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

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

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, © Jacob Benesty 2011

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

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

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

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

$\sigma_x^2 = E\left[x^2(k)\right]$ 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.

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

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

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

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 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

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

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

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

represents the residual interference-plus-noise.

The mean-square error (MSE) criterion is then

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

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

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

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

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

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

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

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

and, obviously,

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

We are only interested in filters for which

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

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

From the two previous expressions, we deduce that

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

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

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

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

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

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

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, $J\,(\mathbf{h})$ [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

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

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

$\mathbf{R}_\mathbf{y}$ can be estimated from the observation signal while the other correlation matrix, $\mathbf{R}_\mathbf{v}$, 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

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

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

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

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

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

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

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

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

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

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

By minimizing $J\left(\mathbf{g}\right)$ 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

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

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

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 $0 \leq \gamma^2 \leq 1$, $\mathbf{h}_{\mathrm{MVDR}}$ 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)

http://www.springer.com/978-3-642-19600-3

Optimal Time-Domain Noise Reduction Filters
A Theoretical Study
Benesty, J.; Chen, J.
2011, VII, 79 p. 1 illus. in color., Softcover
ISBN: 978-3-642-19600-3