OPTIMAL SINGLE-CHANNEL NOISE REDUCTION FILTERING MATRICES FROM THE PEARSON CORRELATION COEFFICIENT PERSPECTIVE

Jiaolong Yu[1], *Jacob Benesty*[2], *Gongping Huang*[1], *and Jingdong Chen*[1]

[1]CIAIC and School of Marine Science and Technology
Northwestern Polytechnical University
127 Youyi West Rd., Xi'an, Shaanxi 710072, China

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

ABSTRACT

This paper studies the problem of single-channel noise reduction in the time domain, where an estimate of a vector of the desired clean speech is achieved by filtering a frame of the noisy signal with a rectangular filtering matrix. The core issue with this problem formulation is then the estimation of the optimal filtering matrix. The squared Pearson correlation coefficient (SPCC) is used. We show that different optimal filtering matrices can be derived by maximizing or minimizing the SPCCs between different signals. For example, maximizing the SPCC between the enhanced signal and the filtered speech gives the reduced-rank Wiener and minimum distortion (MD) filtering matrices while minimizing the SPCC gives the minimum noise (MN) and another reduced-rank Wiener filtering matrices. Simulation results are presented to illustrate the properties of these filtering matrices.

Index Terms— Noise reduction, speech enhancement, single-channel, time-domain filtering, optimal filtering matrices, Pearson correlation coefficient.

1. INTRODUCTION

Noise reduction has been a major challenge in speech signal processing and, as a consequence, lots of efforts have been devoted to this problem over the past few decades [1]– [16]. Among various techniques that have been developed, the filtering technique is perhaps the most straightforward method, which obtains an estimate of the clean speech sample at every time instant by applying a filtering vector to the noisy signal vector [10–12]. Recently, this filtering technique has been extended to a more generic case where an estimate of a block of the desired clean speech is achieved every time by applying a rectangular filtering matrix instead of a filtering vector to the noisy signal [14], [13]. This generalized version of the filtering method does not only improve the noise reduction performance if the block size is properly chosen, but is also computationally more efficient as compared to the sample based method [13]. With this formulation of the noise reduction problem, the core issue is the derivation of optimal filtering matrices.

Typically, the optimal filtering matrices are derived from the mean-squared error (MSE) criterion [7], [8]. Recently, the squared Pearson correlation coefficient (SPCC) has been introduced as the cost function to derive noise reduction filters [7]. Using the SPCC has been shown to have many advantages as compared to the MSE criterion. For instance, it can, on the one hand, provide many new insights into the traditional noise reduction filters derived from the MSE criterion and, on the other hand, help deduce some new filters that were not seen with the MSE criterion. In our previous work [12], we explored the use of the SPCC as the cost function to design optimal filtering vectors. In this work, we extend our previous study to a more general, block-based framework with a rectangular filtering matrix. Combining the SPCC with the block-based filtering framework, we develop a new class of filtering matrices for noise reduction by either maximizing or minimizing the SPCC between different signals. We show that maximizing the SPCC between the enhanced signal and the filtered speech, we can derive two optimal rectangular filtering matrices, i.e., the reduced-rank Wiener and minimum distortion (MD) filtering matrices, which give direct estimates of the clean speech. While minimizing the SPCC, we derive the minimum noise (MN) filtering matrix and another reduced-rank Wiener filtering matrix, which give estimates of the noise from which we deduce estimates of the clean speech.

2. SIGNAL MODEL AND PROBLEM FORMULATION

In the noise reduction problem considered in this paper, the noisy observation or microphone signal is given by [8, 14]

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

where k is the discrete-time index, $x(k)$ is the clean speech signal, and $v(k)$ is the unwanted additive noise, which is assumed to be uncorrelated with $x(k)$. All signals are considered to be zero mean, real, stationary, and broadband.

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

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

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 $\mathbf{y}(k)$. 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

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

where $E[\cdot]$ denotes mathematical expectation, and $\mathbf{R_x} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$ and $\mathbf{R_v} = E\left[\mathbf{v}(k)\mathbf{v}^T(k)\right]$ are the correlation matrices of $\mathbf{x}(k)$ and $\mathbf{v}(k)$, respectively.

Let us define the vector of length M:

$$\widetilde{\mathbf{x}}(k) = \begin{bmatrix} x(k) & x(k-1) & \cdots & x(k-M+1) \end{bmatrix}^T, \tag{5}$$

where $1 \leq M \leq L$. In the same manner, the vector $\widetilde{\mathbf{v}}(k)$ is composed of the first M elements of $\mathbf{v}(k)$. The objective of single-channel noise reduction (or speech enhancement) in the time domain

This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005.

is then to find a "good" estimate of the desired signal vector, $\widetilde{\mathbf{x}}(k)$, 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.

We end this section by defining the input signal-to-noise ratio (SNR):

$$\mathrm{iSNR} = \frac{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}}\right)}{\mathrm{tr}\left(\mathbf{R}_{\mathbf{v}}\right)}, \tag{6}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix. This is one of the most fundamental measures in speech enhancement.

3. LINEAR FILTERING WITH A RECTANGULAR MATRIX AND CORRELATION COEFFICIENT

An estimate of $\widetilde{\mathbf{x}}(k)$ or $\widetilde{\mathbf{v}}(k)$ can be obtained by applying a linear transformation to $\mathbf{y}(k)$ [14], i.e.,

$$\widetilde{\mathbf{z}}(k) = \mathbf{H}\mathbf{y}(k) = \widetilde{\mathbf{x}}_{\mathrm{fd}}(k) + \widetilde{\mathbf{v}}_{\mathrm{fn}}(k), \tag{7}$$

where $\widetilde{\mathbf{z}}(k)$ is an estimate of $\widetilde{\mathbf{x}}(k)$ or $\widetilde{\mathbf{v}}(k)$,

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_M^T \end{bmatrix} \tag{8}$$

is a rectangular filtering matrix of size $M \times L$, $\mathbf{h}_m$, $m = 1, 2, \ldots, M$ are real-valued filters of length L, $\widetilde{\mathbf{x}}_{\mathrm{fd}}(k) = \mathbf{H}\mathbf{x}(k)$ is the filtered desired speech, and $\widetilde{\mathbf{v}}_{\mathrm{fn}}(k) = \mathbf{H}\mathbf{v}(k)$ is the filtered noise.

It is of great importance to know how much of $\widetilde{\mathbf{x}}(k)$ or $\widetilde{\mathbf{v}}(k)$ is contained in the estimator $\widetilde{\mathbf{z}}(k)$. One of the best second-order statistics based measure to evaluate this is via the squared Pearson correlation coefficient (SPCC) [1]. We define the SPCC between $\widetilde{\mathbf{z}}(k)$ and $\widetilde{\mathbf{x}}_{\mathrm{fd}}(k)$ as

$$\begin{aligned} \rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H}) &= \frac{E^2\left[\widetilde{\mathbf{z}}^T(k)\widetilde{\mathbf{x}}_{\mathrm{fd}}(k)\right]}{E\left[\widetilde{\mathbf{z}}^T(k)\widetilde{\mathbf{z}}(k)\right]E\left[\widetilde{\mathbf{x}}_{\mathrm{fd}}^T(k)\widetilde{\mathbf{x}}_{\mathrm{fd}}(k)\right]} \\ &= \frac{\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T\right)}{\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T\right)}. \end{aligned} \tag{9}$$

In the same manner, we define the SPCC between $\widetilde{\mathbf{z}}(k)$ and $\widetilde{\mathbf{v}}_{\mathrm{fn}}(k)$ as

$$\begin{aligned} \rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{v}}_{\mathrm{fn}}}(\mathbf{H}) &= \frac{E^2\left[\widetilde{\mathbf{z}}^T(k)\widetilde{\mathbf{v}}_{\mathrm{fn}}(k)\right]}{E\left[\widetilde{\mathbf{z}}^T(k)\widetilde{\mathbf{z}}(k)\right]E\left[\widetilde{\mathbf{v}}_{\mathrm{fn}}^T(k)\widetilde{\mathbf{v}}_{\mathrm{fn}}(k)\right]} \\ &= \frac{\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T\right)}{\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T\right)}. \end{aligned} \tag{10}$$

It is easy to see that

$$\rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H}) + \rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{v}}_{\mathrm{fn}}}(\mathbf{H}) = 1. \tag{11}$$

We observe that the SPCCs defined above depend explicitly on the filtering matrix, $\mathbf{H}$. This observation suggests that we can use the SPCC as a criterion to derive optimal filtering matrices. In the rest, we focus only on $\rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H})$. The same results can be obtained with $\rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{v}}_{\mathrm{fn}}}(\mathbf{H})$ because of the relation (11).

4. OPTIMAL FILTERING MATRICES

Intuitively, it makes sense to maximize or minimize the SPCC in order to find an estimate of $\widetilde{\mathbf{x}}(k)$ or $\widetilde{\mathbf{v}}(k)$. It is clear that the maximization (resp. minimization) of $\rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H})$ will give a good estimate of $\widetilde{\mathbf{x}}(k)$ [resp. $\widetilde{\mathbf{v}}(k)$], since in this case the SPCC between $\widetilde{\mathbf{z}}(k)$ and $\widetilde{\mathbf{x}}_{\mathrm{fd}}(k)$ will be maximal (resp. minimal), implying that $\widetilde{\mathbf{z}}(k)$ is close to $\widetilde{\mathbf{x}}(k)$ [resp. $\widetilde{\mathbf{v}}(k)$].

The concept of joint diagonalization [17] is going to be useful here. The two Hermitian matrices $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{R}_{\mathbf{y}}$ can be jointly diagonalized as follows [17]:

$$\mathbf{T}^T\mathbf{R}_{\mathbf{x}}\mathbf{T} = \mathbf{\Lambda}, \tag{12}$$

$$\mathbf{T}^T\mathbf{R}_{\mathbf{y}}\mathbf{T} = \mathbf{I}_L, \tag{13}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_L \end{bmatrix} \tag{14}$$

is a full-rank square matrix (of size $L \times L$),

$$\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_L\right) \tag{15}$$

is a diagonal matrix whose main elements are real and nonnegative, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$, and $\mathbf{I}_L$ is the $L \times L$ identity matrix. Furthermore, $\mathbf{\Lambda}$ and $\mathbf{T}$ are the eigenvalue and eigenvector matrices, respectively, of $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$, i.e.,

$$\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}\mathbf{T} = \mathbf{T}\mathbf{\Lambda}. \tag{16}$$

The matrices containing the first P and last Q eigenvectors of $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$ are, respectively,

$$\mathbf{T}_P = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_P \end{bmatrix} \tag{17}$$

and

$$\mathbf{T}_Q = \begin{bmatrix} \mathbf{t}_{L-Q+1} & \mathbf{t}_{L-Q+2} & \cdots & \mathbf{t}_L \end{bmatrix}. \tag{18}$$

These two matrices will be used soon. We deduce from (12) and (13) that $\mathbf{R}_{\mathbf{v}}$ can also be diagonalized as

$$\mathbf{T}^T\mathbf{R}_{\mathbf{v}}\mathbf{T} = \mathbf{I}_L - \mathbf{\Lambda}. \tag{19}$$

Since $\mathbf{R}_{\mathbf{v}}$ is positive semi-definite, it is straightforward to deduce that

$$0 \leq \lambda_l \leq 1,\ l = 1, 2, \ldots, L. \tag{20}$$

It can be shown that

$$\lambda_L \leq \rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H}) \leq \lambda_1. \tag{21}$$

The previous inequalities give tighter bounds as compared to the well-known ones, i.e., $0 \leq \rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H}) \leq 1$, and also give nice links between the SPCC and joint diagonalization.

We define the output SNR as

$$\mathrm{oSNR}\left(\mathbf{H}\right) = \frac{\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T\right)}{\mathrm{tr}\left(\mathbf{H}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T\right)}. \tag{22}$$

Therefore, the SPCC can also be expressed as a function of $\mathrm{oSNR}\left(\mathbf{H}\right)$, i.e.,

$$\rho^2_{\widetilde{\mathbf{z}}\widetilde{\mathbf{x}}_{\mathrm{fd}}}(\mathbf{H}) = \frac{\mathrm{oSNR}\left(\mathbf{H}\right)}{1 + \mathrm{oSNR}\left(\mathbf{H}\right)}. \tag{23}$$

Using (21), we easily deduce the lower and upper bounds for the output SNR:

$$\frac{\lambda_L}{1 - \lambda_L} \leq \mathrm{oSNR}\left(\mathbf{H}\right) \leq \frac{\lambda_1}{1 - \lambda_1}. \tag{24}$$

4.1. Maximization of the SPCC

It should be clear now that the maximization of (9) leads to an estimate of the desired signal. Assume that the largest eigenvalue, λ_1, of the matrix $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$ is of multiplicity P[1]. The corresponding eigenvectors are $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_P$. Let the filtering matrix be of the form:

$$\mathbf{H}_{\mathbf{\Theta}_P} = \mathbf{\Theta}_P \mathbf{T}_P^T, \tag{25}$$

where $\mathbf{\Theta}_P \neq \mathbf{0}$ is an arbitrary matrix of size $M \times P$. It is clear that $\mathbf{H}_{\mathbf{\Theta}_P}$ maximizes (9) and

$$\rho^2_{\mathbf{z}\widetilde{\mathbf{x}}_{\text{fd}}}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right) = \rho^2_{\mathbf{z}\widetilde{\mathbf{x}}_{\text{fd}},\max} = \lambda_1. \tag{26}$$

Therefore, the estimate of $\widetilde{\mathbf{x}}(k)$ is

$$\widetilde{\mathbf{x}}_{\mathbf{\Theta}_P}(k) = \mathbf{H}_{\mathbf{\Theta}_P}\mathbf{y}(k). \tag{27}$$

Since the SPCC is maximized, so is the output SNR. We deduce that

$$\text{oSNR}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right) = \frac{\lambda_1}{1-\lambda_1} \geq \text{iSNR}. \tag{28}$$

Now, we need to determine $\mathbf{\Theta}_P$. The mean-squared error (MSE) between $\widetilde{\mathbf{x}}(k)$ and $\widetilde{\mathbf{x}}_{\mathbf{\Theta}_P}(k)$ is

$$\begin{aligned} J\left(\mathbf{H}_{\mathbf{\Theta}_P}\right) &= \text{tr}\left[E\left\{\left[\widetilde{\mathbf{x}}(k) - \mathbf{H}_{\mathbf{\Theta}_P}\mathbf{y}(k)\right]\left[\widetilde{\mathbf{x}}(k) - \mathbf{H}_{\mathbf{\Theta}_P}\mathbf{y}(k)\right]^T\right\}\right] \\ &= \text{tr}\left(\mathbf{R}_{\widetilde{\mathbf{x}}} - 2\mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T_{\mathbf{\Theta}_P} + \mathbf{H}_{\mathbf{\Theta}_P}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T_{\mathbf{\Theta}_P}\right) \\ &= J_{\text{ds}}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right) + J_{\text{rn}}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right), \end{aligned} \tag{29}$$

where $\mathbf{R}_{\widetilde{\mathbf{x}}}$ is the correlation matrix of $\widetilde{\mathbf{x}}(k)$, $\mathbf{I}_{\text{i}} = \left[\begin{array}{cc} \mathbf{I}_M & \mathbf{0} \end{array}\right]$ is the identity filtering matrix (of size $M \times L$), with $\mathbf{I}_M$ being the $M \times M$ identity matrix,

$$\begin{aligned} J_{\text{ds}}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right) &= \text{tr}\left(\mathbf{R}_{\widetilde{\mathbf{x}}} - 2\mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T_{\mathbf{\Theta}_P} + \mathbf{H}_{\mathbf{\Theta}_P}\mathbf{R}_{\mathbf{x}}\mathbf{H}^T_{\mathbf{\Theta}_P}\right) \\ &= \text{tr}\left(\mathbf{R}_{\widetilde{\mathbf{x}}} - 2\mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\mathbf{\Theta}_P^T + \mathbf{\Theta}_P\mathbf{T}_P^T\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\mathbf{\Theta}_P^T\right) \end{aligned} \tag{30}$$

is the distortion-based MSE, and

$$\begin{aligned} J_{\text{rn}}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right) &= \text{tr}\left(\mathbf{H}_{\mathbf{\Theta}_P}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T_{\mathbf{\Theta}_P}\right) \\ &= \text{tr}\left(\mathbf{\Theta}_P\mathbf{T}_P^T\mathbf{R}_{\mathbf{v}}\mathbf{T}_P\mathbf{\Theta}_P^T\right) \end{aligned} \tag{31}$$

is the power of the residual noise. From (29), we observe that we have at least two obvious options to find $\mathbf{\Theta}_P$.

The first option consists of minimizing $J\left(\mathbf{H}_{\mathbf{\Theta}_P}\right)$. We easily get

$$\mathbf{\Theta}_P = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\left(\mathbf{T}_P^T\mathbf{R}_{\mathbf{y}}\mathbf{T}_P\right)^{-1} = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{T}_P. \tag{32}$$

Then, we deduce the reduced-rank Wiener filtering matrix:

$$\mathbf{H}_{\text{RRW}} = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\mathbf{T}_P^T. \tag{33}$$

For $P = L$, $\mathbf{H}_{\text{RRW}}$ becomes the classical Wiener filtering matrix, i.e.,

$$\mathbf{H}_{\text{W}} = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{y}}^{-1}, \tag{34}$$

since $\mathbf{R}_{\mathbf{y}}^{-1} = \mathbf{T}\mathbf{T}^T$.

[1] In practice, we may consider the P largest eigenvalues of $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$.

In the second option, we minimize $J_{\text{ds}}\left(\mathbf{H}_{\mathbf{\Theta}_P}\right)$. This leads to the minimum distortion (MD) filtering matrix:

$$\mathbf{H}_{\text{MD}} = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\left(\mathbf{T}_P^T\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\right)^{-1}\mathbf{T}_P^T, \tag{35}$$

where it is assumed that the rank of $\mathbf{R}_{\mathbf{x}}$ is at least equal to P. If the rank of $\mathbf{R}_{\mathbf{x}}$ is exactly P, then $\mathbf{H}_{\text{MD}}$ becomes the minimum variance distortionless response (MVDR) filtering matrix. If, indeed, λ_1 is of multiplicity P, (35) simplifies to

$$\mathbf{H}_{\text{MD}} = \frac{1}{\lambda_1}\mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{x}}\mathbf{T}_P\mathbf{T}_P^T. \tag{36}$$

4.2. Minimization of the SPCC

Assume that the smallest eigenvalue, λ_L, of the matrix $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$ is of multiplicity Q[2]. The corresponding eigenvectors are $\mathbf{t}_{L-Q+1}, \mathbf{t}_{L-Q+2}, \ldots, \mathbf{t}_L$. Let the filtering matrix be of the form:

$$\mathbf{H}_{\mathbf{\Theta}_Q} = \mathbf{\Theta}_Q\mathbf{T}_Q^T, \tag{37}$$

where $\mathbf{\Theta}_Q \neq \mathbf{0}$ is an arbitrary matrix of size $M \times Q$. It is clear that $\mathbf{H}_{\mathbf{\Theta}_Q}$ minimizes (9) and

$$\rho^2_{\mathbf{z}\widetilde{\mathbf{x}}_{\text{fd}}}\left(\mathbf{H}_{\mathbf{\Theta}_Q}\right) = \rho^2_{\mathbf{z}\widetilde{\mathbf{x}}_{\text{fd}},\min} = \lambda_L. \tag{38}$$

Therefore, the estimates of $\widetilde{\mathbf{v}}(k)$ and $\widetilde{\mathbf{x}}(k)$ are, respectively,

$$\widetilde{\mathbf{v}}_{\mathbf{\Theta}_Q}(k) = \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{y}(k) \tag{39}$$

and

$$\widetilde{\mathbf{x}}_{\mathbf{\Theta}_Q}(k) = \mathbf{I}_{\text{i}}\mathbf{y}(k) - \widetilde{\mathbf{v}}_{\mathbf{\Theta}_Q}(k) = \mathbf{H}'_{\mathbf{\Theta}_Q}\mathbf{y}(k), \tag{40}$$

where

$$\mathbf{H}'_{\mathbf{\Theta}_Q} = \mathbf{I}_{\text{i}} - \mathbf{H}_{\mathbf{\Theta}_Q} \tag{41}$$

is the equivalent filtering matrix for the estimation of $\widetilde{\mathbf{x}}(k)$.

There are at least two interesting ways to find $\mathbf{\Theta}_Q$. The first one is from the power of the residual noise, i.e.,

$$\begin{aligned} &J_{\text{rn}}\left(\mathbf{H}_{\mathbf{\Theta}_Q}\right) \\ &= \text{tr}\left[E\left\{\left[\widetilde{\mathbf{v}}(k) - \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{v}(k)\right]\left[\widetilde{\mathbf{v}}(k) - \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{v}(k)\right]^T\right\}\right] \\ &= \text{tr}\left(\mathbf{R}_{\widetilde{\mathbf{v}}} - 2\mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T_{\mathbf{\Theta}_Q} + \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T_{\mathbf{\Theta}_Q}\right), \end{aligned} \tag{42}$$

where $\mathbf{R}_{\widetilde{\mathbf{v}}}$ is the correlation matrix of $\widetilde{\mathbf{v}}(k)$. The second possibility is from the MSE between $\widetilde{\mathbf{x}}(k)$ and $\widetilde{\mathbf{x}}_{\mathbf{\Theta}_Q}(k)$, i.e.,

$$\begin{aligned} J\left(\mathbf{H}_{\mathbf{\Theta}_Q}\right) &= \text{tr}\left[E\left\{\left[\widetilde{\mathbf{v}}(k) - \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{y}(k)\right]\left[\widetilde{\mathbf{v}}(k) - \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{y}(k)\right]^T\right\}\right] \\ &= \text{tr}\left(\mathbf{R}_{\widetilde{\mathbf{v}}} - 2\mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T_{\mathbf{\Theta}_Q} + \mathbf{H}_{\mathbf{\Theta}_Q}\mathbf{R}_{\mathbf{y}}\mathbf{H}^T_{\mathbf{\Theta}_Q}\right). \end{aligned} \tag{43}$$

The minimization of $J_{\text{rn}}\left(\mathbf{H}_{\mathbf{\Theta}_Q}\right)$ with respect to $\mathbf{\Theta}_Q$ gives

$$\mathbf{\Theta}_Q = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{v}}\mathbf{T}_Q\left(\mathbf{T}_Q^T\mathbf{R}_{\mathbf{v}}\mathbf{T}_Q\right)^{-1}. \tag{44}$$

As a result,

$$\mathbf{H}_{\mathbf{\Theta}_Q} = \mathbf{I}_{\text{i}}\mathbf{R}_{\mathbf{v}}\mathbf{T}_Q\left(\mathbf{T}_Q^T\mathbf{R}_{\mathbf{v}}\mathbf{T}_Q\right)^{-1}\mathbf{T}_Q^T \tag{45}$$

[2] In practice, we may consider the Q smallest eigenvalues of $\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{x}}$.

and the minimum noise (MN) filter for the estimation of $\widetilde{\mathbf{x}}(k)$ is

$$\mathbf{H}'_{\mathrm{MN}} = \mathbf{I}_{\mathrm{i}} \left[\mathbf{I}_L - \mathbf{R}_{\mathbf{v}} \mathbf{T}_Q \left(\mathbf{T}_Q^T \mathbf{R}_{\mathbf{v}} \mathbf{T}_Q \right)^{-1} \mathbf{T}_Q^T \right]. \quad (46)$$

If, indeed, λ_L is of multiplicity Q, (46) simplifies to

$$\mathbf{H}'_{\mathrm{MN}} = \mathbf{I}_{\mathrm{i}} \left(\mathbf{I}_L - \frac{1}{1-\lambda_L} \mathbf{R}_{\mathbf{v}} \mathbf{T}_Q \mathbf{T}_Q^T \right). \quad (47)$$

By minimizing the MSE, we find another reduced-rank Wiener filtering matrix:

$$\begin{aligned} \mathbf{H}'_{\mathrm{RRW}} &= \mathbf{I}_{\mathrm{i}} \left[\mathbf{I}_L - \mathbf{R}_{\mathbf{v}} \mathbf{T}_Q \left(\mathbf{T}_Q^T \mathbf{R}_{\mathbf{y}} \mathbf{T}_Q \right)^{-1} \mathbf{T}_Q^T \right] \quad (48) \\ &= \mathbf{I}_{\mathrm{i}} \left(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}} \mathbf{T}_Q \mathbf{T}_Q^T \right), \end{aligned}$$

which is different from $\mathbf{H}_{\mathrm{RRW}}$. However, for $Q = L$, (48) becomes

$$\mathbf{H}'_{\mathrm{RRW}} = \mathbf{I}_{\mathrm{i}} \left(\mathbf{I}_L - \mathbf{R}_{\mathbf{v}} \mathbf{R}_{\mathbf{y}}^{-1} \right) = \mathbf{H}_{\mathrm{W}}, \quad (49)$$

which is the conventional Wiener filtering matrix.

5. SIMULATIONS

In this section, we study the noise reduction performance of the deduced filtering matrices through simulations. The clean signal used is a 30-second long speech recorded from a female speaker in a quiet office room with a sampling frequency of 8 kHz. The noise signal is a mixture of white Gaussian noise and a periodic signal (consisting of six harmonics with a fundamental frequency of 200 Hz, and the amplitudes of the harmonics are, respectively, $1, 0.8, 0.5, 0.35, 0.2, 0.1$). The ratio between the intensity of the periodic signal and the white noise is 6 dB. The noisy signal is obtained by adding this noise into the clean speech with an input SNR of 10 d-B.

The correlation matrix $\mathbf{R}_{\mathbf{y}}$ at every time instant k is computed using a short-time average with the most recent 600 samples (75 ms long). The matrix $\mathbf{R}_{\mathbf{v}}$ is computed directly from the noise signal also by a short-time average but with the most recent 960 samples (120 ms long). Then the matrix $\mathbf{R}_{\mathbf{x}}$ is computed according to $\mathbf{R}_{\mathbf{x}} = \mathbf{R}_{\mathbf{y}} - \mathbf{R}_{\mathbf{v}}$ (to ensure that this estimated speech correlation matrix is positive semi-definite, the eigenvalue decomposition is applied to it and all the small eigenvalues are set to zero). We use the output SNR as defined in (22) and the speech distortion index as measures to evaluate performance. The speech distortion index is defined as

$$\upsilon_{\mathrm{sd}} = \frac{\mathrm{tr}\left\{ E\left(\left[\widetilde{\mathbf{x}}_{\mathrm{fd}}(k) - \widetilde{\mathbf{x}}(k) \right] \left[\widetilde{\mathbf{x}}_{\mathrm{fd}}(k) - \widetilde{\mathbf{x}}(k) \right]^T \right) \right\}}{\mathrm{tr}\left(\mathbf{R}_{\widetilde{\mathbf{x}}} \right)}. \quad (50)$$

Several experiments were carried out to evaluate the impact of the values of the parameters L, M, P, and Q on the noise reduction performance. Due to space limit, we present one set of experiments in which we set $M = 10$, $P = L/2$, and $Q = L/2$, and study the performance of different filtering matrices as a function of the filter length L.

Figure 1 plots the results. It is clearly seen that the filter length L plays an important role in noise reduction performance. As the value of L increases from 10 to 80, the output SNR increases while the speech distortion index decreases for all the studied filtering matrices. However, as one can see, for $L \leq 40$, the output SNR increases and the speech distortion index decreases quickly. After that, both performance measures do not change much by further increasing L. Note that as the value of L increases, the computational complexity

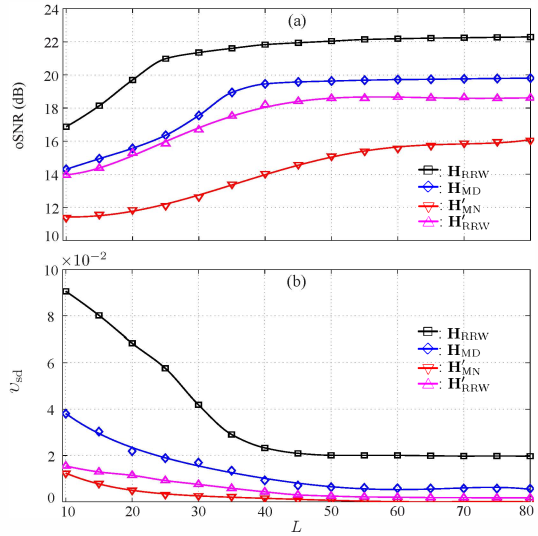

Fig. 1. Performance of the reduced-rank Wiener, MD, MN, and another reduced-rank Wiener filtering matrices as a function of the filter length L: (a) output SNR and (b) speech distortion index. $\mathrm{iSNR} = 10$ dB and $M = 10$.

also increases. As a consequence, the selection of the filter length L is a compromise between the noise reduction performance and the computational complexity. From the results shown, one can see that 40 is a good choice.

It is observed that the output SNR of the reduced-rank Wiener and MD filtering matrices is higher than that of the other two filtering matrices. This is understandable since the reduced-rank Wiener and MD filtering matrices are derived from the maximization of the SPCC, which also maximizes the output SNR. In comparison, the MN and the other reduced-rank Wiener filtering matrices have a s-maller speech distortion index. This is not surprising since these two filtering matrices are derived from the minimization of the SPCC.

6. CONCLUSIONS

This paper studied the single-channel noise reduction problem in the time domain with a filtering matrix. To obtain the optimal filtering matrix, we utilized the SPCC between the enhanced signal and filtered desired signal as the cost function. We showed how to derive the reduced-rank Wiener and minimum distortion (MD) filtering matrices by maximizing the SPCC and the minimum noise (MN) and another reduced-rank Wiener filtering matrices by minimizing the SPCC.

7. RELATION TO PRIOR WORK

Noise reduction is a challenging problem, which has attracted a significant amount of attention over the past decades due to its broad range of applications. Many methods and algorithms have been developed to deal with this challenging problem [1]– [16]. Traditionally, the noise reduction problem in the time domain is achieved with a filtering vector derived from the MSE criterion [7–13]. Recently, the SPCC was introduced as the cost function, which has been very useful in dealing with the noise reduction problem [1, 12]. In this paper, based on the SPCC, we generalized the sample-based filtering technique [12] to a block-based filtering framework and showed how to derive different optimal filtering matrices.

8. REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, pp. 757–765, May 2008.

[2] H. Lev-Ari and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise," *IEEE Trans. Speech, Audio Process.*, vol. 10, pp. 104–106, Apr. 2003.

[3] K. Hermus, P. Wambacq, and H. van Hamme, "A review of signal Subspace speech enhancement and its application to noise robust speech recognition," *EURASIP J. Advances Signal Process.*, vol. 2007(1), p. 15, 2007.

[4] P. C. Hansen and S. H. Jensen, "Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: survey and analysis," *EURASIP J. Advances Signal Process.*, vol. 2007(1), p. 24, 2007.

[5] P. Hansen and S. Jensen, "FIR filter representations of reduced-rank noise reduction," *IEEE Trans. Signal Process.*, vol. 46, pp. 1737–1741, June 1998.

[6] G. Kim, Y. Lu, Y. Hu, and P. Loizou, "An algorithm that improves speech intelligibility in noise," *J. Acoust. Soc. Am.*, vol. 126, pp. 1486–1494, 2009.

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

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

[9] 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.

[10] J. Benesty, J. Chen, Y. Huang, and T. Gaensler, "Time-domain noise reduction based on an orthogonal decomposition for desired signal extraction," *J. Acoust. Soc. Am.*, vol. 132, pp. 452–464, July 2012.

[11] A. Schasse and R. Martin, "Online inter-frame correlation estimation methods for speech enhancement in frequency subbands," in *Proc. IEEE ICASSP*, 2013, pp. 7482–7486.

[12] J. Yu, J. Benesty, G. Huang, and J. Chen, "Examples of optimal noise reduction filters derived from the squared Pearson correlation coefficient," in *Proc. IEEE ICASSP*, 2014, pp. 1571–1575.

[13] T. Long, J. Chen, J. Benesty, and Z. Zhang, "Single-channel noise reduction using optimal rectangular filtering matrices," *J. Acoust. Soc. Am.*, vol. 133, pp. 1090–1101, Feb. 2013.

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

[15] P. C. Hansen, S. H. Jensen, "Prewhitening for rank-deficient noise in subspace methods for noise reduction," *IEEE Trans. Signal Process.*, vol. 53, pp. 3718–3726 , Sept. 2005.

[16] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech, Audio Process.*, vol. 8, pp. 159–167, Mar. 2000.

[17] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1968.