Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 78 (2016) 73–83

www.elsevier.com/locate/specom

Single-channel noise reduction via semi-orthogonal transformations and reduced-rank filtering☆

Wei Zhang[a,*], Jacob Benesty[b], Jingdong Chen[a]

[a] Center of Intelligent Acoustics and Immersive Communications and School of Marine Science and Technology, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, Shaanxi 710072, China

[b] INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, QC H5A 1K6, Canada

Received 10 January 2015; received in revised form 22 November 2015; accepted 2 December 2015

Available online 9 February 2016

Abstract

This paper investigates the problem of single-channel noise reduction in the time domain. The objective is to find a lower dimensional filter that can yield a noise reduction performance as close as possible to or even better than that obtained by the full-rank solution. This is achieved in three steps. First, we transform the observation signal vector sequence, through a semi-orthogonal matrix, into a sequence of transformed signal vectors with a reduced dimension. Second, a reduced-rank filter is applied to get an estimate of the clean speech in the transformed domain. Third, the estimate of the clean speech in the time domain is obtained by an inverse semi-orthogonal transformation. The focus of this paper is on the derivation of semi-orthogonal transformations under certain estimation criteria in the first step and the design of the reduced-rank optimal filters that can be used in the second step. We show how noise reduction using the principle of rank reduction can be cast as an optimal filtering problem, and how different semi-orthogonal transformations affect the noise reduction performance. Simulations are performed under various conditions to validate the deduced filters for noise reduction.

Keywords: Noise reduction; Speech enhancement; Semi-orthogonal transformation; Wiener filter; MVDR filter; Reduced-rank filter.

1. Introduction

The problem of single-channel noise reduction is to recover a clean speech signal of interest from its microphone observations (Benesty and Chen, 2011; Benesty et al., 2009; Loizou, 2007). Due to the importance and broad range of applications, a great deal of efforts have been devoted to this problem over the last decades and many algorithms have been developed e.g., Wiener (1949), Boll (1979), Berouti et al. (1979), Lim and Oppenheim (1979), Ephraim and Malah (1984), Trees and Harry (2001). However, these algorithms achieve noise reduction generally by paying a price of adding speech distortion. One exceptional case is the reduced-rank or subspace method, which has the potential to introduce less distortion if the desired signal correlation matrix is rank deficient and this rank is correctly estimated. This paper is, therefore, devoted to the reduced-rank filtering methods.

The idea of "reduced rank" was first developed in the field of signal estimation (Huffel, 1993; Moor, 1993; Scharf, 1991; Scharf and Tufts, 1987; Tufts and Kumaresan, 1982a, 1982b). It was then applied to the noise reduction problem in the so-called subspace approach (Dendrinos et al., 1991), where the singular value decomposition (SVD) of the noisy data matrix was used to estimate and remove the noise subspace and the estimate of the clean signal was then obtained from the remaining subspace. This approach gained more popularity when Ephraim and Van Trees proposed to decompose the covariance matrix of the noisy observation vector (Ephraim and Trees, 1995). The subspace method was found better than the widely used spectral subtraction (Boll, 1979) for noise reduction in the sense that it has less speech distortion with little music residual noise. Today, the principle has been

☆ This work was supported in part by the NSFC under Grant No. 61425005.

* Corresponding author. Tel.: +86 02988491079.
E-mail addresses: weizhang@mail.nwpu.edu.cn (W. Zhang), benesty@emt.inrs.ca (J. Benesty), jingdongchen@ieee.org (J. Chen).

http://dx.doi.org/10.1016/j.specom.2015.12.007
0167-6393/

studied to deal with not only white (Ephraim and Trees, 1995) but also colored noise (Hu and Loizou, 2003; Huang and Zhao, 1998; 2000; Mittal and Phamdo, 2000; Rezayee and Gazor, 2001). Besides the SVD (Moor, 1993; Scharf, 1991; Scharf and Tufts, 1987; Tufts and Kumaresan, 1982a, 1982b) and the eigenvalue decomposition (EVD) (Ephraim and Trees, 1995; Hu and Loizou, 2003), truncated (Q)SVD (Hansen and Jensen, 1998; Jensen et al., 1995) and triangular decompositions (Hansen and Jensen, 2007) were also investigated in the subspace approach. More recent works on reduced-rank filtering can be found in Hansen and Jensen (2013), Nørholm et al. (2014), Zhang et al. (2014).

This paper is also concerned with the application of reduced-rank principle to noise reduction. But unlike most existing work (e.g., Dendrinos et al., 1991; Ephraim and Trees, 1995; Goldstein et al., 1999; 1998; Scharf, 1991; Scharf and Tufts, 1987), which exploits the structure of either the signal data or covariance matrix to find the signal and noise subspaces, this paper develops a more flexible framework. We choose a semi-orthogonal matrix to do data transformation instead of directly decomposing the subspaces. The semi-orthogonal matrix is not unique, and it can be derived under different criteria. The resulting semi-orthogonal matrices represent the characteristic of both the signal and noise, and thus might be used in various conditions. Another contribution of the paper is the derivation of the optimal filters under the reduced-rank framework.

In this framework, noise reduction is achieved in three steps. We first prefilter the full-length observed vector by a semi-orthogonal matrix, resulting in a reduced-dimension vector. In other words, we apply a linear transformation that transforms the observed data vector to a new coordinate system where the basis are defined by the columns of the semi-orthogonal matrix. This is workable because the dimension of the signal subspace is smaller than that of the observed noisy signal space. The second step is to design an optimal reduced-rank filter and apply this filter to get an estimate of the clean speech in the transformed domain. Note that the optimal filter is matrix-valued and the noisy signal is processed by a vector-by-vector basis. The estimate of the clean speech in the time domain is finally obtained by an inverse semi-orthogonal transformation. We will discuss how to derive different semi-orthogonal transformations under certain estimation criteria and how to design different reduced-rank optimal filters. We will also illustrate the flexibility of this new framework in controlling the compromise between noise reduction and speech distortion.

The rest of the paper is organized as follows. In Section 2, the signal model and problem formulation are presented. Section 3 gives the definition of the semi-orthogonal transformation. Then in Section 4, the principle of linear filtering with a rectangular matrix is discussed. Section 5 presents some performance measures for evaluation and analysis of noise reduction. In Section 6, different optimal filters are derived under a given semi-orthogonal transformation. Different semi-orthogonal transformations are discussed in Section 7. Some simulations are presented in Section 8. Finally, conclusions are drawn in Section 9.

2. Signal model and problem formulation

The noise reduction problem considered in this paper is one of recovering the desired speech signal $x(k)$, k being the discrete-time index, of zero mean from the noisy observation (sensor signal) (Benesty and Chen, 2011; Benesty et al., 2009):

$$y(k) = x(k) + v(k), \tag{1}$$

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.

The signal model given in (1) can be put into a vector form by considering L most recent successive time samples, i.e.,

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k), \tag{2}$$

where

$$\mathbf{y}(k) \triangleq \begin{bmatrix} y(k) & y(k-1) & \cdots & y(k-L+1) \end{bmatrix}^T \tag{3}$$

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

$$\mathbf{R_y} \triangleq E\left[\mathbf{y}(k)\mathbf{y}^T(k)\right] = \mathbf{R_x} + \mathbf{R_v}, \tag{4}$$

where $E[\cdot]$ denotes mathematical expectation, and $\mathbf{R_x} \triangleq E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ and $\mathbf{R_v} \triangleq E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively. The noise correlation matrix, $\mathbf{R_v}$, is assumed to be full rank, i.e., equal to L. Then, the objective of noise reduction in this paper is to find a "good" estimate of the vector $\mathbf{x}(k)$ from the observation signal vector $\mathbf{y}(k)$ in the sense that the additive noise is significantly reduced while the desired signal is not much distorted.

3. Semi-orthogonal transformation

We recall that $\mathbf{x}(k)$ is the desired signal vector that we want to estimate from the observation signal vector, $\mathbf{y}(k)$.

Let

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_0 & \mathbf{t}_1 & \cdots & \mathbf{t}_{P-1} \end{bmatrix} \tag{5}$$

be a semi-orthogonal matrix of size $L \times P$, i.e., $\mathbf{T}^T\mathbf{T} = \mathbf{I}_P$, where $\mathbf{I}_P$ is the $P \times P$ identity matrix and $P \leq L$. We define the transformed desired signal vector of length P as

$$\begin{aligned} \mathbf{x}'(k) &\triangleq \mathbf{T}^T\mathbf{x}(k) \\ &= \begin{bmatrix} x'_0(k) & x'_1(k) & \cdots & x'_{P-1}(k) \end{bmatrix}^T, \end{aligned} \tag{6}$$

where $x'_p(k) = \mathbf{t}_p^T \mathbf{x}(k)$. Now, instead of estimating the vector $\mathbf{x}(k)$ of length L, we will estimate the shorter vector $\mathbf{x}'(k)$ of length P.

In (6), $\mathbf{x}'(k)$ is expressed as a function of $\mathbf{x}(k)$. It is also of interest to write $\mathbf{x}(k)$ as a function of $\mathbf{x}'(k)$. For that, we can decompose $\mathbf{x}(k)$ into two uncorrelated components (Benesty and Chen, 2011):

$$\begin{aligned}\mathbf{x}(k) &= \mathbf{R_x T}\left(\mathbf{T}^T \mathbf{R_x T}\right)^{-1}\mathbf{x}'(k) + \mathbf{x}_\mathrm{u}(k) \\ &= \mathbf{Px}(k) + (\mathbf{I}_L - \mathbf{P})\mathbf{x}(k) \\ &= \mathbf{x}_\mathrm{c}(k) + \mathbf{x}_\mathrm{u}(k),\end{aligned} \tag{7}$$

where $\mathbf{I}_L$ is the $L \times L$ identity matrix, $\mathbf{x}_\mathrm{c}(k)$ and $\mathbf{x}_\mathrm{u}(k)$ are correlated and uncorrelated with $\mathbf{x}'(k)$, respectively,

$$\mathbf{P} \triangleq \mathbf{R_x T}\left(\mathbf{T}^T \mathbf{R_x T}\right)^{-1}\mathbf{T}^T \tag{8}$$

is a projection matrix of rank P, and

$$E\left[\mathbf{x}_\mathrm{c}(k)\mathbf{x}_\mathrm{u}^T(k)\right] = \mathbf{0}_{L\times L}. \tag{9}$$

Ideally, we would like to have $\mathbf{x}_\mathrm{u}(k) = \mathbf{0}_{L\times 1}$ in order to better recover the L desired signal samples from the estimation of the P transformed samples only, but this is not possible in general. The best we can do is to find the semi-orthogonal transformation, $\mathbf{T}$, in such a way that the energy of $\mathbf{x}_\mathrm{u}(k)$ is minimized.

The correlation matrix of $\mathbf{x}_\mathrm{u}(k)$ is

$$\mathbf{R}_{\mathbf{x}_\mathrm{u}} = E\left[\mathbf{x}_\mathrm{u}(k)\mathbf{x}_\mathrm{u}^T(k)\right] = (\mathbf{I}_L - \mathbf{P})\mathbf{R_x}, \tag{10}$$

so the energy of $\mathbf{x}_\mathrm{u}(k)$ is

$$L\sigma_{x_\mathrm{u}}^2 = \mathrm{tr}(\mathbf{R_x}) - \mathrm{tr}(\mathbf{PR_x}), \tag{11}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix. Therefore, the minimization of (11) is equivalent to the maximization of the second term on the right-hand side of (11). The term $\mathrm{tr}(\mathbf{PR_x})$ is maximized when the columns of $\mathbf{T}$, i.e., $\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_{P-1}$, are the eigenvectors corresponding to the P largest eigenvalues, $\lambda_0, \lambda_1, \ldots, \lambda_{P-1}$, of $\mathbf{R_x}$. As a result, (11) simplifies to

$$L\sigma_{x_\mathrm{u}}^2 = \sum_{i=P}^{L-1} \lambda_i, \tag{12}$$

where $\lambda_P, \lambda_{P+1}, \ldots, \lambda_{L-1}$ are the $L-P$ smallest eigenvalues of $\mathbf{R_x}$. In practice, though, some other semi-orthogonal matrices can be used depending on what we want to achieve.

4. Linear filtering with a rectangular matrix

In the general linear filtering approach, we estimate the transformed desired signal vector, $\mathbf{x}'(k)$, by applying a linear transformation to $\mathbf{y}(k)$, i.e.,

$$\begin{aligned}\mathbf{z}'(k) &= \mathbf{H}'\mathbf{y}(k) \\ &= \mathbf{H}'[\mathbf{x}(k) + \mathbf{v}(k)] \\ &= \mathbf{x}'_\mathrm{fd}(k) + \mathbf{v}'_\mathrm{rn}(k),\end{aligned} \tag{13}$$

where $\mathbf{z}'(k)$ is the estimate of $\mathbf{x}'(k)$, $\mathbf{H}'$ is a rectangular filtering matrix of size $P \times L$,

$$\mathbf{x}'_\mathrm{fd}(k) \triangleq \mathbf{H}'\mathbf{x}(k) \tag{14}$$

is the filtered desired signal, and

$$\mathbf{v}'_\mathrm{rn}(k) \triangleq \mathbf{H}'\mathbf{v}(k) \tag{15}$$

is the residual noise. Therefore, the estimate of $\mathbf{x}(k)$ is

$$\begin{aligned}\mathbf{z}(k) &= \mathbf{Tz}'(k) \\ &= \mathbf{TH}'\mathbf{y}(k) \\ &= \mathbf{Hy}(k),\end{aligned} \tag{16}$$

where $\mathbf{H} = \mathbf{TH}'$ is a filtering matrix of size $L \times L$.

We find that the correlation matrix of $\mathbf{z}'(k)$ is

$$\begin{aligned}\mathbf{R}_{\mathbf{z}'} &= \mathbf{H}'\mathbf{R_y}\mathbf{H}'^T \\ &= \mathbf{H}'\mathbf{R_x}\mathbf{H}'^T + \mathbf{H}'\mathbf{R_v}\mathbf{H}'^T.\end{aligned} \tag{17}$$

We deduce that $\mathrm{tr}(\mathbf{R}_{\mathbf{z}'}) = \mathrm{tr}(\mathbf{R_z})$, where $\mathbf{R_z} = \mathbf{HR_y H}^T$ is the correlation matrix of $\mathbf{z}(k)$.

5. Performance measures

In this section, we define two categories of performance measures. The first category evaluates the noise reduction performance while the second one evaluates the signal distortion. We also discuss the very convenient mean-squared error (MSE) criterion and show how it is related to the performance measures.

5.1. Noise reduction

The most important measure in noise reduction is the signal-to-noise ratio (SNR). The input SNR is defined as

$$\mathrm{iSNR} \triangleq \frac{\mathrm{tr}(\mathbf{R_x})}{\mathrm{tr}(\mathbf{R_v})} = \frac{\sigma_x^2}{\sigma_v^2}, \tag{18}$$

where $\sigma_x^2 \triangleq E\left[x^2(k)\right]$ and $\sigma_v^2 \triangleq E\left[v^2(k)\right]$ are the variances of $x(k)$ and $v(k)$, respectively. The output SNR, obtained from (17), helps quantify the SNR after filtering. It is given by

$$\mathrm{oSNR}\left(\mathbf{H}'\right) \triangleq \frac{\mathrm{tr}\left(\mathbf{H}'\mathbf{R_x}\mathbf{H}'^T\right)}{\mathrm{tr}\left(\mathbf{H}'\mathbf{R_v}\mathbf{H}'^T\right)} = \mathrm{oSNR}(\mathbf{H}). \tag{19}$$

The objective of noise reduction is to find an appropriate $\mathbf{H}'$ that will make the output SNR greater than the input SNR. Consequently, the quality of the noisy signal may be enhanced.

The noise reduction factor quantifies the amount of noise being rejected by $\mathbf{H}'$. This quantity is defined as the ratio of the power of the noise at the sensor over the power of the noise remaining after filtering, i.e.,

$$\xi_\mathrm{nr}\left(\mathbf{H}'\right) \triangleq \frac{\mathrm{tr}(\mathbf{R_v})}{\mathrm{tr}\left(\mathbf{H}'\mathbf{R_v}\mathbf{H}'^T\right)} = \xi_\mathrm{nr}(\mathbf{H}). \tag{20}$$

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

5.2. Speech distortion

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

$$\xi_{\rm sr}\left(\mathbf{H}'\right) \triangleq \frac{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}}\right)}{\mathrm{tr}\left(\mathbf{H}'\mathbf{R}_{\mathbf{x}}\mathbf{H}'^{T}\right)} = \xi_{\rm sr}\left(\mathbf{H}\right). \tag{21}$$

We should have $\xi_{\rm sr}(\mathbf{H}') \geq 1$.

A rectangular filtering matrix that does not affect the transformed desired signal vector, $\mathbf{x}'(k)$, requires the constraint:

$$\mathbf{H}'\mathbf{R}_{\mathbf{x}}\mathbf{T}\left(\mathbf{T}^{T}\mathbf{R}_{\mathbf{x}}\mathbf{T}\right)^{-1} = \mathbf{H}'\mathbf{P}\mathbf{T} = \mathbf{I}_{P}. \tag{22}$$

By making the appropriate substitutions, one can derive the relationship among the measures defined so far, i.e.,

$$\frac{\mathrm{oSNR}\left(\mathbf{H}'\right)}{\mathrm{iSNR}} = \frac{\xi_{\rm nr}\left(\mathbf{H}'\right)}{\xi_{\rm sr}\left(\mathbf{H}'\right)}. \tag{23}$$

Another way to measure the distortion of the transformed desired signal due to the filtering operation is via the speech distortion index defined as

$$\begin{aligned}\upsilon_{\rm sd}\left(\mathbf{H}'\right) &= \frac{E\left\{\left[\mathbf{x}'_{\rm fd}(k) - \mathbf{x}'(k)\right]^{T}\left[\mathbf{x}'_{\rm fd}(k) - \mathbf{x}'(k)\right]\right\}}{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}}\right)} \\ &= \upsilon_{\rm sd}\left(\mathbf{H}\right).\end{aligned} \tag{24}$$

The speech distortion index is always greater than or equal to 0; the higher is the value of $\upsilon_{\rm sd}\left(\mathbf{H}'\right)$, the more the transformed desired signal is distorted.

Besides the above performance measures, the perceptual evaluation of speech quality (PESQ) (ITU) is also used in our simulations and experiments.

5.3. MSE criterion

The transformed desired signal is a vector of length P, and so is the error signal. We define the error signal vector between the estimated and desired signals as

$$\begin{aligned}\mathbf{e}'(k) &\triangleq \mathbf{z}'(k) - \mathbf{x}'(k) \\ &= \mathbf{H}'\mathbf{y}(k) - \mathbf{x}'(k),\end{aligned} \tag{25}$$

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

$$\mathbf{e}'(k) = \mathbf{e}'_{\rm ds}(k) + \mathbf{e}'_{\rm rs}(k), \tag{26}$$

where

$$\begin{aligned}\mathbf{e}'_{\rm ds}(k) &\triangleq \mathbf{x}'_{\rm fd}(k) - \mathbf{x}'(k) \\ &= \left(\mathbf{H}' - \mathbf{T}^{T}\right)\mathbf{x}(k)\end{aligned} \tag{27}$$

is the signal distortion due to the rectangular filtering matrix and

$$\mathbf{e}'_{\rm rs}(k) \triangleq \mathbf{v}'_{\rm rn}(k) = \mathbf{H}'\mathbf{v}(k) \tag{28}$$

represents the residual noise. Therefore, the MSE criterion is

$$\begin{aligned}J\left(\mathbf{H}'\right) &\triangleq \mathrm{tr}\left\{E\left[\mathbf{e}'(k)\mathbf{e}'^{T}(k)\right]\right\} \\ &= \mathrm{tr}\left(\mathbf{T}^{T}\mathbf{R}_{\mathbf{x}}\mathbf{T}\right) + \mathrm{tr}\left(\mathbf{H}'\mathbf{R}_{\mathbf{y}}\mathbf{H}'^{T}\right) - 2\mathrm{tr}\left(\mathbf{H}'\mathbf{R}_{\mathbf{x}}\mathbf{T}\right).\end{aligned} \tag{29}$$

Using the fact that $E\left[\mathbf{e}'_{\rm ds}(k)\mathbf{e}'^{T}_{\rm rs}(k)\right] = \mathbf{0}_{P\times P}$, $J(\mathbf{H}')$ can be expressed as the sum of two other MSEs, i.e.,

$$\begin{aligned}J\left(\mathbf{H}'\right) &= \mathrm{tr}\left\{E\left[\mathbf{e}'_{\rm ds}(k)\mathbf{e}'^{T}_{\rm ds}(k)\right]\right\} + \mathrm{tr}\left\{E\left[\mathbf{e}'_{\rm rs}(k)\mathbf{e}'^{T}_{\rm rs}(k)\right]\right\} \\ &= J_{\rm ds}\left(\mathbf{H}'\right) + J_{\rm rs}\left(\mathbf{H}'\right),\end{aligned} \tag{30}$$

where

$$\begin{aligned}J_{\rm ds}\left(\mathbf{H}'\right) &\triangleq \mathrm{tr}\left[\left(\mathbf{H}' - \mathbf{T}^{T}\right)\mathbf{R}_{\mathbf{x}}\left(\mathbf{H}' - \mathbf{T}^{T}\right)^{T}\right] \\ &= \mathrm{tr}\left(\mathbf{R}_{\mathbf{x}}\right)\upsilon_{\rm sd}\left(\mathbf{H}'\right)\end{aligned} \tag{31}$$

and

$$\begin{aligned}J_{\rm rs}\left(\mathbf{H}'\right) &\triangleq \mathrm{tr}\left(\mathbf{H}'\mathbf{R}_{\mathbf{v}}\mathbf{H}'^{T}\right) \\ &= \frac{\mathrm{tr}\left(\mathbf{R}_{\mathbf{v}}\right)}{\xi_{\rm nr}\left(\mathbf{H}'\right)}.\end{aligned} \tag{32}$$

We deduce that

$$\begin{aligned}\frac{J_{\rm ds}\left(\mathbf{H}'\right)}{J_{\rm rs}\left(\mathbf{H}'\right)} &= \mathrm{iSNR} \times \xi_{\rm nr}\left(\mathbf{H}'\right) \times \upsilon_{\rm sd}\left(\mathbf{H}'\right) \\ &= \mathrm{oSNR}\left(\mathbf{H}'\right) \times \xi_{\rm sr}\left(\mathbf{H}'\right) \times \upsilon_{\rm sd}\left(\mathbf{H}'\right).\end{aligned} \tag{33}$$

We observe how the MSEs are related to the performance measures.

6. Optimal rectangular filtering matrices

In this section, we briefly discuss the most important optimal rectangular filtering matrices for noise reduction, which explicitly depend on the semi-orthogonal matrix $\mathbf{T}$.

6.1. Maximum SNR

It can be shown that the maximum SNR filtering matrix is given by Benesty and Chen (2011)

$$\mathbf{H}'_{\max} = \begin{bmatrix} \varsigma_0\mathbf{b}^{T}_{\max} \\ \varsigma_1\mathbf{b}^{T}_{\max} \\ \vdots \\ \varsigma_{P-1}\mathbf{b}^{T}_{\max} \end{bmatrix}, \tag{34}$$

where ς_p, $p = 0, 1, \ldots, P-1$ are arbitrary real numbers with at least one of them different from 0 and $\mathbf{b}_{\max}$ is the eigenvector corresponding to the maximum eigenvalue, $\lambda_{\max}$, of the matrix $\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{R}_{\mathbf{x}}$. Furthermore, we have

$$\mathrm{oSNR}\left(\mathbf{H}'_{\max}\right) = \lambda_{\max} \geq \mathrm{iSNR} \tag{35}$$

and

$$0 \leq \mathrm{oSNR}\left(\mathbf{H}'\right) \leq \mathrm{oSNR}\left(\mathbf{H}'_{\max}\right), \forall\mathbf{H}'. \tag{36}$$

The ς_p's are found in such a way that distortion of the desired signal is minimized. We obtain

$$\varsigma_p = \frac{\mathbf{b}^{T}_{\max}\mathbf{R}_{\mathbf{x}}\mathbf{t}_p}{\lambda_{\max}}, \quad p = 0, 1, \ldots, P-1. \tag{37}$$

Substituting these optimal values into (34), we obtain the optimal maximum SNR filtering matrix with minimum distortion to the transformed desired signal:

$$\mathbf{H}'_{\max} = \mathbf{T}^T \mathbf{R}_{\mathbf{x}} \frac{\mathbf{b}_{\max}\mathbf{b}_{\max}^T}{\lambda_{\max}}. \tag{38}$$

We also deduce that the maximum SNR filtering matrix for the estimation of $\mathbf{x}(k)$ is

$$\mathbf{H}_{\max} = \mathbf{T}\mathbf{T}^T \mathbf{R}_{\mathbf{x}} \frac{\mathbf{b}_{\max}\mathbf{b}_{\max}^T}{\lambda_{\max}}. \tag{39}$$

6.2. 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 rectangular filtering matrix:

$$\begin{aligned}\mathbf{H}'_{\mathrm{W}} &= \mathbf{T}^T \mathbf{R}_{\mathbf{x}} \mathbf{R}_{\mathbf{y}}^{-1} \\ &= \mathbf{T}^T \left(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}}\mathbf{R}_{\mathbf{y}}^{-1}\right).\end{aligned} \tag{40}$$

We deduce that the Wiener square filtering matrix for the estimation of the vector $\mathbf{x}(k)$ is

$$\begin{aligned}\mathbf{H}_{\mathrm{W}} &= \mathbf{T}\mathbf{H}'_{\mathrm{W}} \\ &= \mathbf{T}\mathbf{T}^T \left(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}}\mathbf{R}_{\mathbf{y}}^{-1}\right),\end{aligned} \tag{41}$$

which does not correspond, in general, to the classical Wiener filtering matrix (Benesty et al., 2009). But for $P = L$, (41) becomes the well-known Wiener filter:

$$\mathbf{H}_{\mathrm{W}} = \mathbf{I}_L - \mathbf{R}_{\mathbf{v}}\mathbf{R}_{\mathbf{y}}^{-1}. \tag{42}$$

Obviously, we have

$$\mathrm{oSNR}\left(\mathbf{H}'_{\mathrm{W}}\right) \le \mathrm{oSNR}\left(\mathbf{H}'_{\max}\right) \tag{43}$$

and

$$\upsilon_{\mathrm{sd}}\left(\mathbf{H}'_{\mathrm{W}}\right) \le \upsilon_{\mathrm{sd}}\left(\mathbf{H}'_{\max}\right). \tag{44}$$

6.3. Minimum variance distortionless response

It is possible to derive the minimum variance distortionless response (MVDR) filter [distortionless in the sense that $\mathbf{x}'(k)$ is left intact in the filtering process] by minimizing the MSE of the residual noise, $J_{\mathrm{rs}}(\mathbf{H}')$, with the constraint that the transformed desired signal, $\mathbf{x}'(k)$, is not distorted, see (22). Mathematically, this is equivalent to

$$\min_{\mathbf{H}'} \mathrm{tr}\left(\mathbf{H}'\mathbf{R}_{\mathbf{v}}\mathbf{H}'^T\right) \quad \text{subject to} \quad \mathbf{H}'\mathbf{P}\mathbf{T} = \mathbf{I}_P. \tag{45}$$

The solution to the above optimization problem is

$$\mathbf{H}'_{\mathrm{MVDR}} = \left(\mathbf{T}^T\mathbf{P}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{P}\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{P}^T\mathbf{R}_{\mathbf{v}}^{-1}. \tag{46}$$

We deduce that the MVDR for the estimation of $\mathbf{x}(k)$ is

$$\mathbf{H}_{\mathrm{MVDR}} = \mathbf{T}\left(\mathbf{T}^T\mathbf{P}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{P}\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{P}^T\mathbf{R}_{\mathbf{v}}^{-1}. \tag{47}$$

Of course, for $P = L$, the MVDR filtering matrix degenerates to the identity matrix, i.e., $\mathbf{H}_{\mathrm{MVDR}} = \mathbf{I}_L$.

We should have

$$\mathrm{oSNR}\left(\mathbf{H}'_{\mathrm{MVDR}}\right) \le \mathrm{oSNR}\left(\mathbf{H}'_{\mathrm{W}}\right) \tag{48}$$

and

$$\upsilon_{\mathrm{sd}}\left(\mathbf{H}'_{\mathrm{MVDR}}\right) \le \upsilon_{\mathrm{sd}}\left(\mathbf{H}'_{\mathrm{W}}\right). \tag{49}$$

Another possible MVDR filtering matrix for the estimation of $\mathbf{x}(k)$ can be derived through

$$\min_{\mathbf{H}'} \mathrm{tr}\left(\mathbf{H}'\mathbf{R}_{\mathbf{y}}\mathbf{H}'^T\right) \quad \text{subject to} \quad \mathbf{H}'\mathbf{P}\mathbf{T} = \mathbf{I}_P, \tag{50}$$

and the corresponding solution is

$$\mathbf{H}_{\mathrm{MVDR},2} = \mathbf{T}\left(\mathbf{T}^T\mathbf{P}^T\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{P}\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{P}^T\mathbf{R}_{\mathbf{y}}^{-1}. \tag{51}$$

6.4. 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

$$\min_{\mathbf{H}'} J_{\mathrm{ds}}\left(\mathbf{H}'\right) \quad \text{subject to} \quad J_{\mathrm{rs}}\left(\mathbf{H}'\right) = \beta\mathrm{tr}\left(\mathbf{R}_{\mathbf{v}}\right), \tag{52}$$

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 filtering matrix:

$$\mathbf{H}'_{\mathrm{T},\mu} = \mathbf{T}^T\mathbf{R}_{\mathbf{x}}\left(\mathbf{R}_{\mathbf{x}} + \mu\mathbf{R}_{\mathbf{v}}\right)^{-1}, \tag{53}$$

which can be rewritten, for the estimation of $\mathbf{x}(k)$, as

$$\mathbf{H}_{\mathrm{T},\mu} = \mathbf{T}\mathbf{T}^T\mathbf{R}_{\mathbf{x}}\left(\mathbf{R}_{\mathbf{x}} + \mu\mathbf{R}_{\mathbf{v}}\right)^{-1}. \tag{54}$$

For $P = L$, (54) degenerates to the classical tradeoff filter:

$$\mathbf{H}_{\mathrm{T},\mu} = \mathbf{R}_{\mathbf{x}}\left(\mathbf{R}_{\mathbf{x}} + \mu\mathbf{R}_{\mathbf{v}}\right)^{-1}. \tag{55}$$

Usually, μ is chosen in a heuristic way, so that for

- $\mu = 1$, we have $\mathbf{H}'_{\mathrm{T},1} = \mathbf{H}'_{\mathrm{W}}$, which is the Wiener filtering matrix;
- $\mu > 1$, we obtain a filtering matrix with low residual noise at the expense of high transformed desired signal distortion (as compared to Wiener), and
- $\mu < 1$, we obtain a filtering matrix with high residual noise and low transformed desired signal distortion (as compared to Wiener).

7. Examples of semi-orthogonal matrices

In Section 3, we gave an example for the choice of $\mathbf{T}$. In this section, we show some other important possibilities depending on what we want to achieve.

7.1. Minimum MSE

Let us take the Wiener filtering matrix [Eq. (40)]. The minimum MSE (MMSE) is

$$J\left(\mathbf{H}_{\mathrm{W}}'\right) = \mathrm{tr}\left(\mathbf{T}^T\mathbf{R}_{\mathbf{x}}\mathbf{T}\right) - \mathrm{tr}\left(\mathbf{T}^T\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}\mathbf{T}\right). \tag{56}$$

This MMSE, of course, depends on $\mathbf{T}$. It is easy to check that the smallest value for (56) is obtained when the columns of $\mathbf{T}$ are the eigenvectors corresponding to the P largest eigenvalues of $\mathbf{R}_1 = \mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$. We denote this semi-orthogonal matrix by $\mathbf{T}_1$. In this case, we obtain the well-known reduced-rank Wiener filtering matrix (Goldstein and Reed, 1997; Scharf, 1991; Scharf and Tufts, 1987):

$$\mathbf{H}_{\mathrm{W}} = \mathbf{T}_1\mathbf{T}_1^T\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}. \tag{57}$$

We also deduce the other reduced-rank optimal filtering matrices:

$$\mathbf{H}_{\max} = \mathbf{T}_1\mathbf{T}_1^T\mathbf{R}_{\mathbf{x}}\frac{\mathbf{b}_{\max}\mathbf{b}_{\max}^T}{\lambda_{\max}}, \tag{58}$$

$$\mathbf{H}_{\mathrm{MVDR}} = \mathbf{T}_1\left(\mathbf{T}_1^T\mathbf{P}^T\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{P}\mathbf{T}_1\right)^{-1}\mathbf{T}_1^T\mathbf{P}^T\mathbf{R}_{\mathbf{v}}^{-1}, \tag{59}$$

and

$$\mathbf{H}_{\mathrm{T},\mu} = \mathbf{T}_1\mathbf{T}_1^T\mathbf{R}_{\mathbf{x}}(\mathbf{R}_{\mathbf{x}} + \mu\mathbf{R}_{\mathbf{v}})^{-1}. \tag{60}$$

7.2. Minimum distortion

The distortion-based MSE with Wiener is

$$J_{\mathrm{ds}}\left(\mathbf{H}_{\mathrm{W}}'\right) = \mathrm{tr}\left(\mathbf{T}^T\mathbf{R}_{\mathbf{x}}\mathbf{T}\right) - \mathrm{tr}\left(\mathbf{T}^T\mathbf{R}_2\mathbf{T}\right), \tag{61}$$

where

$$\mathbf{R}_2 = 2\mathbf{R}_1 - \mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_1. \tag{62}$$

One can verify that the smallest value for (61) is obtained when the columns of $\mathbf{T}$ are the eigenvectors corresponding to the P largest eigenvalues of $\mathbf{R}_2$. Let us denote by $\mathbf{T}_2$ this semi-orthogonal matrix. By simply replacing $\mathbf{T}_2$ by $\mathbf{T}_1$ in (57)–(60), we obtain the new optimal filtering matrices.

7.3. Minimum residual noise

Let us consider, again, the Wiener filtering matrix. The residual noise corresponding to this optimal matrix is

$$J_{\mathrm{rs}}\left(\mathbf{H}_{\mathrm{W}}'\right) = \mathrm{tr}\left(\mathbf{T}^T\mathbf{R}_3\mathbf{T}\right), \tag{63}$$

where

$$\mathbf{R}_3 = \mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{v}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}. \tag{64}$$

The smallest value for (63) is found when the columns of $\mathbf{T}$ are the eigenvectors corresponding to the P smallest eigenvalues of $\mathbf{R}_3$. If we denote by $\mathbf{T}_3$ this semi-orthogonal matrix and replacing $\mathbf{T}_3$ by $\mathbf{T}_1$ in (57)–(60), we obtain the new optimal filtering matrices. This approach may lead to large distortions.

8. Simulations and experiments

In this section, we study, using simulations and experiments, the impact of some important parameters on the noise reduction performance of the optimal filters derived in Section 6.

8.1. Clean speech signal and noise

The clean speech signal used in most of our simulations is recorded from a female speaker in a quiet office room. It is sampled at 8 kHz. The overall length of the signal is approximately 30-s long. The noisy signal is generated by adding noise to the desired signal, where noise signal is properly scaled to control the input SNR. We study three different types of noise, i.e., white Gaussian, car, and babble noise, which are representative noise samples from white and stationary to highly nonstationary. The car noise is recorded from a sedan car running at 50 miles/hour on a highway with all the windows closed; this noise is still close to stationary, but it is colored. The babble noise is recorded in a New York Stock Exchange (NYSE) room; it consists of sounds from various sources such as speakers, telephone rings, electric fans, etc. and is highly nonstationary.

We also present a set of experiments with real recorded signals where a pre-recorded speech signal is played back through a loudspeaker in a noisy but light reverberant room and we use a microphone to record the noisy signal.

8.2. Estimation of the signal statistics

The implementation of all the noise reduction filters derived in the previous sections requires the estimation of the correlation matrices $\mathbf{R}_{\mathbf{y}}$, $\mathbf{R}_{\mathbf{x}}$, and $\mathbf{R}_{\mathbf{v}}$. In most of our simulations, we compute the $\mathbf{R}_{\mathbf{y}}$ and $\mathbf{R}_{\mathbf{v}}$ matrices using the most recent 320 samples (40-ms long) of the noisy and noise signals, respectively, with a short-time average at every time instant k. An estimate of the $\mathbf{R}_{\mathbf{x}}$ matrix is then obtained by subtracting $\mathbf{R}_{\mathbf{v}}$ from $\mathbf{R}_{\mathbf{y}}$. To ensure that the $\mathbf{R}_{\mathbf{x}}$ estimate is positive semi-definite, we replace all the negative eigenvalues of this estimated matrix with a very small threshold.

8.3. Noise estimation

In some of our simulations, we assume that the noise signal is accessible and its correlation matrix is computed with a short time average. This provides a fair way to compare the performance of different noise reduction filters. In reality, however, the noise is not accessible and has to be estimated from the noisy signal. So, in some simulations and experiments, we also adopt the improved minima controlled recursive averaging (IMCRA) approach (Cohen, 2003), which has been proved to be a robust noise estimator in a broad range of noise environments.

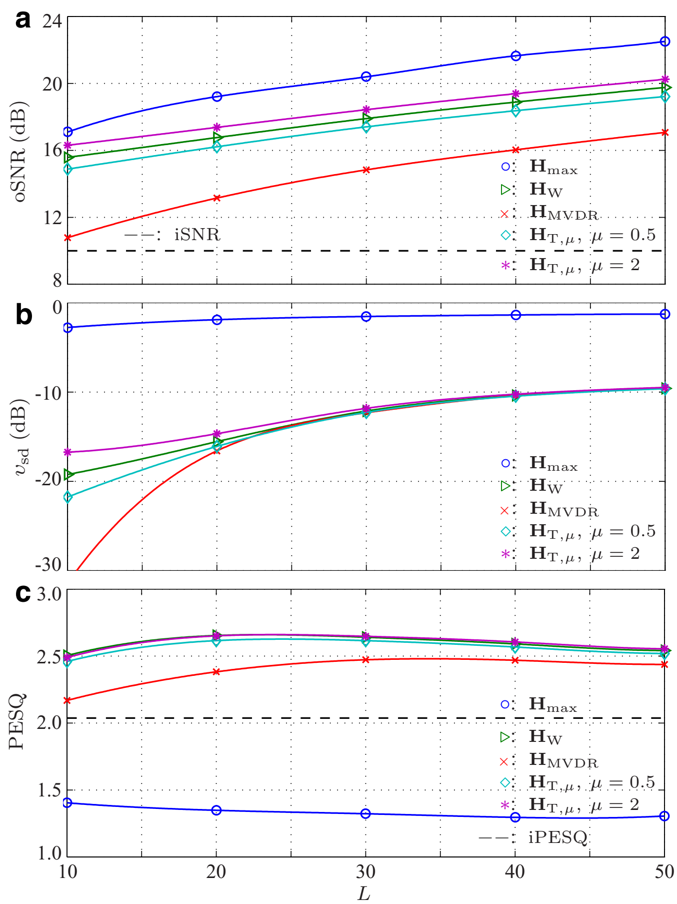

Fig. 1. Performance of the maximum SNR, Wiener, MVDR, and the tradeoff filters as a function of the filter length, L, in white Gaussian noise: (a) output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: iSNR = 10˜dB and $P = 10$.

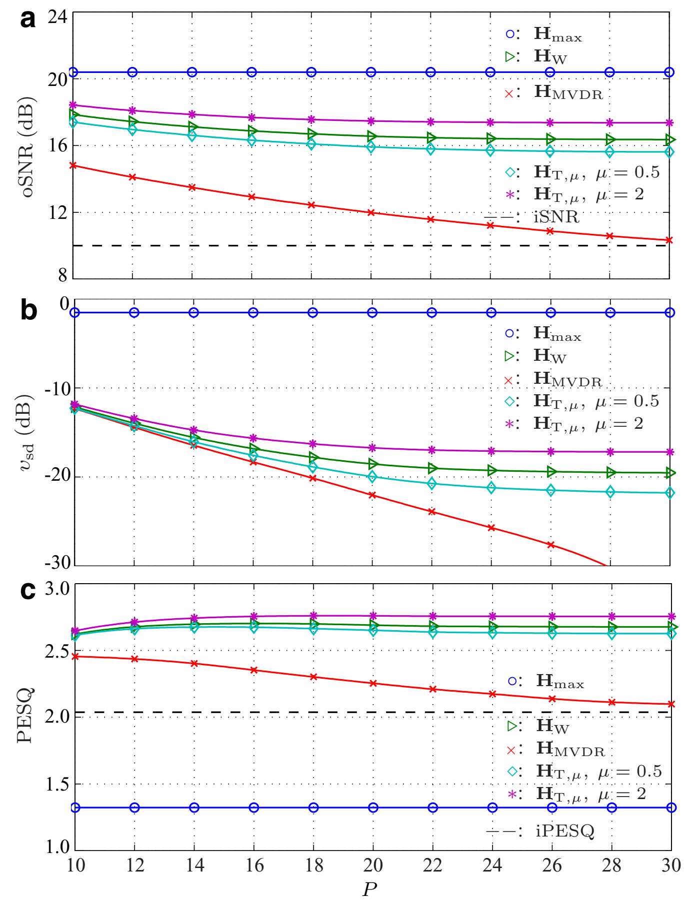

Fig. 2. Performance of the maximum SNR, Wiener, MVDR, and tradeoff filters as a function of the rank parameter, P, in white Gaussian noise: (a) output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: iSNR = 10 dB and $L = 30$.

8.4. Performance of optimal rectangular filtering matrices

In this section, we present the evaluation of the noise reduction filters with three performance measures: the output SNR defined in (19), the speech distortion index defined in (24), and the PESQ (ITU). The semi-orthogonal matrix $\mathbf{T}$ used here is obtained by minimizing the energy of the signal that is uncorrelated with the transformed signal, and its columns consists of the eigenvectors corresponding to the P largest eigenvalues of $\mathbf{R}_{\mathbf{x}}$.

Fig. 1 plots the simulation results of the noise reduction filters as a function of the filter length, L. The background noise is white Gaussian, the input SNR is 10 dB, P is set to 10, and the noise statistics are computed directly from the noise signal. It is seen that all filters can improve the SNR with a similar degree of speech distortion except the for maximum SNR filter. The output SNR and the speech distortion index both increase with the filter length, L. The maximum SNR filter expectedly has overwhelming superiority in output SNR, as its name indicates; but it also introduces tremendous speech distortion. The tradeoff filter with $\mu = 2$ achieves higher output SNR and speech distortion than the Wiener filter, while that with $\mu = 0.5$ behaves the opposite, which is in good agreement with what was analyzed in Section 6. The SNR improvement of the MVDR filter is the least among all the studied filters, but the value of its speech distortion index is also the smallest. It is also seen that the influence of the filter length on the speech distortion index is dramatic when it is small, but less remarkable as the filter length is large. Finally, the PESQ results indicate that: (1) the best performance generally appears when the filter length is slightly larger than the signal dimension P; (2) all the noise reduction filters except for the maximum SNR one improve the quality. The PESQ score of the maximum SNR filter is even worse than the original noisy speech. These results corroborate with what was observed in the literature of noise reduction (Benesty et al., 2009).

Fig. 2 plots the performance of different noise reduction filters as a function of the rank parameter, P, in white Gaussian noise. The input SNR is 10 dB and the filter length is set to 30 based on the previous simulation. Again, the noise statistics are computed directly from the noise signal. It is seen that the output SNR decreases as P increases in the studied range of P (note that the smallest value of P is set to 10 in our simulations since we found that the level of speech distortion is acceptable with $P \geq 10$), and so is the speech distortion index. The PESQ scores, however, do not change much with P for all the filters except for the MVDR one. Note that as the value of P approaches the value of L, the reduced-rank filters approaches the full-rank solutions. So, the results in this simulation show that the reduced-rank noise reduction filters can yield a similar PESQ performance as their full-rank counterparts as long as the rank parameter, P, is in a reasonable range; but the reduced-rank solution provides an extra degree of freedom to compromise between noise reduction and speech distortion. For the MVDR filter, more constraint is applied to ensure no speech distortion as the value of P increases and, as a result, there is less noise reduction. When $P = L$, the MVDR filter degenerates to the identity matrix, so there is neither noise reduction nor speech distortion. This is validated by the results in Fig. 2.

Simulations were also conducted to evaluate the performance of the noise reduction filters in different input SNR conditions. Again, the background noise is white Gaussian and the noise statistics are computed directly from the noise signal. The filter length $L = 30$ and the rank parameter $P = 20$. The results are plotted in Fig. 3. In all studied input SNR conditions, the Wiener, tradeoff, and MVDR filters can improve both the SNR and the PESQ score. However, the maximum SNR filter achieves the highest output SNR and its speech distortion index does not change much with the input SNR. Taking into account the previous simulations, we can conclude that the maximum SNR filter should be regarded as the filter setting an upper bound on the output SNR rather than a practical solution in real applications.

In this simulation, we study the performance of different filters with the use of the IMCRA method (Cohen, 2003) for noise estimation. The filter length L is set to 30, the noise is white Gaussian and the input SNR is 10 dB. The corresponding performance of different noise reduction filters as a function of the rank parameter, P, is plotted in Fig. 4. Comparing Figs. 4 and 2, one can see that the performance with noise being estimated and that with *a priori* assumed known noise are similar. Fig. 5 plots the noisy speech, enhanced speech, and their spectrograms of the Wiener filter with $L = 30$ and $P = 14$. It can be clearly seen that the reduced-rank Wiener filter with the IMCRA noise estimation method achieves tremendous noise reduction, as seen from both the waveforms and spectrograms. To further validate the noise reduction performance, subjective evaluation is performed where 18 listeners are asked to rate speech quality according to the mean opinion score (MOS) standard in Table 1. The 18 listeners are graduate students and professors at Northwestern Polytechnical University. Nine of those listeners are familiar with speech quality and speech processing while the rest are not familiar with speech evaluation and they were just taught how to rate MOS scores. The results are as follows. The MOS score of the noisy speech is 2.44. With the Wiener filter ($L = 30$ and $P = 14$), the MOS score of the enhanced speech is 3.11. The MOS improvement is approximately 0.7, which is significant.

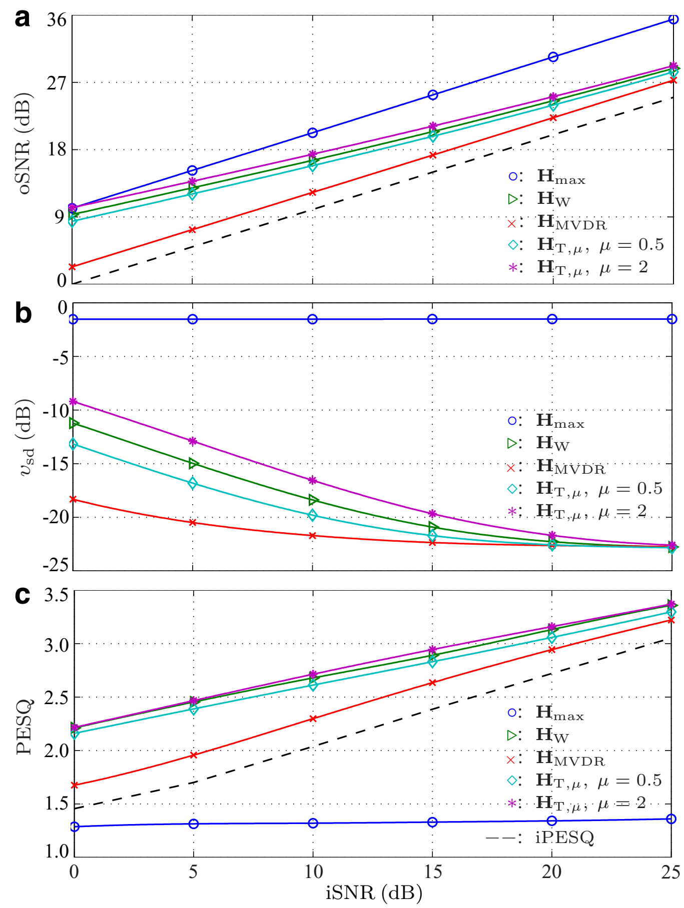

Fig. 3. Performance of the maximum SNR, Wiener, MVDR, and tradeoff filters as a function of input SNR in white Gaussian noise: (a) output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: $L = 30$ and $P = 20$.

Table 1
Mean Opinion Score (MOS) Standard.

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

In this experiment, we study the noise reduction performance with real recorded speech. Since neither noise nor clean speech is accessible in this situation, we use the IMCRA method to estimate the noise statistics and subjective evaluation to assess the quality. The outputs of the Wiener filter with $L = 30$ and $P = 14$ for this experiment are plotted in Fig. 6. It can be seen that much noise has been reduced. Using the MOS test with 18 listeners, the MOS of the noisy speech is 2.38, and that for the enhanced speech is 2.72. One may notice that the improvement in MOS for real recorded speech is less than that for the simulated speech. The underlying reasons are as follows: (1) the noise statistics are more difficult to estimate and (2) the playback-recording system

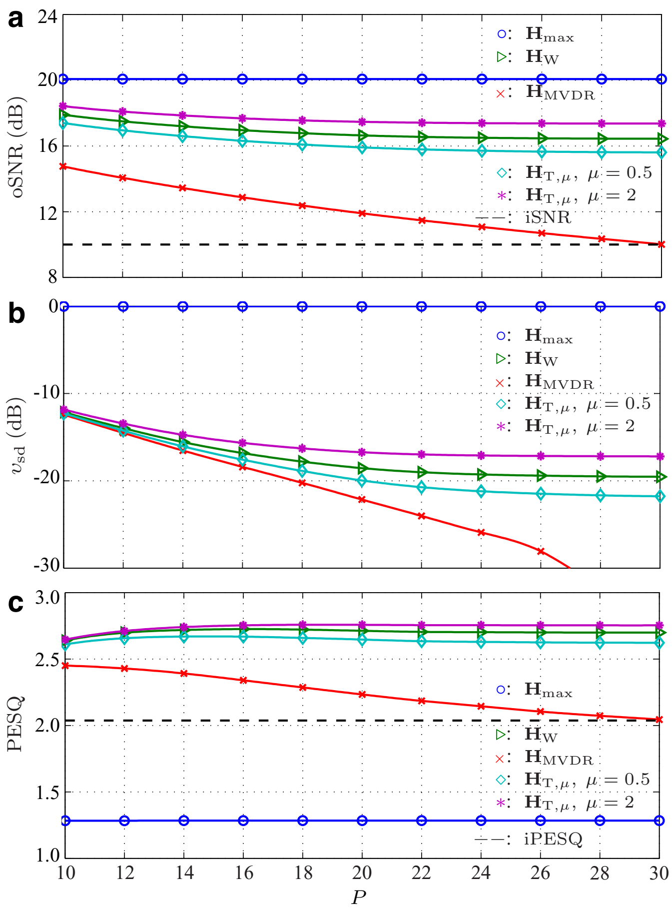

Fig. 4. Performance of the maximum SNR, Wiener, MVDR, and tradeoff filters as a function of the rank parameter, P, with noise estimation using IMCRA: (a) output SNR, (b) speech distortion index, and (c) PESQ score. Simulation conditions: white noise, $\mathrm{iSNR} = 10$˜dB, and $L = 30$.

also introduces some distortion, which is not additive and cannot be efficiently dealt with noise reduction techniques.

8.5. Impact of semi-orthogonal transformations on noise reduction performance

Section 7 presented three important semi-orthogonal transformations. They were derived from three different criteria, i.e., minimum correlation, minimum MSE (reduced-rank), and minimum distortion. In this section, we study the performance of the Wiener filter with different semi-orthogonal transformations and compare it to that of the Wiener filter in Section 6. Three noise environments (white Gaussian, car, and NYSE) are investigated and the rank parameter, P, is set to 10. The PESQ results as a function of the filter length, L, are plotted in Fig. 7. Note that the performance of the transformation derived from the minimum-residual-noise criterion was found much worse than that of the other two transformations. The underlying reason is that this criterion does not apply any constraint to speech distortion, so the results of this transformation are not plotted in Fig. 7.

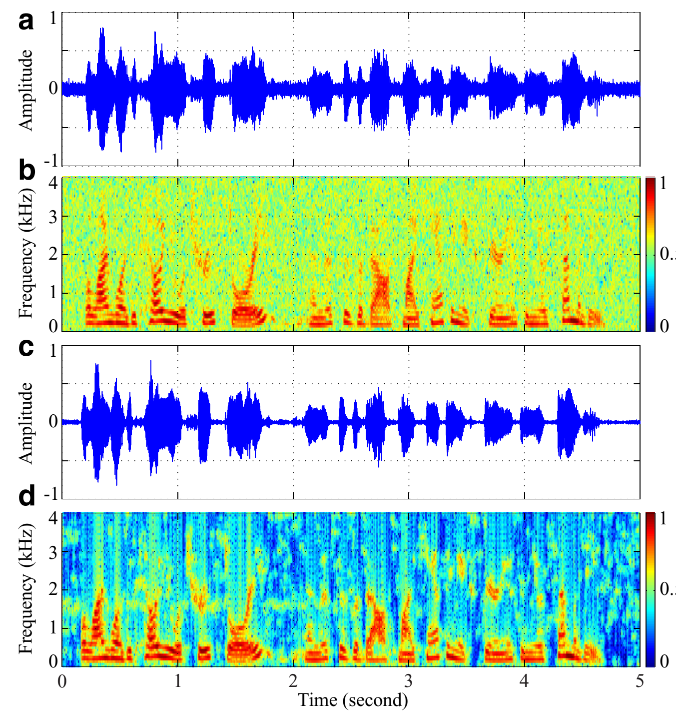

Fig. 5. Outputs of the Wiener filter in white Gaussian noise with noise estimation using IMCRA: (a) waveform of the noisy signal,(b) spectrogram of the noisy signal, (c) waveform of the filtered signal, and (d) spectrogram of the filtered signal. Simulation conditions: $L = 30$, $P = 14$, and $\mathrm{iSNR} = 10\,\mathrm{dB}$.

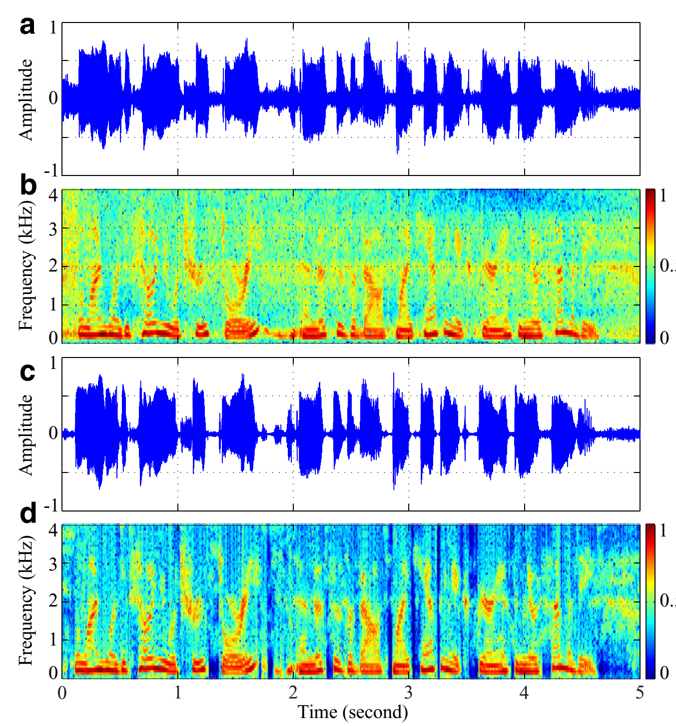

Fig. 6. Outputs of the Wiener filter for real recorded signal with noise estimation using IMCRA: (a) waveform of the noisy signal, (b) spectrogram of the noisy signal, (c) waveform of the filtered signal, and (d) spectrogram of the filtered signal. Simulation conditions: $L = 30$ and $P = 14$.

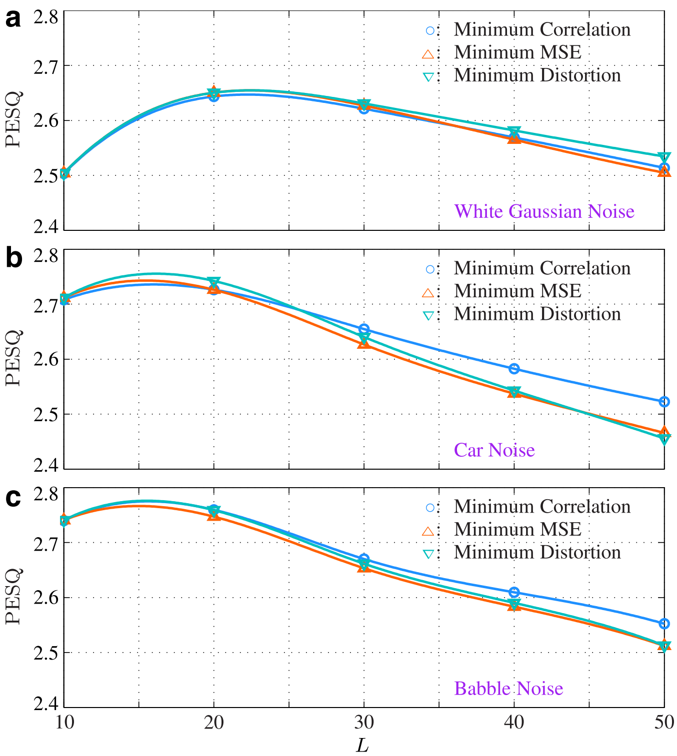

Fig. 7. PESQ scores of the Wiener filter as a function of the filter length, L, in different noise conditions: (a) white Gaussian noise, (b) car noise, and (c) babble noise. Simulation conditions: $P = 10$ and iSNR $= 10$˜dB. The minimum correlation refers to the method in Section 6, where the semi-orthogonal transformation is constructed using the eigenvectors corresponding to the P largest eigenvalues of $\mathbf{R_x}$.

It is seen from Fig. 7 that the PESQ results of the Wiener filter with the two transformations derived from the minimum MSE and minimum distortion are similar, which is slightly worse than that of the transformation used in Section 6 when L is large.

9. Conclusion

This paper investigated the use of the reduced-rank principle to the problem of single-channel noise reduction in the framework of semi-orthogonal transformations. Under this framework, we derived the maximum SNR, the Wiener, the MVDR, and the tradeoff filters. Simulation results showed that these reduced-rank optimal filters can yield better output SNR and similar PESQ performance as compared to their full-rank counterparts if the rank parameter is properly chosen. Furthermore, these reduced-rank filter are more flexible in controlling the compromise between the amount of noise reduction and the degree of speech distortion than their full-rank counterparts. We also discussed how to derive different semi-orthogonal transformations though these transformations may not affect much the noise reduction performance in terms of PESQ score if the rank parameter is properly chosen, but they affect the amount of noise reduction and speech distortion.

References

Benesty, J., Chen, J., 2011. Optimal Time-domain Noise Reduction Filters: A Theoretical Study. Springer-Verlag, Berlin, Germany.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Noise Reduction in Speech Processing. Springer-Verlag, Berlin, Germany.

Berouti, M., Schwartz, R., Makhoul, J., 1979. Enhancement of speech corrupted by acoustic noise estimation using linear prediction. In: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP), pp. 208–211.

Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust., Speech, Signal Process 27 (2), 113–120.

Cohen, I., 2003. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. IEEE Trans. Signal Process 11 (5), 466–475.

Dendrinos, M., Bakamidis, S., Carayannis, G., 1991. Speech enhancement from noise: A regenerative approach. Speech Commun. 10 (1), 45–57.

Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust., Speech, Signal Process. 32 (6), 1109–1121.

Ephraim, Y., Trees, H.L.V., 1995. A signal subspace approach for speech enhancement. IEEE Trans. Speech, Audio Process 3 (4), 251–266.

Goldstein, J.S., Reed, I.S., 1997. Reduced-rank adaptive filtering. IEEE Trans. Signal Process 45 (2), 492–496.

Goldstein, J.S., Reed, I.S., Dudgeon, D., Guerci, J.R., 1999. A multistage matrix Wiener filter for subspace detection. In: Proceedings of the IT Workshop on Detection, Estimation and Classification and Imaging, pp. 243–258.

Goldstein, J.S., Reed, I.S., Scharf, L.L., 1998. A multistage representation of the Wiener filter based on orthogonal projections. IEEE Trans. Inf. Theory 44 (7), 2943–2959.

Hansen, P.C., Jensen, S.H., 1998. FIR filter representations of reduced-rank noise reduction. IEEE Trans. Signal Process 46 (6), 1737–1741.

Hansen, P.C., Jensen, S.H., 2007. Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: Survey and analysis. EURASIP J. Adv. Signal Process 2007 (1), 1–24.

Hansen, P.C., Jensen, S.H., 2013. A class of optimal rectangular filtering matrices for single-channel signal enhancement in the time domain. IEEE Trans. Audio, Speech Lang. Process 21 (12), 2595–2606.

Hu, Y., Loizou, P.C., 2003. A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans. Speech, Audio Process 11 (4), 334–341.

Huang, J., Zhao, Y., 1998. An energy-constrained signal subspace method for speech enhancement and recognition in white and colored noises. Speech Commun. 26 (3), 165–181.

Huang, J., Zhao, Y., 2000. A DCT-based fast signal subspace technique for robust speech recognition. IEEE Trans. Speech, Audio Process. 8 (6), 747–751.

Huffel, S.V., 1993. Enhanced resolution based on minimum variance estimation and exponential data modeling. IEEE Trans. Signal Process. 33 (3), 333–355.

ITU-T P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, ITU-T Recommendation P.862.

Jensen, S.H., Hansen, P.C., Hansen, S.D., Sorensen, J.A., 1995. Reduction of broad-band noise in speech by truncated QSVD. IEEE Trans. Speech, Audio Process. 3 (6), 439–448.

Lim, J.S., Oppenheim, A.V., 1979. Enhancement and bandwidth compression of noisy speech. Proc. IEEE 67 (12), 1586–1604.

Loizou, P.C., 2007. Speech Enhancement: Theory and Practice. CRC Press, Boca Raton Florida.

Mittal, U., Phamdo, N., 2000. Signal/noise KLT based approach for enhancing speech degraded by colored noise. IEEE Trans. Speech, Audio Process 8 (2), 159–167.

Moor, B.D., 1993. The singular value decomposition and long and short spaces of noisy matrices. IEEE Trans. Signal Process 41 (9), 2826–2838.

Nørholm, S.M., Benesty, J., Jensen, J.R., Christensen, M.G., 2014. Single-channel noise reduction using unified joint diagonalization and optimal filtering. EURASIP J. Adv. Signal Process 2014 (37), 1–11.

Rezayee, A., Gazor, S., 2001. An adaptive KLT approach for speech enhancement. IEEE Trans. Speech, Audio Process 9 (2), 87–95.

Scharf, L.L., 1991. The SVD and reduced rank signal processing. Signal Process. 25 (2), 113–133.

Scharf, L.L., Tufts, D.W., 1987. Rank reduction for modeling stationary signals. IEEE Trans. Acoust., Speech, Signal Process. 35 (3), 350–355.

Trees, V., Harry, L., 2001. Detection, Estimation, and Modulation Theory-Part l-Detection, Estimation, and Linear Modulation Theory. John Wiley & Sons Ltd, England.

Tufts, D.W., Kumaresan, R., 1982a. Data adaptive signal estimation by singular value decomposition of a data matrix. Proc. IEEE 70 (6), 684–685.

Tufts, D.W., Kumaresan, R., 1982b. Singular value decomposition and improved frequency estimation using linear prediction. IEEE Trans. Acoust., Speech, Signal Process. 30 (4), 671–675.

Wiener, N., 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The MIT Press, Cambridge, MA.

Zhang, W., Chen, J., Benesty, J., 2014. A reduced-rank approach to single-channel noise reduction. In: Processing of the International Workshop on Acoustic Echo and Noise Control (IWAENC), pp. 293–297.