OPTIMAL RECTANGULAR FILTERING MATRIX FOR NOISE REDUCTION IN THE TIME DOMAIN

Chao Li[1], *Jacob Benesty*[2], *and Jingdong Chen*[3]

[1]: NLPR, IA, Chinese Academy of Sciences
95 Zhongguancun East Road
Beijing 100190, China

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

[3]: Northwestern Polytechnical University
127 Youyi West Road
Xi'an, Shaanxi 710072, China

ABSTRACT

In this paper, we study the noise reduction problem in the time domain and present a frame-based method to decompose the clean speech vector into two orthogonal components: one correlated and the other uncorrelated with the current desired speech vector to be estimated. In comparison with the sample-based decomposition developed in the previous research that uses only forward prediction, this new decomposition exploits both the forward prediction and interpolation. Based on this new decomposition, we formulate different optimization cost functions and address the issue of how to design Wiener and minimum variance distortionless response (MVDR) filtering matrices by optimizing these new cost functions. We also discuss the relationship between the Wiener and MVDR filtering matrices and show that the MVDR filtering matrix can achieve noise reduction without adding speech distortion; but it reduces less noise than the Wiener filtering matrix. Compared with the sample-based algorithms developed in the previous study, the proposed frame-based algorithms can achieve better noise reduction performance. Furthermore, they are computationally more efficient, and therefore, more suitable for practical implementation.

Index Terms— Noise reduction, time domain, orthogonal decomposition, rectangular filtering matrix, Wiener filter, minimum variance distortionless response (MVDR) filter.

1. INTRODUCTION

In many areas of speech processing an effective noise reduction algorithm is required. Over the past several decades, many algorithms have been developed and improved [1], [2], [3], [4], [5]. However, it is well known that these algorithms in the single-channel case achieve noise reduction at a price of adding some distortion into the speech signal. In general, the more the noise is reduced, the more the speech is distorted.

Recently, it has been shown that by decomposing the clean speech signal vector into two orthogonal components, i.e., the desired speech and interference, we can design many new noise reduction filters [6], [7]. Particularly, due to this orthogonal decomposition, the minimum variance distortionless response (MVDR) filter can be designed, which can achieve noise reduction without introducing speech distortion in the single-channel case.

The algorithms developed in [6] are sample-based, which estimate one sample at a time where the speech signal at the current time instant is always predicted using the past samples (i.e., via forward prediction). In this paper, we extend the sample-based approach into a frame-based framework and deduce some frame-based algorithms. These algorithms estimate more than one sample at a time where both forward prediction and interpolation are utilized. It is well known that speech is highly correlated with its neighboring (both previous and future) samples. This correlation can be well exploited in the frame-based formulation to improve noise reduction as will be demonstrated in Section 4 about experiments. Another advantage of the frame-based framework as compared to the sample-based one is the computational complexity. As we know, the largest computational burden of a time-domain noise reduction algorithm is from the matrix inversion. In the sample-based method, matrix inversion is needed for every sample. But in the frame-based implementation it is only needed for each frame. So, the larger the frame length, the lower the computational burden.

In the frame-based approach, we deal with a rectangular filtering matrix instead of a filtering vector as described in [6]. As will become clearer soon, this frame-based approach is more general and all the results from [6] can be viewed as particular cases of the results derived in this paper by just setting the frame length to 1.

Effort of the third author is partially supported by the Anhui Science and Technology Project (11010202191).

2. SIGNAL MODEL AND PROBLEM FORMULATION

In the time domain, we assume that the observed signal, $y(k)$, is an additive mixture of the clean speech, $x(k)$, and the noise, $v(k)$, i.e.,

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

where $x(k)$ and $v(k)$ are assumed to be uncorrelated and zero-mean random processes, and k is the discrete-time index. All signals are considered to be real and broadband.

The signal model given in (1) can be put into a vector form:

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

where

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

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 $\mathbf{y}(k)$.

In this paper, we estimate more than one sample at a time. Therefore, we define the vector of length M:

$$\mathbf{x}^M(k) = \begin{bmatrix} x(k) & x(k-1) & \cdots & x(k-M+1) \end{bmatrix}^T, \tag{4}$$

where $M \leq L$. In the general linear filtering approach, we estimate the desired signal vector, $\mathbf{x}^M(k)$, by applying a linear transformation to $\mathbf{y}(k)$ [1], [7], i.e.,

$$\begin{aligned} \mathbf{z}^M(k) &= \mathbf{H}\mathbf{y}(k) = \mathbf{H}\left[\mathbf{x}(k) + \mathbf{v}(k)\right] \\ &= \mathbf{x}_{\mathrm{f}}^M(k) + \mathbf{v}_{\mathrm{rn}}^M(k), \end{aligned} \tag{5}$$

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

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_M \end{bmatrix}^T \tag{6}$$

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

$$\mathbf{h}_m = [h_{m,0} \;\; h_{m,1} \;\; \cdots \;\; h_{m,L-1}]^T, \; m = 1, 2, \ldots, M \quad (7)$$

are FIR filters of length L, and

$$\mathbf{x}_{\mathrm{f}}^M(k) = \mathbf{H}\mathbf{x}(k) \quad (8)$$

$$\mathbf{v}_{\mathrm{rn}}^M(k) = \mathbf{H}\mathbf{v}(k) \quad (9)$$

are the filtered speech and residual noise respectively. Depending on the value of M, there are two important particular cases of (5) as described below.

- $M = 1$. In this situation, $\mathbf{z}^1(k) = z(k)$ is a scalar and $\mathbf{H}$ degenerates to an FIR filter $\mathbf{h}^T$ of length L. This case has been well studied in [6].
- $M = L$. In this situation, $\mathbf{z}^L(k) = \mathbf{z}(k)$ is a vector of length L and $\mathbf{H} = \mathbf{H}_{\mathrm{S}}$ is a square matrix of size $L \times L$. This scenario has been widely covered in [1], [4], [5] and in many other papers.

By definition, our desired signal is the vector $\mathbf{x}^M(k)$. Therefore, we need to extract $\mathbf{x}^M(k)$ from $\mathbf{x}(k)$. For that, we need to decompose $\mathbf{x}(k)$ into two orthogonal components: one that is correlated with (or is a linear transformation of) the desired signal $\mathbf{x}^M(k)$ and another that is orthogonal to $\mathbf{x}^M(k)$ and, hence, will be considered as the interference component. Specifically, the vector $\mathbf{x}(k)$ is decomposed into the following form:

$$\mathbf{x}(k) = \mathbf{R}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}^{-1}\mathbf{x}^M(k) + \mathbf{x}_{\mathrm{i}}(k) = \mathbf{x}_{\mathrm{d}}(k) + \mathbf{x}_{\mathrm{i}}(k), \quad (10)$$

where

$$\mathbf{x}_{\mathrm{d}}(k) = \mathbf{R}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}^{-1}\mathbf{x}^M(k) = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{x}^M(k) \quad (11)$$

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

$$\mathbf{x}_{\mathrm{i}}(k) = \mathbf{x}(k) - \mathbf{x}_{\mathrm{d}}(k) \quad (12)$$

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

$$E\left[\mathbf{x}_{\mathrm{d}}(k)\mathbf{x}_{\mathrm{i}}^T(k)\right] = \mathbf{0}_{L\times L}. \quad (13)$$

For the particular case $M = L$, we have $\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}} = \mathbf{I}_L$, which is the identity matrix (of size $L \times L$), and $\mathbf{x}_{\mathrm{d}}(k)$ coincides with $\mathbf{x}(k)$, which obviously makes sense. For $M = 1$, $\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^1}$ simplifies to the normalized correlation vector [6]

$$\boldsymbol{\gamma}_{\mathbf{x}} = \frac{E\left[\mathbf{x}(k)x(k)\right]}{E\left[x^2(k)\right]}. \quad (14)$$

Substituting (10) into (5), we get

$$\begin{aligned}\mathbf{z}^M(k) &= \mathbf{H}\left[\mathbf{x}_{\mathrm{d}}(k) + \mathbf{x}_{\mathrm{i}}(k) + \mathbf{v}(k)\right] \\ &= \mathbf{x}_{\mathrm{fd}}^M(k) + \mathbf{x}_{\mathrm{ri}}^M(k) + \mathbf{v}_{\mathrm{rn}}^M(k), \end{aligned} \quad (15)$$

where

$$\mathbf{x}_{\mathrm{fd}}^M(k) = \mathbf{H}\mathbf{x}_{\mathrm{d}}(k) \quad (16)$$

$$\mathbf{x}_{\mathrm{ri}}^M(k) = \mathbf{H}\mathbf{x}_{\mathrm{i}}(k) \quad (17)$$

are the filtered desired signal and the residual interference respectively. It can be checked that the three terms $\mathbf{x}_{\mathrm{fd}}^M(k)$, $\mathbf{x}_{\mathrm{ri}}^M(k)$, and $\mathbf{v}_{\mathrm{rn}}^M(k)$ are mutually uncorrelated. Therefore, the correlation matrix of $\mathbf{z}^M(k)$ is

$$\mathbf{R}_{\mathbf{z}^M} = E\left[\mathbf{z}^M(k)\mathbf{z}^{MT}(k)\right] = \mathbf{R}_{\mathbf{x}_{\mathrm{fd}}^M} + \mathbf{R}_{\mathbf{x}_{\mathrm{ri}}^M} + \mathbf{R}_{\mathbf{v}_{\mathrm{rn}}^M}, \quad (18)$$

where

$$\mathbf{R}_{\mathbf{x}_{\mathrm{fd}}^M} = \mathbf{H}\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}\mathbf{H}^T, \quad (19)$$

$$\mathbf{R}_{\mathbf{x}_{\mathrm{ri}}^M} = \mathbf{H}\mathbf{R}_{\mathbf{x}_{\mathrm{i}}}\mathbf{H}^T = \mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T - \mathbf{H}\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}\mathbf{H}^T, \quad (20)$$

$$\mathbf{R}_{\mathbf{v}_{\mathrm{rn}}^M} = \mathbf{H}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T, \quad (21)$$

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

Now, the error signal between the estimated and desired signals can be defined as a vector of length M:

$$\mathbf{e}^M(k) = \mathbf{z}^M(k) - \mathbf{x}^M(k) = \mathbf{e}_{\mathrm{d}}^M(k) + \mathbf{e}_{\mathrm{r}}^M(k), \quad (22)$$

where

$$\mathbf{e}_{\mathrm{d}}^M(k) = \mathbf{x}_{\mathrm{fd}}^M(k) - \mathbf{x}^M(k) = \left(\mathbf{H}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M} - \mathbf{I}_M\right)\mathbf{x}^M(k) \quad (23)$$

is the signal distortion due to the rectangular filtering matrix with $\mathbf{I}_M$ being the $M \times M$ identity matrix and

$$\mathbf{e}_{\mathrm{r}}^M(k) = \mathbf{x}_{\mathrm{ri}}^M(k) + \mathbf{v}_{\mathrm{rn}}^M(k) = \mathbf{H}\mathbf{x}_{\mathrm{i}}(k) + \mathbf{H}\mathbf{v}(k) \quad (24)$$

represents the residual interference-plus-noise.

Having defined the error signal, we can now write the mean-square error (MSE) criterion:

$$\begin{aligned} J(\mathbf{H}) &= \frac{1}{M}\cdot \mathrm{tr}\left\{E\left[\mathbf{e}^M(k)\mathbf{e}^{MT}(k)\right]\right\} \quad (25) \\ &= \frac{1}{M}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}^M}\right) + \mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T\right) - 2\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}\mathbf{x}^M}\right)\right] \\ &= \frac{1}{M}\left[\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}^M}\right) + \mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T\right) - 2\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{x}\mathbf{x}^M}\right)\right], \end{aligned}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a square matrix and

$$\begin{aligned}\mathbf{R}_{\mathbf{y}\mathbf{x}^M} &= E\left[\mathbf{y}(k)\mathbf{x}^{MT}(k)\right] \\ &= E\left[\mathbf{x}(k)\mathbf{x}^{MT}(k)\right] \\ &= \mathbf{R}_{\mathbf{x}\mathbf{x}^M}, \end{aligned} \quad (26)$$

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

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

$$J(\mathbf{H}) = J_{\mathrm{d}}(\mathbf{H}) + J_{\mathrm{r}}(\mathbf{H}), \quad (27)$$

where

$$J_{\mathrm{d}}(\mathbf{H}) = \frac{1}{M}\cdot \mathrm{tr}\left\{E\left[\mathbf{e}_{\mathrm{d}}^M(k)\mathbf{e}_{\mathrm{d}}^{MT}(k)\right]\right\}, \quad (28)$$

$$J_{\mathrm{r}}(\mathbf{H}) = \frac{1}{M}\cdot \mathrm{tr}\left\{E\left[\mathbf{e}_{\mathrm{r}}^M(k)\mathbf{e}_{\mathrm{r}}^{MT}(k)\right]\right\}. \quad (29)$$

3. OPTIMAL RECTANGULAR FILTERING MATRICES

In this section, we are going to derive two important filtering matrices that can help reduce the noise in the microphone signal.

3.1. Wiener

If we differentiate the MSE criterion, $J(\mathbf{H})$, defined in (25), with respect to $\mathbf{H}$ and equate the result to zero, we find the Wiener filtering matrix

$$\mathbf{H}_{\mathrm{W}} = \mathbf{R}_{\mathbf{y}\mathbf{x}^M}^T \mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{R}_{\mathbf{x}\mathbf{x}^M}^T \mathbf{R}_{\mathbf{y}}^{-1}. \tag{30}$$

Using the identity filtering matrix $\mathbf{I}_{\mathrm{i}} = \begin{bmatrix} \mathbf{I}_M & \mathbf{0}_{M\times(L-M)} \end{bmatrix}$, we can rewrite the Wiener filtering matrix as

$$\mathbf{H}_{\mathrm{W}} = \mathbf{I}_{\mathrm{i}}\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{I}_{\mathrm{i}}\left(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}}\mathbf{R}_{\mathbf{y}}^{-1}\right). \tag{31}$$

Since

$$\mathbf{R}_{\mathbf{x}\mathbf{x}^M} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}, \tag{32}$$

we can rewrite (30) as

$$\mathbf{H}_{\mathrm{W}} = \mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}. \tag{33}$$

By exploiting the decomposition of $\mathbf{x}(k)$ in (10), we can decompose the correlation matrix of $\mathbf{y}(k)$ as

$$\mathbf{R}_{\mathbf{y}} = \mathbf{R}_{\mathbf{x}_{\mathrm{d}}} + \mathbf{R}_{\mathrm{in}} = \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T + \mathbf{R}_{\mathrm{in}}, \tag{34}$$

where

$$\mathbf{R}_{\mathrm{in}} = \mathbf{R}_{\mathbf{x}_{\mathrm{i}}} + \mathbf{R}_{\mathbf{v}} \tag{35}$$

is the interference-plus-noise correlation matrix.

Determining the inverse of $\mathbf{R}_{\mathbf{y}}$ from (34) with the Woodbury's identity

$$\mathbf{R}_{\mathbf{y}}^{-1} = \tag{36}$$
$$\mathbf{R}_{\mathrm{in}}^{-1} - \mathbf{R}_{\mathrm{in}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\left(\mathbf{R}_{\mathbf{x}^M}^{-1} + \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}$$

and substituting (36) into (33), we get another interesting formulation of the Wiener filtering matrix

$$\begin{aligned}\mathbf{H}_{\mathrm{W}} &= \left(\mathbf{I}_M + \mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{R}_{\mathbf{x}^M}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1} \\ &= \left(\mathbf{R}_{\mathbf{x}^M}^{-1} + \mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}. \end{aligned} \tag{37}$$

3.2. MVDR

The celebrated MVDR approach, requiring no distortion to the desired signal, is usually derived in the multichannel case. Interestingly, with the new formulation, we can also derive the MVDR in the single-channel case, just like in [6], [7]. The corresponding rectangular filtering matrix is obtained by minimizing the MSE of the residual interference-plus-noise, $J_{\mathrm{r}}(\mathbf{H})$, with the constraint that the desired signal is not distorted. Mathematically, this is equivalent to

$$\min_{\mathbf{H}} \frac{1}{M}\cdot \mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathrm{in}}\mathbf{H}^T\right) \quad \text{subject to} \quad \mathbf{H}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M} = \mathbf{I}_M. \tag{38}$$

The solution to the above optimization problem is

$$\mathbf{H}_{\mathrm{MVDR}} = \left(\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathrm{in}}^{-1}, \tag{39}$$

which is interesting to compare to $\mathbf{H}_{\mathrm{W}}$ in (37).

Obviously, by using the Woodbury's identity of $\mathbf{R}_{\mathrm{y}}^{-1}$ in (36) we can rewrite (39) as

$$\mathbf{H}_{\mathrm{MVDR}} = \left(\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}^T\mathbf{R}_{\mathbf{y}}^{-1}. \tag{40}$$

From (33) and (40), we deduce the relationship between the MVDR and Wiener filtering matrices:

$$\mathbf{H}_{\mathrm{MVDR}} = \left(\mathbf{H}_{\mathrm{W}}\mathbf{\Gamma}_{\mathbf{x}\mathbf{x}^M}\right)^{-1}\mathbf{H}_{\mathrm{W}}. \tag{41}$$

4. EXPERIMENTAL RESULTS

In this section, we use experiments to evaluate the performance of the two proposed optimal rectangular filtering matrices, i.e., Wiener and MVDR.

The clean speech signal used in the experiments was recorded from a male talker in a quiet office room. It was sampled at 8 kHz. The overall length of the signal is 30 seconds. White Gaussian noise is added into the clean speech and the input signal-to-noise ratio (SNR) of the noisy speech is 10 dB.

Implementation of the noise reduction filtering matrices derived in Section 3 require the estimation of the correlation matrices $\mathbf{R}_{\mathbf{y}}$, $\mathbf{R}_{\mathbf{x}}$, and $\mathbf{R}_{\mathbf{v}}$. To avoid the complicated voice activity detection (VAD) issue and make the performance study simple, we compute the three matrices $\mathbf{R}_{\mathbf{y}}$, $\mathbf{R}_{\mathbf{x}}$, and $\mathbf{R}_{\mathbf{v}}$ directly from the corresponding signals. Specifically, at each frame, the matrices $\mathbf{R}_{\mathbf{y}}$ and $\mathbf{R}_{\mathbf{x}}$ are computed, respectively, using the most recent 600 samples (75-ms long) of the noisy and clean speech signals with a short-time average. Since noise is stationary, we estimate $\mathbf{R}_{\mathbf{v}}$ using 960 samples (120-ms long) with a short-time average during periods of silence. With this implementation, our experiments should demonstrate the upper limit performance of the optimal rectangular filtering matrices.

To evaluate the amount of noise reduction, the output SNR is adopted as the objective performance measure. It is computed according to [6], [7]

$$\mathrm{oSNR}(\mathbf{H}) = \frac{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_{\mathrm{fd}}^M}\right)}{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}_{\mathrm{ri}}^M} + \mathbf{R}_{\mathbf{v}_{\mathrm{rn}}^M}\right)}. \tag{42}$$

The higher is the value of $\mathrm{oSNR}(\mathbf{H})$, the more the noise is reduced.

We also evaluate the amount of speech distortion using the speech distortion index [6], [7]

$$\upsilon_{\mathrm{sd}}(\mathbf{H}) = \frac{\mathrm{tr}\{E\left[\mathbf{e}_{\mathrm{d}}^M(k)\mathbf{e}_{\mathrm{d}}^{MT}(k)\right]\}}{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}^M}\right)}. \tag{43}$$

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

In order to calculate the above objective measures, the clean speech signal was also processed for each filtering matrix in addition to the noisy signal. That is to say, the filtering matrix was calculated on the noisy signal and then also applied to the clean speech. In this way, two signals are available at the output: the enhanced noisy signal $\mathbf{z}^M(k)$ and the processed desired speech $\mathbf{x}_{\mathrm{fd}}^M(k)$. Moreover, both measures are computed based on the 30-s long signals using a long-time average.

The first experiment investigates the influence of the filter length L on the noise reduction performance. Figure 1(a) shows that the output SNR of the Wiener filtering matrix increases with L if the value of M is small, i.e., ($M \leq 2$), while it does not change much with L if the value of M is large ($M \geq 4$). In comparison, the MVDR filtering matrix yields a higher output SNR with a larger L in all the studied cases. Figure 1(b) shows that the speech distortion index of the Wiener filtering matrix decreases linearly with L, while such index of the MVDR filtering matrix is approximately 0 for all the different values of M and L.

The second experiment tests the noise reduction performance as a function of the frame length M. Based on the previous experiment, we set $L = 48$. We see from Figure 2(a) that the output SNR of the Wiener filtering matrix grows quickly as M is increased up to 4, and then continues to grow but with a slower rate, while the

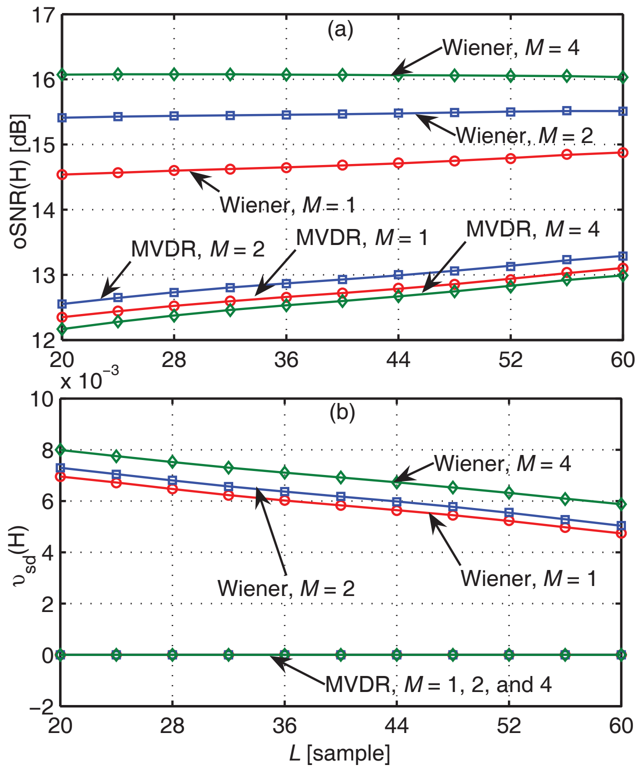

Fig. 1. Performance of the Wiener and MVDR filtering matrices as a function of the filter length L: (a) output SNR and (b) speech distortion index. White Gaussian noise is used, the input SNR is 10 dB, and $M = 1, 2$, and 4.

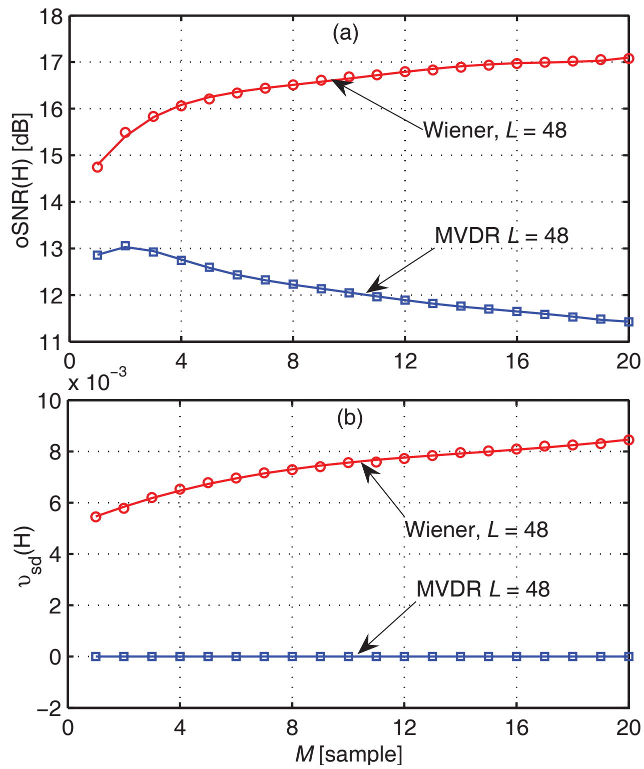

Fig. 2. Performance of the Wiener and MVDR filtering matrices as a function of the frame length M: (a) output SNR and (b) speech distortion index. White Gaussian noise is used, input SNR is 10 dB, and $L = 48$.

output SNR of the MVDR filtering matrix first increases to its maximum, and then decreases as M is increased. Figure 2(b) shows that the speech distortion index of the Wiener filtering matrix increases linearly with M, while such index of the MVDR filtering matrix is always small (approximately 0). One can see that the MVDR filtering matrix achieves less noise reduction as compared to the Wiener filtering matrix but it does not introduce speech distortion, which is a strong advantage.

All the experiments demonstrated that both the Wiener and MVDR filtering matrices can yield a better performance when $M > 1$ as compared to $M = 1$ (if M is not too large). This shows the advantage of using filtering matrices over using filtering vectors developed in [6]. The underlying reason, as we have explained in Section 1, is that speech is highly correlated with its neighboring (either previous or future) samples. Another benefit of using a filtering matrix instead of a filtering vector is that the computational burden reduces with M, which indicates that the frame-based algorithms are more suitable for real-time implementation.

5. CONCLUSIONS

In this paper, we proposed a frame-based signal orthogonal decomposition to resolve the noise reduction problem in the time domain. By decomposing the clean speech vector into two orthogonal components, i.e., the desired speech and interference, we formulated different optimization cost functions and deduced two filtering matrices, i.e., Wiener and MVDR. The experimental results demonstrated that the Wiener filtering matrix can remove more noise, while the MVDR filtering matrix can achieve noise reduction without introducing any speech distortion. Furthermore, compared to the sample-based method developed previously, the frame-based approach can lead to a better performance in terms of noise reduction. Another side benefit of the frame-based approach is that it is computationally more efficient than the sample-based one.

6. REFERENCES

[1] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.

[2] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC Press, 2007.

[3] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Chichester, England: John Wiley & Sons Ltd, 2006.

[4] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 3, pp. 251–266, July 1995.

[5] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, pp. 2230–2244, Sept. 2002.

[6] J. Chen, J. Benesty, Y. Huang, and T. Gaensler, "On single-channel noise reduction in the time domain," in *Proc. IEEE ICASSP*, 2011, pp. 277–280.

[7] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study*. Springer Briefs in Electrical and Computer Engineering, 2011.