{i}}) \leq J_{\text{d}}(\mathbf{h}) < J_{\text{d}}(\mathbf{0}_{L\times 1}), \tag{2.55}$$

$$J_{\text{r}}(\mathbf{0}_{L\times 1}) < J_{\text{r}}(\mathbf{h}) < J_{\text{r}}(\mathbf{i}_{\text{i}}). \tag{2.56}$$

From the two previous expressions, we deduce that

$$0 \leq \upsilon_{\text{sd}}(\mathbf{h}) < 1, \tag{2.57}$$

$$1 < \xi_{\text{nr}}(\mathbf{h}) < \infty. \tag{2.58}$$

It is clear that the objective of noise reduction is to find optimal filtering vectors that would either minimize $J(\mathbf{h})$ or minimize $J_{\text{d}}(\mathbf{h})$ or $J_{\text{r}}(\mathbf{h})$ subject to some constraint.

2.4 Optimal Filtering Vectors

In this section, we are going to derive the most important filtering vectors that can help mitigate the level of the noise picked up by the microphone signal.

2.4.1 Maximum Signal-to-Noise Ratio (SNR)

The maximum SNR filter, $\mathbf{h}_{\max}$, is obtained by maximizing the output SNR as given in (2.22) from which, we recognize the generalized Rayleigh quotient [7]. It is well known that this quotient is maximized with the maximum eigenvector of the matrix $\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\mathbf{x}_{\text{d}}}$. Let us denote by $\lambda_{\max}$ the maximum eigenvalue corresponding to this maximum eigenvector. Since the rank of the mentioned matrix is equal to 1, we have

$$\begin{aligned}\lambda_{\max} &= \text{tr}\left(\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\mathbf{x}_{\text{d}}}\right) \\ &= \sigma_x^2 \cdot \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x},\end{aligned} \tag{2.59}$$

where tr $(\cdot)$ denotes the trace of a square matrix. As a result,

$$\begin{aligned}\text{oSNR}\left(\mathbf{h}_{\max}\right) &= \lambda_{\max} \\ &= \sigma_x^2 \cdot \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x},\end{aligned} \tag{2.60}$$

which corresponds to the maximum possible output SNR, i.e., $\text{oSNR}_{\max}$.

Obviously, we also have

$$\mathbf{h}_{\max} = \varsigma \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}, \tag{2.61}$$

where ς is an arbitrary non-zero scaling factor. While this factor has no effect on the output SNR, it may have on the speech distortion. In fact, all filters (except for the LCMV) derived in the rest of this section are equivalent up to this scaling factor. These filters also try to find the respective scaling factors depending on what we optimize.

2.4.2 Wiener

The Wiener filter is easily derived by taking the gradient of the MSE, $J\left(\mathbf{h}\right)$ [Eq. (2.42)], with respect to $\mathbf{h}$ and equating the result to zero:

$$\mathbf{h}_{\text{W}} = \sigma_x^2 \mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}.$$

The Wiener filter can also be expressed as

$$\begin{aligned}\mathbf{h}_{\text{W}} &= \mathbf{R}_{\mathbf{y}}^{-1} E\left[\mathbf{x}(k)x(k)\right] \\ &= \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}\mathbf{i}_{\text{i}} \\ &= \left(\mathbf{I}_L - \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{v}}\right)\mathbf{i}_{\text{i}},\end{aligned} \tag{2.62}$$

where $\mathbf{I}_L$ is the identity matrix of size $L \times L$. The above formulation depends on the second-order statistics of the observation and noise signals. The correlation matrix

$\mathbf{R_y}$ can be estimated from the observation signal while the other correlation matrix, $\mathbf{R_v}$, can be estimated during noise-only intervals assuming that the statistics of the noise do not change much with time.

We now propose to write the general form of the Wiener filter in another way that will make it easier to compare to other optimal filters. We can verify that

$$\mathbf{R_y} = \sigma_x^2 \boldsymbol{\rho}_{\mathbf{x}x} \boldsymbol{\rho}_{\mathbf{x}x}^T + \mathbf{R}_{\text{in}}. \tag{2.63}$$

Determining the inverse of $\mathbf{R_y}$ from the previous expression with the Woodbury's identity, we get

$$\mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{R}_{\text{in}}^{-1} - \frac{\mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x} \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1}}{\sigma_x^{-2} + \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}}. \tag{2.64}$$

Substituting (2.64) into (2.62), leads to another interesting formulation of the Wiener filter:

$$\mathbf{h}_{\text{W}} = \frac{\sigma_x^2 \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}}{1 + \sigma_x^2 \boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x}}, \tag{2.65}$$

that we can rewrite as

$$\begin{aligned} \mathbf{h}_{\text{W}} &= \frac{\sigma_x^2 \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x} \boldsymbol{\rho}_{\mathbf{x}x}^T}{1 + \lambda_{\max}} \mathbf{i}_{\text{i}} \\ &= \frac{\mathbf{R}_{\text{in}}^{-1} \left(\mathbf{R_y} - \mathbf{R}_{\text{in}} \right)}{1 + \text{tr} \left[\mathbf{R}_{\text{in}}^{-1} \left(\mathbf{R_y} - \mathbf{R}_{\text{in}} \right) \right]} \mathbf{i}_{\text{i}} \\ &= \frac{\mathbf{R}_{\text{in}}^{-1} \mathbf{R_y} - \mathbf{I}_L}{1 - L + \text{tr}\left(\mathbf{R}_{\text{in}}^{-1} \mathbf{R_y} \right)} \mathbf{i}_{\text{i}}. \end{aligned} \tag{2.66}$$

From (2.66), we deduce that the output SNR is

$$\begin{aligned} \text{oSNR} \left(\mathbf{h}_{\text{W}} \right) &= \lambda_{\max} \\ &= \text{tr}\left(\mathbf{R}_{\text{in}}^{-1} \mathbf{R_y} \right) - L. \end{aligned} \tag{2.67}$$

We observe from (2.67) that the more the amount of noise, the smaller is the output SNR.

The speech distortion index is an explicit function of the output SNR:

$$\upsilon_{\text{sd}} \left(\mathbf{h}_{\text{W}} \right) = \frac{1}{\left[1 + \text{oSNR} \left(\mathbf{h}_{\text{W}} \right) \right]^2} \leq 1. \tag{2.68}$$

The higher the value of $\text{oSNR} \left(\mathbf{h}_{\text{W}} \right)$, the less the desired signal is distorted.

Clearly,

$$\text{oSNR}\left(\mathbf{h}_{\text{W}}\right) \geq \text{iSNR}, \tag{2.69}$$

since the Wiener filter maximizes the output SNR.

It is of interest to observe that the two filters $\mathbf{h}_{\max}$ and $\mathbf{h}_{\text{W}}$ are equivalent up to a scaling factor. Indeed, taking

$$\varsigma = \frac{\sigma_x^2}{1+\lambda_{\max}} \tag{2.70}$$

in (2.61) (maximum SNR filter), we find (2.66) (Wiener filter).

With the Wiener filter, the noise and speech reduction factors are

$$\begin{aligned} \xi_{\text{nr}}\left(\mathbf{h}_{\text{W}}\right) &= \frac{\left(1+\lambda_{\max}\right)^2}{\text{iSNR}\cdot\lambda_{\max}} \\ &\geq \left(1+\frac{1}{\lambda_{\max}}\right)^2, \end{aligned} \tag{2.71}$$

$$\xi_{\text{sr}}\left(\mathbf{h}_{\text{W}}\right) = \left(1+\frac{1}{\lambda_{\max}}\right)^2. \tag{2.72}$$

Finally, we give the minimum NMSEs (MNMSEs):

$$\widetilde{J}\left(\mathbf{h}_{\text{W}}\right) = \frac{\text{iSNR}}{1+\text{oSNR}\left(\mathbf{h}_{\text{W}}\right)} \leq 1, \tag{2.73}$$

$$\overline{J}\left(\mathbf{h}_{\text{W}}\right) = \frac{1}{1+\text{oSNR}\left(\mathbf{h}_{\text{W}}\right)} \leq 1. \tag{2.74}$$

2.4.3 Minimum Variance Distortionless Response (MVDR)

The celebrated minimum variance distortionless response (MVDR) filter proposed by Capon [8, 9] is usually derived in a context where we have at least two sensors (or microphones) available. Interestingly, with the linear model proposed in this chapter, we can also derive the MVDR (with one sensor only) by minimizing the MSE of the residual interference-plus-noise, $J_{\text{r}}\left(\mathbf{h}\right)$, with the constraint that the desired signal is not distorted. Mathematically, this is equivalent to

$$\min_{\mathbf{h}} \mathbf{h}^T \mathbf{R}_{\text{in}} \mathbf{h} \quad \text{subject to} \quad \mathbf{h}^T \boldsymbol{\rho}_{\mathbf{x}x} = 1, \tag{2.75}$$

for which the solution is

$$\mathbf{h}_{\text{MVDR}} = \frac{\mathbf{R}_{\text{in}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\boldsymbol{\rho}_{\mathbf{x}x}^T\mathbf{R}_{\text{in}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}, \tag{2.76}$$

that we can rewrite as

$$\begin{aligned}\mathbf{h}_{\text{MVDR}} &= \frac{\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\mathbf{y}} - \mathbf{I}_L}{\text{tr}\left(\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\mathbf{y}}\right) - L}\mathbf{i}_{\text{i}} \\ &= \frac{\sigma_x^2\mathbf{R}_{\text{in}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\lambda_{\max}}.\end{aligned} \tag{2.77}$$

Alternatively, we can express the MVDR as

$$\mathbf{h}_{\text{MVDR}} = \frac{\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\boldsymbol{\rho}_{\mathbf{x}x}^T\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}. \tag{2.78}$$

The Wiener and MVDR filters are simply related as follows:

$$\mathbf{h}_{\text{W}} = \varsigma_0\mathbf{h}_{\text{MVDR}}, \tag{2.79}$$

where

$$\begin{aligned}\varsigma_0 &= \mathbf{h}_{\text{W}}^T\boldsymbol{\rho}_{\mathbf{x}x} \\ &= \frac{\lambda_{\max}}{1+\lambda_{\max}}.\end{aligned} \tag{2.80}$$

So, the two filters $\mathbf{h}_{\text{W}}$ and $\mathbf{h}_{\text{MVDR}}$ are equivalent up to a scaling factor. From a theoretical point of view, this scaling is not significant. But from a practical point of view it can be important. Indeed, the signals are usually nonstationary and the estimations are done frame by frame, so it is essential to have this scaling factor right from one frame to another in order to avoid large distortions. Therefore, it is recommended to use the MVDR filter rather than the Wiener filter in speech enhancement applications.

It is clear that we always have

$$\text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right) = \text{oSNR}\left(\mathbf{h}_{\text{W}}\right), \tag{2.81}$$

$$\upsilon_{\text{sd}}\left(\mathbf{h}_{\text{MVDR}}\right) = 0, \tag{2.82}$$

$$\xi_{\text{sr}}\left(\mathbf{h}_{\text{MVDR}}\right) = 1, \tag{2.83}$$

$$\xi_{\text{nr}}\left(\mathbf{h}_{\text{MVDR}}\right) = \frac{\text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right)}{\text{iSNR}} \leq \xi_{\text{nr}}\left(\mathbf{h}_{\text{W}}\right), \tag{2.84}$$

and

$$1 \geq \widetilde{J}(\mathbf{h}_{\mathrm{MVDR}}) = \frac{\mathrm{iSNR}}{\mathrm{oSNR}(\mathbf{h}_{\mathrm{MVDR}})} \geq \widetilde{J}(\mathbf{h}_{\mathrm{W}}), \tag{2.85}$$

$$\overline{J}(\mathbf{h}_{\mathrm{MVDR}}) = \frac{1}{\mathrm{oSNR}(\mathbf{h}_{\mathrm{MVDR}})} \geq \overline{J}(\mathbf{h}_{\mathrm{W}}). \tag{2.86}$$

2.4.4 Prediction

Assume that we can find a simple prediction filter $\mathbf{g}$ of length L in such a way that

$$\mathbf{x}(k) \approx x(k)\mathbf{g}. \tag{2.87}$$

In this case, we can derive a distortionless filter for noise reduction as follows:

$$\min_{\mathbf{h}} \mathbf{h}^T \mathbf{R}_{\mathbf{y}} \mathbf{h} \quad \text{subject to} \quad \mathbf{h}^T \mathbf{g} = 1. \tag{2.88}$$

We deduce the solution

$$\mathbf{h}_{\mathrm{P}} = \frac{\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{g}}{\mathbf{g}^T \mathbf{R}_{\mathbf{y}}^{-1}\mathbf{g}}. \tag{2.89}$$

Now, we can find the optimal $\mathbf{g}$ in the Wiener sense. For that, we need to define the error signal vector

$$\mathbf{e}_{\mathrm{P}}(k) = \mathbf{x}(k) - x(k)\mathbf{g} \tag{2.90}$$

and form the MSE

$$J(\mathbf{g}) = E\left[\mathbf{e}_{\mathrm{P}}^T(k)\mathbf{e}_{\mathrm{P}}(k)\right]. \tag{2.91}$$

By minimizing $J(\mathbf{g})$ with respect to $\mathbf{g}$, we easily find the optimal filter

$$\mathbf{g}_{\mathrm{o}} = \boldsymbol{\rho}_{\mathbf{x}x}. \tag{2.92}$$

It is interesting to observe that the error signal vector with the optimal filter, $\mathbf{g}_{\mathrm{o}}$, corresponds to the interference signal, i.e.,

$$\begin{aligned} \mathbf{e}_{\mathrm{P,o}}(k) &= \mathbf{x}(k) - x(k)\boldsymbol{\rho}_{\mathbf{x}x} \\ &= \mathbf{x}_{\mathrm{i}}(k). \end{aligned} \tag{2.93}$$

This result is obviously expected because of the orthogonality principle.

Substituting (2.92) into (2.89), we find that

$$\mathbf{h}_{\mathrm{P}} = \frac{\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\boldsymbol{\rho}_{\mathbf{x}x}^{T}\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}. \tag{2.94}$$

Clearly, the two filters $\mathbf{h}_{\mathrm{MVDR}}$ and $\mathbf{h}_{\mathrm{P}}$ are identical. Therefore, the prediction approach can be seen as another way to derive the MVDR. This approach is also an intuitive manner to justify the decomposition given in (2.5).

Left multiplying both sides of (2.93) by $\mathbf{h}_{\mathrm{P}}^{T}$ results in

$$x(k) = \mathbf{h}_{\mathrm{P}}^{T}\mathbf{x}(k) - \mathbf{h}_{\mathrm{P}}^{T}\mathbf{e}_{\mathrm{P,o}}(k). \tag{2.95}$$

Therefore, the filter $\mathbf{h}_{\mathrm{P}}$ can also be interpreted as a temporal prediction filter that is less noisy than the one that can be obtained from the noisy signal, $y(k)$, directly.

2.4.5 Tradeoff

In the tradeoff approach, we try to compromise between noise reduction and speech distortion. Instead of minimizing the MSE to find the Wiener filter or minimizing the filter output with a distortionless constraint to find the MVDR as we already did in the preceding subsections, we could minimize the speech distortion index with the constraint that the noise reduction factor is equal to a positive value that is greater than 1. Mathematically, this is equivalent to

$$\min_{\mathbf{h}} J_{\mathrm{d}}(\mathbf{h}) \quad \text{subject to} \quad J_{\mathrm{r}}(\mathbf{h}) = \beta\sigma_v^2, \tag{2.96}$$

where $0 < \beta < 1$ to insure that we get some noise reduction. By using a Lagrange multiplier, $\mu > 0$, to adjoin the constraint to the cost function and assuming that the matrix $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}} + \mu\mathbf{R}_{\mathrm{in}}$ is invertible, we easily deduce the tradeoff filter

$$\begin{aligned}
\mathbf{h}_{\mathrm{T},\mu} &= \sigma_x^2\left[\mathbf{R}_{\mathbf{x}_{\mathrm{d}}} + \mu\mathbf{R}_{\mathrm{in}}\right]^{-1}\boldsymbol{\rho}_{\mathbf{x}x} \\
&= \frac{\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}}{\mu\sigma_x^{-2} + \boldsymbol{\rho}_{\mathbf{x}x}^{T}\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\mathbf{x}x}} \\
&= \frac{\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\mathbf{y}} - \mathbf{I}_L}{\mu - L + \mathrm{tr}\left(\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\mathbf{y}}\right)}\mathbf{i}_{\mathrm{i}},
\end{aligned} \tag{2.97}$$

where the Lagrange multiplier, μ, satisfies

$$J_{\mathrm{r}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \beta\sigma_v^2. \tag{2.98}$$

However, in practice it is not easy to determine the optimal μ. Therefore, when this parameter is chosen in an ad hoc way, we can see that for

- $\mu = 1$, $\mathbf{h}_{\mathrm{T},1} = \mathbf{h}_{\mathrm{W}}$, which is the Wiener filter;
- $\mu = 0$, $\mathbf{h}_{\mathrm{T},0} = \mathbf{h}_{\mathrm{MVDR}}$, which is the MVDR filter;
- $\mu > 1$, results in a filter with low residual noise (compared with the Wiener filter) at the expense of high speech distortion;
- $\mu < 1$, results in a filter with high residual noise and low speech distortion.

Note that the MVDR cannot be derived from the first line of (2.97) since by taking $\mu = 0$, we have to invert a matrix that is not full rank.

Again, we observe here as well that the tradeoff, Wiener, and maximum SNR filters are equivalent up to a scaling factor. As a result, the output SNR of the tradeoff filter is independent of μ and is identical to the output SNR of the Wiener filter, i.e.,

$$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right), \quad \forall \mu \geq 0. \tag{2.99}$$

We have

$$\upsilon_{\mathrm{sd}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \left(\frac{\mu}{\mu + \lambda_{\max}}\right)^2, \tag{2.100}$$

$$\xi_{\mathrm{sr}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \left(1 + \frac{\mu}{\lambda_{\max}}\right)^2, \tag{2.101}$$

$$\xi_{\mathrm{nr}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \frac{(\mu + \lambda_{\max})^2}{\mathrm{iSNR} \cdot \lambda_{\max}}, \tag{2.102}$$

and

$$\widetilde{J}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \mathrm{iSNR}\frac{\mu^2 + \lambda_{\max}}{(\mu + \lambda_{\max})^2} \geq \widetilde{J}\left(\mathbf{h}_{\mathrm{W}}\right), \tag{2.103}$$

$$\overline{J}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \frac{\mu^2 + \lambda_{\max}}{(\mu + \lambda_{\max})^2} \geq \overline{J}\left(\mathbf{h}_{\mathrm{W}}\right). \tag{2.104}$$

2.4.6 Linearly Constrained Minimum Variance (LCMV)

We can derive a linearly constrained minimum variance (LCMV) filter [10, 11], which can handle more than one linear constraint, by exploiting the structure of the noise signal.

In Sect. 2.1, we decomposed the vector $\mathbf{x}(k)$ into two orthogonal components to extract the desired signal, $x(k)$. We can also decompose (but for a different objective as explained below) the noise signal vector, $\mathbf{v}(k)$, into two orthogonal vectors:

$$\mathbf{v}(k) = \boldsymbol{\rho}_{\mathbf{v}v} \cdot v(k) + \mathbf{v}_{\mathrm{u}}(k), \tag{2.105}$$

where $\boldsymbol{\rho}_{\mathbf{v}v}$ is defined in a similar way to $\boldsymbol{\rho}_{\mathbf{x}x}$ and $\mathbf{v}_{\mathrm{u}}(k)$ is the noise signal vector that is uncorrelated with $v(k)$.

Our problem this time is the following. We wish to perfectly recover our desired signal, $x(k)$, and completely remove the correlated components of the noise signal, $\boldsymbol{\rho}_{\mathbf{v}v} \cdot v(k)$. Thus, the two constraints can be put together in a matrix form as

$$\mathbf{C}_{xv}^T \mathbf{h} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{2.106}$$

where

$$\mathbf{C}_{xv} = \begin{bmatrix} \boldsymbol{\rho}_{\mathbf{x}x} & \boldsymbol{\rho}_{\mathbf{v}v} \end{bmatrix} \tag{2.107}$$

is our constraint matrix of size $L \times 2$. Then, our optimal filter is obtained by minimizing the energy at the filter output, with the constraints that the correlated noise components are cancelled and the desired speech is preserved, i.e.,

$$\mathbf{h}_{\mathrm{LCMV}} = \arg\min_{\mathbf{h}} \mathbf{h}^T \mathbf{R}_{\mathbf{y}} \mathbf{h} \quad \text{subject to} \quad \mathbf{C}_{xv}^T \mathbf{h} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.108}$$

The solution to (2.108) is given by

$$\mathbf{h}_{\mathrm{LCMV}} = \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{C}_{xv} \left(\mathbf{C}_{xv}^T \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{C}_{xv} \right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.109}$$

By developing (2.109), it can easily be shown that the LCMV can be written as a function of the MVDR:

$$\mathbf{h}_{\mathrm{LCMV}} = \frac{1}{1-\gamma^2} \mathbf{h}_{\mathrm{MVDR}} - \frac{\gamma^2}{1-\gamma^2} \mathbf{t}, \tag{2.110}$$

where

$$\gamma^2 = \frac{\left(\boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{v}v} \right)^2}{\left(\boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{x}x} \right) \left(\boldsymbol{\rho}_{\mathbf{v}v}^T \mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{v}v} \right)}, \tag{2.111}$$

with $0 \le \gamma^2 \le 1$, $\mathbf{h}_{\mathrm{MVDR}}$ is defined in (2.78), and

$$\mathbf{t} = \frac{\mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{v}v}}{\boldsymbol{\rho}_{\mathbf{x}x}^T \mathbf{R}_{\mathbf{y}}^{-1} \boldsymbol{\rho}_{\mathbf{v}v}}. \tag{2.112}$$

We observe from (2.110) that when $\gamma^2 = 0$, the LCMV filter becomes the MVDR filter; however, when γ^2 tends to 1, which happens if and only if $\boldsymbol{\rho}_{\mathbf{x}x} = \boldsymbol{\rho}_{\mathbf{v}v}$, we have no solution since we have conflicting requirements.

Obviously, we always have

$$\text{oSNR}\left(\mathbf{h}_{\text{LCMV}}\right) \leq \text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right), \tag{2.113}$$

$$\upsilon_{\text{sd}}\left(\mathbf{h}_{\text{LCMV}}\right) = 0, \tag{2.114}$$

$$\xi_{\text{sr}}\left(\mathbf{h}_{\text{LCMV}}\right) = 1, \tag{2.115}$$

and

$$\xi_{\text{nr}}\left(\mathbf{h}_{\text{LCMV}}\right) \leq \xi_{\text{nr}}\left(\mathbf{h}_{\text{MVDR}}\right) \leq \xi_{\text{nr}}\left(\mathbf{h}_{\text{W}}\right). \tag{2.116}$$

The LCMV filter is able to remove all the correlated noise; however, its overall noise reduction is lower than that of the MVDR filter.

2.4.7 Practical Considerations

All the algorithms presented in the preceding subsections can be implemented from the second-order statistics estimates of the noise and noisy signals. Let us take the MVDR as an example. In this filter, we need the estimates of $\mathbf{R}_{\mathbf{y}}$ and $\boldsymbol{\rho}_{\mathbf{x}x}$. The correlation matrix, $\mathbf{R}_{\mathbf{y}}$, can be easily estimated from the observations. However, the correlation vector, $\boldsymbol{\rho}_{\mathbf{x}x}$, cannot be estimated directly since $x(k)$ is not accessible but it can be rewritten as

$$\begin{aligned}\boldsymbol{\rho}_{\mathbf{x}x} &= \frac{E\left[\mathbf{y}(k)y(k)\right] - E\left[\mathbf{v}(k)v(k)\right]}{\sigma_y^2 - \sigma_v^2} \\ &= \frac{\sigma_y^2\boldsymbol{\rho}_{\mathbf{y}y} - \sigma_v^2\boldsymbol{\rho}_{\mathbf{v}v}}{\sigma_y^2 - \sigma_v^2},\end{aligned} \tag{2.117}$$

which now depends on the statistics of $y(k)$ and $v(k)$. However, a voice activity detector (VAD) is required in order to be able to estimate the statistics of the noise signal during silences [i.e., when $x(k) = 0$]. Nowadays, more and more sophisticated VADs are developed [12] since a VAD is an integral part of most speech enhancement algorithms. A good VAD will obviously improve the performance of a noise reduction filter since the estimates of the signals statistics will be more reliable. A system integrating an optimal filter and a VAD may not be easy to design but much progress has been made recently in this area of research [13].

2.5 Summary

In this chapter, we revisited the single-channel noise reduction problem in the time domain. We showed how to extract the desired signal sample from a vector containing its past samples. Thanks to the orthogonal decomposition that results from this, the presentation of the problem is simplified. We defined several interesting performance measures in this context and deduced optimal noise reduction filters: maximum SNR, Wiener, MVDR, prediction, tradeoff, and LCMV. Interestingly, all these filters (except for the LCMV) are equivalent up to a scaling factor. Consequently, their performance in terms of SNR improvement is the same given the same statistics estimates.

References

1. J. Benesty, J. Chen, Y. Huang, I. Cohen, *Noise Reduction in Speech Processing* (Springer, Berlin, 2009)
2. P. Vary, R. Martin, *Digital Speech Transmission*: *Enhancement*, *Coding and Error Concealment* (Wiley, Chichester, 2006)
3. P. Loizou, *Speech Enhancement*: *Theory and Practice* (CRC Press, Boca Raton, 2007)
4. J. Benesty, J. Chen, Y. Huang, S. Doclo, Study of the Wiener filter for noise reduction, in *Speech Enhancement*, Chap. 2, ed. by J. Benesty, S. Makino, J. Chen (Springer, Berlin, 2005)
5. J. Chen, J. Benesty, Y. Huang, S. Doclo, New insights into the noise reduction Wiener filter. IEEE Trans. Audio Speech Language Process. **14**, 1218–1234 (2006)
6. S. Haykin, *Adaptive Filter Theory*, 4th edn. (Prentice-Hall, Upper Saddle River, 2002)
7. J.N. Franklin, *Matrix Theory* (Prentice-Hall, Englewood Cliffs, 1968)
8. J. Capon, High resolution frequency-wavenumber spectrum analysis. Proc. IEEE **57**, 1408–1418 (1969)
9. R.T. Lacoss, Data adaptive spectral analysis methods. Geophysics **36**, 661–675 (1971)
10. O. Frost, An algorithm for linearly constrained adaptive array processing. Proc. IEEE **60**, 926–935 (1972)
11. M. Er, A. Cantoni, Derivative constraints for broad-band element space antenna array processors. IEEE Trans. Acoust. Speech Signal Process. **31**, 1378–1393 (1983)
12. I. Cohen, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans. Speech Audio Process. **11**, 466–475 (2003)
13. I. Cohen, J. Benesty, S. Gannot (eds.), *Speech Processing in Modern Communication—Challenges and Perspectives* (Springer, Berlin, 2010)

Chapter 3
Single-Channel Noise Reduction with a Rectangular Filtering Matrix

In the previous chapter, we tried to estimate one sample only at a time from the observation signal vector. In this part, we are going to estimate more than one sample at a time. As a result, we now deal with a rectangular filtering matrix instead of a filtering vector. If M is the number of samples to be estimated and L is the length of the observation signal vector, then the size of the filtering matrix is $M \times L$. Also, this approach is more general and all the results from Chap. 2 are particular cases of the results derived in this chapter by just setting $M = 1$. The signal model is the same as in Sect. 2.1; so we start by explaining the principle of linear filtering with a rectangular matrix.

3.1 Linear Filtering with a Rectangular Matrix

Define the vector of length M:

$$\mathbf{x}^M(k) = [x(k)\ x(k-1) \cdots x(k-M+1)]^T, \tag{3.1}$$

where $M \le L$. In the general linear filtering approach, we estimate the desired signal vector, $\mathbf{x}^M(k)$, by applying a linear transformation to $\mathbf{y}(k)$ [1–4], i.e.,

$$\begin{aligned} \mathbf{z}^M(k) &= \mathbf{H}\mathbf{y}(k) \\ &= \mathbf{H}[\mathbf{x}(k) + \mathbf{v}(k)] \\ &= \mathbf{x}_{\mathrm{f}}^M(k) + \mathbf{v}_{\mathrm{rn}}^M(k), \end{aligned} \tag{3.2}$$

where $\mathbf{z}^M(k)$ is the estimate of $\mathbf{x}^M(k)$,

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_M^T \end{bmatrix} \tag{3.3}$$

J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters*,
SpringerBriefs in Electrical and Computer Engineering, 1,
DOI: 10.1007/978-3-642-19601-0_3, © Jacob Benesty 2011

is a rectangular filtering matrix of size $M \times L$,

$$\mathbf{h}_m = \left[h_{m,0}\ h_{m,1} \cdots h_{m,L-1}\right]^T, \quad m = 1, 2, \ldots, M \tag{3.4}$$

are FIR filters of length L,

$$\mathbf{x}_{\mathrm{f}}^M(k) = \mathbf{H}\mathbf{x}(k) \tag{3.5}$$

is the filtered speech, and

$$\mathbf{v}_{\mathrm{rn}}^M(k) = \mathbf{H}\mathbf{v}(k) \tag{3.6}$$

is the residual noise.

Two important particular cases of (3.2) are immediate.

- $M = 1$. In this situation, $\mathbf{z}^1(k) = z(k)$ is a scalar and $\mathbf{H}$ simplifies to an FIR filter $\mathbf{h}^T$ of length L. This case was well studied in Chap. 2.
- $M = L$. In this situation, $\mathbf{z}^L(k) = \mathbf{z}(k)$ is a vector of length L and $\mathbf{H} = \mathbf{H}_{\mathrm{S}}$ is a square matrix of size $L \times L$. This scenario has been widely covered in [1–5] and in many other papers. We will get back to this case a bit later in this chapter.

By definition, our desired signal is the vector $\mathbf{x}^M(k)$. The filtered speech, $\mathbf{x}_{\mathrm{f}}^M(k)$, depends on $\mathbf{x}(k)$ but our desired signal after noise reduction should explicitly depends on $\mathbf{x}^M(k)$. Therefore, we need to extract $\mathbf{x}^M(k)$ from $\mathbf{x}(k)$. For that, we need to decompose $\mathbf{x}(k)$ into two orthogonal components: one that is correlated with (or is a linear transformation of) the desired signal $\mathbf{x}^M(k)$ and the other one that is orthogonal to $\mathbf{x}^M(k)$ and, hence, will be considered as the interference signal. Specifically, the vector $\mathbf{x}(k)$ is decomposed into the following form:

$$\begin{aligned}\mathbf{x}(k) &= \mathbf{R}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}^{-1}\mathbf{x}^M(k) + \mathbf{x}_{\mathrm{i}}(k) \\ &= \mathbf{x}_{\mathrm{d}}(k) + \mathbf{x}_{\mathrm{i}}(k),\end{aligned} \tag{3.7}$$

where

$$\begin{aligned}\mathbf{x}_{\mathrm{d}}(k) &= \mathbf{R}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}^{-1}\mathbf{x}^M(k) \\ &= \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{x}^M(k)\end{aligned} \tag{3.8}$$

is a linear transformation of the desired signal, $\mathbf{R}_{\mathbf{x}^M} = E\left[\mathbf{x}^M(k)\mathbf{x}^{MT}(k)\right]$ is the correlation matrix (of size $M \times M$) of $\mathbf{x}^M(k)$, $\mathbf{R}_{\mathbf{x}\mathbf{x}^M} = E\left[\mathbf{x}(k)\mathbf{x}^{MT}(k)\right]$ is the cross-correlation matrix (of size $L \times M$) between $\mathbf{x}(k)$ and $\mathbf{x}^M(k)$, $\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M} = \mathbf{R}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}^{-1}$, and

$$\mathbf{x}_{\mathrm{i}}(k) = \mathbf{x}(k) - \mathbf{x}_{\mathrm{d}}(k) \tag{3.9}$$

is the interference signal. It is easy to see that $\mathbf{x}_\mathrm{d}(k)$ and $\mathbf{x}_\mathrm{i}(k)$ are orthogonal, i.e.,

$$E\left[\mathbf{x}_\mathrm{d}(k)\mathbf{x}_\mathrm{i}^T(k)\right] = \mathbf{0}_{L\times L}. \tag{3.10}$$

For the particular case $M = L$, we have $\mathbf{\Gamma}_{\mathbf{xx}} = \mathbf{I}_L$, which is the identity matrix (of size $L \times L$), and $\mathbf{x}_\mathrm{d}(k)$ coincides with $\mathbf{x}(k)$, which obviously makes sense. For $M = 1$, $\mathbf{\Gamma}_{\mathbf{xx}^1}$ simplifies to the normalized correlation vector (see Chap. 2)

$$\boldsymbol{\rho}_{\mathbf{x}x} = \frac{E\left[\mathbf{x}(k)x(k)\right]}{E\left[x^2(k)\right]}. \tag{3.11}$$

Substituting (3.7) into (3.2), we get

$$\begin{aligned}\mathbf{z}^M(k) &= \mathbf{H}\left[\mathbf{x}_\mathrm{d}(k) + \mathbf{x}_\mathrm{i}(k) + \mathbf{v}(k)\right] \\ &= \mathbf{x}_\mathrm{fd}^M(k) + \mathbf{x}_\mathrm{ri}^M(k) + \mathbf{v}_\mathrm{rn}^M(k),\end{aligned} \tag{3.12}$$

where

$$\mathbf{x}_\mathrm{fd}^M(k) = \mathbf{H}\mathbf{x}_\mathrm{d}(k) \tag{3.13}$$

is the filtered desired signal,

$$\mathbf{x}_\mathrm{ri}^M(k) = \mathbf{H}\mathbf{x}_\mathrm{i}(k) \tag{3.14}$$

is the residual interference, and $\mathbf{v}_\mathrm{rn}^M(k) = \mathbf{H}\mathbf{v}(k)$, again, represents the residual noise. It can be checked that the three terms $\mathbf{x}_\mathrm{fd}^M(k)$, $\mathbf{x}_\mathrm{ri}^M(k)$, and $\mathbf{v}_\mathrm{rn}^M(k)$ are mutually orthogonal. Therefore, the correlation matrix of $\mathbf{z}^M(k)$ is

$$\begin{aligned}\mathbf{R}_{\mathbf{z}^M} &= E\left[\mathbf{z}^M(k)\mathbf{z}^{MT}(k)\right] \\ &= \mathbf{R}_{\mathbf{x}_\mathrm{fd}^M} + \mathbf{R}_{\mathbf{x}_\mathrm{ri}^M} + \mathbf{R}_{\mathbf{v}_\mathrm{rn}^M},\end{aligned} \tag{3.15}$$

where

$$\mathbf{R}_{\mathbf{x}_\mathrm{fd}^M} = \mathbf{H}\mathbf{R}_{\mathbf{x}_\mathrm{d}}\mathbf{H}^T, \tag{3.16}$$

$$\begin{aligned}\mathbf{R}_{\mathbf{x}_\mathrm{ri}^M} &= \mathbf{H}\mathbf{R}_{\mathbf{x}_\mathrm{i}}\mathbf{H}^T \\ &= \mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T - \mathbf{H}\mathbf{R}_{\mathbf{x}_\mathrm{d}}\mathbf{H}^T,\end{aligned} \tag{3.17}$$

$$\mathbf{R}_{\mathbf{v}_\mathrm{rn}^M} = \mathbf{H}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T, \tag{3.18}$$

$\mathbf{R}_{\mathbf{x}_\mathrm{d}} = \mathbf{\Gamma}_{\mathbf{xx}^M}\mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{xx}^M}^T$ is the correlation matrix (whose rank is equal to M) of $\mathbf{x}_\mathrm{d}(k)$, and $\mathbf{R}_{\mathbf{x}_\mathrm{i}} = E\left[\mathbf{x}_\mathrm{i}(k)\mathbf{x}_\mathrm{i}^T(k)\right]$ is the correlation matrix of $\mathbf{x}_\mathrm{i}(k)$. The correlation matrix of $\mathbf{z}^M(k)$ is helpful in defining meaningful performance measures.

3.2 Joint Diagonalization

By exploiting the decomposition of $\mathbf{x}(k)$, we can decompose the correlation matrix of $\mathbf{y}(k)$ as

$$\begin{aligned}\mathbf{R}_{\mathbf{y}} &= \mathbf{R}_{\mathbf{x}_\mathrm{d}} + \mathbf{R}_{\mathrm{in}} \\ &= \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M} \mathbf{R}_{\mathbf{x}^M} \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T + \mathbf{R}_{\mathrm{in}},\end{aligned} \tag{3.19}$$

where

$$\mathbf{R}_{\mathrm{in}} = \mathbf{R}_{\mathbf{x}_\mathrm{i}} + \mathbf{R}_{\mathbf{v}} \tag{3.20}$$

is the interference-plus-noise correlation matrix. It is interesting to observe from (3.19) that the noisy signal correlation matrix is the sum of two other correlation matrices: the linear transformation of the desired signal correlation matrix of rank M and the interference-plus-noise correlation matrix of rank L.

The two symmetric matrices $\mathbf{R}_{\mathbf{x}_\mathrm{d}}$ and $\mathbf{R}_{\mathrm{in}}$ can be jointly diagonalized as follows [6, 7]:

$$\mathbf{B}^T \mathbf{R}_{\mathbf{x}_\mathrm{d}} \mathbf{B} = \mathbf{\Lambda}, \tag{3.21}$$

$$\mathbf{B}^T \mathbf{R}_{\mathrm{in}} \mathbf{B} = \mathbf{I}_L, \tag{3.22}$$

where $\mathbf{B}$ is a full-rank square matrix (of size $L \times L$) and $\mathbf{\Lambda}$ is a diagonal matrix whose main elements are real and nonnegative. Furthermore, $\mathbf{\Lambda}$ and $\mathbf{B}$ are the eigenvalue and eigenvector matrices, respectively, of $\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\mathbf{x}_\mathrm{d}}$, i.e.,

$$\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\mathbf{x}_\mathrm{d}} \mathbf{B} = \mathbf{B}\mathbf{\Lambda}. \tag{3.23}$$

Since the rank of the matrix $\mathbf{R}_{\mathbf{x}_\mathrm{d}}$ is equal to M, the eigenvalues of $\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\mathbf{x}_\mathrm{d}}$ can be ordered as $\lambda_1^M \geq \lambda_2^M \geq \cdots \geq \lambda_M^M > \lambda_{M+1}^M = \cdots = \lambda_L^M = 0$. In other words, the last $L - M$ eigenvalues of $\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\mathbf{x}_\mathrm{d}}$ are exactly zero while its first M eigenvalues are positive, with λ_1^M being the maximum eigenvalue. We also denote by $\mathbf{b}_1^M, \mathbf{b}_2^M, \ldots, \mathbf{b}_M^M, \mathbf{b}_{M+1}^M, \ldots, \mathbf{b}_L^M$, the corresponding eigenvectors. Therefore, the noisy signal covariance matrix can also be diagonalized as

$$\mathbf{B}^T \mathbf{R}_{\mathbf{y}} \mathbf{B} = \mathbf{\Lambda} + \mathbf{I}_L. \tag{3.24}$$

Note that the same diagonalization was proposed in [8] but for the classical subspace approach [2].

Now, we have all the necessary ingredients to define the performance measures and derive the most well-known optimal filtering matrices.

3.3 Performance Measures

In this section, the performance measures tailored for linear filtering with a rectangular matrix are defined.

3.3.1 Noise Reduction

The input SNR was already defined in Chap. 2; but it can be rewritten as

$$\begin{aligned} \text{iSNR} &= \frac{\sigma_x^2}{\sigma_v^2} \\ &= \frac{\text{tr}\left(\mathbf{R}_{\mathbf{x}}\right)}{\text{tr}\left(\mathbf{R}_{\mathbf{v}}\right)}. \end{aligned} \tag{3.25}$$

Taking the trace of the filtered desired signal correlation matrix from the right-hand side of (3.15) over the trace of the two other correlation matrices gives the output SNR:

$$\begin{aligned} \text{oSNR}\left(\mathbf{H}\right) &= \frac{\text{tr}\left(\mathbf{R}_{\mathbf{x}_{\text{fd}}^M}\right)}{\text{tr}\left(\mathbf{R}_{\mathbf{x}_{\text{ri}}^M} + \mathbf{R}_{\mathbf{v}_{\text{rn}}^M}\right)} \\ &= \frac{\text{tr}\left(\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{H}^T\right)}{\text{tr}\left(\mathbf{H}\mathbf{R}_{\text{in}}\mathbf{H}^T\right)}. \end{aligned} \tag{3.26}$$

The obvious objective is to find an appropriate $\mathbf{H}$ in such a way that $\text{oSNR}\left(\mathbf{H}\right) \geq \text{iSNR}$.

For the particular filtering matrix

$$\mathbf{H} = \mathbf{I}_{\text{i}} = \left[\mathbf{I}_M \mathbf{0}_{M\times(L-M)}\right], \tag{3.27}$$

called the identity filtering matrix, where $\mathbf{I}_M$ is the $M \times M$ identity matrix, we have

$$\text{oSNR}\left(\mathbf{I}_{\text{i}}\right) = \text{iSNR}. \tag{3.28}$$

With $\mathbf{I}_{\text{i}}$, the SNR cannot be improved.

The maximum output SNR cannot be derived from a simple inequality as it was done in the previous chapter in the particular case of $M = 1$. We will see how to find this value when we derive the maximum SNR filter.

The noise reduction factor is

$$\begin{aligned} \xi_{\text{nr}}\left(\mathbf{H}\right) &= M \cdot \frac{\sigma_v^2}{\text{tr}\left(\mathbf{R}_{\mathbf{x}_{\text{ri}}^M} + \mathbf{R}_{\mathbf{v}_{\text{rn}}^M}\right)} \\ &= M \cdot \frac{\sigma_v^2}{\text{tr}\left(\mathbf{H}\mathbf{R}_{\text{in}}\mathbf{H}^T\right)}. \end{aligned} \tag{3.29}$$

Any good choice of $\mathbf{H}$ should lead to $\xi_{\text{nr}}\left(\mathbf{H}\right) \geq 1$.

3.3.2 Speech Distortion

The desired speech signal can be distorted by the rectangular filtering matrix. Therefore, the speech reduction factor is defined as

$$\xi_{\rm sr}(\mathbf{H}) = M \cdot \frac{\sigma_x^2}{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_{\rm fd}^M}\right)} = M \cdot \frac{\sigma_x^2}{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{H}^T\right)}. \tag{3.30}$$

A rectangular filtering matrix that does not affect the desired signal requires the constraint

$$\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M} = \mathbf{I}_M. \tag{3.31}$$

Hence, $\xi_{\rm sr}(\mathbf{H}) = 1$ in the absence of distortion and $\xi_{\rm sr}(\mathbf{H}) > 1$ in the presence of distortion.

By making the appropriate substitutions, one can derive the relationship among the measures defined so far:

$$\frac{\mathrm{oSNR}(\mathbf{H})}{\mathrm{iSNR}} = \frac{\xi_{\rm nr}(\mathbf{H})}{\xi_{\rm sr}(\mathbf{H})}. \tag{3.32}$$

When no distortion occurs, the gain in SNR coincides with the noise reduction factor.

We can also quantify the distortion with the speech distortion index:

$$\upsilon_{\rm sd}(\mathbf{H}) = \frac{1}{M} \cdot \frac{E\left\{\left[\mathbf{x}_{\rm fd}^M(k) - \mathbf{x}^M(k)\right]^T\left[\mathbf{x}_{\rm fd}^M(k) - \mathbf{x}^M(k)\right]\right\}}{\sigma_x^2} = \frac{1}{M} \cdot \frac{\mathrm{tr}\left[\left(\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M} - \mathbf{I}_M\right)\mathbf{R}_{\mathbf{x}^M}\left(\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M} - \mathbf{I}_M\right)^T\right]}{\sigma_x^2}. \tag{3.33}$$

The speech distortion index is always greater than or equal to 0 and should be upper bounded by 1 for optimal filtering matrices; so the higher is the value of $\upsilon_{\rm sd}(\mathbf{H})$, the more the desired signal is distorted.

3.3.3 MSE Criterion

Since the desired signal is a vector of length M, so is the error signal. We define the error signal vector between the estimated and desired signals as

$$\mathbf{e}^M(k) = \mathbf{z}^M(k) - \mathbf{x}^M(k) = \mathbf{H}\mathbf{y}(k) - \mathbf{x}^M(k), \tag{3.34}$$

which can also be expressed as the sum of two orthogonal error signal vectors:

$$\mathbf{e}^M(k) = \mathbf{e}_\mathrm{d}^M(k) + \mathbf{e}_\mathrm{r}^M(k), \tag{3.35}$$

where

$$\begin{aligned}\mathbf{e}_\mathrm{d}^M(k) &= \mathbf{x}_\mathrm{fd}^M(k) - \mathbf{x}^M(k) \\ &= \left(\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M} - \mathbf{I}_M\right)\mathbf{x}^M(k)\end{aligned} \tag{3.36}$$

is the signal distortion due to the rectangular filtering matrix and

$$\begin{aligned}\mathbf{e}_\mathrm{r}^M(k) &= \mathbf{x}_\mathrm{ri}^M(k) + \mathbf{v}_\mathrm{rn}^M(k) \\ &= \mathbf{H}\mathbf{x}_\mathrm{i}(k) + \mathbf{H}\mathbf{v}(k)\end{aligned} \tag{3.37}$$

represents the residual interference-plus-noise.

Having defined the error signal, we can now write the MSE criterion:

$$\begin{aligned}J(\mathbf{H}) &= \frac{1}{M}\cdot \mathrm{tr}\left\{E\left[\mathbf{e}^M(k)\mathbf{e}^{MT}(k)\right]\right\} \\ &= \frac{1}{M}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}^M}\right) + \mathrm{tr}\left(\mathbf{H}\mathbf{R}_\mathbf{y}\mathbf{H}^T\right) - 2\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}\mathbf{x}^M}\right)\right] \\ &= \frac{1}{M}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}^M}\right) + \mathrm{tr}\left(\mathbf{H}\mathbf{R}_\mathbf{y}\mathbf{H}^T\right) - 2\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}\right)\right],\end{aligned} \tag{3.38}$$

where

$$\begin{aligned}\mathbf{R}_{\mathbf{y}\mathbf{x}^M} &= E\left[\mathbf{y}(k)\mathbf{x}^{MT}(k)\right] \\ &= \boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}\end{aligned}$$

is the cross-correlation matrix between $\mathbf{y}(k)$ and $\mathbf{x}^M(k)$.

Using the fact that $E\left[\mathbf{e}_\mathrm{d}^M(k)\mathbf{e}_\mathrm{r}^{MT}(k)\right] = \mathbf{0}_{M\times M}$, $J(\mathbf{H})$ can be expressed as the sum of two other MSEs, i.e.,

$$\begin{aligned}J(\mathbf{H}) &= \frac{1}{M}\cdot \mathrm{tr}\left\{E\left[\mathbf{e}_\mathrm{d}^M(k)\mathbf{e}_\mathrm{d}^{MT}(k)\right]\right\} + \frac{1}{M}\cdot \mathrm{tr}\left\{E\left[\mathbf{e}_\mathrm{r}^M(k)\mathbf{e}_\mathrm{r}^{MT}(k)\right]\right\} \\ &= J_\mathrm{d}(\mathbf{H}) + J_\mathrm{r}(\mathbf{H}).\end{aligned} \tag{3.39}$$

Two particular filtering matrices are of great importance: $\mathbf{H} = \mathbf{I}_\mathrm{i}$ and $\mathbf{H} = \mathbf{0}_{M\times L}$. With the first one (identity filtering matrix), we have neither noise reduction nor speech distortion and with the second one (zero filtering matrix), we have maximum noise reduction and maximum speech distortion (i.e., the desired speech signal is completely nulled out). For both filtering matrices, however, it can be verified that the output SNR is equal to the input SNR. For these two particular filtering matrices, the MSEs are

$$J\left(\mathbf{I}_{\mathrm{i}}\right) = J_{\mathrm{r}}\left(\mathbf{I}_{\mathrm{i}}\right) = \sigma_v^2, \tag{3.40}$$

$$J\left(\mathbf{0}_{M\times L}\right) = J_{\mathrm{d}}\left(\mathbf{0}_{M\times L}\right) = \sigma_x^2. \tag{3.41}$$

As a result,

$$\mathrm{iSNR} = \frac{J\left(\mathbf{0}_{M\times L}\right)}{J\left(\mathbf{I}_{\mathrm{i}}\right)}. \tag{3.42}$$

We define the NMSE with respect to $J\left(\mathbf{I}_{\mathrm{i}}\right)$ as

$$\begin{aligned} \widetilde{J}\left(\mathbf{H}\right) &= \frac{J\left(\mathbf{H}\right)}{J\left(\mathbf{I}_{\mathrm{i}}\right)} \\ &= \mathrm{iSNR}\cdot\upsilon_{\mathrm{sd}}\left(\mathbf{H}\right) + \frac{1}{\xi_{\mathrm{nr}}\left(\mathbf{H}\right)} \\ &= \mathrm{iSNR}\left[\upsilon_{\mathrm{sd}}\left(\mathbf{H}\right) + \frac{1}{\mathrm{oSNR}\left(\mathbf{H}\right)\cdot\xi_{\mathrm{sr}}\left(\mathbf{H}\right)}\right], \end{aligned} \tag{3.43}$$

where

$$\upsilon_{\mathrm{sd}}\left(\mathbf{H}\right) = \frac{J_{\mathrm{d}}\left(\mathbf{H}\right)}{J_{\mathrm{d}}\left(\mathbf{0}_{M\times L}\right)}, \tag{3.44}$$

$$\mathrm{iSNR}\cdot\upsilon_{\mathrm{sd}}\left(\mathbf{H}\right) = \frac{J_{\mathrm{d}}\left(\mathbf{H}\right)}{J_{\mathrm{r}}\left(\mathbf{I}_{\mathrm{i}}\right)}, \tag{3.45}$$

$$\xi_{\mathrm{nr}}\left(\mathbf{H}\right) = \frac{J_{\mathrm{r}}\left(\mathbf{I}_{\mathrm{i}}\right)}{J_{\mathrm{r}}\left(\mathbf{H}\right)}, \tag{3.46}$$

$$\mathrm{oSNR}\left(\mathbf{H}\right)\cdot\xi_{\mathrm{sr}}\left(\mathbf{H}\right) = \frac{J_{\mathrm{d}}\left(\mathbf{0}_{M\times L}\right)}{J_{\mathrm{r}}\left(\mathbf{H}\right)}. \tag{3.47}$$

This shows how this NMSE and the different MSEs are related to the performance measures.

We define the NMSE with respect to $J\left(\mathbf{0}_{M\times L}\right)$ as

$$\begin{aligned} \overline{J}\left(\mathbf{H}\right) &= \frac{J\left(\mathbf{H}\right)}{J\left(\mathbf{0}_{M\times L}\right)} \\ &= \upsilon_{\mathrm{sd}}\left(\mathbf{H}\right) + \frac{1}{\mathrm{oSNR}\left(\mathbf{H}\right)\cdot\xi_{\mathrm{sr}}\left(\mathbf{H}\right)} \end{aligned} \tag{3.48}$$

and, obviously,

$$\widetilde{J}\left(\mathbf{H}\right) = \mathrm{iSNR}\cdot\overline{J}\left(\mathbf{H}\right). \tag{3.49}$$

We are only interested in filtering matrices for which

$$J_{\mathrm{d}}\left(\mathbf{I}_{\mathrm{i}}\right) \leq J_{\mathrm{d}}(\mathbf{H}) < J_{\mathrm{d}}\left(\mathbf{0}_{M\times L}\right), \tag{3.50}$$

$$J_{\mathrm{r}}\left(\mathbf{0}_{M\times L}\right) < J_{\mathrm{r}}(\mathbf{H}) < J_{\mathrm{r}}\left(\mathbf{I}_{\mathrm{i}}\right). \tag{3.51}$$

From the two previous expressions, we deduce that

$$0 \leq \upsilon_{\mathrm{sd}}(\mathbf{H}) < 1, \tag{3.52}$$

$$1 < \xi_{\mathrm{nr}}(\mathbf{H}) < \infty. \tag{3.53}$$

The optimal filtering matrices are obtained by minimizing $J(\mathbf{H})$ or minimizing $J_{\mathrm{r}}(\mathbf{H})$ or $J_{\mathrm{d}}(\mathbf{H})$ subject to some constraint.

3.4 Optimal Rectangular Filtering Matrices

In this section, we are going to derive the most important filtering matrices that can help reduce the noise picked up by the microphone signal.

3.4.1 Maximum SNR

Our first optimal filtering matrix is not derived from the MSE criterion but from the output SNR defined in (3.26) that we can rewrite as

$$\mathrm{oSNR}(\mathbf{H}) = \frac{\sum_{m=1}^{M} \mathbf{h}_m^T \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} \mathbf{h}_m}{\sum_{m=1}^{M} \mathbf{h}_m^T \mathbf{R}_{\mathrm{in}} \mathbf{h}_m}. \tag{3.54}$$

It is then natural to try to maximize this SNR with respect to $\mathbf{H}$. Let us first give the following lemma.

Lemma 3.1 *We have*

$$\mathrm{oSNR}(\mathbf{H}) \leq \max_m \frac{\mathbf{h}_m^T \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} \mathbf{h}_m}{\mathbf{h}_m^T \mathbf{R}_{\mathrm{in}} \mathbf{h}_m} = \chi. \tag{3.55}$$

Proof Let us define the positive reals $a_m = \mathbf{h}_m^T \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} \mathbf{h}_m$ and $b_m = \mathbf{h}_m^T \mathbf{R}_{\mathrm{in}} \mathbf{h}_m$. We have

$$\frac{\sum_{m=1}^{M} a_m}{\sum_{m=1}^{M} b_m} = \sum_{m=1}^{M} \left(\frac{a_m}{b_m} \cdot \frac{b_m}{\sum_{i=1}^{M} b_i} \right). \tag{3.56}$$

Now, define the following two vectors:

$$\mathbf{u} = \left[\frac{a_1}{b_1} \frac{a_2}{b_2} \cdots \frac{a_M}{b_M} \right]^T, \tag{3.57}$$

$$\mathbf{u}' = \left[\frac{b_1}{\sum_{i=1}^M b_i} \frac{b_2}{\sum_{i=1}^M b_i} \cdots \frac{b_M}{\sum_{i=1}^M b_i} \right]^T. \tag{3.58}$$

Using the Holder's inequality, we see that

$$\begin{aligned} \frac{\sum_{m=1}^M a_m}{\sum_{m=1}^M b_m} &= \mathbf{u}^T \mathbf{u}' \\ &\leq \|\mathbf{u}\|_\infty \left\|\mathbf{u}'\right\|_1 = \max_m \frac{a_m}{b_m}, \end{aligned} \tag{3.59}$$

which ends the proof. □

Theorem 3.1 *The maximum SNR filtering matrix is given by*

$$\mathbf{H}_{\max} = \begin{bmatrix} \beta_1 \mathbf{b}_1^{MT} \\ \beta_2 \mathbf{b}_1^{MT} \\ \vdots \\ \beta_m \mathbf{b}_1^{MT} \end{bmatrix}, \tag{3.60}$$

where β_m, $m = 1, 2, \ldots, M$ *are real numbers with at least one of them different from* 0. *The corresponding output SNR is*

$$\text{oSNR}\left(\mathbf{H}_{\max}\right) = \lambda_1^M. \tag{3.61}$$

We recall that λ_1^M *is the maximum eigenvalue of the matrix* $\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\mathbf{x}_\text{d}}$ *and its corresponding eigenvector is* $\mathbf{b}_1^M$.

Proof From Lemma 3.1, we know that the output SNR is upper bounded by χ whose maximum value is clearly λ_1^M. On the other hand, it can be checked from (3.54) that $\text{oSNR}\left(\mathbf{H}_{\max}\right) = \lambda_1^M$. Since this output SNR is maximal, $\mathbf{H}_{\max}$ is indeed the maximum SNR filtering matrix. □

Property 3.1 *The output SNR with the maximum SNR filtering matrix is always greater than or equal to the input SNR, i.e.,* $\text{oSNR}\left(\mathbf{H}_{\max}\right) \geq \text{iSNR}$.

It is interesting to see that we have these bounds:

$$0 \leq \text{oSNR}\left(\mathbf{H}\right) \leq \lambda_1^M, \forall \mathbf{H}, \tag{3.62}$$

but, obviously, we are only interested in filtering matrices that can improve the output SNR, i.e., $\text{oSNR}\left(\mathbf{H}\right) \geq \text{iSNR}$.

For a fixed L, increasing the value of M (from 1 to L) will, in principle, increase the output SNR of the maximum SNR filtering matrix since more and more information is taken into account. The distortion should also increase significantly as M is increased.

3.4.2 Wiener

If we differentiate the MSE criterion, $J(\mathbf{H})$, with respect to $\mathbf{H}$ and equate the result to zero, we find the Wiener filtering matrix

$$\begin{aligned}\mathbf{H}_{\mathrm{W}} &= \mathbf{R}_{\mathbf{x}^M}\boldsymbol{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}\\ &= \mathbf{I}_{\mathrm{i}}\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}\\ &= \mathbf{I}_{\mathrm{i}}\left(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}}\mathbf{R}_{\mathbf{y}}^{-1}\right).\end{aligned} \tag{3.63}$$

This matrix depends only on the second-order statistics of the noise and observation signals. Note that the first line of $\mathbf{H}_{\mathrm{W}}$ is exactly $\mathbf{h}_{\mathrm{W}}^T$.

Lemma 3.2 *We can rewrite the Wiener filtering matrix as*

$$\begin{aligned}\mathbf{H}_{\mathrm{W}} &= \left(\mathbf{I}_M + \mathbf{R}_{\mathbf{x}^M}\boldsymbol{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\Gamma}_{\mathbf{xx}^M}\right)^{-1}\mathbf{R}_{\mathbf{x}^M}\boldsymbol{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\\ &= \left(\mathbf{R}_{\mathbf{x}^M}^{-1} + \boldsymbol{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\Gamma}_{\mathbf{xx}^M}\right)^{-1}\boldsymbol{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}.\end{aligned} \tag{3.64}$$

Proof This expression is easy to show by applying the Woodbury's identity in (3.19) and then substituting the result in (3.63). □

The form of the Wiener filtering matrix presented in (3.64) is interesting because it shows an obvious link with some other optimal filtering matrices as it will be verified later.

Another way to express Wiener is

$$\begin{aligned}\mathbf{H}_{\mathrm{W}} &= \mathbf{I}_{\mathrm{i}}\boldsymbol{\Gamma}_{\mathbf{xx}^M}\mathbf{R}_{\mathbf{x}^M}\boldsymbol{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}\\ &= \mathbf{I}_{\mathrm{i}}\left(\mathbf{I}_L - \mathbf{R}_{\mathrm{in}}\mathbf{R}_{\mathbf{y}}^{-1}\right).\end{aligned} \tag{3.65}$$

Using the joint diagonalization, we can rewrite Wiener as a subspace-type approach:

$$\begin{aligned}\mathbf{H}_{\mathrm{W}} &= \mathbf{I}_{\mathrm{i}}\mathbf{B}^{-T}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda} + \mathbf{I}_L\right)^{-1}\mathbf{B}^T\\ &= \mathbf{I}_{\mathrm{i}}\mathbf{B}^{-T}\begin{bmatrix}\boldsymbol{\Sigma} & \mathbf{0}_{M\times(L-M)}\\ \mathbf{0}_{(L-M)\times M} & \mathbf{0}_{(L-M)\times(L-M)}\end{bmatrix}\mathbf{B}^T\\ &= \mathbf{T}\begin{bmatrix}\boldsymbol{\Sigma} & \mathbf{0}_{M\times(L-M)}\\ \mathbf{0}_{(L-M)\times M} & \mathbf{0}_{(L-M)\times(L-M)}\end{bmatrix}\mathbf{B}^T,\end{aligned} \tag{3.66}$$

where

$$\mathbf{T} = \begin{bmatrix}\mathbf{t}_1^T\\ \mathbf{t}_2^T\\ \vdots\\ \mathbf{t}_M^T\end{bmatrix} = \mathbf{I}_{\mathrm{i}}\mathbf{B}^{-T} \tag{3.67}$$

and

$$\mathbf{\Sigma} = \text{diag}\left(\frac{\lambda_1^M}{\lambda_1^M + 1}, \frac{\lambda_2^M}{\lambda_2^M + 1}, \ldots, \frac{\lambda_M^M}{\lambda_M^M + 1}\right) \tag{3.68}$$

is an $M \times M$ diagonal matrix. Expression (3.66) is also

$$\mathbf{H}_{\text{W}} = \mathbf{I}_{\text{i}}\mathbf{M}_{\text{W}}, \tag{3.69}$$

where

$$\mathbf{M}_{\text{W}} = \mathbf{B}^{-T} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0}_{M\times(L-M)} \\ \mathbf{0}_{(L-M)\times M} & \mathbf{0}_{(L-M)\times(L-M)} \end{bmatrix} \mathbf{B}^T. \tag{3.70}$$

We see that $\mathbf{H}_{\text{W}}$ is the product of two other matrices: the rectangular identity filtering matrix and a square matrix of size $L \times L$ whose rank is equal to M.

For $M = 1$, (3.66) degenerates to

$$\mathbf{h}_{\text{W}} = \mathbf{B} \begin{bmatrix} \lambda_{\max} & \mathbf{0}_{1\times(L-1)} \\ \mathbf{0}_{(L-1)\times 1} & \mathbf{0}_{(L-1)\times(L-1)} \end{bmatrix} \mathbf{B}^{-1}\mathbf{i}_{\text{i}}. \tag{3.71}$$

With the joint diagonalization, the input SNR and the output SNR with Wiener can be expressed as

$$\text{iSNR} = \frac{\text{tr}\left(\mathbf{T}\mathbf{\Lambda}\mathbf{T}^T\right)}{\text{tr}\left(\mathbf{T}\mathbf{T}^T\right)}, \tag{3.72}$$

$$\text{oSNR}\left(\mathbf{H}_{\text{W}}\right) = \frac{\text{tr}\left[\mathbf{T}\mathbf{\Lambda}^3\left(\mathbf{\Lambda} + \mathbf{I}_L\right)^{-2}\mathbf{T}^T\right]}{\text{tr}\left[\mathbf{T}\mathbf{\Lambda}^2\left(\mathbf{\Lambda} + \mathbf{I}_L\right)^{-2}\mathbf{T}^T\right]}. \tag{3.73}$$

Property 3.2 *The output SNR with the Wiener filtering matrix is always greater than or equal to the input SNR, i.e.,* $\text{oSNR}\left(\mathbf{H}_{\text{W}}\right) \geq \text{iSNR}$.

Proof This property can be proven by induction, exactly as in [9]. □

Obviously, we have

$$\text{oSNR}\left(\mathbf{H}_{\text{W}}\right) \leq \text{oSNR}\left(\mathbf{H}_{\max}\right). \tag{3.74}$$

Same as for the maximum SNR filtering matrix, for a fixed L, a higher value of M in the Wiener filtering matrix should give a higher value of the output SNR.

We can easily deduce that

$$\xi_{\text{nr}}\left(\mathbf{H}_{\text{W}}\right) = \frac{\text{tr}\left(\mathbf{T}\mathbf{T}^T\right)}{\text{tr}\left[\mathbf{T}\mathbf{\Lambda}^2\left(\mathbf{\Lambda} + \mathbf{I}_L\right)^{-2}\mathbf{T}^T\right]}, \tag{3.75}$$

$$\xi_{\mathrm{sr}}\left(\mathbf{H}_{\mathrm{W}}\right)=\frac{\mathrm{tr}\left(\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T\right)}{\mathrm{tr}\left[\mathbf{T}\boldsymbol{\Lambda}^3\left(\boldsymbol{\Lambda}+\mathbf{I}_L\right)^{-2}\mathbf{T}^T\right]}, \tag{3.76}$$

$$\upsilon_{\mathrm{sd}}\left(\mathbf{H}_{\mathrm{W}}\right)=\frac{\mathrm{tr}\left[\mathbf{T}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}+\mathbf{I}_L\right)^{-1}\mathbf{T}^T\mathbf{R}_{\mathbf{x}^M}^{-1}\mathbf{T}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}+\mathbf{I}_L\right)^{-1}\mathbf{T}^T\right]}{\mathrm{tr}\left(\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T\right)}. \tag{3.77}$$

3.4.3 MVDR

We recall that the MVDR approach requires no distortion to the desired signal. Therefore, the corresponding rectangular filtering matrix is obtained by minimizing the MSE of the residual interference-plus-noise, $J_{\mathrm{r}}\left(\mathbf{H}\right)$, with the constraint that the desired signal is not distorted. Mathematically, this is equivalent to

$$\min_{\mathbf{H}}\frac{1}{M}\cdot\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathrm{in}}\mathbf{H}^T\right) \quad \text{subject to} \quad \mathbf{H}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}=\mathbf{I}_M. \tag{3.78}$$

The solution to the above optimization problem is

$$\mathbf{H}_{\mathrm{MVDR}}=\left(\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}, \tag{3.79}$$

which is interesting to compare to $\mathbf{H}_{\mathrm{W}}$ (Eq. 3.64).

Obviously, with the MVDR filtering matrix, we have no distortion, i.e.,

$$\xi_{\mathrm{sr}}\left(\mathbf{H}_{\mathrm{MVDR}}\right)=1, \tag{3.80}$$

$$\upsilon_{\mathrm{sd}}\left(\mathbf{H}_{\mathrm{MVDR}}\right)=0. \tag{3.81}$$

Lemma 3.3 *We can rewrite the MVDR filtering matrix as*

$$\mathbf{H}_{\mathrm{MVDR}}=\left(\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}. \tag{3.82}$$

Proof This expression is easy to show by using the Woodbury's identity in $\mathbf{R}_{\mathbf{y}}^{-1}$. □

From (3.82), we deduce the relationship between the MVDR and Wiener filtering matrices:

$$\mathbf{H}_{\mathrm{MVDR}}=\left(\mathbf{H}_{\mathrm{W}}\boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{H}_{\mathrm{W}}. \tag{3.83}$$

Property 3.3 *The output SNR with the MVDR filtering matrix is always greater than or equal to the input SNR, i.e.,* $\mathrm{oSNR}\left(\mathbf{H}_{\mathrm{MVDR}}\right)\geq\mathrm{iSNR}$.

Proof We can prove this property by induction. □

We should have

$$\text{oSNR}\left(\mathbf{H}_{\text{MVDR}}\right) \leq \text{oSNR}\left(\mathbf{H}_{\text{W}}\right) \leq \text{oSNR}\left(\mathbf{H}_{\max}\right). \tag{3.84}$$

Contrary to $\mathbf{H}_{\max}$ and $\mathbf{H}_{\text{W}}$, for a fixed L, a higher value of M in the MVDR filtering matrix implies a lower value of the output SNR.

3.4.4 Prediction

Let $\mathbf{G}$ be a temporal prediction matrix of size $M \times L$ so that

$$\mathbf{x}(k) \approx \mathbf{G}^T \mathbf{x}^M(k). \tag{3.85}$$

The distortionless filtering matrix for noise reduction is derived by

$$\min_{\mathbf{H}} \text{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T\right) \quad \text{subject to} \quad \mathbf{H}\mathbf{G}^T = \mathbf{I}_M, \tag{3.86}$$

from which we deduce the solution

$$\mathbf{H}_{\text{P}} = \left(\mathbf{G}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{G}^T\right)^{-1}\mathbf{G}\mathbf{R}_{\mathbf{y}}^{-1}. \tag{3.87}$$

The best way to find $\mathbf{G}$ is in the Wiener sense. Indeed, define the error signal vector

$$\mathbf{e}_{\text{P}}(k) = \mathbf{x}(k) - \mathbf{G}^T \mathbf{x}^M(k) \tag{3.88}$$

and form the MSE

$$J\left(\mathbf{G}\right) = E\left[\mathbf{e}_{\text{P}}^T(k)\mathbf{e}_{\text{P}}(k)\right]. \tag{3.89}$$

The minimization of $J\left(\mathbf{G}\right)$ with respect to $\mathbf{G}$ leads to

$$\mathbf{G}_{\text{o}} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T \tag{3.90}$$

and substituting this result into (3.87) gives

$$\mathbf{H}_{\text{P}} = \left(\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}, \tag{3.91}$$

which corresponds to the MVDR.

It is interesting to observe that the error signal vector with the optimal matrix, $\mathbf{G}_{\text{o}}$, corresponds to the interference signal vector, i.e.,

$$\begin{aligned} \mathbf{e}_{\text{P,o}}(k) &= \mathbf{x}(k) - \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{x}^M(k) \\ &= \mathbf{x}_{\text{i}}(k). \end{aligned} \tag{3.92}$$

This result is a consequence of the orthogonality principle.

3.4.5 Tradeoff

In the tradeoff approach, we minimize the speech distortion index with the constraint that the noise reduction factor is equal to a positive value that is greater than 1. Mathematically, this is equivalent to

$$\min_{\mathbf{H}} J_{\rm d}(\mathbf{H}) \quad \text{subject to} \quad J_{\rm r}(\mathbf{H}) = \beta J_{\rm r}(\mathbf{I}_{\rm i}), \tag{3.93}$$

where $0 < \beta < 1$ to insure that we get some noise reduction. By using a Lagrange multiplier, $\mu > 0$, to adjoin the constraint to the cost function and assuming that the matrix $\mathbf{\Gamma}_{\mathbf{xx}^M}\mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{xx}^M}^T + \mu\mathbf{R}_{\rm in}$ is invertible, we easily deduce the tradeoff filtering matrix

$$\mathbf{H}_{\mathrm{T},\mu} = \mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{xx}^M}^T\left(\mathbf{\Gamma}_{\mathbf{xx}^M}\mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{xx}^M}^T + \mu\mathbf{R}_{\rm in}\right)^{-1}, \tag{3.94}$$

which can be rewritten, thanks to the Woodbury's identity, as

$$\mathbf{H}_{\mathrm{T},\mu} = \left(\mu\mathbf{R}_{\mathbf{x}^M}^{-1} + \mathbf{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\rm in}^{-1}\mathbf{\Gamma}_{\mathbf{xx}^M}\right)^{-1}\mathbf{\Gamma}_{\mathbf{xx}^M}^T\mathbf{R}_{\rm in}^{-1}, \tag{3.95}$$

where μ satisfies $J_{\rm r}\left(\mathbf{H}_{\mathrm{T},\mu}\right) = \beta J_{\rm r}(\mathbf{I}_{\rm i})$. Usually, μ is chosen in an ad-hoc way, so that for

- $\mu = 1$, $\mathbf{H}_{\mathrm{T},1} = \mathbf{H}_{\rm W}$, which is the Wiener filtering matrix;
- $\mu = 0$ [from (3.95)], $\mathbf{H}_{\mathrm{T},0} = \mathbf{H}_{\rm MVDR}$, which is the MVDR filtering matrix;
- $\mu > 1$, results in a filter with low residual noise (compared with the Wiener filter) at the expense of high speech distortion;
- $\mu < 1$, results in a filter with high residual noise and low speech distortion.

Property 3.4 *The output SNR with the tradeoff filtering matrix is always greater than or equal to the input SNR, i.e.,* $\mathrm{oSNR}\left(\mathbf{H}_{\mathrm{T},\mu}\right) \geq \mathrm{iSNR}$, $\forall\mu \geq 0$.

Proof We can prove this property by induction. □

We should have for $\mu \geq 1$,

$$\mathrm{oSNR}(\mathbf{H}_{\rm MVDR}) \leq \mathrm{oSNR}(\mathbf{H}_{\rm W}) \leq \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{T},\mu}\right) \leq \mathrm{oSNR}(\mathbf{H}_{\max}) \tag{3.96}$$

and for $\mu \leq 1$,

$$\mathrm{oSNR}(\mathbf{H}_{\rm MVDR}) \leq \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{T},\mu}\right) \leq \mathrm{oSNR}(\mathbf{H}_{\rm W}) \leq \mathrm{oSNR}(\mathbf{H}_{\max}). \tag{3.97}$$

We can write the tradeoff filtering matrix as a subspace-type approach. Indeed, from (3.94), we get

$$\mathbf{H}_{\mathrm{T},\mu} = \mathbf{T}\begin{bmatrix} \mathbf{\Sigma}_\mu & \mathbf{0}_{M\times(L-M)} \\ \mathbf{0}_{(L-M)\times M} & \mathbf{0}_{(L-M)\times(L-M)} \end{bmatrix}\mathbf{B}^T, \tag{3.98}$$

where

$$\mathbf{\Sigma}_\mu = \text{diag}\left(\frac{\lambda_1^M}{\lambda_1^M + \mu}, \frac{\lambda_2^M}{\lambda_2^M + \mu}, \ldots, \frac{\lambda_M^M}{\lambda_M^M + \mu}\right) \tag{3.99}$$

is an $M \times M$ diagonal matrix. Expression (3.98) is also

$$\mathbf{H}_{\text{T},\mu} = \mathbf{I}_\text{i}\mathbf{M}_{\text{T},\mu}, \tag{3.100}$$

where

$$\mathbf{M}_{\text{T},\mu} = \mathbf{B}^{-T} \begin{bmatrix} \mathbf{\Sigma}_\mu & \mathbf{0}_{M\times(L-M)} \\ \mathbf{0}_{(L-M)\times M} & \mathbf{0}_{(L-M)\times(L-M)} \end{bmatrix} \mathbf{B}^T. \tag{3.101}$$

We see that $\mathbf{H}_{\text{T},\mu}$ is the product of two other matrices: the rectangular identity filtering matrix and an adjustable square matrix of size $L \times L$ whose rank is equal to M. Note that $\mathbf{H}_{\text{T},\mu}$ as presented in (3.98) is not, in principle, defined for $\mu = 0$ as this expression was derived from (3.94), which is clearly not defined for this particular case. Although it is possible to have $\mu = 0$ in (3.98), this does not lead to the MVDR.

3.4.6 Particular Case: $M = L$

For $M = L$, the rectangular matrix $\mathbf{H}$ becomes a square matrix $\mathbf{H}_\text{S}$ of size $L \times L$. It can be verified that $\mathbf{x}_\text{i}(k) = \mathbf{0}_{L\times 1}$; as a result, $\mathbf{R}_\text{in} = \mathbf{R}_\mathbf{v}$, $\mathbf{R}_{\mathbf{x}_\text{i}} = \mathbf{0}_{L\times L}$, and $\mathbf{R}_{\mathbf{x}_\text{d}} = \mathbf{R}_\mathbf{x}$. Therefore, the optimal filtering matrices are

$$\mathbf{H}_{\text{S,max}} = \begin{bmatrix} \beta_1 \mathbf{b}_1^{LT} \\ \beta_2 \mathbf{b}_1^{LT} \\ \vdots \\ \beta_L \mathbf{b}_1^{LT} \end{bmatrix}, \tag{3.102}$$

$$\begin{aligned} \mathbf{H}_{\text{S,W}} &= \mathbf{R}_\mathbf{x}\mathbf{R}_\mathbf{y}^{-1} \\ &= \mathbf{I}_L - \mathbf{R}_\mathbf{v}\mathbf{R}_\mathbf{y}^{-1}, \end{aligned} \tag{3.103}$$

$$\mathbf{H}_{\text{S,MVDR}} = \mathbf{I}_L, \tag{3.104}$$

$$\begin{aligned} \mathbf{H}_{\text{S,T},\mu} &= \mathbf{R}_\mathbf{x}\left(\mathbf{R}_\mathbf{x} + \mu\mathbf{R}_\mathbf{v}\right)^{-1} \\ &= \left(\mathbf{R}_\mathbf{y} - \mathbf{R}_\mathbf{v}\right)\left[\mathbf{R}_\mathbf{y} + (\mu - 1)\mathbf{R}_\mathbf{v}\right]^{-1}, \end{aligned} \tag{3.105}$$

where $\mathbf{b}_1^L$ is the eigenvector corresponding to the maximum eigenvalue of the matrix $\mathbf{R}_\mathbf{v}^{-1}\mathbf{R}_\mathbf{x}$. In this case, all filtering matrices are very much different and the MVDR is the identity matrix.

Applying the joint diagonalization in (3.105), we get

$$\mathbf{H}_{\mathrm{S,T},\mu} = \mathbf{B}^{-T} \boldsymbol{\Lambda} \left(\boldsymbol{\Lambda} + \mathbf{I}_L\right)^{-1} \mathbf{B}^T. \tag{3.106}$$

It is believed that a speech signal can be modelled as a linear combination of a number of some (linearly independent) basis vectors smaller than the dimension of these vectors [2, 4, 10–13]. As a result, the vector space of the noisy signal can be decomposed in two subspaces: the signal-plus-noise subspace of length L_{s} and the null subspace of length L_{n}, with $L = L_{\mathrm{s}} + L_{\mathrm{n}}$. This implies that the last L_{n} eigenvalues of the matrix $\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{R}_{\mathbf{x}}$ are equal to zero. Therefore, we can rewrite (3.106) to obtain the subspace-type filter:

$$\mathbf{H}_{\mathrm{S,T},\mu} = \mathbf{B}^{-T} \begin{bmatrix} \boldsymbol{\Sigma}_\mu & \mathbf{0}_{L_{\mathrm{s}}\times L_{\mathrm{n}}} \\ \mathbf{0}_{L_{\mathrm{n}}\times L_{\mathrm{s}}} & \mathbf{0}_{L_{\mathrm{n}}\times L_{\mathrm{n}}} \end{bmatrix} \mathbf{B}^T, \tag{3.107}$$

where now

$$\boldsymbol{\Sigma}_\mu = \mathrm{diag}\left(\frac{\lambda_1^L}{\lambda_1^L + \mu}, \frac{\lambda_2^L}{\lambda_2^L + \mu}, \ldots, \frac{\lambda_{L_{\mathrm{s}}}^L}{\lambda_{L_{\mathrm{s}}}^L + \mu}\right) \tag{3.108}$$

is an $L_{\mathrm{s}} \times L_{\mathrm{s}}$ diagonal matrix. This algorithm is often referred to as the generalized subspace approach. One should note, however, that there is no noise-only subspace with this formulation. Therefore, noise reduction can only be achieved by modifying the speech-plus-noise subspace by setting μ to a positive number.

It can be shown that for $\mu \geq 1$,

$$\begin{aligned} \mathrm{iSNR} =& \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,MVDR}}\right) \leq \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,W}}\right) \leq \\ & \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,T},\mu}\right) \leq \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,max}}\right) = \lambda_1^L \end{aligned} \tag{3.109}$$

and for $0 \leq \mu \leq 1$,

$$\begin{aligned} \mathrm{iSNR} =& \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,MVDR}}\right) \leq \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,T},\mu}\right) \leq \\ & \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,W}}\right) \leq \mathrm{oSNR}\left(\mathbf{H}_{\mathrm{S,max}}\right) = \lambda_1^L, \end{aligned} \tag{3.110}$$

where λ_1^L is the maximum eigenvalue of the matrix $\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{R}_{\mathbf{x}}$.

The results derived in the preceding subsections are not surprising because the optimal filtering matrices derived so far in this chapter are related as follows:

$$\mathbf{H}_{\mathrm{o}} = \mathbf{A}_{\mathrm{o}} \boldsymbol{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T \mathbf{R}_{\mathrm{in}}^{-1}, \tag{3.111}$$

where $\mathbf{A}_{\mathrm{o}}$ is a square matrix of size $M \times M$. Therefore, depending on how we choose $\mathbf{A}_{\mathrm{o}}$, we obtain the different optimal filtering matrices. In other words, these optimal filtering matrices are equivalent up to the matrix $\mathbf{A}_{\mathrm{o}}$. For $M = 1$, the matrix $\mathbf{A}_{\mathrm{o}}$ degenerates to a scalar and the filters derived in Chap. 2 are obtained, which are basically equivalent.

3.4.7 LCMV

The LCMV beamformer is able to handle other constraints than the distortionless ones.

We can exploit the structure of the noise signal in the same manner as we did it in Chap. 2. Indeed, in the proposed LCMV, we will not only perfectly recover the desired signal vector, $\mathbf{x}^M(k)$, but we will also completely remove the coherent noise signal. Therefore, our constraints are

$$\mathbf{H}\mathbf{C}_{\mathbf{x}^M v} = \left[\mathbf{I}_M \ \mathbf{0}_{M \times 1} \right], \tag{3.112}$$

where

$$\mathbf{C}_{\mathbf{x}^M v} = \left[\Gamma_{\mathbf{x}\mathbf{x}^M} \ \rho_{\mathbf{v}v} \right] \tag{3.113}$$

is our constraint matrix of size $L \times (M+1)$.

Our optimization problem is now

$$\min_{\mathbf{H}} \operatorname{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T\right) \quad \text{subject to} \quad \mathbf{H}\mathbf{C}_{\mathbf{x}^M v} = \left[\mathbf{I}_M \ \mathbf{0}_{M \times 1} \right], \tag{3.114}$$

from which we find the LCMV filtering matrix

$$\mathbf{H}_{\text{LCMV}} = \left[\mathbf{I}_M \ \mathbf{0}_{M \times 1} \right] \left(\mathbf{C}_{\mathbf{x}^M v}^T \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{C}_{\mathbf{x}^M v} \right)^{-1} \mathbf{C}_{\mathbf{x}^M v}^T \mathbf{R}_{\mathbf{y}}^{-1}. \tag{3.115}$$

If the coherent noise is the main issue, then the LCMV is perhaps the most interesting solution.

3.5 Summary

The ideas of single-channel noise reduction in the time domain of Chap. 2 were generalized in this chapter. In particular, we were able to derive the same noise reduction algorithms but for the estimation of M samples at a time with a rectangular filtering matrix. This can lead to a potential better performance in terms of noise reduction for most of the optimization criteria. However, this time, the optimal filtering matrices are very much different from one to another since the corresponding output SNRs are not equal.

References

1. J. Benesty, J. Chen, Y. Huang, I. Cohen, *Noise Reduction in Speech Processing*. (Springer, Berlin, 2009)
2. Y. Ephraim, H.L. Van Trees, A signal subspace approach for speech enhancement. IEEE Trans. Speech Audio Process. **3**, 251–266 (1995)

3. P.S.K. Hansen, *Signal subspace methods for speech enhancement*. Ph. D. dissertation, Technical University of Denmark, Lyngby, Denmark (1997)
4. S.H. Jensen, P.C. Hansen, S.D. Hansen, J.A. Sørensen, Reduction of broad-band noise in speech by truncated QSVD. IEEE Trans. Speech Audio Process. **3**, 439–448 (1995)
5. S. Doclo, M. Moonen, GSVD-based optimal filtering for single and multimicrophone speech enhancement. IEEE Trans. Signal Process. **50**, 2230–2244 (2002)
6. S.B. Searle, *Matrix Algebra Useful for Statistics*. (Wiley, New York, 1982)
7. G. Strang, *Linear Algebra and its Applications*. 3rd edn (Harcourt Brace Jovanonich, Orlando, 1988)
8. Y. Hu, P.C. Loizou, A subspace approach for enhancing speech corrupted by colored noise. IEEE Signal Process. Lett. **9**, 204–206 (2002)
9. J. Chen, J. Benesty, Y. Huang, S. Doclo, New insights into the noise reduction Wiener filter. IEEE Trans. Audio Speech Lang. Process. **14**, 1218–1234 (2006)
10. M. Dendrinos, S. Bakamidis, G. Carayannis, Speech enhancement from noise: a regenerative approach. Speech Commun. **10**, 45–57 (1991)
11. Y. Hu, P.C. Loizou, A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans. Speech Audio Process. **11**, 334–341 (2003)
12. F. Jabloun, B. Champagne, Signal subspace techniques for speech enhancement. In: J. Benesty, S. Makino, J. Chen (eds) *Speech Enhancement*, (Springer, Berlin, 2005) pp. 135–159. Chap. 7
13. K. Hermus, P. Wambacq, H. Van hamme, A review of signal subspace speech enhancement and its application to noise robust speech recognition. EURASIP J Adv. Signal Process. **2007**, 15 pages, Article ID 45821 (2007)

Chapter 4
Multichannel Noise Reduction with a Filtering Vector

In the previous two chapters, we exploited the temporal correlation information from a single microphone signal to derive different filtering vectors and matrices for noise reduction. In this chapter and the next one, we will exploit both the temporal and spatial information available from signals picked up by a determined number of microphones at different positions in the acoustics space in order to mitigate the noise effect. The focus of this chapter is on optimal filtering vectors.

4.1 Signal Model

We consider the conventional signal model in which a microphone array with N sensors captures a convolved source signal in some noise field. The received signals are expressed as [1, 2]

$$\begin{aligned} y_n(k) &= g_n(k) * s(k) + v_n(k) \\ &= x_n(k) + v_n(k), \quad n = 1, 2, \ldots, N, \end{aligned} \tag{4.1}$$

where $g_n(k)$ is the acoustic impulse response from the unknown speech source, $s(k)$, location to the nth microphone, $*$ stands for linear convolution, and $v_n(k)$ is the additive noise at microphone n. We assume that the signals $x_n(k) = g_n(k) * s(k)$ and $v_n(k)$ are uncorrelated, zero mean, real, and broadband. By definition, $x_n(k)$ is coherent across the array. The noise signals, $v_n(k)$, are typically only partially coherent across the array. To simplify the development and analysis of the main ideas of this work, we further assume that the signals are Gaussian and stationary.

By processing the data by blocks of L samples, the signal model given in (4.1) can be put into a vector form as

$$\mathbf{y}_n(k) = \mathbf{x}_n(k) + \mathbf{v}_n(k), \quad n = 1, 2, \ldots, N, \tag{4.2}$$

J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters*, SpringerBriefs in Electrical and Computer Engineering, 1,
DOI: 10.1007/978-3-642-19601-0_4, © Jacob Benesty 2011

where

$$\mathbf{y}_n(k) = [y_n(k)\ y_n(k-1)\ \cdots\ y_n(k-L+1)]^T \tag{4.3}$$

is a vector of length L, and $\mathbf{x}_n(k)$ and $\mathbf{v}_n(k)$ are defined similarly to $\mathbf{y}_n(k)$. It is more convenient to concatenate the N vectors $\mathbf{y}_n(k)$ together as

$$\begin{aligned}\underline{\mathbf{y}}(k) &= \left[\mathbf{y}_1^T(k)\ \mathbf{y}_2^T(k)\ \cdots\ \mathbf{y}_N^T(k)\right]^T \\ &= \underline{\mathbf{x}}(k) + \underline{\mathbf{v}}(k),\end{aligned} \tag{4.4}$$

where vectors $\underline{\mathbf{x}}(k)$ and $\underline{\mathbf{v}}(k)$ of length NL are defined in a similar way to $\underline{\mathbf{y}}(k)$. Since $x_n(k)$ and $v_n(k)$ are uncorrelated by assumption, the correlation matrix (of size $NL \times NL$) of the microphone signals is

$$\begin{aligned}\mathbf{R}_{\underline{\mathbf{y}}} &= E\left[\underline{\mathbf{y}}(k)\underline{\mathbf{y}}^T(k)\right] \\ &= \mathbf{R}_{\underline{\mathbf{x}}} + \mathbf{R}_{\underline{\mathbf{v}}},\end{aligned} \tag{4.5}$$

where $\mathbf{R}_{\underline{\mathbf{x}}} = E\left[\underline{\mathbf{x}}(k)\underline{\mathbf{x}}^T(k)\right]$ and $\mathbf{R}_{\underline{\mathbf{v}}} = E\left[\underline{\mathbf{v}}(k)\underline{\mathbf{v}}^T(k)\right]$ are the correlation matrices of $\underline{\mathbf{x}}(k)$ and $\underline{\mathbf{v}}(k)$, respectively.

In this work, our desired signal is designated by the clean (but convolved) speech signal received at microphone 1, namely $x_1(k)$. Obviously, any signal $x_n(k)$ could be taken as the reference. Our problem then may be stated as follows [3]: given N mixtures of two uncorrelated signals $x_n(k)$ and $v_n(k)$, our aim is to preserve $x_1(k)$ while minimizing the contribution of the noise terms, $v_n(k)$, at the array output.

Since $x_1(k)$ is the signal of interest, it is important to write the vector $\underline{\mathbf{y}}(k)$ as a function of $x_1(k)$. For that, we need first to decompose $\underline{\mathbf{x}}(k)$ into two orthogonal components: one proportional to the desired signal, $x_1(k)$, and the other corresponding to the interference. Indeed, it is easy to see that this decomposition is

$$\underline{\mathbf{x}}(k) = \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \cdot x_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k), \tag{4.6}$$

where

$$\begin{aligned}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} &= \left[\boldsymbol{\rho}_{\mathbf{x}_1x_1}^T\ \boldsymbol{\rho}_{\mathbf{x}_2x_1}^T\ \cdots\ \boldsymbol{\rho}_{\mathbf{x}_Nx_1}^T\right]^T \\ &= \frac{E\left[\underline{\mathbf{x}}(k)x_1(k)\right]}{E\left[x_1^2(k)\right]}\end{aligned} \tag{4.7}$$

is the partially normalized [with respect to $x_1(k)$] cross-correlation vector (of length NL) between $\underline{\mathbf{x}}(k)$ and $x_1(k)$,

$$\begin{aligned}\boldsymbol{\rho}_{\mathbf{x}_nx_1} &= \left[\rho_{x_nx_1}(0)\ \rho_{x_nx_1}(1)\ \cdots\ \rho_{x_nx_1}(L-1)\right]^T \\ &= \frac{E[\mathbf{x}_n(k)x_1(k)]}{E[x_1^2(k)]}, \quad n = 1, 2, \ldots, N\end{aligned} \tag{4.8}$$

is the partially normalized [with respect to $x_1(k)$] cross-correlation vector (of length L) between $\mathbf{x}_n(k)$ and $x_1(k)$,

$$\rho_{x_n x_1}(l) = \frac{E\left[x_n(k-l)x_1(k)\right]}{E\left[x_1^2(k)\right]}, \quad n = 1, 2, \ldots, N, \; l = 0, 1, \ldots, L-1 \tag{4.9}$$

is the partially normalized [with respect to $x_1(k)$] cross-correlation coefficient between $x_n(k-l)$ and $x_1(k)$,

$$\underline{\mathbf{x}}_{\mathrm{i}}(k) = \underline{\mathbf{x}}(k) - \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \cdot x_1(k) \tag{4.10}$$

is the interference signal vector, and

$$E\left[\underline{\mathbf{x}}_{\mathrm{i}}(k)x_1(k)\right] = \mathbf{0}_{NL\times 1}. \tag{4.11}$$

Substituting (4.6) into (4.4), we get the signal model for noise reduction in the time domain:

$$\begin{aligned} \underline{\mathbf{y}}(k) &= \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \cdot x_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) + \underline{\mathbf{v}}(k) \\ &= \underline{\mathbf{x}}_{\mathrm{d}}(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) + \underline{\mathbf{v}}(k), \end{aligned} \tag{4.12}$$

where $\underline{\mathbf{x}}_{\mathrm{d}}(k) = \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \cdot x_1(k)$ is the desired signal vector. The vector $\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}$ is clearly a general definition in the time domain of the steering vector [4, 5] for noise reduction since it determines the direction of the desired signal, $x_1(k)$.

4.2 Linear Filtering with a Vector

The array processing, beamforming, or multichannel noise reduction is performed by applying a temporal filter to each microphone signal and summing the filtered signals. Thus, the clear objective is to estimate the sample $x_1(k)$ from the vector $\underline{\mathbf{y}}(k)$ of length NL. Let us denote by $z(k)$ this estimate. We have

$$\begin{aligned} z(k) &= \sum_{n=1}^{N} \mathbf{h}_n^T \mathbf{y}_n(k) \\ &= \underline{\mathbf{h}}^T \underline{\mathbf{y}}(k), \end{aligned} \tag{4.13}$$

where $\mathbf{h}_n,\ n = 1, 2, \ldots, N$ are N FIR filters of length L and

$$\underline{\mathbf{h}} = \left[\mathbf{h}_1^T \; \mathbf{h}_2^T \; \cdots \; \mathbf{h}_N^T\right]^T \tag{4.14}$$

is a long filtering vector of length NL.

Using the formulation of $\underline{\mathbf{y}}(k)$ that is explicitly a function of the steering vector, we can rewrite (4.13) as

$$\begin{aligned} z(k) &= \underline{\mathbf{h}}^T \left[\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \cdot x_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) + \underline{\mathbf{v}}(k) \right] \\ &= x_{\mathrm{fd}}(k) + x_{\mathrm{ri}}(k) + v_{\mathrm{rn}}(k), \end{aligned} \tag{4.15}$$

where

$$x_{\mathrm{fd}}(k) = x_1(k) \underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \tag{4.16}$$

is the filtered desired signal,

$$x_{\mathrm{ri}}(k) = \underline{\mathbf{h}}^T \underline{\mathbf{x}}_{\mathrm{i}}(k) \tag{4.17}$$

is the residual interference, and

$$v_{\mathrm{rn}}(k) = \underline{\mathbf{h}}^T \underline{\mathbf{v}}(k) \tag{4.18}$$

is the residual noise.

Since the estimate of the desired signal at time k is the sum of three terms that are mutually uncorrelated, the variance of $z(k)$ is

$$\begin{aligned} \sigma_z^2 &= \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{y}}} \underline{\mathbf{h}} \\ &= \sigma_{x_{\mathrm{fd}}}^2 + \sigma_{x_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2, \end{aligned} \tag{4.19}$$

where

$$\sigma_{x_{\mathrm{fd}}}^2 = \sigma_{x_1}^2 \left(\underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \right)^2, \tag{4.20}$$

$$\begin{aligned} \sigma_{x_{\mathrm{ri}}}^2 &= \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{i}}} \underline{\mathbf{h}} \\ &= \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{x}}} \underline{\mathbf{h}} - \sigma_{x_1}^2 \left(\underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \right)^2, \end{aligned} \tag{4.21}$$

$$\sigma_{v_{\mathrm{rn}}}^2 = \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{v}}} \underline{\mathbf{h}}, \tag{4.22}$$

$\sigma_{x_1}^2 = E\left[x_1^2(k)\right]$ and $\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{i}}} = E\left[\underline{\mathbf{x}}_{\mathrm{i}}(k)\underline{\mathbf{x}}_{\mathrm{i}}^T(k)\right]$. The variance of $z(k)$ will be extensively used in the coming sections.

4.3 Performance Measures

In this section, we define some fundamental measures that fit well in the multiple microphone case and with a linear filtering vector. We recall that microphone 1 is the reference; therefore, all measures are derived with respect to this microphone.

4.3.1 Noise Reduction

The input SNR is

$$\mathrm{iSNR} = \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2}, \tag{4.23}$$

where $\sigma_{v_1}^2 = E\left[v_1^2(k)\right]$ is the variance of the noise at microphone 1.

The output SNR is obtained from (4.19):

$$\begin{aligned}\mathrm{oSNR}\left(\underline{\mathbf{h}}\right) &= \frac{\sigma_{x_{\mathrm{fd}}}^2}{\sigma_{x_{\mathrm{ri}}}^2 + \sigma_{v_{\mathrm{rn}}}^2} \\ &= \frac{\sigma_{x_1}^2\left(\underline{\mathbf{h}}^T\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\right)^2}{\underline{\mathbf{h}}^T\mathbf{R}_{\mathrm{in}}\underline{\mathbf{h}}},\end{aligned} \tag{4.24}$$

where

$$\mathbf{R}_{\mathrm{in}} = \mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{i}}} + \mathbf{R}_{\underline{\mathbf{v}}} \tag{4.25}$$

is the interference-plus-noise covariance matrix. We observe from (4.24) that the output SNR is defined as the variance of the first signal (filtered desired) from the right-hand side of (4.19) over the variance of the two other signals (filtered interference-plus-noise).

For the particular filtering vector

$$\underline{\mathbf{h}} = \underline{\mathbf{i}}_{\mathrm{i}} = [1\ 0\ \cdots\ 0]^T \tag{4.26}$$

of length NL, we have

$$\mathrm{oSNR}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right) = \mathrm{iSNR}. \tag{4.27}$$

With the identity filtering vector $\underline{\mathbf{i}}_{\mathrm{i}}$, the SNR cannot be improved.

For any two vectors $\underline{\mathbf{h}}$ and $\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}$ and a positive definite matrix $\mathbf{R}_{\mathrm{in}}$, we have

$$\left(\underline{\mathbf{h}}^T\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\right)^2 \leq \left(\underline{\mathbf{h}}^T\mathbf{R}_{\mathrm{in}}\underline{\mathbf{h}}\right)\left(\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\right), \tag{4.28}$$

with equality if and only if $\underline{\mathbf{h}} = \varsigma\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x}$, where $\varsigma(\neq 0)$ is a real number. Using the previous inequality in (4.24), we deduce an upper bound for the output SNR:

$$\mathrm{oSNR}\left(\underline{\mathbf{h}}\right) \leq \sigma_{x_1}^2 \cdot \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}, \quad \forall\underline{\mathbf{h}} \tag{4.29}$$

and clearly,

$$\mathrm{oSNR}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right) \leq \sigma_{x_1}^2 \cdot \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}, \tag{4.30}$$

which implies that

$$\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \geq \frac{1}{\sigma_{v_1}^2}. \tag{4.31}$$

The role of the beamformer is to produce a signal whose SNR is higher than that of the received signal. This is measured by the array gain:

$$\mathcal{A}\left(\underline{\mathbf{h}}\right) = \frac{\text{oSNR}(\underline{\mathbf{h}})}{\text{iSNR}}. \tag{4.32}$$

From (4.29), we deduce that the maximum array gain is

$$\mathcal{A}_{\max} = \sigma_{v_1}^2 \cdot \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \geq 1. \tag{4.33}$$

Taking the ratio of the power of the noise at the reference microphone over the power of the interference-plus-noise remaining at the beamformer output, we get the noise reduction factor:

$$\xi_{\text{nr}}\left(\underline{\mathbf{h}}\right) = \frac{\sigma_{v_1}^2}{\underline{\mathbf{h}}^T \mathbf{R}_{\text{in}} \underline{\mathbf{h}}}, \tag{4.34}$$

which should be lower bounded by 1 for optimal filtering vectors.

4.3.2 Speech Distortion

The speech reduction factor defined as

$$\begin{aligned} \xi_{\text{sr}}\left(\underline{\mathbf{h}}\right) &= \frac{\sigma_{x_1}^2}{\sigma_{x_{\text{fd}}}^2} \\ &= \frac{1}{\left(\underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\right)^2}, \end{aligned} \tag{4.35}$$

measures the distortion of the desired speech signal. It is supposed to be equal to 1 if there is no distortion and expected to be greater than 1 when distortion happens.

The speech distortion index is

$$\begin{aligned} \upsilon_{\text{sd}}\left(\underline{\mathbf{h}}\right) &= \frac{E\left\{\left[x_{\text{fd}}(k) - x_1(k)\right]^2\right\}}{E\left[x_1^2(k)\right]} \\ &= \left(\mathbf{h}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} - 1\right)^2 \\ &= \left[\xi_{\text{sr}}^{-1/2}\left(\underline{\mathbf{h}}\right) - 1\right]^2. \end{aligned} \tag{4.36}$$

For optimal beamformers, we should have $0 \leq \upsilon_{\text{sd}}\left(\underline{\mathbf{h}}\right) \leq 1$.

It is easy to verify that we have the following fundamental relation:

$$\mathcal{A}\left(\underline{\mathbf{h}}\right) = \frac{\xi_{\text{nr}}\left(\underline{\mathbf{h}}\right)}{\xi_{\text{sr}}\left(\underline{\mathbf{h}}\right)}. \tag{4.37}$$

This expression indicates the equivalence between array gain/loss and distortion.

4.3.3 MSE Criterion

In the multichannel case, we define the error signal between the estimated and desired signals as

$$\begin{aligned} e(k) &= z(k) - x_1(k) \\ &= x_{\text{fd}}(k) + x_{\text{ri}}(k) + v_{\text{rn}}(k) - x_1(k). \end{aligned} \tag{4.38}$$

This error can be expressed as the sum of two other uncorrelated errors:

$$e(k) = e_{\text{d}}(k) + e_{\text{r}}(k), \tag{4.39}$$

where

$$\begin{aligned} e_{\text{d}}(k) &= x_{\text{fd}}(k) - x_1(k) \\ &= \left(\underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} - 1\right) x_1(k) \end{aligned} \tag{4.40}$$

is the signal distortion due to the filtering vector and

$$\begin{aligned} e_{\text{r}}(k) &= x_{\text{ri}}(k) + v_{\text{rn}}(k) \\ &= \underline{\mathbf{h}}^T \underline{\mathbf{x}}_{\text{i}}(k) + \underline{\mathbf{h}}^T \underline{\mathbf{v}}(k) \end{aligned} \tag{4.41}$$

represents the residual interference-plus-noise.

The MSE criterion, which is formed from the error (4.38), is given by

$$\begin{aligned} J\left(\underline{\mathbf{h}}\right) &= E\left[e^2(k)\right] \\ &= \sigma_{x_1}^2 + \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{y}}} \underline{\mathbf{h}} - 2\underline{\mathbf{h}}^T E\left[\underline{\mathbf{x}}(k) x_1(k)\right] \\ &= \sigma_{x_1}^2 + \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{y}}} \underline{\mathbf{h}} - 2\sigma_{x_1}^2 \underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x} \\ &= J_{\text{d}}\left(\underline{\mathbf{h}}\right) + J_{\text{r}}\left(\underline{\mathbf{h}}\right), \end{aligned} \tag{4.42}$$

where

$$\begin{aligned} J_{\text{d}}\left(\underline{\mathbf{h}}\right) &= E\left[e_{\text{d}}^2(k)\right] \\ &= \sigma_{x_1}^2 \left(\underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} - 1\right)^2 \end{aligned} \tag{4.43}$$

and

$$J_{\mathrm{r}}\left(\underline{\mathbf{h}}\right) = E\left[e_{\mathrm{r}}^2(k)\right] \\ = \underline{\mathbf{h}}^T \mathbf{R}_{\mathrm{in}} \underline{\mathbf{h}}. \tag{4.44}$$

We are interested in two particular filtering vectors: $\underline{\mathbf{h}} = \underline{\mathbf{i}}_{\mathrm{i}}$ and $\underline{\mathbf{h}} = \mathbf{0}_{NL\times 1}$. With the first one (identity filtering vector), we have neither noise reduction nor speech distortion and with the second one (zero filtering vector), we have maximum noise reduction and maximum speech distortion. For both filters, however, it can be verified that the output SNR is equal to the input SNR. For these two particular filters, the MSEs are

$$J\left(\underline{\mathbf{i}}_{\mathrm{i}}\right) = J_{\mathrm{r}}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right) = \sigma_{v_1}^2, \tag{4.45}$$

$$J\left(\mathbf{0}_{NL\times 1}\right) = J_{\mathrm{d}}\left(\mathbf{0}_{NL\times 1}\right) = \sigma_{x_1}^2. \tag{4.46}$$

As a result,

$$\mathrm{iSNR} = \frac{J\left(\mathbf{0}_{NL\times 1}\right)}{J\left(\underline{\mathbf{i}}_{\mathrm{i}}\right)}. \tag{4.47}$$

We define the NMSE with respect to $J\left(\underline{\mathbf{i}}_{\mathrm{i}}\right)$ as

$$\widetilde{J}\left(\underline{\mathbf{h}}\right) = \frac{J\left(\underline{\mathbf{h}}\right)}{J\left(\underline{\mathbf{i}}_{\mathrm{i}}\right)} \\ = \mathrm{iSNR} \cdot \upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}\right) + \frac{1}{\xi_{\mathrm{nr}}\left(\underline{\mathbf{h}}\right)} \\ = \mathrm{iSNR}\left[\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}\right) + \frac{1}{\mathrm{oSNR}\left(\underline{\mathbf{h}}\right) \cdot \xi_{\mathrm{sr}}\left(\underline{\mathbf{h}}\right)}\right], \tag{4.48}$$

where

$$\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}\right) = \frac{J_{\mathrm{d}}\left(\underline{\mathbf{h}}\right)}{J_{\mathrm{d}}\left(\mathbf{0}_{NL\times 1}\right)}, \tag{4.49}$$

$$\mathrm{iSNR} \cdot \upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}\right) = \frac{J_{\mathrm{d}}\left(\underline{\mathbf{h}}\right)}{J_{\mathrm{r}}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right)}, \tag{4.50}$$

$$\xi_{\mathrm{nr}}\left(\underline{\mathbf{h}}\right) = \frac{J_{\mathrm{r}}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right)}{J_{\mathrm{r}}\left(\underline{\mathbf{h}}\right)}, \tag{4.51}$$

$$\mathrm{oSNR}\left(\underline{\mathbf{h}}\right) \cdot \xi_{\mathrm{sr}}\left(\underline{\mathbf{h}}\right) = \frac{J_{\mathrm{d}}\left(\mathbf{0}_{NL\times 1}\right)}{J_{\mathrm{r}}\left(\underline{\mathbf{h}}\right)}. \tag{4.52}$$

This shows how this NMSE and the different MSEs are related to the performance measures.

We define the NMSE with respect to $J\left(\mathbf{0}_{NL\times 1}\right)$ as

$$\begin{aligned} \overline{J}\left(\underline{\mathbf{h}}\right) &= \frac{J\left(\underline{\mathbf{h}}\right)}{J\left(\mathbf{0}_{NL\times 1}\right)} \\ &= \upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}\right) + \frac{1}{\mathrm{oSNR}\left(\underline{\mathbf{h}}\right)\cdot \xi_{\mathrm{sr}}\left(\underline{\mathbf{h}}\right)} \end{aligned} \tag{4.53}$$

and, obviously,

$$\widetilde{J}\left(\underline{\mathbf{h}}\right) = \mathrm{iSNR}\cdot \overline{J}\left(\underline{\mathbf{h}}\right). \tag{4.54}$$

We are only interested in beamformers for which

$$J_{\mathrm{d}}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right) \leq J_{\mathrm{d}}\left(\underline{\mathbf{h}}\right) < J_{\mathrm{d}}\left(\mathbf{0}_{NL\times 1}\right), \tag{4.55}$$

$$J_{\mathrm{r}}\left(\mathbf{0}_{NL\times 1}\right) < J_{\mathrm{r}}\left(\underline{\mathbf{h}}\right) < J_{\mathrm{r}}\left(\underline{\mathbf{i}}_{\mathrm{i}}\right). \tag{4.56}$$

From the two previous expressions, we deduce that

$$0 \leq \upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}\right) < 1, \tag{4.57}$$

$$1 < \xi_{\mathrm{nr}}\left(\underline{\mathbf{h}}\right) < \infty. \tag{4.58}$$

It is clear that the objective of multichannel noise reduction in the time domain is to find optimal beamformers that would either minimize $J\left(\underline{\mathbf{h}}\right)$ or minimize $J_{\mathrm{d}}\left(\underline{\mathbf{h}}\right)$ or $J_{\mathrm{r}}\left(\underline{\mathbf{h}}\right)$ subject to some constraint.

4.4 Optimal Filtering Vectors

In this section, we derive many well-known time-domain beamformers. Obviously, taking $N = 1$ (single-channel case), we find all the optimal filtering vectors derived in Chap. 2.

4.4.1 Maximum SNR

Let us rewrite the output SNR:

$$\mathrm{oSNR}\left(\underline{\mathbf{h}}\right) = \frac{\sigma_{x_1}^2 \underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \mathbf{R}_{\mathrm{in}} \underline{\mathbf{h}}}. \tag{4.59}$$

The maximum SNR filter, $\underline{\mathbf{h}}_{\max}$, is obtained by maximizing the output SNR as given above. In (4.59), we recognize the generalized Rayleigh quotient [6]. It is well known that this quotient is maximized with the maximum eigenvector of the matrix $\sigma_{x_1}^2 \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T$. Let us denote by $\underline{\lambda}_{\max}$ the maximum eigenvalue corresponding to this maximum eigenvector. Since the rank of the mentioned matrix is equal to 1, we have

$$\begin{aligned} \underline{\lambda}_{\max} &= \text{tr}\left(\sigma_{x_1}^2 \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\right) \\ &= \sigma_{x_1}^2 \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}. \end{aligned} \tag{4.60}$$

As a result,

$$\text{oSNR}\left(\underline{\mathbf{h}}_{\max}\right) = \sigma_{x_1}^2 \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}, \tag{4.61}$$

which corresponds to the maximum possible SNR and

$$\mathcal{A}\left(\underline{\mathbf{h}}_{\max}\right) = \mathcal{A}_{\max}. \tag{4.62}$$

Let us denote by $\mathcal{A}_{\max}^{(n)}$ the maximum array gain of a microphone array with n sensors. By virtue of the inclusion principle [6] for the matrix $\sigma_{x_1}^2 \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T$, we have

$$\mathcal{A}_{\max}^{(N)} \geq \mathcal{A}_{\max}^{(N-1)} \geq \cdots \geq \mathcal{A}_{\max}^{(2)} \geq \mathcal{A}_{\max}^{(1)} \geq 1. \tag{4.63}$$

This shows that by increasing the number of microphones, we necessarily increase the gain.

Obviously, we also have

$$\underline{\mathbf{h}}_{\max} = \varsigma \mathbf{R}_{\text{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}, \tag{4.64}$$

where ς is an arbitrary scaling factor different from zero. While this factor has no effect on the output SNR, it may have on the speech distortion. In fact, all filters (except for the LCMV) derived in the rest of this section are equivalent up to this scaling factor. These filters also try to find the respective scaling factors depending on what we optimize.

4.4.2 Wiener

By minimizing $J\left(\underline{\mathbf{h}}\right)$ with respect to $\underline{\mathbf{h}}$, we find the Wiener filter

$$\underline{\mathbf{h}}_{\text{W}} = \sigma_{x_1}^2 \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}.$$

The Wiener filter can also be expressed as

$$\begin{aligned}\underline{\mathbf{h}}_{\mathrm{W}} &= \mathbf{R}_{\underline{\mathbf{y}}}^{-1} E\left[\underline{\mathbf{x}}(k)x_1(k)\right] \\ &= \mathbf{R}_{\underline{\mathbf{y}}}^{-1}\mathbf{R}_{\underline{\mathbf{x}}}\underline{\mathbf{i}}_{\mathrm{i}} \\ &= \left(\mathbf{I}_{NL} - \mathbf{R}_{\underline{\mathbf{y}}}^{-1}\mathbf{R}_{\underline{\mathbf{v}}}\right)\underline{\mathbf{i}}_{\mathrm{i}},\end{aligned} \tag{4.65}$$

where $\mathbf{I}_{NL}$ is the identity matrix of size $NL \times NL$. The above formulation depends on the second-order statistics of the observation and noise signals. The correlation matrix $\mathbf{R}_{\underline{\mathbf{y}}}$ can be estimated during speech-and-noise periods while the other correlation matrix, $\mathbf{R}_{\underline{\mathbf{v}}}$, can be estimated during noise-only intervals assuming that the statistics of the noise do not change much with time.

Determining the inverse of $\mathbf{R}_{\underline{\mathbf{y}}}$ from

$$\mathbf{R}_{\underline{\mathbf{y}}} = \sigma_{x_1}^2 \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T + \mathbf{R}_{\mathrm{in}} \tag{4.66}$$

with the Woodbury's identity, we get

$$\mathbf{R}_{\underline{\mathbf{y}}}^{-1} = \mathbf{R}_{\mathrm{in}}^{-1} - \frac{\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\mathrm{in}}^{-1}}{\sigma_{x_1}^{-2} + \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}. \tag{4.67}$$

Substituting (4.67) into (4.65) leads to another interesting formulation of the Wiener filter:

$$\underline{\mathbf{h}}_{\mathrm{W}} = \frac{\sigma_{x_1}^2\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}{1 + \sigma_{x_1}^2\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}, \tag{4.68}$$

that we can rewrite as

$$\begin{aligned}\underline{\mathbf{h}}_{\mathrm{W}} &= \frac{\sigma_{x_1}^2\mathbf{R}_{\mathrm{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T}{1 + \underline{\lambda}_{\max}}\underline{\mathbf{i}}_{\mathrm{i}} \\ &= \frac{\mathbf{R}_{\mathrm{in}}^{-1}\left(\mathbf{R}_{\underline{\mathbf{y}}} - \mathbf{R}_{\mathrm{in}}\right)}{1 + \mathrm{tr}\left[\mathbf{R}_{\mathrm{in}}^{-1}\left(\mathbf{R}_{\underline{\mathbf{y}}} - \mathbf{R}_{\mathrm{in}}\right)\right]}\underline{\mathbf{i}}_{\mathrm{i}} \\ &= \frac{\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}} - \mathbf{I}_{NL}}{1 - NL + \mathrm{tr}\left(\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}}\right)}\underline{\mathbf{i}}_{\mathrm{i}}.\end{aligned} \tag{4.69}$$

From (4.69), we deduce that the output SNR is

$$\begin{aligned}\mathrm{oSNR}\left(\underline{\mathbf{h}}_{\mathrm{W}}\right) &= \underline{\lambda}_{\max} \\ &= \mathrm{tr}\left(\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}}\right) - NL.\end{aligned} \tag{4.70}$$

We observe from (4.70) that the more noise, the smaller is the output SNR. However, the more the number of sensors, the higher is the value of $\text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right)$.

The speech distortion index is an explicit function of the output SNR:

$$\upsilon_{\text{sd}}\left(\underline{\mathbf{h}}_{\text{W}}\right) = \frac{1}{\left[1 + \text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right)\right]^2} \leq 1. \tag{4.71}$$

The higher the value of $\text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right)$ or the number of microphones, the less the desired signal is distorted.

Clearly,

$$\text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right) \geq \text{iSNR}, \tag{4.72}$$

since the Wiener filter maximizes the output SNR.

It is of interest to observe that the two filters $\underline{\mathbf{h}}_{\max}$ and $\underline{\mathbf{h}}_{\text{W}}$ are equivalent up to a scaling factor. Indeed, taking

$$\varsigma = \frac{\sigma_{x_1}^2}{1 + \underline{\lambda}_{\max}} \tag{4.73}$$

in (4.64) (maximum SNR filter), we find (4.69) (Wiener filter).

With the Wiener filter, the noise and speech reduction factors are

$$\begin{aligned} \xi_{\text{nr}}\left(\underline{\mathbf{h}}_{\text{W}}\right) &= \frac{\left(1 + \underline{\lambda}_{\max}\right)^2}{\text{iSNR} \cdot \underline{\lambda}_{\max}} \\ &\geq \left(1 + \frac{1}{\underline{\lambda}_{\max}}\right)^2, \end{aligned} \tag{4.74}$$

$$\xi_{\text{sr}}\left(\underline{\mathbf{h}}_{\text{W}}\right) = \left(1 + \frac{1}{\underline{\lambda}_{\max}}\right)^2. \tag{4.75}$$

Finally, we give the minimum NMSEs (MNMSEs):

$$\widetilde{J}\left(\underline{\mathbf{h}}_{\text{W}}\right) = \frac{\text{iSNR}}{1 + \text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right)} \leq 1, \tag{4.76}$$

$$\overline{J}\left(\underline{\mathbf{h}}_{\text{W}}\right) = \frac{1}{1 + \text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right)} \leq 1. \tag{4.77}$$

As the number of microphones increases, the values of these MNMSEs decrease.

4.4.3 MVDR

By minimizing the MSE of the residual interference-plus-noise, $J_{\mathrm{r}}\left(\underline{\mathbf{h}}\right)$, with the constraint that the desired signal is not distorted, i.e.,

$$\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^T \mathbf{R}_{\mathrm{in}} \underline{\mathbf{h}} \quad \text{subject to} \quad \underline{\mathbf{h}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} = 1, \tag{4.78}$$

we find the MVDR filter

$$\underline{\mathbf{h}}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}{\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}, \tag{4.79}$$

that we can rewrite as

$$\begin{aligned} \underline{\mathbf{h}}_{\mathrm{MVDR}} &= \frac{\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\underline{\mathbf{y}}} - \mathbf{I}_{NL}}{\mathrm{tr}\left(\mathbf{R}_{\mathrm{in}}^{-1} \mathbf{R}_{\underline{\mathbf{y}}}\right) - NL} \underline{\mathbf{i}}_{\mathrm{i}} \\ &= \frac{\sigma_{x_1}^2 \mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}{\underline{\lambda}_{\max}}. \end{aligned} \tag{4.80}$$

Alternatively, we can express the MVDR as

$$\underline{\mathbf{h}}_{\mathrm{MVDR}} = \frac{\mathbf{R}_{\underline{\mathbf{y}}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}{\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}. \tag{4.81}$$

The Wiener and MVDR filters are simply related as follows:

$$\underline{\mathbf{h}}_{\mathrm{W}} = \varsigma_0 \underline{\mathbf{h}}_{\mathrm{MVDR}}, \tag{4.82}$$

where

$$\begin{aligned} \varsigma_0 &= \underline{\mathbf{h}}_{\mathrm{W}}^T \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \\ &= \frac{\underline{\lambda}_{\max}}{1 + \underline{\lambda}_{\max}}. \end{aligned} \tag{4.83}$$

So, the two filters $\underline{\mathbf{h}}_{\mathrm{W}}$ and $\underline{\mathbf{h}}_{\mathrm{MVDR}}$ are equivalent up to a scaling factor. However, as explained in Chap. 2, in real-time applications, it is more appropriate to use the MVDR beamformer than the Wiener one.

It is clear that we always have

$$\mathrm{oSNR}\left(\underline{\mathbf{h}}_{\mathrm{MVDR}}\right) = \mathrm{oSNR}\left(\underline{\mathbf{h}}_{\mathrm{W}}\right), \tag{4.84}$$

$$\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{h}}_{\mathrm{MVDR}}\right) = 0, \tag{4.85}$$

$$\xi_{\rm sr}\left(\underline{\mathbf{h}}_{\rm MVDR}\right) = 1, \tag{4.86}$$

$$\xi_{\rm nr}\left(\underline{\mathbf{h}}_{\rm MVDR}\right) = \mathcal{A}\left(\underline{\mathbf{h}}_{\rm MVDR}\right) \leq \xi_{\rm nr}\left(\underline{\mathbf{h}}_{\rm W}\right), \tag{4.87}$$

and

$$1 \geq \widetilde{J}\left(\underline{\mathbf{h}}_{\rm MVDR}\right) = \frac{1}{\mathcal{A}\left(\underline{\mathbf{h}}_{\rm MVDR}\right)} \geq \widetilde{J}\left(\underline{\mathbf{h}}_{\rm W}\right), \tag{4.88}$$

$$\overline{J}\left(\underline{\mathbf{h}}_{\rm MVDR}\right) = \frac{1}{\mathrm{oSNR}\left(\underline{\mathbf{h}}_{\rm MVDR}\right)} \geq \overline{J}\left(\underline{\mathbf{h}}_{\rm W}\right). \tag{4.89}$$

4.4.4 Space–Time Prediction

In the space–time (ST) prediction approach, we find a distortionless filter in two steps [1, 7, 8].

Assume that we can find a simple ST prediction filter $\underline{\mathbf{g}}$ of length NL in such a way that

$$\underline{\mathbf{x}}(k) \approx x_1(k)\underline{\mathbf{g}}. \tag{4.90}$$

The distortionless filter with the ST approach is then obtained by

$$\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{v}}} \underline{\mathbf{h}} \quad \text{subject to} \quad \underline{\mathbf{h}}^T \underline{\mathbf{g}} = 1. \tag{4.91}$$

We deduce the solution

$$\underline{\mathbf{h}}_{\rm ST} = \frac{\mathbf{R}_{\underline{\mathbf{v}}}^{-1}\underline{\mathbf{g}}}{\underline{\mathbf{g}}^T \mathbf{R}_{\underline{\mathbf{v}}}^{-1}\underline{\mathbf{g}}}. \tag{4.92}$$

The second step consist of finding the optimal $\mathbf{g}$ in the Wiener sense. For that, we need to define the error signal vector

$$\underline{\mathbf{e}}_{\rm ST}(k) = \underline{\mathbf{x}}(k) - x_1(k)\underline{\mathbf{g}} \tag{4.93}$$

and form the MSE

$$J\left(\underline{\mathbf{g}}\right) = E\left[\underline{\mathbf{e}}_{\rm ST}^T(k)\underline{\mathbf{e}}_{\rm ST}(k)\right]. \tag{4.94}$$

By minimizing $J\left(\underline{\mathbf{g}}\right)$ with respect to $\underline{\mathbf{g}}$, we easily find the optimal filter

$$\underline{\mathbf{g}}_{\rm o} = \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}. \tag{4.95}$$

It is interesting to observe that the error signal vector with the optimal filter, $\underline{\mathbf{g}}_{\text{o}}$, corresponds to the interference signal, i.e.,

$$\begin{aligned}\underline{\mathbf{e}}_{\text{ST,o}}(k) &= \underline{\mathbf{x}}(k) - x_1(k)\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} \\ &= \underline{\mathbf{x}}_{\text{i}}(k).\end{aligned} \tag{4.96}$$

This result is obviously expected because of the orthogonality principle.

Substituting (4.95) into (4.92), we find that

$$\underline{\mathbf{h}}_{\text{ST}} = \frac{\mathbf{R}_{\underline{\mathbf{v}}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}{\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\underline{\mathbf{v}}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}. \tag{4.97}$$

Comparing $\underline{\mathbf{h}}_{\text{MVDR}}$ with $\underline{\mathbf{h}}_{\text{ST}}$, we see that the latter is an approximation of the former. Indeed, in the ST approach, the interference signal is neglected: instead of using the correlation matrix of the interference-plus-noise, i.e., $\mathbf{R}_{\text{in}}$, only the correlation matrix of the noise is used, i.e., $\mathbf{R}_{\underline{\mathbf{v}}}$. However, identical expressions of the MVDR and ST-prediction filters can be obtained if we consider minimizing the overall mixture energy subject to the no distortion constraint.

4.4.5 Tradeoff

Following the ideas from Chap. 2, we can derive the multichannel tradeoff beamformer, which is given by

$$\begin{aligned}\underline{\mathbf{h}}_{\text{T},\mu} &= \frac{\mathbf{R}_{\text{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}}{\mu\sigma_{x_1}^{-2} + \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}^T\mathbf{R}_{\text{in}}^{-1}\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}} \\ &= \frac{\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}} - \mathbf{I}_{NL}}{\mu - NL + \text{tr}\left(\mathbf{R}_{\text{in}}^{-1}\mathbf{R}_{\underline{\mathbf{y}}}\right)}\underline{\mathbf{i}}_{\text{i}},\end{aligned} \tag{4.98}$$

where $\mu \geq 0$.

We have

$$\text{oSNR}\left(\underline{\mathbf{h}}_{\text{T},\mu}\right) = \text{oSNR}\left(\underline{\mathbf{h}}_{\text{W}}\right), \quad \forall \mu \geq 0, \tag{4.99}$$

$$\upsilon_{\text{sd}}\left(\underline{\mathbf{h}}_{\text{T},\mu}\right) = \left(\frac{\mu}{\mu + \underline{\lambda}_{\max}}\right)^2, \tag{4.100}$$

$$\xi_{\text{sr}}\left(\underline{\mathbf{h}}_{\text{T},\mu}\right) = \left(1 + \frac{\mu}{\underline{\lambda}_{\max}}\right)^2, \tag{4.101}$$

$$\xi_{\text{nr}}\left(\underline{\mathbf{h}}_{\text{T},\mu}\right) = \frac{\left(\mu + \underline{\lambda}_{\max}\right)^2}{\text{iSNR} \cdot \underline{\lambda}_{\max}}, \tag{4.102}$$

and

$$\widetilde{J}\left(\underline{\mathbf{h}}_{\mathrm{T},\mu}\right) = \mathrm{iSNR}\frac{\mu^2 + \underline{\lambda}_{\max}}{\left(\mu + \underline{\lambda}_{\max}\right)^2} \geq \overline{J}\left(\underline{\mathbf{h}}_{\mathrm{W}}\right), \tag{4.103}$$

$$\overline{J}\left(\underline{\mathbf{h}}_{\mathrm{T},\mu}\right) = \frac{\mu^2 + \underline{\lambda}_{\max}}{\left(\mu + \underline{\lambda}_{\max}\right)^2} \geq \overline{J}\left(\underline{\mathbf{h}}_{\mathrm{W}}\right). \tag{4.104}$$

4.4.6 LCMV

We can decompose the noise signal vector, $\underline{\mathbf{v}}(k)$, into two orthogonal vectors:

$$\underline{\mathbf{v}}(k) = \boldsymbol{\rho}_{\underline{\mathbf{v}}v_1} \cdot v_1(k) + \underline{\mathbf{v}}_{\mathrm{u}}(k), \tag{4.105}$$

where $\boldsymbol{\rho}_{\underline{\mathbf{v}}v_1}$ is defined in a similar way to $\boldsymbol{\rho}_{\underline{\mathbf{x}}x_1}$ and $\underline{\mathbf{v}}_{\mathrm{u}}(k)$ is the noise signal vector that is uncorrelated with $v_1(k)$.

In the LCMV beamformer that will be derived in this subsection, we wish to perfectly recover our desired signal, $x_1(k)$, and completely remove the correlated components of the noise signal at the reference microphone, $\boldsymbol{\rho}_{\underline{\mathbf{v}}v_1} \cdot v_1(k)$. Thus, the two constraints can be put together in a matrix form as

$$\mathbf{C}_{x_1 v_1}^T \underline{\mathbf{h}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{4.106}$$

where

$$\mathbf{C}_{x_1 v_1} = \begin{bmatrix} \boldsymbol{\rho}_{\underline{\mathbf{x}}x_1} & \boldsymbol{\rho}_{\underline{\mathbf{v}}v_1} \end{bmatrix} \tag{4.107}$$

is our constraint matrix of size $NL \times 2$. Then, our optimal filter is obtained by minimizing the energy at the filter output, with the constraints that the correlated noise components are cancelled and the desired speech is preserved, i.e.,

$$\underline{\mathbf{h}}_{\mathrm{LCMV}} = \arg\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{y}}} \underline{\mathbf{h}} \quad \text{subject to} \quad \mathbf{C}_{x_1 v_1}^T \underline{\mathbf{h}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{4.108}$$

The solution to (4.108) is given by

$$\underline{\mathbf{h}}_{\mathrm{LCMV}} = \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \mathbf{C}_{x_1 v_1} \left(\mathbf{C}_{x_1 v_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \mathbf{C}_{x_1 v_1}\right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{4.109}$$

The LCMV beamformer can be useful when the noise is mostly coherent.

All beamformers presented in this section can be implemented by estimating the second-order statistics of the noise and observation signals, as in the single-channel case. The statistics of the noise can be estimated during silences with the help of a VAD (see Chap. 2).

4.5 Summary

We started this chapter by explaining the signal model for multichannel noise reduction with an array of N microphones. With this model, we showed how to achieve noise reduction (or beamforming) with a long filtering vector of length NL in order to recover the desired signal sample, which is defined as the convolved speech at microphone 1. We then gave all important performance measures in this context. Finally, we derived the most useful beamforming algorithms. With the proposed framework, we see that the single- and multichannel cases look very similar. This approach simplifies the understanding and analysis of the time-domain noise reduction problem.

References

1. J. Benesty, J. Chen, Y. Huang, *Microphone Array Signal Processing* (Springer, Berlin, 2008)
2. M. Brandstein, D.B. Ward (eds), *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, Berlin, 2001)
3. J.P. Dmochowski, J. Benesty, Microphone arrays: fundamental concepts, in *Speech Processing in Modern Communication—Challenges and Perspectives*, Chap. 8, pp. 199–223, ed. by I. Cohen, J. Benesty, S. Gannot (Springer, Berlin, 2010)
4. L.C. Godara, Application of antenna arrays to mobile communications, part II: beam-forming and direction-of-arrival considerations. Proc. IEEE **85**, 1195–1245 (1997)
5. B.D. Van Veen, K.M. Buckley, Beamforming: a versatile approach to spatial filtering. IEEE Acoust. Speech Signal Process. Mag. **5**, 4–24 (1988)
6. J.N. Franklin, *Matrix Theory* (Prentice-Hall, Englewood Cliffs, 1968)
7. J. Benesty, J. Chen, Y. Huang, A minimum speech distortion multichannel algorithm for noise reduction, in *Proceedings of the IEEE ICASSP*, pp. 321–324 (2008)
8. J. Chen, J. Benesty, Y. Huang, A minimum distortion noise reduction algorithm with multiple microphones. IEEE Trans. Audio Speech Language Process. **16**, 481–493 (2008)

Chapter 5
Multichannel Noise Reduction with a Rectangular Filtering Matrix

In this last chapter, we are going to estimate L samples of the desired signal from NL observations, where N is the number of microphones and L is the number of samples from each microphone signal. This time, a rectangular filtering matrix of size $L \times NL$ is required for the estimation of the desired signal vector. The signal model is the same as in Sect. 4.1; so we start by explaining the principle of multichannel linear filtering with a rectangular matrix.

5.1 Linear Filtering with a Rectangular Matrix

In this chapter, the desired signal is the whole vector $\mathbf{x}_1(k)$ of length L. Therefore, multichannel noise reduction or beamforming is performed by applying a linear transformation to each microphone signal and summing the transformed signals [1, 2]. We have

$$\begin{aligned}\mathbf{z}(k) &= \sum_{n=1}^{N} \mathbf{H}_n \mathbf{y}_n(k) \\ &= \underline{\mathbf{H}}\underline{\mathbf{y}}(k) \\ &= \underline{\mathbf{H}}[\underline{\mathbf{x}}(k) + \underline{\mathbf{v}}(k)],\end{aligned} \tag{5.1}$$

where $\mathbf{z}(k)$ is the estimate of $\mathbf{x}_1(k)$, $\mathbf{H}_n,\ n = 1, 2, \ldots, N$ are N filtering matrices of size $L \times L$, and

$$\underline{\mathbf{H}} = \begin{bmatrix}\mathbf{H}_1 \ \mathbf{H}_2 \ \cdots \ \mathbf{H}_N\end{bmatrix} \tag{5.2}$$

is a rectangular filtering matrix of size $L \times NL$.

Since $\mathbf{x}_1(k)$ is the desired signal vector, we need to extract it from $\underline{\mathbf{x}}(k)$. Specifically, the vector $\underline{\mathbf{x}}(k)$ is decomposed into the following form:

J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters*, SpringerBriefs in Electrical and Computer Engineering, 1,
DOI: 10.1007/978-3-642-19601-0_5, © Jacob Benesty 2011

$$\begin{aligned}\underline{\mathbf{x}}(k) &= \mathbf{R}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}^{-1}\mathbf{x}_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) \\ &= \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \cdot \mathbf{x}_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k),\end{aligned} \tag{5.3}$$

where

$$\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} = \mathbf{R}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}^{-1} \tag{5.4}$$

is the time-domain steering matrix, $\mathbf{R}_{\underline{\mathbf{x}}\mathbf{x}_1} = E\left[\underline{\mathbf{x}}(k)\mathbf{x}_1^T(k)\right]$ is the cross-correlation matrix of size $NL \times L$ between $\underline{\mathbf{x}}(k)$ and $\mathbf{x}_1(k)$, $\mathbf{R}_{\mathbf{x}_1} = E\left[\mathbf{x}_1(k)\mathbf{x}_1^T(k)\right]$ is the correlation matrix of $\mathbf{x}_1(k)$, and $\underline{\mathbf{x}}_{\mathrm{i}}(k)$ is the interference signal vector. It is easy to check that $\underline{\mathbf{x}}_{\mathrm{d}}(k) = \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \cdot \mathbf{x}_1(k)$ and $\underline{\mathbf{x}}_{\mathrm{i}}(k)$ are orthogonal, i.e.,

$$E\left[\underline{\mathbf{x}}_{\mathrm{d}}(k)\underline{\mathbf{x}}_{\mathrm{i}}^T(k)\right] = \mathbf{0}_{NL\times NL}. \tag{5.5}$$

Using (5.3), we can rewrite $\underline{\mathbf{y}}(k)$ as

$$\begin{aligned}\underline{\mathbf{y}}(k) &= \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \cdot \mathbf{x}_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) + \underline{\mathbf{v}}(k) \\ &= \underline{\mathbf{x}}_{\mathrm{d}}(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) + \underline{\mathbf{v}}(k).\end{aligned} \tag{5.6}$$

Substituting (5.3) into (5.1), we get

$$\begin{aligned}\mathbf{z}(k) &= \underline{\mathbf{H}}\left[\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \cdot \mathbf{x}_1(k) + \underline{\mathbf{x}}_{\mathrm{i}}(k) + \underline{\mathbf{v}}(k)\right] \\ &= \mathbf{x}_{\mathrm{fd}}(k) + \mathbf{x}_{\mathrm{ri}}(k) + \mathbf{v}_{\mathrm{rn}}(k),\end{aligned} \tag{5.7}$$

where

$$\mathbf{x}_{\mathrm{fd}}(k) = \underline{\mathbf{H}}\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \cdot \mathbf{x}_1(k) \tag{5.8}$$

is the filtered desired signal vector,

$$\mathbf{x}_{\mathrm{ri}}(k) = \underline{\mathbf{H}}\,\underline{\mathbf{x}}_{\mathrm{i}}(k) \tag{5.9}$$

is the residual interference vector, and

$$\mathbf{v}_{\mathrm{rn}}(k) = \underline{\mathbf{H}}\,\underline{\mathbf{v}}(k) \tag{5.10}$$

is the residual noise vector.

The three terms $\mathbf{x}_{\mathrm{fd}}(k)$, $\mathbf{x}_{\mathrm{ri}}(k)$, and $\mathbf{v}_{\mathrm{rn}}(k)$ are mutually orthogonal; therefore, the correlation matrix of $\mathbf{z}(k)$ is

$$\begin{aligned}\mathbf{R}_{\mathbf{z}} &= E\left[\mathbf{z}(k)\mathbf{z}^T(k)\right] \\ &= \mathbf{R}_{\mathbf{x}_{\mathrm{fd}}} + \mathbf{R}_{\mathbf{x}_{\mathrm{ri}}} + \mathbf{R}_{\mathbf{v}_{\mathrm{rn}}},\end{aligned} \tag{5.11}$$

where

$$\mathbf{R}_{\mathbf{x}_{\mathrm{fd}}} = \underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\underline{\mathbf{H}}^T, \tag{5.12}$$

$$\begin{aligned}\mathbf{R}_{\mathbf{x}_{\mathrm{ri}}} &= \underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{i}}}\underline{\mathbf{H}}^T \\ &= \underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{x}}}\underline{\mathbf{H}}^T - \underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\underline{\mathbf{H}}^T,\end{aligned} \tag{5.13}$$

$$\mathbf{R}_{\mathbf{v}_{\mathrm{rn}}} = \underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{v}}}\underline{\mathbf{H}}^T. \tag{5.14}$$

The correlation matrix of $\mathbf{z}(k)$ is useful in the definitions of the performance measures.

5.2 Joint Diagonalization

The correlation matrix of $\underline{\mathbf{y}}(k)$ is

$$\begin{aligned}\mathbf{R}_{\underline{\mathbf{y}}} &= \mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}} + \mathbf{R}_{\mathrm{in}} \\ &= \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T + \mathbf{R}_{\mathrm{in}},\end{aligned} \tag{5.15}$$

where

$$\mathbf{R}_{\mathrm{in}} = \mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{i}}} + \mathbf{R}_{\underline{\mathbf{v}}} \tag{5.16}$$

is the interference-plus-noise correlation matrix. It is interesting to observe from (5.15) that the noisy signal correlation matrix is the sum of two other correlation matrices: the linear transformation of the desired signal correlation matrix of rank L and the interference-plus-noise correlation matrix of rank NL.

The two symmetric matrices $\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}$ and $\mathbf{R}_{\mathrm{in}}$ can be jointly diagonalized as follows [3, 4]:

$$\underline{\mathbf{B}}^T\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}\underline{\mathbf{B}} = \underline{\boldsymbol{\Lambda}}, \tag{5.17}$$

$$\underline{\mathbf{B}}^T\mathbf{R}_{\mathrm{in}}\underline{\mathbf{B}} = \mathbf{I}_{NL}, \tag{5.18}$$

where $\underline{\mathbf{B}}$ is a full-rank square matrix (of size $NL \times NL$) and $\underline{\boldsymbol{\Lambda}}$ is a diagonal matrix whose main elements are real and nonnegative. Furthermore, $\underline{\boldsymbol{\Lambda}}$ and $\underline{\mathbf{B}}$ are the eigenvalue and eigenvector matrices, respectively, of $\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}$, i.e.,

$$\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}\underline{\mathbf{B}} = \underline{\mathbf{B}}\,\underline{\boldsymbol{\Lambda}}. \tag{5.19}$$

Since the rank of the matrix $\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}$ is equal to L, the eigenvalues of $\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}$ can be ordered as $\underline{\lambda}_1 \geq \underline{\lambda}_2 \geq \cdots \geq \underline{\lambda}_L > \underline{\lambda}_{L+1} = \cdots = \underline{\lambda}_{NL} = 0$. In other words, the last $NL - L$ eigenvalues of $\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{R}_{\underline{\mathbf{x}}_{\mathrm{d}}}$ are exactly zero while its first L eigenvalues are positive, with $\underline{\lambda}_1$ being the maximum eigenvalue. We also denote by $\underline{\mathbf{b}}_1, \underline{\mathbf{b}}_2, \ldots, \underline{\mathbf{b}}_{NL}$, the corresponding eigenvectors. Therefore, the noisy signal covariance matrix can also be diagonalized as

$$\underline{\mathbf{B}}^T \mathbf{R}_{\underline{\mathbf{y}}} \underline{\mathbf{B}} = \underline{\mathbf{\Lambda}} + \mathbf{I}_{NL}. \tag{5.20}$$

This joint diagonalization is very helpful in the analysis of the beamformers for noise reduction.

5.3 Performance Measures

We derive the performance measures in the context of a multichannel linear filtering matrix with microphone 1 as the reference.

5.3.1 Noise Reduction

The input SNR was already defined in Chap. 4 but we can also express it as

$$\text{iSNR} = \frac{\text{tr}\left(\mathbf{R}_{\mathbf{x}_1}\right)}{\text{tr}\left(\mathbf{R}_{\mathbf{v}_1}\right)}, \tag{5.21}$$

where $\mathbf{R}_{\mathbf{v}_1} = E\left[\mathbf{v}_1(k)\mathbf{v}_1^T(k)\right]$.

We define the output SNR as

$$\begin{aligned} \text{oSNR}\left(\underline{\mathbf{H}}\right) &= \frac{\text{tr}\left(\mathbf{R}_{\mathbf{x}_{\text{fd}}}\right)}{\text{tr}\left(\mathbf{R}_{\mathbf{x}_{\text{ri}}} + \mathbf{R}_{\mathbf{v}_{\text{rn}}}\right)} \\ &= \frac{\text{tr}\left(\underline{\mathbf{H}}\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\underline{\mathbf{H}}^T\right)}{\text{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\text{in}}\underline{\mathbf{H}}^T\right)}. \end{aligned} \tag{5.22}$$

This definition is obtained from (5.11). Consequently, the array gain is

$$\mathcal{A}\left(\underline{\mathbf{H}}\right) = \frac{\text{oSNR}\left(\underline{\mathbf{H}}\right)}{\text{iSNR}}. \tag{5.23}$$

For the particular filtering matrix

$$\underline{\mathbf{H}} = \underline{\mathbf{I}}_{\text{i}} = \left[\mathbf{I}_L \ \mathbf{0}_{L\times(N-1)L}\right] \tag{5.24}$$

of size $L \times NL$, called the identity filtering matrix, we have

$$\mathcal{A}\left(\underline{\mathbf{I}}_{\text{i}}\right) = 1 \tag{5.25}$$

and no improvement in gain is possible in this scenario.

The noise reduction factor is

$$\xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}\right) = \frac{\mathrm{tr}\left(\mathbf{R}_{\mathbf{v}_1}\right)}{\mathrm{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\mathrm{in}}\underline{\mathbf{H}}^T\right)}. \tag{5.26}$$

Any good choice of $\underline{\mathbf{H}}$ should lead to $\xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}\right) \geq 1$.

5.3.2 Speech Distortion

We can quantify speech distortion with the speech reduction factor

$$\xi_{\mathrm{sr}}\left(\underline{\mathbf{H}}\right) = \frac{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_1}\right)}{\mathrm{tr}\left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\mathbf{H}^T\right)} \tag{5.27}$$

or with the speech distortion index

$$\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{H}}\right) = \frac{\mathrm{tr}\left[\left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} - \mathbf{I}_L\right)\mathbf{R}_{\mathbf{x}_1}\left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} - \mathbf{I}_L\right)^T\right]}{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_1}\right)}. \tag{5.28}$$

We observe from the two previous expressions that the design of beamformers that do not cancel the desired signal requires the constraint

$$\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} = \mathbf{I}_L. \tag{5.29}$$

In this case $\xi_{\mathrm{sr}}\left(\underline{\mathbf{H}}\right) = 1$ and $\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{H}}\right) = 0$.

It is easy to verify that we have the following fundamental relation:

$$\mathcal{A}\left(\underline{\mathbf{H}}\right) = \frac{\xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}\right)}{\xi_{\mathrm{sr}}\left(\underline{\mathbf{H}}\right)}. \tag{5.30}$$

This expression indicates the equivalence between array gain/loss and distortion.

5.3.3 MSE Criterion

The error signal vector between the estimated and desired signals is

$$\begin{aligned}\mathbf{e}(k) &= \mathbf{z}(k) - \mathbf{x}_1(k) \\ &= \mathbf{x}_{\mathrm{fd}}(k) + \mathbf{x}_{\mathrm{ri}}(k) + \mathbf{v}_{\mathrm{rn}}(k) - \mathbf{x}_1(k),\end{aligned} \tag{5.31}$$

which can also be written as the sum of two orthogonal error signal vectors:

$$\mathbf{e}(k) = \mathbf{e}_\mathrm{d}(k) + \mathbf{e}_\mathrm{r}(k), \tag{5.32}$$

where

$$\begin{aligned}\mathbf{e}_\mathrm{d}(k) &= \mathbf{x}_\mathrm{fd}(k) - \mathbf{x}_1(k) \\ &= \left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} - \mathbf{I}_L\right)\mathbf{x}_1(k)\end{aligned} \tag{5.33}$$

is the signal distortion due to the linear transformation and

$$\begin{aligned}\mathbf{e}_\mathrm{r}(k) &= \mathbf{x}_\mathrm{ri}(k) + \mathbf{v}_\mathrm{rn}(k) \\ &= \underline{\mathbf{H}}\,\underline{\mathbf{x}}_\mathrm{i}(k) + \underline{\mathbf{H}}\,\underline{\mathbf{v}}(k)\end{aligned} \tag{5.34}$$

represents the residual interference-plus-noise.

Having defined the error signal, we can now write the MSE criterion as

$$\begin{aligned}J\left(\underline{\mathbf{H}}\right) &= \mathrm{tr}\left\{E\left[\mathbf{e}(k)\mathbf{e}^T(k)\right]\right\} \\ &= \mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_1}\right) + \mathrm{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{y}}}\underline{\mathbf{H}}^T\right) - 2\mathrm{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{x}}\mathbf{x}_1}\right) \\ &= J_\mathrm{d}\left(\underline{\mathbf{H}}\right) + J_\mathrm{r}\left(\underline{\mathbf{H}}\right),\end{aligned} \tag{5.35}$$

where

$$\begin{aligned}J_\mathrm{d}\left(\underline{\mathbf{H}}\right) &= \mathrm{tr}\left\{E\left[\mathbf{e}_\mathrm{d}(k)\mathbf{e}_\mathrm{d}^T(k)\right]\right\} \\ &= \mathrm{tr}\left[\left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} - \mathbf{I}_L\right)\mathbf{R}_{\mathbf{x}_1}\left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} - \mathbf{I}_L\right)^T\right]\end{aligned} \tag{5.36}$$

and

$$\begin{aligned}J_\mathrm{r}\left(\underline{\mathbf{H}}\right) &= \mathrm{tr}\left\{E\left[\mathbf{e}_\mathrm{r}(k)\mathbf{e}_\mathrm{r}^T(k)\right]\right\} \\ &= \mathrm{tr}\left(\underline{\mathbf{H}}\mathbf{R}_\mathrm{in}\underline{\mathbf{H}}^T\right).\end{aligned} \tag{5.37}$$

For the particular filtering matrices $\underline{\mathbf{H}} = \underline{\mathbf{I}}_\mathrm{i}$ and $\underline{\mathbf{H}} = \mathbf{0}_{L\times NL}$, the MSEs are

$$J\left(\underline{\mathbf{I}}_\mathrm{i}\right) = J_\mathrm{r}\left(\underline{\mathbf{I}}_\mathrm{i}\right) = \mathrm{tr}\left(\mathbf{R}_{\mathbf{v}_1}\right), \tag{5.38}$$

$$J\left(\mathbf{0}_{L\times NL}\right) = J_\mathrm{d}\left(\mathbf{0}_{L\times NL}\right) = \mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_1}\right). \tag{5.39}$$

As a result,

$$\mathrm{iSNR} = \frac{J\left(\mathbf{0}_{L\times NL}\right)}{J\left(\underline{\mathbf{I}}_\mathrm{i}\right)} \tag{5.40}$$

and the NMSEs are

$$\widetilde{J}\left(\underline{\mathbf{H}}\right) = \frac{J\left(\underline{\mathbf{H}}\right)}{J\left(\underline{\mathbf{I}}_{\mathrm{i}}\right)} = \mathrm{iSNR} \cdot \upsilon_{\mathrm{sd}}\left(\underline{\mathbf{H}}\right) + \frac{1}{\xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}\right)}, \tag{5.41}$$

$$\overline{J}\left(\underline{\mathbf{H}}\right) = \frac{J\left(\underline{\mathbf{H}}\right)}{J\left(\mathbf{0}_{L\times NL}\right)} = \upsilon_{\mathrm{sd}}\left(\underline{\mathbf{H}}\right) + \frac{1}{\mathrm{oSNR}\left(\underline{\mathbf{H}}\right) \cdot \xi_{\mathrm{sr}}\left(\underline{\mathbf{H}}\right)}, \tag{5.42}$$

where

$$\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{H}}\right) = \frac{J_{\mathrm{d}}\left(\underline{\mathbf{H}}\right)}{J\left(\mathbf{0}_{L\times NL}\right)}, \tag{5.43}$$

$$\xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}\right) = \frac{J\left(\underline{\mathbf{I}}_{\mathrm{i}}\right)}{J_{\mathrm{r}}\left(\underline{\mathbf{H}}\right)}, \tag{5.44}$$

$$\mathrm{oSNR}\left(\underline{\mathbf{H}}\right) \cdot \xi_{\mathrm{sr}}\left(\underline{\mathbf{H}}\right) = \frac{J\left(\mathbf{0}_{L\times NL}\right)}{J_{\mathrm{r}}\left(\underline{\mathbf{H}}\right)}, \tag{5.45}$$

and

$$\widetilde{J}\left(\underline{\mathbf{H}}\right) = \mathrm{iSNR} \cdot \overline{J}\left(\underline{\mathbf{H}}\right). \tag{5.46}$$

We obtain again fundamental relations between the NMSEs, speech distortion index, noise reduction factor, speech reduction factor, and output SNR.

5.4 Optimal Filtering Matrices

In this section, we derive all obvious time-domain beamformers with a rectangular filtering matrix.

5.4.1 Maximum SNR

We can write the filtering matrix as

$$\underline{\mathbf{H}} = \begin{bmatrix} \underline{\mathbf{h}}_1^T \\ \underline{\mathbf{h}}_2^T \\ \vdots \\ \underline{\mathbf{h}}_L^T \end{bmatrix}, \tag{5.47}$$

where $\underline{\mathbf{h}}_l,\ l = 1, 2, \ldots, L$ are FIR filters of length NL. As a result, the output SNR can be expressed as a function of the $\underline{\mathbf{h}}_l,\ l = 1, 2, \ldots, L$, i.e.,

$$\begin{aligned}\text{oSNR}\left(\underline{\mathbf{H}}\right) &= \frac{\text{tr}\left(\underline{\mathbf{H}}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\underline{\mathbf{H}}^T\right)}{\text{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\text{in}}\underline{\mathbf{H}}^T\right)} \\ &= \frac{\sum_{l=1}^{L}\underline{\mathbf{h}}_l^T\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\underline{\mathbf{h}}_l}{\sum_{l=1}^{L}\underline{\mathbf{h}}_l^T\mathbf{R}_{\text{in}}\underline{\mathbf{h}}_l}.\end{aligned} \tag{5.48}$$

It is then natural to try to maximize this SNR with respect to $\underline{\mathbf{H}}$. Let us first give the following lemma.

Lemma 5.1 *We have*

$$\text{oSNR}\left(\underline{\mathbf{H}}\right) \le \max_l \frac{\underline{\mathbf{h}}_l^T\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\underline{\mathbf{h}}_l}{\underline{\mathbf{h}}_l^T\mathbf{R}_{\text{in}}\underline{\mathbf{h}}_l} = \chi. \tag{5.49}$$

Proof This proof is similar to the one given in Chap. 3

Theorem 5.1 *The maximum SNR filtering matrix is given by*

$$\underline{\mathbf{H}}_{\max} = \begin{bmatrix} \beta_1\underline{\mathbf{b}}_1^T \\ \beta_2\underline{\mathbf{b}}_1^T \\ \vdots \\ \beta_L\underline{\mathbf{b}}_1^T \end{bmatrix}, \tag{5.50}$$

where $\beta_l,\ l = 1, 2, \ldots, L$ are real numbers with at least one of them different from 0. *The corresponding output SNR is*

$$\text{oSNR}\left(\underline{\mathbf{H}}_{\max}\right) = \underline{\lambda}_1. \tag{5.51}$$

We recall that $\underline{\lambda}_1$ is the maximum eigenvalue of the matrix $\mathbf{R}_{\text{in}}^{-1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{R}_{\mathbf{x}_1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T$ and its corresponding eigenvector is $\underline{\mathbf{b}}_1$.

Proof From Lemma 5.1, we know that the output SNR is upper bounded by χ whose maximum value is clearly $\underline{\lambda}_1$. On the other hand, it can be checked from (5.48) that $\text{oSNR}\left(\underline{\mathbf{H}}_{\max}\right) = \underline{\lambda}_1$. Since this output SNR is maximal, $\underline{\mathbf{H}}_{\max}$ is indeed the maximum SNR filter.

Property 5.1 The output SNR with the maximum SNR filtering matrix is always greater than or equal to the input SNR, i.e., $\text{oSNR}\left(\underline{\mathbf{H}}_{\max}\right) \ge \text{iSNR}$.

It is interesting to observe that we have these bounds:

$$0 \le \text{oSNR}\left(\underline{\mathbf{H}}\right) \le \underline{\lambda}_1,\ \forall\ \underline{\mathbf{H}}, \tag{5.52}$$

but, obviously, we are only interested in filtering matrices that can improve the output SNR, i.e., $\text{oSNR}\left(\underline{\mathbf{H}}\right) \ge \text{iSNR}$.

5.4.2 Wiener

If we differentiate the MSE criterion, $J\left(\underline{\mathbf{H}}\right)$, with respect to $\underline{\mathbf{H}}$ and equate the result to zero, we find the Wiener filtering matrix

$$\begin{aligned}\underline{\mathbf{H}}_{\mathrm{W}} &= \mathbf{R}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \\ &= \mathbf{R}_{\mathbf{x}_1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1}.\end{aligned} \tag{5.53}$$

It is easy to verify that $\underline{\mathbf{h}}_{\mathrm{W}}^T$ (see Chap. 4) corresponds to the first line of $\underline{\mathbf{H}}_{\mathrm{W}}$.

The Wiener filtering matrix can be rewritten as

$$\begin{aligned}\underline{\mathbf{H}}_{\mathrm{W}} &= \mathbf{R}_{\mathbf{x}_1\underline{\mathbf{x}}} \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \\ &= \underline{\mathbf{I}}_{\mathrm{i}} \mathbf{R}_{\underline{\mathbf{x}}} \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \\ &= \underline{\mathbf{I}}_{\mathrm{i}} \left(\mathbf{I}_{NL} - \mathbf{R}_{\underline{\mathbf{v}}} \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \right).\end{aligned} \tag{5.54}$$

This matrix depends only on the second-order statistics of the noise and observation signals.

Using the Woodbury's identity, it can be shown that Wiener is also

$$\begin{aligned}\underline{\mathbf{H}}_{\mathrm{W}} &= \left(\mathbf{I}_{NL} + \mathbf{R}_{\mathbf{x}_1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \right)^{-1} \mathbf{R}_{\mathbf{x}_1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\mathrm{in}}^{-1} \\ &= \left(\mathbf{R}_{\mathbf{x}_1}^{-1} + \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \right)^{-1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\mathrm{in}}^{-1}.\end{aligned} \tag{5.55}$$

Another way to express Wiener is

$$\begin{aligned}\underline{\mathbf{H}}_{\mathrm{W}} &= \underline{\mathbf{I}}_{\mathrm{i}} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \mathbf{R}_{\mathbf{x}_1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \\ &= \underline{\mathbf{I}}_{\mathrm{i}} - \underline{\mathbf{I}}_{\mathrm{i}} \mathbf{R}_{\mathrm{in}} \mathbf{R}_{\underline{\mathbf{y}}}^{-1}.\end{aligned} \tag{5.56}$$

Using the joint diagonalization, we can rewrite Wiener as a subspace-type approach:

$$\begin{aligned}\underline{\mathbf{H}}_{\mathrm{W}} &= \underline{\mathbf{I}}_{\mathrm{i}} \underline{\mathbf{B}}^{-T} \underline{\boldsymbol{\Lambda}} \left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL} \right)^{-1} \underline{\mathbf{B}}^T \\ &= \underline{\mathbf{I}}_{\mathrm{i}} \underline{\mathbf{B}}^{-T} \begin{bmatrix} \underline{\boldsymbol{\Sigma}} & \mathbf{0}_{L \times (NL-L)} \\ \mathbf{0}_{(NL-L) \times L} & \mathbf{0}_{(NL-L) \times (NL-L)} \end{bmatrix} \underline{\mathbf{B}}^T \\ &= \underline{\mathbf{T}} \begin{bmatrix} \underline{\boldsymbol{\Sigma}} & \mathbf{0}_{L \times (NL-L)} \\ \mathbf{0}_{(NL-L) \times L} & \mathbf{0}_{(NL-L) \times (NL-L)} \end{bmatrix} \underline{\mathbf{B}}^T,\end{aligned} \tag{5.57}$$

where

$$\underline{\mathbf{T}} = \begin{bmatrix} \underline{\mathbf{t}}_1^T \\ \underline{\mathbf{t}}_2^T \\ \vdots \\ \underline{\mathbf{t}}_L^T \end{bmatrix} = \underline{\mathbf{I}}_{\mathrm{i}} \underline{\mathbf{B}}^{-T} \tag{5.58}$$

and

$$\underline{\boldsymbol{\Sigma}} = \text{diag}\left(\frac{\underline{\lambda}_1}{\underline{\lambda}_1 + 1}, \frac{\underline{\lambda}_2}{\underline{\lambda}_2 + 1}, \ldots, \frac{\underline{\lambda}_L}{\underline{\lambda}_L + 1}\right) \tag{5.59}$$

is an $L \times L$ diagonal matrix. We recall that $\underline{\mathbf{I}}_{\text{i}}$ is the identity filtering matrix (which replicates the reference microphone signal). Expression (5.57) is also

$$\underline{\mathbf{H}}_{\text{W}} = \underline{\mathbf{I}}_{\text{i}}\underline{\mathbf{M}}_{\text{W}}, \tag{5.60}$$

where

$$\underline{\mathbf{M}}_{\text{W}} = \underline{\mathbf{B}}^{-T} \begin{bmatrix} \underline{\boldsymbol{\Sigma}} & \mathbf{0}_{L\times(NL-L)} \\ \mathbf{0}_{(NL-L)\times L} & \mathbf{0}_{(NL-L)\times(NL-L)} \end{bmatrix} \underline{\mathbf{B}}^{T}. \tag{5.61}$$

We see that $\underline{\mathbf{H}}_{\text{W}}$ is the product of two other matrices: the rectangular identity filtering matrix and a square matrix of size $NL \times NL$ whose rank is equal to L.

With the joint diagonalization, the input SNR and output SNR with Wiener are

$$\text{iSNR} = \frac{\text{tr}\left(\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}\,\underline{\mathbf{T}}^T\right)}{\text{tr}\left(\underline{\mathbf{T}}\,\underline{\mathbf{T}}^T\right)}, \tag{5.62}$$

$$\text{oSNR}\left(\underline{\mathbf{H}}_{\text{W}}\right) = \frac{\text{tr}\left[\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}^3\left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL}\right)^{-2}\underline{\mathbf{T}}^T\right]}{\text{tr}\left[\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}^2\left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL}\right)^{-2}\underline{\mathbf{T}}^T\right]}. \tag{5.63}$$

Property 5.2 The output SNR with the Wiener filtering matrix is always greater than or equal to the input SNR, i.e., $\text{oSNR}\left(\underline{\mathbf{H}}_{\text{W}}\right) \geq \text{iSNR}$.

Proof This property can be shown by induction.

Obviously, we have

$$\text{oSNR}\left(\underline{\mathbf{H}}_{\text{W}}\right) \leq \text{oSNR}\left(\underline{\mathbf{H}}_{\max}\right). \tag{5.64}$$

We can easily deduce that

$$\xi_{\text{nr}}\left(\underline{\mathbf{H}}_{\text{W}}\right) = \frac{\text{tr}\left(\underline{\mathbf{T}}\,\underline{\mathbf{T}}^T\right)}{\text{tr}\left[\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}^2\left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL}\right)^{-2}\underline{\mathbf{T}}^T\right]}, \tag{5.65}$$

$$\xi_{\text{sr}}\left(\underline{\mathbf{H}}_{\text{W}}\right) = \frac{\text{tr}\left(\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}\,\underline{\mathbf{T}}^T\right)}{\text{tr}\left[\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}^3\left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL}\right)^{-2}\underline{\mathbf{T}}^T\right]}, \tag{5.66}$$

$$\upsilon_{\text{sd}}\left(\underline{\mathbf{H}}_{\text{W}}\right) = \frac{\text{tr}\left[\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}\left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL}\right)^{-1}\underline{\mathbf{T}}^T\mathbf{R}_{\mathbf{x}_1}^{-1}\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}\left(\underline{\boldsymbol{\Lambda}} + \mathbf{I}_{NL}\right)^{-1}\underline{\mathbf{T}}^T\right]}{\text{tr}\left(\underline{\mathbf{T}}\,\underline{\boldsymbol{\Lambda}}\,\underline{\mathbf{T}}^T\right)}. \tag{5.67}$$

5.4.3 MVDR

The MVDR beamformer is derived from the constrained minimization problem:

$$\min_{\underline{\mathbf{H}}} \operatorname{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\text{in}}\underline{\mathbf{H}}^T\right) \quad \text{subject to} \quad \underline{\mathbf{H}}\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} = \mathbf{I}_L. \tag{5.68}$$

The solution to this optimization is

$$\underline{\mathbf{H}}_{\text{MVDR}} = \left(\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\text{in}}^{-1} \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\right)^{-1} \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\text{in}}^{-1}. \tag{5.69}$$

Obviously, with the MVDR filtering matrix, we have no distortion, i.e.,

$$\xi_{\text{sr}}\left(\underline{\mathbf{H}}_{\text{MVDR}}\right) = 1, \tag{5.70}$$

$$\upsilon_{\text{sd}}\left(\underline{\mathbf{H}}_{\text{MVDR}}\right) = 0. \tag{5.71}$$

Using the Woodbury's identity, it can be shown that the MVDR is also

$$\underline{\mathbf{H}}_{\text{MVDR}} = \left(\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\right)^{-1} \mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1}. \tag{5.72}$$

From (5.72), it is easy to deduce the relationship between the MVDR and Wiener beamformers:

$$\underline{\mathbf{H}}_{\text{MVDR}} = \left(\underline{\mathbf{H}}_{\text{W}}\mathbf{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\right)^{-1} \underline{\mathbf{H}}_{\text{W}}. \tag{5.73}$$

The two are equivalent up to an $L \times L$ filtering matrix.

Property 5.3 The output SNR with the MVDR filtering matrix is always greater than or equal to the input SNR, i.e., $\text{oSNR}\left(\underline{\mathbf{H}}_{\text{MVDR}}\right) \geq \text{iSNR}$.

Proof We can prove this property by induction.

We should have

$$\text{oSNR}\left(\underline{\mathbf{H}}_{\text{MVDR}}\right) \leq \text{oSNR}\left(\underline{\mathbf{H}}_{\text{W}}\right) \leq \text{oSNR}\left(\underline{\mathbf{H}}_{\max}\right). \tag{5.74}$$

5.4.4 Space–Time Prediction

The ST approach tries to find a distortionless filtering matrix (different from $\underline{\mathbf{I}}_{\text{i}}$) in two steps.

First, we assume that we can find an ST filtering matrix $\underline{\mathbf{G}}$ of size $L \times NL$ in such a way that

$$\underline{\mathbf{x}}(k) \approx \underline{\mathbf{G}}^T \mathbf{x}_1(k). \tag{5.75}$$

This filtering matrix extracts from $\underline{\mathbf{x}}(k)$ the correlated components to $\mathbf{x}_1(k)$.

The distortionless filter with the ST approach is then obtained by

$$\min_{\underline{\mathbf{H}}} \operatorname{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{y}}}\underline{\mathbf{H}}^T\right) \quad \text{subject to} \quad \underline{\mathbf{H}}\,\underline{\mathbf{G}}^T = \mathbf{I}_L. \tag{5.76}$$

We deduce the solution

$$\underline{\mathbf{H}}_{\text{ST}} = \left(\underline{\mathbf{G}}\mathbf{R}_{\underline{\mathbf{y}}}^{-1}\underline{\mathbf{G}}^T\right)^{-1}\underline{\mathbf{G}}\mathbf{R}_{\underline{\mathbf{y}}}^{-1}. \tag{5.77}$$

The second step consists of finding the optimal $\underline{\mathbf{G}}$ in the Wiener sense. For that, we need to define the error signal vector

$$\underline{\mathbf{e}}_{\text{ST}}(k) = \underline{\mathbf{x}}(k) - \underline{\mathbf{G}}^T\mathbf{x}_1(k) \tag{5.78}$$

and form the MSE

$$J\left(\underline{\mathbf{G}}\right) = E\left[\underline{\mathbf{e}}_{\text{ST}}^T(k)\underline{\mathbf{e}}_{\text{ST}}(k)\right]. \tag{5.79}$$

By minimizing $J\left(\underline{\mathbf{G}}\right)$ with respect to $\underline{\mathbf{G}}$, we easily find the optimal ST filtering matrix

$$\underline{\mathbf{G}}_{\text{o}} = \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T. \tag{5.80}$$

It is interesting to observe that the error signal vector with the optimal ST filtering matrix corresponds to the interference signal, i.e.,

$$\begin{aligned}\underline{\mathbf{e}}_{\text{ST,o}}(k) &= \underline{\mathbf{x}}(k) - \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\mathbf{x}_1(k) \\ &= \underline{\mathbf{x}}_{\text{i}}(k).\end{aligned} \tag{5.81}$$

This result is obviously expected because of the orthogonality principle.

Substituting (5.80) into (5.77), we finally find that

$$\underline{\mathbf{H}}_{\text{ST}} = \left(\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\mathbf{R}_{\underline{\mathbf{y}}}^{-1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}\right)^{-1}\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T\mathbf{R}_{\underline{\mathbf{y}}}^{-1}. \tag{5.82}$$

Obviously, the two filters $\underline{\mathbf{H}}_{\text{MVDR}}$ and $\underline{\mathbf{H}}_{\text{ST}}$ are strictly equivalent.

5.4.5 Tradeoff

In the tradeoff approach, we minimize the speech distortion index with the constraint that the noise reduction factor is equal to a positive value that is greater than 1, i.e.,

$$\min_{\underline{\mathbf{H}}} J_{\text{d}}\left(\underline{\mathbf{H}}\right) \quad \text{subject to} \quad J_{\text{r}}\left(\underline{\mathbf{H}}\right) = \beta J_{\text{r}}\left(\underline{\mathbf{I}}_{\text{i}}\right), \tag{5.83}$$

where $0 < \beta < 1$ to insure that we get some noise reduction. By using a Lagrange multiplier, $\mu > 0$, to adjoin the constraint to the cost function, we easily deduce the tradeoff filter:

$$\underline{\mathbf{H}}_{\mathrm{T},\mu} = \mathbf{R}_{\mathbf{x}_1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \left(\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \mathbf{R}_{\mathbf{x}_1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T + \mu \mathbf{R}_{\mathrm{in}} \right)^{-1}, \tag{5.84}$$

which can be rewritten, thanks to the Woodbury's identity, as

$$\underline{\mathbf{H}}_{\mathrm{T},\mu} = \left(\mu \mathbf{R}_{\mathbf{x}_1}^{-1} + \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\mathrm{in}}^{-1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \right)^{-1} \boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1}^T \mathbf{R}_{\mathrm{in}}^{-1}, \tag{5.85}$$

where μ satisfies $J_{\mathrm{r}} \left(\underline{\mathbf{H}}_{\mathrm{T},\mu} \right) = \beta J_{\mathrm{r}} \left(\underline{\mathbf{I}}_{\mathrm{i}} \right)$. Usually, μ is chosen in an ad hoc way, so that for

- $\mu = 1$, $\underline{\mathbf{H}}_{\mathrm{T},1} = \underline{\mathbf{H}}_{\mathrm{W}}$, which is the Wiener filtering matrix;
- $\mu = 0$ [from (5.85)], $\underline{\mathbf{H}}_{\mathrm{T},0} = \underline{\mathbf{H}}_{\mathrm{MVDR}}$, which is the MVDR beamformer;
- $\mu > 1$, results in a filtering matrix with low residual noise at the expense of high speech distortion;
- $\mu < 1$, results in a filtering matrix with high residual noise and low speech distortion.

Property 5.4 The output SNR with the tradeoff filtering matrix as given in (5.85) is always greater than or equal to the input SNR, i.e., $\mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{T},\mu} \right) \geq \mathrm{iSNR}$, $\forall \mu \geq 0$.

Proof This property can be shown by induction.

We should have for $\mu \geq 1$,

$$\mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{MVDR}} \right) \leq \mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{W}} \right) \leq \mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{T},\mu} \right) \leq \mathrm{oSNR} \left(\underline{\mathbf{H}}_{\max} \right) \tag{5.86}$$

and for $0 \leq \mu \leq 1$,

$$\mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{MVDR}} \right) \leq \mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{T},\mu} \right) \leq \mathrm{oSNR} \left(\underline{\mathbf{H}}_{\mathrm{W}} \right) \leq \mathrm{oSNR} \left(\underline{\mathbf{H}}_{\max} \right). \tag{5.87}$$

We can write the tradeoff beamformer as a subspace-type approach. Indeed, from (5.84), we get

$$\underline{\mathbf{H}}_{\mathrm{T},\mu} = \underline{\mathbf{T}} \begin{bmatrix} \underline{\boldsymbol{\Sigma}}_{\mu} & \mathbf{0}_{L \times (NL-L)} \\ \mathbf{0}_{(NL-L) \times L} & \mathbf{0}_{(NL-L) \times (NL-L)} \end{bmatrix} \underline{\mathbf{B}}^T, \tag{5.88}$$

where

$$\underline{\boldsymbol{\Sigma}}_{\mu} = \mathrm{diag} \left(\frac{\underline{\lambda}_1}{\underline{\lambda}_1 + \mu}, \frac{\underline{\lambda}_2}{\underline{\lambda}_2 + \mu}, \ldots, \frac{\underline{\lambda}_L}{\underline{\lambda}_L + \mu} \right) \tag{5.89}$$

is an $L \times L$ diagonal matrix. Expression (5.88) is also

$$\underline{\mathbf{H}}_{\mathrm{T},\mu} = \underline{\mathbf{I}}_{\mathrm{i}} \underline{\mathbf{M}}_{\mathrm{T},\mu}, \tag{5.90}$$

where

$$\underline{\mathbf{M}}_{\mathrm{T},\mu} = \underline{\mathbf{B}}^{-T} \begin{bmatrix} \underline{\boldsymbol{\Sigma}}_{\mu} & \mathbf{0}_{L\times(NL-L)} \\ \mathbf{0}_{(NL-L)\times L} & \mathbf{0}_{(NL-L)\times(NL-L)} \end{bmatrix} \underline{\mathbf{B}}^{T}. \tag{5.91}$$

We see that $\underline{\mathbf{H}}_{\mathrm{T},\mu}$ is the product of two other matrices: the rectangular identity filtering matrix and an adjustable square matrix of size $NL \times NL$ whose rank is equal to L. Note that $\underline{\mathbf{H}}_{\mathrm{T},\mu}$ as presented in (5.88) is not, in principle, defined for $\mu = 0$ as this expression was derived from (5.84), which is clearly not defined for this particular case. Although it is possible to have $\mu = 0$ in (5.88), this does not lead to the MVDR.

5.4.6 LCMV

The LCMV beamformer is able to handle as many constraints as we desire.

We can exploit the structure of the noise signal. Indeed, in the proposed LCMV, we will not only perfectly recover the desired signal vector, $\mathbf{x}_1(k)$, but we will also completely remove the noise components at microphones $i = 2, 3, \ldots, N$ that are correlated with the noise signal at microphone 1 [i.e., $v_1(k)$]. Therefore, our constraints are

$$\underline{\mathbf{H}}\mathbf{C}_{\mathbf{x}_1 v_1} = \left[\mathbf{I}_L \ \mathbf{0}_{L\times 1}\right], \tag{5.92}$$

where

$$\mathbf{C}_{\mathbf{x}_1 v_1} = \left[\boldsymbol{\Gamma}_{\underline{\mathbf{x}}\mathbf{x}_1} \ \boldsymbol{\rho}_{\underline{\mathbf{v}}v_1}\right] \tag{5.93}$$

is our constraint matrix of size $NL \times (L+1)$.

Our optimization problem is now

$$\min_{\underline{\mathbf{H}}} \operatorname{tr}\left(\underline{\mathbf{H}}\mathbf{R}_{\underline{\mathbf{y}}}\underline{\mathbf{H}}^T\right) \quad \text{subject to} \quad \underline{\mathbf{H}}\mathbf{C}_{\mathbf{x}_1 v_1} = \left[\mathbf{I}_L \ \mathbf{0}_{L\times 1}\right], \tag{5.94}$$

from which we find the LCMV beamformer

$$\underline{\mathbf{H}}_{\mathrm{LCMV}} = \left[\mathbf{I}_L \ \mathbf{0}_{L\times 1}\right] \left(\mathbf{C}_{\mathbf{x}_1 v_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1} \mathbf{C}_{\mathbf{x}_1 v_1}\right)^{-1} \mathbf{C}_{\mathbf{x}_1 v_1}^T \mathbf{R}_{\underline{\mathbf{y}}}^{-1}. \tag{5.95}$$

Clearly, we always have

$$\mathrm{oSNR}\left(\underline{\mathbf{H}}_{\mathrm{LCMV}}\right) \leq \mathrm{oSNR}\left(\underline{\mathbf{H}}_{\mathrm{MVDR}}\right), \tag{5.96}$$

$$\upsilon_{\mathrm{sd}}\left(\underline{\mathbf{H}}_{\mathrm{LCMV}}\right) = 0, \tag{5.97}$$

$$\xi_{\mathrm{sr}}\left(\underline{\mathbf{H}}_{\mathrm{LCMV}}\right) = 1, \tag{5.98}$$

and

$$\xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}_{\mathrm{LCMV}}\right) \leq \xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}_{\mathrm{MVDR}}\right) \leq \xi_{\mathrm{nr}}\left(\underline{\mathbf{H}}_{\mathrm{W}}\right). \tag{5.99}$$

5.5 Summary

In this chapter, we showed how to derive different noise reduction (or beamforming) algorithms in the time domain with a rectangular filtering matrix. This approach is very general and encompasses all the cases studied in the previous chapters and in the literature. It can be quite powerful and the same ideas can be generalized to dereverberation as well.

References

1. S. Doclo, M. Moonen, GSVD-based optimal filtering for single and multimicrophone speech enhancement. IEEE Trans. Signal Process. **50**, 2230–2244 (2002)
2. J. Benesty, J. Chen, Y. Huang, *Microphone Array Signal Processing* (Springer, Berlin, 2008)
3. S.B. Searle, *Matrix Algebra Useful for Statistics* (Wiley, New York, 1982)
4. G. Strang, *Linear Algebra and Its Applications*, 3rd edn. (Harcourt Brace Jovanonich, Orlando, 1988)

Index