

Aalborg Universitet



AALBORG
UNIVERSITY

A Class of Optimal Rectangular Filtering Matrices for Single-Channel Signal Enhancement in the Time Domain

Jensen, Jesper Rindom; Benesty, Jacob; Christensen, Mads Græsbøll; Chen, Jingdong

Published in:

I E E Transactions on Audio, Speech and Language Processing

DOI (link to publication from Publisher):

[10.1109/TASL.2013.2280215](https://doi.org/10.1109/TASL.2013.2280215)

Publication date:

2013

Document Version

Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Jensen, J. R., Benesty, J., Christensen, M. G., & Chen, J. (2013). A Class of Optimal Rectangular Filtering Matrices for Single-Channel Signal Enhancement in the Time Domain. *I E E Transactions on Audio, Speech and Language Processing*, 21(12), 2595-2606. <https://doi.org/10.1109/TASL.2013.2280215>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

A Class of Optimal Rectangular Filtering Matrices for Single-Channel Signal Enhancement in the Time Domain

Jesper Rindom Jensen*, *Member, IEEE*, Jacob Benesty, Mads Græsbøll Christensen, *Senior Member, IEEE*, and Jingdong Chen, *Senior Member, IEEE*

Abstract—In this paper, we introduce a new class of optimal rectangular filtering matrices for single-channel speech enhancement. The new class of filters exploits the fact that the dimension of the signal subspace is lower than that of the full space. By doing this, extra degrees of freedom in the filters, that are otherwise reserved for preserving the signal subspace, can be used for achieving an improved output signal-to-noise ratio (SNR). Moreover, the filters allow for explicit control of the tradeoff between noise reduction and speech distortion via the chosen rank of the signal subspace. An interesting aspect is that the framework in which the filters are derived unifies the ideas of optimal filtering and subspace methods. A number of different optimal filter designs are derived in this framework, and the properties and performance of these are studied using both synthetic, periodic signals and real signals. The results show a number of interesting things. Firstly, they show how speech distortion can be traded for noise reduction and vice versa in a seamless manner. Moreover, the introduced filter designs are capable of achieving both the upper and lower bounds for the output SNR via the choice of a single parameter.

Index Terms—Noise reduction, signal enhancement, time-domain filtering, maximum SNR filtering matrix, Wiener filtering matrix, MVDR filtering matrix, tradeoff filtering matrix.

I. INTRODUCTION

The problem of speech enhancement, namely that of estimating a desired speech signal from noisy observations [1]–[3], is one of the oldest problems of our community, with a history that dates back to the dawn of signal processing, and it remains a widely studied problem today. It occurs in many systems and devices, including voice over IP, hearing aids, teleconferencing, mobile telephony, etc. There are primarily two reasons for this. Firstly, noise has a detrimental impact on the perceived quality and intelligibility of speech signals and causes listener fatigue under extended exposure. Secondly, many speech processing systems or components are designed under the premise that only one, clean signal is present at the time. This is, most often, done to simplify the design of these, like in the codebooks used in speech coders and in

the statistical models used in automatic speech recognizers. Even though more and more systems are now using multiple channels obtained using, for example, microphone arrays, many systems today are still based on only a single channel, and this is also the context in which we will study the speech enhancement problem.

The speech enhancement problem can be posed as a filtering problem, wherein an estimate of the desired speech signal is obtained via filtering of the observed, noisy signal. An example of this is the classical Wiener filter. Such filtering approaches often require that either an estimate of the speech statistics or the noise statistics be found or known, and in the past decade, most efforts in improving speech enhancement algorithms has been devoted to the problem of estimating the noise statistics, with some examples being [4]–[7]. Recently, a number of important advances have, however, been made formulating different kinds of optimal filters. These include the adaptation of the linearly constrained minimum variance (LCMV) and the minimum variance distortionless response (MVDR) principles to speech enhancement [3], [8] in combination with the orthogonal [3] and harmonic decompositions [9], as well as the extension of these to non-causal filters [10].

An alternative approach to speech enhancement is so-called subspace methods [11], [12], wherein bases of the signal and noise subspaces are obtained from the eigenvalue decomposition of the covariance matrix. Then enhancement is performed by modifying the eigenvalues corresponding to the signal and noise subspaces after which an estimate of the clean signal can be obtained. In the literature, the subspace methods are usually described as a competing approach to speech enhancement, although some interpretations of these approaches as filtering exist [13]. For an up-to-date and complete overview of subspace methods for speech enhancement, we refer the interested reader to [14].

In this paper, we introduce a new class of optimal filters that combines the notion of subspace-based enhancement with classical filtering approaches. As such, the proposed approach unifies subspace and filtering methods in a common framework. More specifically, we show how to exploit the nullspace of the desired signal correlation matrix to derive a class of optimal rectangular filtering matrices for single-channel signal enhancement in the time domain. In this framework, we show that it is clear how the output SNR is bounded, how we can design a filter to reach this bound, and how we can design

This work was supported by the Villum foundation.

J. R. Jensen and M. G. Christensen are with the Audio Analysis Lab, Dept. of Architecture, Design & Media Technology, Aalborg University, Denmark (e-mail: {jrj,mgc}@create.aau.dk).

J. Benesty is with INRS-EMT, University of Quebec, Montreal, Canada (e-mail: benesty@emt.inrs.ca) and was with the Audio Analysis Lab, Dept. of Architecture, Design & Media Technology, Aalborg University, Denmark when this work was carried out.

J. Chen is with Northwestern Polytechnical University, Xi'an, Shaanxi, China (e-mail: jingdongchen@ieee.org).

filters with lower output SNRs that instead give lower or no distortion of the desired signal. In some of the filter designs, a tuning parameter is available, which directly enables trading off noise reduction for a lower distortion of the desired signal.

The remainder of this paper is organized as follows. In Section II, the basic signal model is introduced and the speech enhancement problem is stated, after which the linear filtering approach with a rectangular filtering matrix is introduced in Section III. Then, in Section IV, some performance measures are introduced and used to analyze and bound the performance of the enhancement filters. In Section V, various optimal rectangular filtering matrices are derived. These include the maximum SNR, Wiener, and MVDR filters as well as two tradeoff filters. The performance and properties of these filters are then studied in Section VI for the case of periodic signals, a class of signals to which voiced speech belongs. Finally, some results obtained for real speech signals are presented in Section VII, and Section VIII concludes on the work.

II. SIGNAL MODEL AND PROBLEM FORMULATION

The signal enhancement (or noise reduction) problem considered in this work is one of recovering the desired signal (or clean signal) $x(k)$, with k being the discrete-time index, from the noisy observation (sensor signal):

$$y(k) = x(k) + v(k), \quad (1)$$

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

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

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

where

$$\mathbf{y}(k) = [y(k) \quad y(k-1) \quad \cdots \quad y(k-L+1)]^T \quad (3)$$

is a vector of length L , $(\cdot)^T$ denotes the transpose of a vector or a matrix, and $\mathbf{x}(k)$ and $\mathbf{v}(k)$ are defined in a similar way to $\mathbf{y}(k)$ from (3). 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

$$\mathbf{R}_y = E[\mathbf{y}(k)\mathbf{y}^T(k)] = \mathbf{R}_x + \mathbf{R}_v, \quad (4)$$

where $E[\cdot]$ denotes the mathematical expectation, and $\mathbf{R}_x = E[\mathbf{x}(k)\mathbf{x}^T(k)]$ and $\mathbf{R}_v = E[\mathbf{v}(k)\mathbf{v}^T(k)]$ 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., its rank is equal to L . In the rest, we assume that the rank of the desired signal correlation matrix, \mathbf{R}_x , is equal to P , where P is smaller than L . This assumption is reasonable in several applications such as speech enhancement, where the speech signal can be modeled as the sum of a small number of sinusoids. In any case, we can always choose L much greater than P . Then, the objective of signal enhancement (or noise reduction) is to estimate the desired signal vector, $\mathbf{x}(k)$, or any known linear transformation of it from $\mathbf{y}(k)$. This should be done in such a

way that the noise is reduced as much as possible with little or no distortion of the desired signal.

Using the well-known eigenvalue decomposition, the desired signal correlation matrix can be diagonalized as [15]

$$\mathbf{R}_x = \mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^T, \quad (5)$$

where

$$\mathbf{Q}_x = [\mathbf{q}_{x,1} \quad \mathbf{q}_{x,2} \quad \cdots \quad \mathbf{q}_{x,L}] \quad (6)$$

is an orthogonal matrix, i.e., $\mathbf{Q}_x^T \mathbf{Q}_x = \mathbf{Q}_x \mathbf{Q}_x^T = \mathbf{I}_L$, with \mathbf{I}_L being the $L \times L$ identity matrix, and

$$\mathbf{\Lambda}_x = \text{diag}(\lambda_{x,1}, \lambda_{x,2}, \dots, \lambda_{x,L}) \quad (7)$$

is a diagonal matrix. The orthonormal vectors $\mathbf{q}_{x,1}, \mathbf{q}_{x,2}, \dots, \mathbf{q}_{x,L}$ are the eigenvectors corresponding, respectively, to the eigenvalues $\lambda_{x,1}, \lambda_{x,2}, \dots, \lambda_{x,L}$ of the matrix \mathbf{R}_x , where $\lambda_{x,1} \geq \lambda_{x,2} \geq \cdots \geq \lambda_{x,P} > 0$ and $\lambda_{x,P+1} = \lambda_{x,P+2} = \cdots = \lambda_{x,L} = 0$. Let

$$\mathbf{Q}_x = [\mathbf{T}_x \quad \mathbf{\Upsilon}_x], \quad (8)$$

where the $L \times P$ matrix \mathbf{T}_x contains the eigenvectors corresponding to the nonzero eigenvalues of \mathbf{R}_x and the $L \times (L-P)$ matrix $\mathbf{\Upsilon}_x$ contains the eigenvectors corresponding to the null eigenvalues of \mathbf{R}_x . It can be verified that

$$\mathbf{I}_L = \mathbf{T}_x \mathbf{T}_x^T + \mathbf{\Upsilon}_x \mathbf{\Upsilon}_x^T. \quad (9)$$

Notice that $\mathbf{T}_x \mathbf{T}_x^T$ and $\mathbf{\Upsilon}_x \mathbf{\Upsilon}_x^T$ are two orthogonal projection matrices of rank P and $L-P$, respectively. Hence, $\mathbf{T}_x \mathbf{T}_x^T$ is the orthogonal projector onto the desired signal subspace where all the energy of the desired signal is concentrated and $\mathbf{\Upsilon}_x \mathbf{\Upsilon}_x^T$ is the orthogonal projector onto the null subspace. Using (9), we can write the desired signal vector as

$$\mathbf{x}(k) = \mathbf{Q}_x \mathbf{Q}_x^T \mathbf{x}(k) = \mathbf{T}_x \tilde{\mathbf{x}}(k), \quad (10)$$

where $\tilde{\mathbf{x}}(k) = \mathbf{T}_x^T \mathbf{x}(k)$ is the transformed desired signal vector of length P . Therefore, the signal model for noise reduction becomes

$$\mathbf{y}(k) = \mathbf{T}_x \tilde{\mathbf{x}}(k) + \mathbf{v}(k). \quad (11)$$

Fundamentally, from the L observations, we wish to estimate the P components of the transformed desired signal, i.e., $\tilde{\mathbf{x}}(k)$. Thanks to this transformation and the nullspace of \mathbf{R}_x , we are able to reduce the dimension of the desired signal vector that we want to estimate. Indeed, there is no need to use the subspace $\mathbf{\Upsilon}_x$ since it contains no desired signal information. From (11), we give another form of the correlation matrix of $\mathbf{y}(k)$:

$$\mathbf{R}_y = \mathbf{T}_x \mathbf{R}_{\tilde{\mathbf{x}}} \mathbf{T}_x^T + \mathbf{R}_v = \mathbf{T}_x \mathbf{\Lambda}_{\tilde{\mathbf{x}}} \mathbf{T}_x^T + \mathbf{R}_v, \quad (12)$$

where

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{x}}} &= E[\tilde{\mathbf{x}}(k)\tilde{\mathbf{x}}^T(k)] = \mathbf{T}_x^T \mathbf{R}_x \mathbf{T}_x \\ &= \mathbf{T}_x^T \mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^T \mathbf{T}_x \\ &= \text{diag}(\lambda_{x,1}, \lambda_{x,2}, \dots, \lambda_{x,P}) = \mathbf{\Lambda}_{\tilde{\mathbf{x}}} \end{aligned} \quad (13)$$

and, obviously, $\mathbf{R}_x = \mathbf{T}_x \mathbf{R}_{\tilde{\mathbf{x}}} \mathbf{T}_x^T = \mathbf{T}_x \mathbf{\Lambda}_{\tilde{\mathbf{x}}} \mathbf{T}_x^T$.

III. LINEAR FILTERING WITH A RECTANGULAR MATRIX

From the general linear filtering approach [1], [3], [11], [16], [12], we can estimate the desired signal vector, $\tilde{\mathbf{x}}(k)$, by applying a linear transformation to the observation signal vector, $\mathbf{y}(k)$, i.e.,

$$\begin{aligned}\tilde{\mathbf{z}}(k) &= \tilde{\mathbf{H}}\mathbf{y}(k) = \tilde{\mathbf{H}}[\mathbf{x}(k) + \mathbf{v}(k)] \\ &= \tilde{\mathbf{x}}_{\text{fd}}(k) + \tilde{\mathbf{v}}_{\text{rn}}(k),\end{aligned}\quad (14)$$

where $\tilde{\mathbf{z}}(k)$ is supposed to be the estimate of $\tilde{\mathbf{x}}(k)$,

$$\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{h}}_1 & \tilde{\mathbf{h}}_2 & \dots & \tilde{\mathbf{h}}_P \end{bmatrix}^T \quad (15)$$

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

$$\tilde{\mathbf{h}}_p = \begin{bmatrix} \tilde{h}_{p,0} & \tilde{h}_{p,1} & \dots & \tilde{h}_{p,L-1} \end{bmatrix}^T, \quad p = 1, 2, \dots, P \quad (16)$$

are finite-impulse-response (FIR) filters of length L ,

$$\tilde{\mathbf{x}}_{\text{fd}}(k) = \tilde{\mathbf{H}}\mathbf{x}(k) = \tilde{\mathbf{H}}\mathbf{T}_x\tilde{\mathbf{x}}(k) \quad (17)$$

is the filtered transformed desired signal, and

$$\tilde{\mathbf{v}}_{\text{rn}}(k) = \tilde{\mathbf{H}}\mathbf{v}(k) \quad (18)$$

is the residual noise. As a result, the estimate of $\mathbf{x}(k)$ is supposed to be

$$\mathbf{z}(k) = \mathbf{T}_x\tilde{\mathbf{z}}(k) = \mathbf{T}_x\tilde{\mathbf{H}}\mathbf{y}(k) = \mathbf{H}\mathbf{y}(k), \quad (19)$$

where

$$\mathbf{H} = \mathbf{T}_x\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \dots & \mathbf{h}_L \end{bmatrix}^T \quad (20)$$

is the filtering matrix of size $L \times L$ that leads to the estimation of $\mathbf{x}(k)$. The correlation matrix of $\tilde{\mathbf{z}}(k)$ is then

$$\mathbf{R}_{\tilde{\mathbf{z}}} = E[\tilde{\mathbf{z}}(k)\tilde{\mathbf{z}}^T(k)] = \mathbf{R}_{\tilde{\mathbf{x}}_{\text{fd}}} + \mathbf{R}_{\tilde{\mathbf{v}}_{\text{rn}}}, \quad (21)$$

where

$$\mathbf{R}_{\tilde{\mathbf{x}}_{\text{fd}}} = \tilde{\mathbf{H}}\mathbf{R}_x\tilde{\mathbf{H}}^T = \tilde{\mathbf{H}}\mathbf{T}_x\Lambda_{\tilde{\mathbf{x}}}\mathbf{T}_x^T\tilde{\mathbf{H}}^T, \quad (22)$$

$$\mathbf{R}_{\tilde{\mathbf{v}}_{\text{rn}}} = \tilde{\mathbf{H}}\mathbf{R}_v\tilde{\mathbf{H}}^T. \quad (23)$$

We also observe that $\mathbf{R}_z = \mathbf{T}_x\mathbf{R}_{\tilde{\mathbf{z}}}\mathbf{T}_x^T$ and $\text{tr}(\mathbf{R}_z) = \text{tr}(\mathbf{R}_{\tilde{\mathbf{z}}})$, where $\text{tr}(\cdot)$ denotes the trace of a square matrix. The correlation matrix of $\tilde{\mathbf{z}}(k)$ or $\mathbf{z}(k)$ is helpful in defining meaningful performance measures.

IV. PERFORMANCE MEASURES

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

A. Noise Reduction

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

$$\text{iSNR} = \frac{\text{tr}(\mathbf{R}_x)}{\text{tr}(\mathbf{R}_v)} = \frac{\sigma_x^2}{\sigma_v^2},$$

where $\sigma_x^2 = E[x^2(k)]$ and $\sigma_v^2 = E[v^2(k)]$ are the variances of $x(k)$ and $v(k)$, respectively.

The output SNR, obtained from (21), helps quantify the SNR after filtering. It is given by

$$\begin{aligned}\text{oSNR}(\tilde{\mathbf{H}}) &= \frac{\text{tr}(\mathbf{R}_{\tilde{\mathbf{x}}_{\text{fd}}})}{\text{tr}(\mathbf{R}_{\tilde{\mathbf{v}}_{\text{rn}}})} = \frac{\text{tr}(\tilde{\mathbf{H}}\mathbf{R}_x\tilde{\mathbf{H}}^T)}{\text{tr}(\tilde{\mathbf{H}}\mathbf{R}_v\tilde{\mathbf{H}}^T)} \\ &= \frac{\sum_{p=1}^P \tilde{\mathbf{h}}_p^T \mathbf{R}_x \tilde{\mathbf{h}}_p}{\sum_{p=1}^P \tilde{\mathbf{h}}_p^T \mathbf{R}_v \tilde{\mathbf{h}}_p}.\end{aligned}\quad (24)$$

The objective is to find an appropriate $\tilde{\mathbf{H}}$ to make the output SNR greater than the input SNR. Consequently, the quality of the noisy signal will be enhanced. It can be shown that [3]

$$\text{oSNR}(\tilde{\mathbf{H}}) \leq \max_p \frac{\tilde{\mathbf{h}}_p^T \mathbf{R}_x \tilde{\mathbf{h}}_p}{\tilde{\mathbf{h}}_p^T \mathbf{R}_v \tilde{\mathbf{h}}_p}, \quad (25)$$

which implies that

$$\text{oSNR}(\tilde{\mathbf{H}}) \leq \lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x), \quad (26)$$

where $\lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x)$ is the maximum eigenvalue of the matrix $\mathbf{R}_v^{-1}\mathbf{R}_x$. This shows how the output SNR is upper bounded. It is easy to check that

$$\text{oSNR}(\mathbf{H}) = \frac{\text{tr}(\mathbf{H}\mathbf{R}_x\mathbf{H}^T)}{\text{tr}(\mathbf{H}\mathbf{R}_v\mathbf{H}^T)} = \text{oSNR}(\tilde{\mathbf{H}}) \quad (27)$$

and

$$\text{oSNR}(\mathbf{H}) \leq \max_l \frac{\mathbf{h}_l^T \mathbf{R}_x \mathbf{h}_l}{\mathbf{h}_l^T \mathbf{R}_v \mathbf{h}_l}. \quad (28)$$

Fundamentally, there is no difference between $\tilde{\mathbf{H}}$ and \mathbf{H} . Both matrices lead to the same result as we should expect.

The noise reduction factor quantifies the amount of noise being rejected by $\tilde{\mathbf{H}}$. 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.,

$$\xi_{\text{nr}}(\tilde{\mathbf{H}}) = \frac{\text{tr}(\mathbf{R}_v)}{\text{tr}(\tilde{\mathbf{H}}\mathbf{R}_v\tilde{\mathbf{H}}^T)} = \xi_{\text{nr}}(\mathbf{H}). \quad (29)$$

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

B. Desired Signal Distortion

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

$$\begin{aligned} \xi_{\text{sr}}(\tilde{\mathbf{H}}) &= \frac{\text{tr}(\mathbf{R}_{\tilde{\mathbf{x}}})}{\text{tr}(\mathbf{R}_{\tilde{\mathbf{x}}_{\text{fd}}})} = \frac{\text{tr}(\Lambda_{\tilde{\mathbf{x}}})}{\text{tr}(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}}\Lambda_{\tilde{\mathbf{x}}}\mathbf{T}_{\mathbf{x}}^T\tilde{\mathbf{H}}^T)} \quad (30) \\ &= \xi_{\text{sr}}(\mathbf{H}). \end{aligned}$$

Clearly, a rectangular filtering matrix that does not affect the desired signal requires the constraint:

$$\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}} = \mathbf{I}_P, \quad (31)$$

where \mathbf{I}_P is the $P \times P$ identity matrix. Hence, $\xi_{\text{sr}}(\tilde{\mathbf{H}}) = 1$ in the absence of distortion and $\xi_{\text{sr}}(\tilde{\mathbf{H}}) > 1$ in the presence of distortion. Taking the minimum ℓ_2 -norm solution of (31), we get

$$\tilde{\mathbf{H}} = (\mathbf{T}_{\mathbf{x}}^T\mathbf{T}_{\mathbf{x}})^{-1}\mathbf{T}_{\mathbf{x}}^T = \mathbf{T}_{\mathbf{x}}^T. \quad (32)$$

This solution corresponds to the MVDR filter for the white noise case (see Subsection V-C).

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

$$\frac{\text{oSNR}(\tilde{\mathbf{H}})}{\text{iSNR}} = \frac{\xi_{\text{nr}}(\tilde{\mathbf{H}})}{\xi_{\text{sr}}(\tilde{\mathbf{H}})}. \quad (33)$$

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

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

$$\begin{aligned} v_{\text{sd}}(\tilde{\mathbf{H}}) &= \frac{E\left\{[\tilde{\mathbf{x}}_{\text{fd}}(k) - \tilde{\mathbf{x}}(k)]^T[\tilde{\mathbf{x}}_{\text{fd}}(k) - \tilde{\mathbf{x}}(k)]\right\}}{\text{tr}(\mathbf{R}_{\tilde{\mathbf{x}}})} \quad (34) \\ &= \frac{\text{tr}\left[\left(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}} - \mathbf{I}_P\right)\Lambda_{\tilde{\mathbf{x}}}\left(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}} - \mathbf{I}_P\right)^T\right]}{\text{tr}(\Lambda_{\tilde{\mathbf{x}}})} \\ &= v_{\text{sd}}(\mathbf{H}). \end{aligned}$$

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

C. MSE Criterion

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

$$\tilde{\mathbf{e}}(k) = \tilde{\mathbf{z}}(k) - \tilde{\mathbf{x}}(k) = \tilde{\mathbf{H}}\mathbf{y}(k) - \tilde{\mathbf{x}}(k),$$

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

$$\tilde{\mathbf{e}}(k) = \tilde{\mathbf{e}}_{\text{ds}}(k) + \tilde{\mathbf{e}}_{\text{rs}}(k), \quad (35)$$

where

$$\tilde{\mathbf{e}}_{\text{ds}}(k) = \tilde{\mathbf{x}}_{\text{fd}}(k) - \tilde{\mathbf{x}}(k) = \left(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}} - \mathbf{I}_P\right)\tilde{\mathbf{x}}(k) \quad (36)$$

is the signal distortion due to the rectangular filtering matrix and

$$\tilde{\mathbf{e}}_{\text{rs}}(k) = \tilde{\mathbf{v}}_{\text{rn}}(k) = \tilde{\mathbf{H}}\mathbf{v}(k) \quad (37)$$

represents the residual noise. Therefore, the MSE criterion is

$$\begin{aligned} J(\tilde{\mathbf{H}}) &= \text{tr}\{E[\tilde{\mathbf{e}}(k)\tilde{\mathbf{e}}^T(k)]\} \quad (38) \\ &= \text{tr}(\Lambda_{\tilde{\mathbf{x}}}) + \text{tr}(\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{y}}\tilde{\mathbf{H}}^T) - 2\text{tr}(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}}\Lambda_{\tilde{\mathbf{x}}}). \end{aligned}$$

Using the fact that $E[\tilde{\mathbf{e}}_{\text{ds}}(k)\tilde{\mathbf{e}}_{\text{rs}}^T(k)] = \mathbf{0}$, $J(\tilde{\mathbf{H}})$ can be expressed as the sum of two other MSEs, i.e.,

$$\begin{aligned} J(\tilde{\mathbf{H}}) &= \text{tr}\{E[\tilde{\mathbf{e}}_{\text{ds}}(k)\tilde{\mathbf{e}}_{\text{ds}}^T(k)]\} + \text{tr}\{E[\tilde{\mathbf{e}}_{\text{rs}}(k)\tilde{\mathbf{e}}_{\text{rs}}^T(k)]\} \\ &= J_{\text{ds}}(\tilde{\mathbf{H}}) + J_{\text{rs}}(\tilde{\mathbf{H}}), \quad (39) \end{aligned}$$

where

$$\begin{aligned} J_{\text{ds}}(\tilde{\mathbf{H}}) &= \text{tr}\left[\left(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}} - \mathbf{I}_P\right)\Lambda_{\tilde{\mathbf{x}}}\left(\tilde{\mathbf{H}}\mathbf{T}_{\mathbf{x}} - \mathbf{I}_P\right)^T\right] \\ &= \text{tr}(\Lambda_{\tilde{\mathbf{x}}})v_{\text{sd}}(\tilde{\mathbf{H}}) \quad (40) \end{aligned}$$

and

$$J_{\text{rs}}(\tilde{\mathbf{H}}) = \text{tr}(\tilde{\mathbf{H}}\mathbf{R}_{\mathbf{v}}\tilde{\mathbf{H}}^T) = \frac{\text{tr}(\mathbf{R}_{\mathbf{v}})}{\xi_{\text{nr}}(\tilde{\mathbf{H}})}. \quad (41)$$

We deduce that

$$\begin{aligned} \frac{J_{\text{ds}}(\tilde{\mathbf{H}})}{J_{\text{rs}}(\tilde{\mathbf{H}})} &= \text{iSNR} \cdot \xi_{\text{nr}}(\tilde{\mathbf{H}}) \cdot v_{\text{sd}}(\tilde{\mathbf{H}}) \\ &= \text{oSNR}(\tilde{\mathbf{H}}) \cdot \xi_{\text{sr}}(\tilde{\mathbf{H}}) \cdot v_{\text{sd}}(\tilde{\mathbf{H}}). \quad (42) \end{aligned}$$

From (40)–(42), we observe how the MSEs are related to the performance measures.

V. OPTIMAL RECTANGULAR FILTERING MATRICES

In this section, we derive the most important rectangular filtering matrices that can help mitigate the level of the noise picked up by the sensor signal. We will see how these optimal matrices depend explicitly on the desired signal subspace and, in some cases, how the nullspace of $\mathbf{R}_{\mathbf{x}}$ is exploited.

A. Maximum SNR

From Subsection IV-A, we know that the output SNR is upper bounded by $\lambda_{\text{max}}(\mathbf{R}_{\mathbf{v}}^{-1}\mathbf{R}_{\mathbf{x}})$, which we can consider as the maximum possible output SNR. Then, it is easy to verify that with

$$\tilde{\mathbf{H}}_{\text{max}} = \begin{bmatrix} \varsigma_1 \mathbf{b}_{\text{max}}^T \\ \varsigma_2 \mathbf{b}_{\text{max}}^T \\ \vdots \\ \varsigma_P \mathbf{b}_{\text{max}}^T \end{bmatrix}, \quad (43)$$

where ς_p , $p = 1, 2, \dots, P$ are arbitrary real numbers with at least one of them different from 0, and \mathbf{b}_{max} is the eigenvector

of the matrix $\mathbf{R}_v^{-1}\mathbf{R}_x$ corresponding to $\lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x)$, we have

$$\text{oSNR}(\tilde{\mathbf{H}}_{\max}) = \lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x). \quad (44)$$

As a consequence, $\tilde{\mathbf{H}}_{\max}$ can be considered as the maximum SNR filtering matrix. Clearly,

$$\text{oSNR}(\tilde{\mathbf{H}}_{\max}) \geq \text{iSNR} \quad (45)$$

and

$$0 \leq \text{oSNR}(\tilde{\mathbf{H}}) \leq \text{oSNR}(\tilde{\mathbf{H}}_{\max}), \forall \tilde{\mathbf{H}}. \quad (46)$$

The choice of the values of ς_p , $p = 1, 2, \dots, P$ is extremely important in practice; with a poor choice of these values, the transformed desired signal vector can be highly distorted. Therefore, the ς_p 's should be found in such a way that distortion is minimized. We can rewrite the distortion-based MSE as

$$\begin{aligned} J_{\text{ds}}(\tilde{\mathbf{H}}) &= \text{tr}(\Lambda_{\tilde{\mathbf{x}}}) + \text{tr}(\tilde{\mathbf{H}}\mathbf{R}_x\tilde{\mathbf{H}}^T) - 2\text{tr}(\tilde{\mathbf{H}}\mathbf{T}_x\Lambda_{\tilde{\mathbf{x}}}) \\ &= \text{tr}(\Lambda_{\tilde{\mathbf{x}}}) + \sum_{p=1}^P \tilde{\mathbf{h}}_p^T \mathbf{R}_x \tilde{\mathbf{h}}_p - 2 \sum_{p=1}^P \lambda_{x,p} \tilde{\mathbf{h}}_p^T \mathbf{q}_{x,p}. \end{aligned} \quad (47)$$

Substituting (43) into (47), we get

$$\begin{aligned} J_{\text{ds}}(\tilde{\mathbf{H}}_{\max}) &= \text{tr}(\Lambda_{\tilde{\mathbf{x}}}) + \mathbf{b}_{\max}^T \mathbf{R}_x \mathbf{b}_{\max} \sum_{p=1}^P \varsigma_p^2 \\ &\quad - 2 \sum_{p=1}^P \varsigma_p \lambda_{x,p} \mathbf{b}_{\max}^T \mathbf{q}_{x,p} \end{aligned} \quad (48)$$

and minimizing this expression with respect to the ς_p 's, we find

$$\varsigma_p = \frac{\lambda_{x,p} \mathbf{b}_{\max}^T \mathbf{q}_{x,p}}{\mathbf{b}_{\max}^T \mathbf{R}_x \mathbf{b}_{\max}} = \frac{\lambda_{x,p} \mathbf{b}_{\max}^T \mathbf{q}_{x,p}}{\lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x)}, \quad (49)$$

where $\lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x) = \mathbf{b}_{\max}^T \mathbf{R}_x \mathbf{b}_{\max}$. Substituting these optimal values in (43), we obtain the optimal maximum SNR filtering matrix with minimum desired signal distortion:

$$\tilde{\mathbf{H}}_{\max} = \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T \frac{\mathbf{b}_{\max} \mathbf{b}_{\max}^T}{\lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x)}. \quad (50)$$

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

$$\mathbf{H}_{\max} = \mathbf{T}_x \tilde{\mathbf{H}}_{\max} = \mathbf{R}_x \frac{\mathbf{b}_{\max} \mathbf{b}_{\max}^T}{\lambda_{\max}(\mathbf{R}_v^{-1}\mathbf{R}_x)}. \quad (51)$$

B. Wiener

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

$$\begin{aligned} \tilde{\mathbf{H}}_W &= \mathbf{R}_{\tilde{\mathbf{x}}} \mathbf{T}_x^T \mathbf{R}_y^{-1} = \mathbf{T}_x^T \mathbf{R}_x \mathbf{R}_y^{-1} \\ &= \mathbf{T}_x^T (\mathbf{I}_L - \mathbf{R}_v \mathbf{R}_y^{-1}). \end{aligned} \quad (52)$$

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

$$\begin{aligned} \mathbf{H}_W &= \mathbf{T}_x \tilde{\mathbf{H}}_W = \mathbf{T}_x \mathbf{R}_x \mathbf{T}_x^T \mathbf{R}_y^{-1} \\ &= \mathbf{R}_x \mathbf{R}_y^{-1} = \mathbf{I}_L - \mathbf{R}_v \mathbf{R}_y^{-1}, \end{aligned} \quad (53)$$

which corresponds to the classical Wiener filtering matrix [1]. It is extremely important to observe that, thanks to the eigenvalue decomposition and the nullspace of \mathbf{R}_x , the size ($P \times L$) of the proposed Wiener filtering matrix is smaller than the size ($L \times L$) of the classical Wiener filtering matrix, for the estimation of the desired signal vector $\mathbf{x}(k)$, while the two methods lead to the exact same result. We deduce that the optimal Wiener filter for the estimation of $x(k-l)$, $l = 0, 1, \dots, L-1$ is

$$\begin{aligned} \mathbf{h}_{W,l+1} &= \mathbf{R}_y^{-1} \mathbf{R}_x \mathbf{i}_{l+1} \\ &= (\mathbf{I}_L - \mathbf{R}_v^{-1} \mathbf{R}_x) \mathbf{i}_{l+1}, \end{aligned} \quad (54)$$

where \mathbf{i}_{l+1} is the $(l+1)$ th column of \mathbf{I}_L .

By applying the Woodbury's identity in (12) and then substituting the result in (52), we easily deduce another form of the Wiener filtering matrix:

$$\begin{aligned} \tilde{\mathbf{H}}_W &= (\mathbf{I}_P + \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T \mathbf{R}_v^{-1} \\ &= (\Lambda_{\tilde{\mathbf{x}}}^{-1} + \mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \mathbf{T}_x^T \mathbf{R}_v^{-1}. \end{aligned} \quad (55)$$

The expression is interesting because it shows an obvious link with some other optimal rectangular filtering matrices as it will be verified later. We also have

$$\mathbf{H}_W = \mathbf{T}_x (\mathbf{I}_P + \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T \mathbf{R}_v^{-1}. \quad (56)$$

If \mathbf{R}_v is diagonal, i.e., $\mathbf{R}_v = \sigma_{\text{wn}}^2 \mathbf{I}_L$, the previous expression simplifies to

$$\mathbf{H}_W = \mathbf{T}_x (\mathbf{I}_P + \Lambda_{\tilde{\mathbf{x}}})^{-1} \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T. \quad (57)$$

This shows how the desired signal subspace is modified to get a good estimate of $\mathbf{x}(k)$ from $\mathbf{y}(k)$ with Wiener.

Property 5.1: The output SNR with the Wiener filtering matrix is always greater than or equal to the input SNR, i.e., $\text{oSNR}(\tilde{\mathbf{H}}_W) \geq \text{iSNR}$.

Obviously, we have

$$\text{oSNR}(\tilde{\mathbf{H}}_W) \leq \text{oSNR}(\tilde{\mathbf{H}}_{\max}) \quad (58)$$

and, in general,

$$v_{\text{sd}}(\tilde{\mathbf{H}}_W) \leq v_{\text{sd}}(\tilde{\mathbf{H}}_{\max}). \quad (59)$$

C. Minimum Variance Distortionless Response

The celebrated minimum variance distortionless response (MVDR) filter proposed by Capon [17], [18] is usually derived in a context where we have at least two sensors available. Interestingly, with the signal model proposed in this work, we can also derive the MVDR with one sensor only by minimizing the MSE of the residual noise, $J_{\text{rs}}(\tilde{\mathbf{H}})$, with the constraint that the desired signal is not distorted. Mathematically, this is equivalent to

$$\min_{\tilde{\mathbf{H}}} \text{tr}(\tilde{\mathbf{H}} \mathbf{R}_v \tilde{\mathbf{H}}^T) \quad \text{subject to} \quad \tilde{\mathbf{H}} \mathbf{T}_x = \mathbf{I}_P. \quad (60)$$

The solution to the above optimization problem is

$$\tilde{\mathbf{H}}_{\text{MVDR}} = (\mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \mathbf{T}_x^T \mathbf{R}_v^{-1}, \quad (61)$$

which is interesting to compare to $\tilde{\mathbf{H}}_W$ [eq. (55)]. We deduce that the MVDR filter for the estimation of $\mathbf{x}(k)$ is

$$\mathbf{H}_{\text{MVDR}} = \mathbf{T}_x (\mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \mathbf{T}_x^T \mathbf{R}_v^{-1}. \quad (62)$$

Of course, for $P = L$, the MVDR filtering matrix degenerates to the identity matrix, i.e., $\mathbf{H}_{\text{MVDR}} = \mathbf{I}_L$. As a consequence, we can state that the higher is the dimension of the nullspace of \mathbf{R}_x , the more the MVDR is efficient in terms of noise reduction. The best scenario corresponds to $P = 1$. If $\mathbf{R}_v = \sigma_{\text{wn}}^2 \mathbf{I}_L$, the MVDR simplifies to [19], [11]

$$\mathbf{H}_{\text{MVDR}} = \mathbf{T}_x \mathbf{T}_x^T. \quad (63)$$

In this case, signal enhancement consists of projecting $\mathbf{y}(k)$ onto the desired signal subspace. Obviously, with the MVDR filtering matrix, we have no distortion, i.e.,

$$\xi_{\text{sr}}(\tilde{\mathbf{H}}_{\text{MVDR}}) = 1 \quad \text{and} \quad v_{\text{sd}}(\tilde{\mathbf{H}}_{\text{MVDR}}) = 0. \quad (64)$$

Using the Woodbury's identity, we can rewrite the MVDR filtering matrix as

$$\tilde{\mathbf{H}}_{\text{MVDR}} = (\mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \mathbf{T}_x^T \mathbf{R}_v^{-1}. \quad (65)$$

From (65), we deduce the relationship between the MVDR and Wiener filtering matrices:

$$\tilde{\mathbf{H}}_{\text{MVDR}} = (\tilde{\mathbf{H}}_W \mathbf{T}_x)^{-1} \tilde{\mathbf{H}}_W. \quad (66)$$

Expression (65) can also be derived from the following reasoning. We know that

$$\mathbf{x}(k) = \mathbf{T}_x \tilde{\mathbf{x}}(k), \quad (67)$$

where \mathbf{T}_x can be seen as a temporal prediction matrix. Left multiplying the previous expression by $\tilde{\mathbf{H}}$, we see that the distortionless constraint is $\tilde{\mathbf{H}} \mathbf{T}_x = \mathbf{I}_P$. Now, by minimizing the energy at the output of the filtering matrix, i.e., $\text{tr}(\tilde{\mathbf{H}} \mathbf{R}_y \tilde{\mathbf{H}}^T)$, with the distortionless constraint, we find (65).

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

Moreover, we have

$$\text{oSNR}(\tilde{\mathbf{H}}_{\text{MVDR}}) \leq \text{oSNR}(\tilde{\mathbf{H}}_W) \leq \text{oSNR}(\tilde{\mathbf{H}}_{\text{max}}). \quad (68)$$

D. Tradeoff I

In the tradeoff approach [1], [3], 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

$$\min_{\tilde{\mathbf{H}}} J_{\text{ds}}(\tilde{\mathbf{H}}) \quad \text{subject to} \quad J_{\text{rs}}(\tilde{\mathbf{H}}) = \beta \text{tr}(\mathbf{R}_v), \quad (69)$$

where $0 < \beta < 1$ to insure that we get some noise reduction. By using a Lagrange multiplier, $\mu > 0$, to adjoin the constraint to the cost function and assuming that the matrix

$\mathbf{T}_x \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T + \mu \mathbf{R}_v$ is invertible, we easily deduce the tradeoff filtering matrix:

$$\tilde{\mathbf{H}}_{T,\mu} = \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T (\mathbf{T}_x \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T + \mu \mathbf{R}_v)^{-1}, \quad (70)$$

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

$$\tilde{\mathbf{H}}_{T,\mu} = (\mu \Lambda_{\tilde{\mathbf{x}}}^{-1} + \mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \mathbf{T}_x^T \mathbf{R}_v^{-1}, \quad (71)$$

where μ satisfies $J_{\text{rs}}(\tilde{\mathbf{H}}_{T,\mu}) = \beta \text{tr}(\mathbf{R}_v)$. Usually, μ is chosen in a heuristic way, so that for

- $\mu = 1$, $\tilde{\mathbf{H}}_{T,1} = \tilde{\mathbf{H}}_W$, which is the Wiener filtering matrix;
- $\mu = 0$, the problem in (69) does not have a solution since $(\mathbf{T}_x \Lambda_{\tilde{\mathbf{x}}} \mathbf{T}_x^T + \mu \mathbf{R}_v)$ is not invertible but one can obtain from (71) that $\tilde{\mathbf{H}}_{T,0} = \tilde{\mathbf{H}}_{\text{MVDR}}$, which is the MVDR filtering matrix;
- $\mu > 1$, results in a filtering matrix with low residual noise at the expense of high desired signal distortion (as compared to Wiener); and
- $\mu < 1$, results in a filtering matrix with high residual noise and low desired signal distortion (as compared to Wiener).

Property 5.3: The output SNR with the tradeoff filtering matrix is always greater than or equal to the input SNR, i.e., $\text{oSNR}(\tilde{\mathbf{H}}_{T,\mu}) \geq \text{iSNR}$, $\forall \mu \geq 0$.

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

$$\text{oSNR}(\tilde{\mathbf{H}}_{\text{MVDR}}) \leq \text{oSNR}(\tilde{\mathbf{H}}_W) \quad (72)$$

$$\leq \text{oSNR}(\tilde{\mathbf{H}}_{T,\mu}) \leq \text{oSNR}(\tilde{\mathbf{H}}_{\text{max}}),$$

$$0 = v_{\text{sd}}(\tilde{\mathbf{H}}_{\text{MVDR}}) \leq v_{\text{sd}}(\tilde{\mathbf{H}}_W) \leq v_{\text{sd}}(\tilde{\mathbf{H}}_{T,\mu}), \quad (73)$$

and for $\mu \leq 1$,

$$\text{oSNR}(\tilde{\mathbf{H}}_{\text{MVDR}}) \leq \text{oSNR}(\tilde{\mathbf{H}}_{T,\mu}) \quad (74)$$

$$\leq \text{oSNR}(\tilde{\mathbf{H}}_W) \leq \text{oSNR}(\tilde{\mathbf{H}}_{\text{max}}),$$

$$0 = v_{\text{sd}}(\tilde{\mathbf{H}}_{\text{MVDR}}) \leq v_{\text{sd}}(\tilde{\mathbf{H}}_{T,\mu}) \quad (75)$$

$$\leq v_{\text{sd}}(\tilde{\mathbf{H}}_W) \leq v_{\text{sd}}(\tilde{\mathbf{H}}_{\text{max}}).$$

Let us end this subsection by writing the tradeoff filtering matrix for the estimation of $\mathbf{x}(k)$:

$$\mathbf{H}_{T,\mu} = \mathbf{T}_x (\mu \Lambda_{\tilde{\mathbf{x}}}^{-1} + \mathbf{T}_x^T \mathbf{R}_v^{-1} \mathbf{T}_x)^{-1} \mathbf{T}_x^T \Phi_v^{-1}, \quad (76)$$

which clearly shows how the desired signal subspace should be modified in order to make a compromise between noise reduction and desired signal distortion.

E. Tradeoff II

We can also come up with another, and maybe more useful, tradeoff filter than the classical one by inheriting the principle behind the MVDR filter in Section V-C. Here, the principle is used to obtain a filter that minimizes the MSE of the residual noise, $J_{\text{rs}}(\tilde{\mathbf{H}})$, with the constraint that the filter should be

distortionless with respect to the $(P - q)$ th most dominant subspace components, i.e.,

$$\min_{\tilde{\mathbf{H}}_q} \text{tr} \left(\tilde{\mathbf{H}}_q \mathbf{R}_v \tilde{\mathbf{H}}_q^T \right) \quad \text{subject to} \quad \tilde{\mathbf{H}}_q \mathbf{T}_{\mathbf{x},q} = \begin{bmatrix} \mathbf{I}_{P-q} \\ \mathbf{0}_{q \times (P-q)} \end{bmatrix}, \quad (77)$$

where

$$\mathbf{T}_{\mathbf{x},q} = [\mathbf{q}_{\mathbf{x},1} \quad \mathbf{q}_{\mathbf{x},2} \quad \cdots \quad \mathbf{q}_{\mathbf{x},P-q}] \quad (78)$$

and $0 \leq q \leq P$. Obviously, q needs to be an integer, as it refers to a certain number of columns in $\mathbf{T}_{\mathbf{x}}$. Solving (77) wrt. the unknown filter response, $\tilde{\mathbf{H}}_q$ yields

$$\tilde{\mathbf{H}}_{\text{T-II},q} = \begin{bmatrix} \mathbf{I}_{P-q} \\ \mathbf{0}_{q \times (P-q)} \end{bmatrix} \left(\mathbf{T}_{\mathbf{x},q}^T \mathbf{R}_v^{-1} \mathbf{T}_{\mathbf{x},q} \right)^{-1} \mathbf{T}_{\mathbf{x},q}^T \mathbf{R}_v^{-1}. \quad (79)$$

We can then deduce that the tradeoff filter for the estimation of $\mathbf{x}(k)$ is given by

$$\mathbf{H}_{\text{T-II},q} = \mathbf{T}_{\mathbf{x},q} \left(\mathbf{T}_{\mathbf{x},q}^T \mathbf{R}_v^{-1} \mathbf{T}_{\mathbf{x},q} \right)^{-1} \mathbf{T}_{\mathbf{x},q}^T \mathbf{R}_v^{-1}. \quad (80)$$

We can then obtain different filters by using different values of q which enable us to trade off signal distortion for noise reduction. Moreover, we observe the following:

- if $q = 1$ and the noise is white, the tradeoff filter in (80) resembles the maximum SNR filter in (51), i.e., $\mathbf{H}_{\text{T-II},1} = \mathbf{H}_{\text{max}}$;
- if $q = P$, the tradeoff filter in (80) resembles the MVDR filter in (62), i.e., $\mathbf{H}_{\text{T-II},P} = \mathbf{H}_{\text{MVDR}}$; and
- if $1 < q < P$, a tradeoff filter, $\mathbf{H}_{\text{T-II},q}$, is obtained that has noise reduction and signal distortion measures in between those of the maximum SNR and MVDR filters, respectively.

The tradeoff filter proposed in this section exhibits a smooth and always increasing/decreasing behaviour in terms of output SNR and signal distortion index as a function of q . That is,

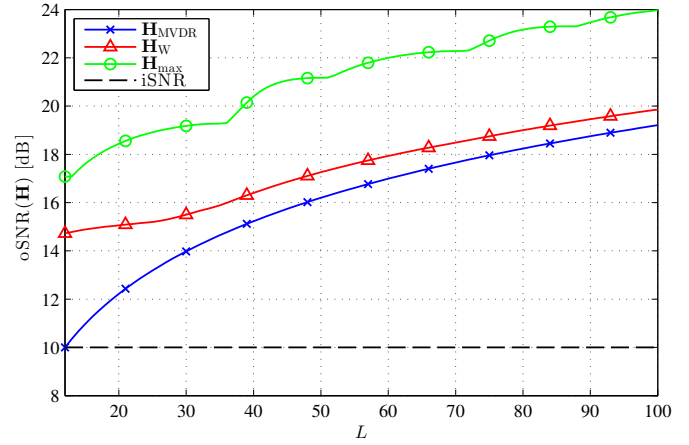
$$\text{oSNR}(\tilde{\mathbf{H}}_{\text{T-II},P}) < \text{oSNR}(\tilde{\mathbf{H}}_{\text{T-II},P-1}) < \cdots < \text{oSNR}(\tilde{\mathbf{H}}_{\text{T-II},1}), \quad (81)$$

$$0 = v(\tilde{\mathbf{H}}_{\text{T-II},P}) < v(\tilde{\mathbf{H}}_{\text{T-II},P-1}) < \cdots < v(\tilde{\mathbf{H}}_{\text{T-II},1}). \quad (82)$$

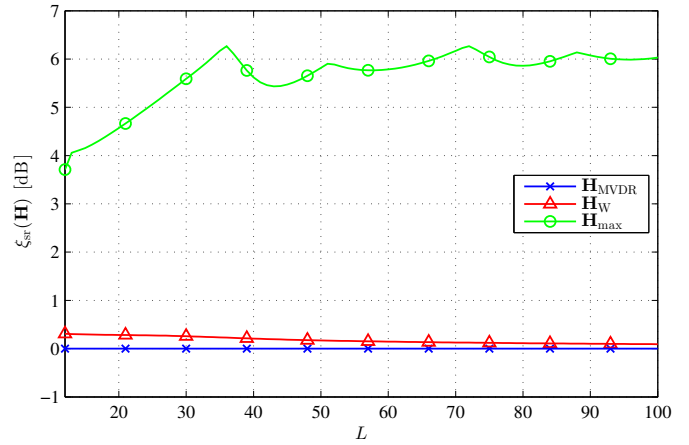
We note that the tradeoff filter, $\tilde{\mathbf{H}}_{\text{T-II},q}$, can attain the maximum output SNR with a signal distortion bounded by the distortion of the maximum SNR filter in white Gaussian noise scenarios. This is opposed to the tradeoff filter in Section V-D which may never reach the maximum SNR, and it will most likely introduce much more signal distortion than the maximum SNR filter. More details and observations on the comparison of the tradeoff filters can be found in the experimental part of the paper.

VI. CASE STUDY: PERIODIC SIGNALS

Then, we proceed with a case study of the rectangular filtering methods proposed in Sec. V. In this study, the desired signal is assumed to be periodic, which is a valid assumption for short segments of, e.g., recordings of voiced speech and musical instruments. As it becomes clear later, the periodicity assumption enables us to derive closed-form expressions for the performance measures of the filters that,



(a)



(b)

Fig. 1. Plots of (a) the output SNR and (b) the signal reduction factor for the \mathbf{H}_{MVDR} , \mathbf{H}_{W} , and \mathbf{H}_{max} filters as functions of the filter length, L .

eventually, facilitates evaluation of the filters' performance without having to estimate any statistics. This is an important observation since we can then conduct evaluations of the filters that are not disturbed by estimation errors in the statistics. On a side note, the resemblance between the filters proposed herein and previously proposed filtering methods for periodic signals [20], [21] also becomes clear from this case study.

When the desired signal is periodic, we can rewrite the signal model in (1) as

$$y(k) = \sum_{c=1}^C (\alpha_c e^{j c \omega_0 k} + \alpha_c^* e^{-j c \omega_0 k}) + v(k), \quad (83)$$

where C is the number of harmonics constituting the periodic signal, ω_0 is the fundamental frequency relating the harmonics, $\alpha_c = A_c e^{j \phi_c}$, $A_c > 0$ and ϕ_c are the complex amplitude, the real amplitude and the phase of the c th harmonic, respectively, and $(\cdot)^*$ denotes the elementwise conjugate of a scalar, vector or matrix. The single snapshot, signal model in (83) can be extended to a vector model as

$$\mathbf{y}(k) = \mathbf{Z}(\omega_0) \boldsymbol{\alpha} + \mathbf{v}(k), \quad (84)$$

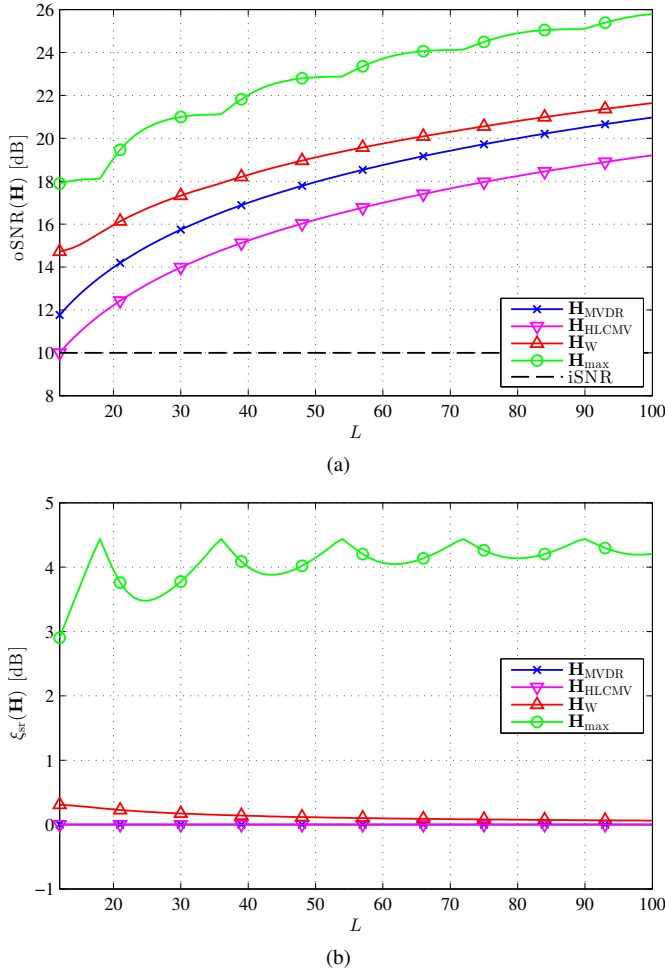


Fig. 2. Plots of (a) the output SNR and (b) the signal reduction factor for the \mathbf{H}_{MVDR} , $\mathbf{H}_{\text{HLCMV}}$, \mathbf{H}_{W} , and \mathbf{H}_{max} filters as functions of the filter length, L , when two harmonics are missing.

where

$$\mathbf{Z}(\omega_0) = [\bar{\mathbf{Z}}(\omega_0) \quad \bar{\mathbf{Z}}^*(\omega_0)], \quad (85)$$

$$[\bar{\mathbf{Z}}(\omega_0)]_{:c} = [1 \quad e^{-j\omega_0} \quad \dots \quad e^{-j\omega_0(K-1)}]^T, \quad (86)$$

$$\boldsymbol{\alpha} = [\bar{\boldsymbol{\alpha}}^T \quad \bar{\boldsymbol{\alpha}}^H]^T, \quad (87)$$

$$\bar{\boldsymbol{\alpha}} = [\alpha_1 \quad \dots \quad \alpha_C]^T, \quad (88)$$

with $(\cdot)_{:c}$ denoting the c th column of a matrix, and $(\cdot)^H$ denoting the complex conjugate transpose of a vector or matrix.

A. Link between MVDR and Harmonic LCMV Filters

In cases where the desired signal is indeed periodic and the above-mentioned model holds, the matrix $\mathbf{Z}(\omega_0)$ spans the signal subspace, i.e., $\text{range}\{\mathbf{Z}(\omega_0)\} = \text{range}\{\mathbf{T}_{\mathbf{x}}\}$ and we have that [21]

$$\mathbf{T}_{\mathbf{x}} = \mathbf{Z}(\omega_0)\mathbf{Q}, \quad (89)$$

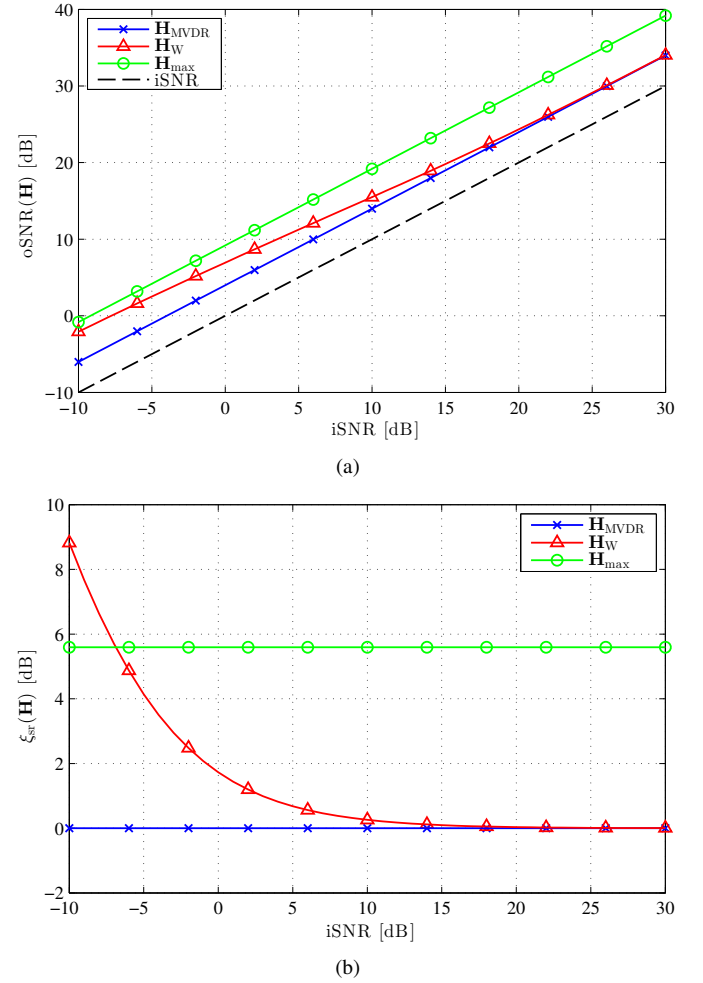


Fig. 3. Plots of (a) the output SNR and (b) the signal reduction factor for the \mathbf{H}_{MVDR} , \mathbf{H}_{W} , and \mathbf{H}_{max} filters as functions of the input SNR.

with

$$\mathbf{Q} = \mathbf{P}\mathbf{Z}^H(\omega_0)\mathbf{T}_{\mathbf{x}}\boldsymbol{\Lambda}_{\mathbf{x}}^{-1}, \quad (90)$$

$$[\mathbf{P}]_{ij} = \begin{cases} |[\boldsymbol{\alpha}]_k|^2, & \text{for } k = i = j \\ 0, & \text{for } i \neq j \end{cases}. \quad (91)$$

Substituting (89) and (90) into, e.g., the expression for the MVDR filter in (61), we get

$$\tilde{\mathbf{H}}_{\text{MVDR}} = \mathbf{Q}^{-1} [\mathbf{Z}^H(\omega_0)\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{Z}(\omega_0)]^{-1} \mathbf{Z}^H \mathbf{R}_{\mathbf{x}}^{-1}. \quad (92)$$

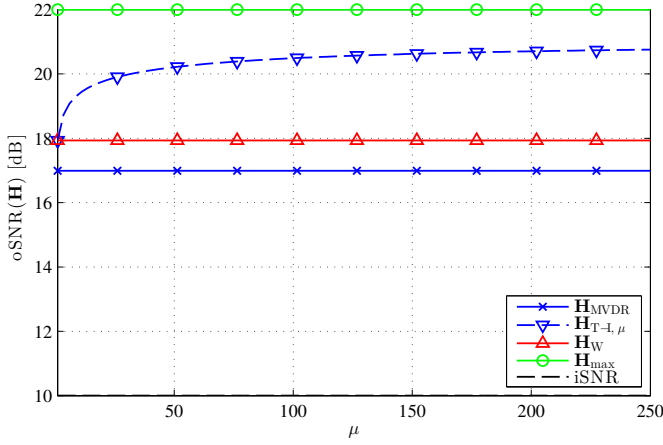
This is clearly related to the harmonic LCMV filterbank, $\tilde{\mathbf{H}}_{\text{HLCMV}}$, proposed in [20], [21] for fundamental frequency estimation as

$$\tilde{\mathbf{H}}_{\text{MVDR}} = \mathbf{Q}^{-1} \tilde{\mathbf{H}}_{\text{HLCMV}}. \quad (93)$$

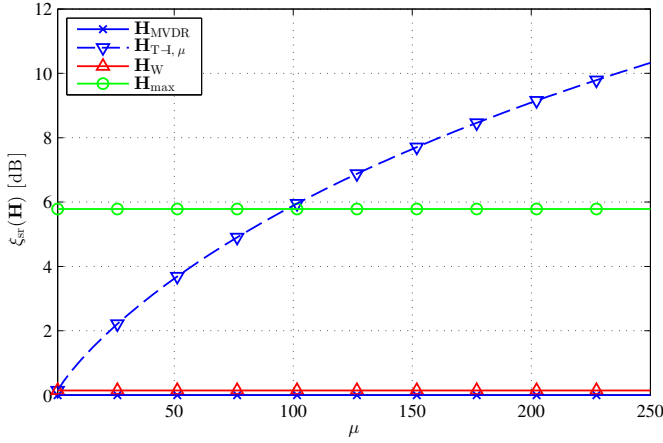
By means of the framework considered in this paper, the harmonic LCMV filterbank can be interpreted as a filterbank estimating the amplitudes of the harmonics in a transform domain where the inverse transform is $\mathbf{Z}(\omega_0)$:

$$\hat{\boldsymbol{\alpha}} = \tilde{\mathbf{H}}_{\text{HLCMV}}\mathbf{y}(k). \quad (94)$$

Adopting the idea of estimating parameters in a transform domain and applying an inverse transform on those to get an



(a)



(b)

Fig. 4. Plots of (a) the output SNR and (b) the signal reduction factor for the \mathbf{H}_{MVDR} , $\mathbf{H}_{\text{T-I}, \mu}$, \mathbf{H}_{W} , and \mathbf{H}_{max} filters as functions of μ .

estimate of $\mathbf{x}(k)$ yields the following version of the harmonic LCMV filterbank:

$$\mathbf{H}_{\text{HLCMV}} = \mathbf{Z}(\omega_0) [\mathbf{Z}^H(\omega_0) \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{Z}(\omega_0)]^{-1} \mathbf{Z}^H(\omega_0) \mathbf{R}_{\mathbf{y}}^{-1}. \quad (95)$$

Interestingly, it can be shown that this filterbank is identical to the corresponding version of the MVDR filterbank, i.e.,

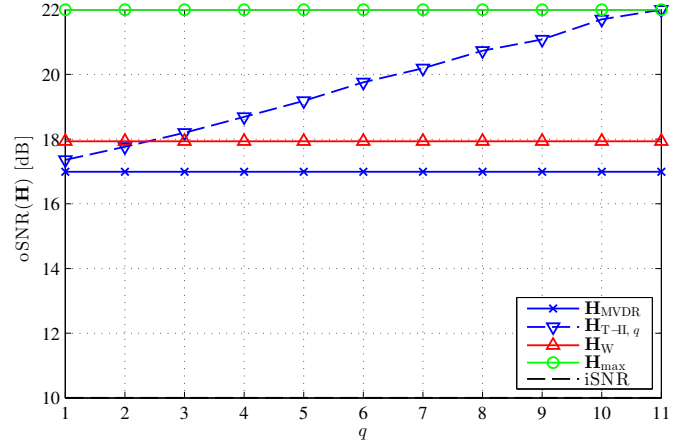
$$\mathbf{H}_{\text{MVDR}} = \mathbf{H}_{\text{HLCMV}}. \quad (96)$$

B. Performance Evaluation for Periodic Signals

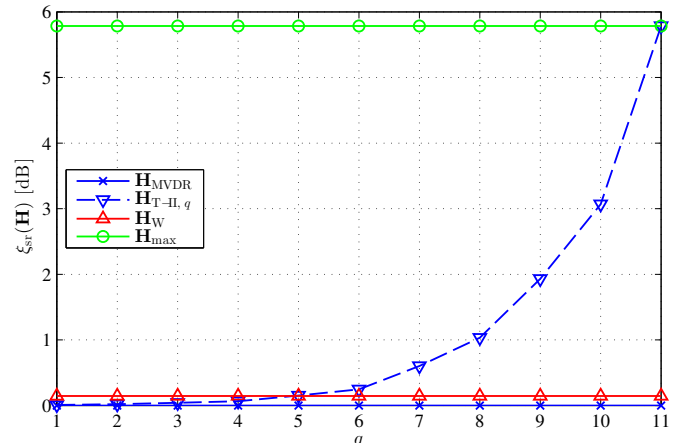
We can also further specify the model of the covariance matrix of the desired signal, when the desired signal is periodic. In that case, $\mathbf{R}_{\mathbf{x}}$ is given by [22]

$$\mathbf{R}_{\mathbf{x}} = \mathbf{Z}(\omega_0) \mathbf{P} \mathbf{Z}^H(\omega_0). \quad (97)$$

That is, the covariance matrix of the desired signal is fully specified by the fundamental frequency, the model order, and the amplitudes of the harmonics in cases with periodic, desired signals. If the covariance matrix of the noise is also known as in, e.g., the white Gaussian noise case where $\mathbf{R}_{\mathbf{v}} = \sigma_v^2 \mathbf{I}$, these expressions for the covariance matrices can be inserted in the expressions for the performance measures of the different



(a)



(b)

Fig. 5. Plots of (a) the output SNR and (b) the signal reduction factor for the \mathbf{H}_{MVDR} , $\mathbf{H}_{\text{T-II}, q}$, \mathbf{H}_{W} , and \mathbf{H}_{max} filters as functions of q .

filter designs proposed herein to get closed-form performance measure expressions.

In this way, we evaluated the filters in different scenarios with periodic signals as described in the following. The so-obtained results provide insight into how the filters would perform for enhancement of, e.g., speech and musical instrument recordings, as most of such signals can be assumed periodic for short segments. In these scenarios, we assumed that the desired signal was periodic, having a fundamental frequency of $\omega_0 = 0.175$ and $C = 6$ harmonics. The amplitudes of the harmonics were assumed to be $|\alpha| = [1, 0.8, 0.6, 0.3, 0.15, 0.1]^T$. Using this setup, we first evaluated the MVDR, Wiener, and maximum SNR filters for different filter lengths, L , and the results are depicted in Fig. 1. From the figure, we see that the maximum SNR filter expectedly has the highest output SNR, but also the highest signal reduction factor, for all different filter lengths. The Wiener filter outperforms the MVDR filter in terms of output SNR, but at the expense of signal distortion. At high filter lengths, the Wiener and MVDR filters have similar performances. Then, we again investigated the filters' performance versus the filter lengths, but with two missing harmonics, i.e., the second and fourth. In this case the rank of the signal subspace is only $P = 8$, whereas it was

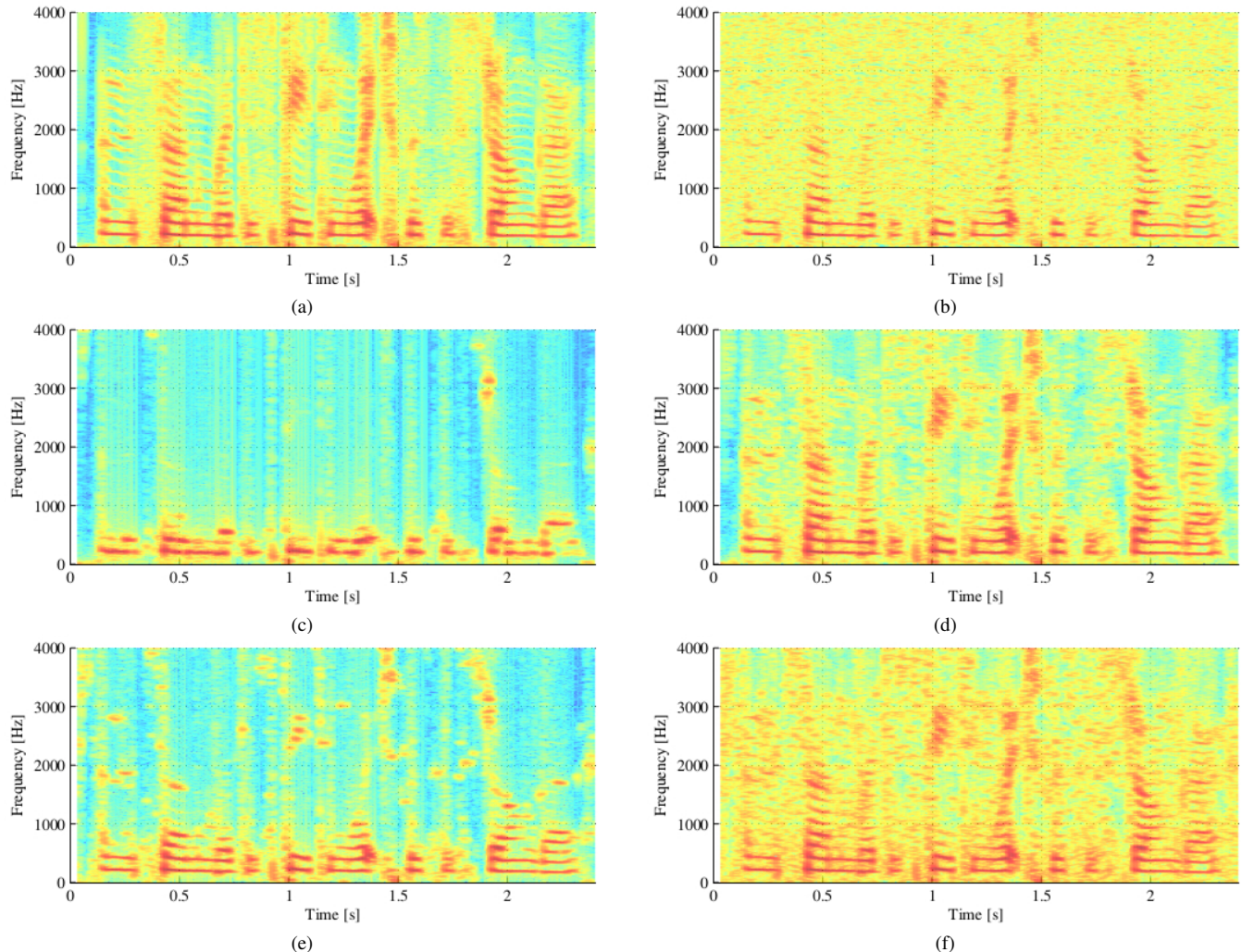


Fig. 6. Spectrograms for (a) a clean speech signal, (b) the speech signal in noise, and the noisy signals enhanced using the (c) maximum SNR, (d) Wiener, and (e)–(f) MVDR filters. The MVDR filters were applied with two different assumed model orders, i.e., (e) $P = 10$ and (f) $P = 50$.

$P = 12$ in the previous setup. This means that the MVDR filter can be designed with fewer constraints compared to the HLCMV filter, while still being distortionless. Effectively, this should leave more degrees of freedom in the filter for noise reduction. This was also confirmed by our experimental results in Fig. 2, where the MVDR filter is shown to outperform the HLCMV filter in terms of output SNR, while both filters are distortionless. We then proceeded to evaluate the filters versus different input SNRs as shown in Fig. 3. An interesting observation from this experiment is that the Wiener filter has a higher signal reduction factor than the maximum SNR filter at low iSNRs, while it also has a lower output SNR. Furthermore, the MVDR and Wiener filters asymptotically yield the same performance. Finally, we investigated the performance of the different tradeoff filters. Both filters are indeed able to trade off the signal reduction factor for a higher output SNR (see Fig. 4 and 5). The second tradeoff filter, $\mathbf{H}_{T-II,q}$, seems more efficient in doing this, though, as both its output SNR and signal reduction factor are bounded by those of the maximum SNR and MVDR filters. This is opposed to the first, classical tradeoff filter, which never attains the output SNR of the

maximum SNR filter, and it introduces even more distortion than the maximum SNR filter.

VII. EXPERIMENTAL STUDY

In this section, we present the evaluation of the maximum SNR, Wiener, and MVDR filters on real-life speech. This is to verify that the filters are indeed applicable on real-life signals, and that the relations between the performance measures of the different filters hold. For this experiment, we used a 2.4 seconds long, female, speech excerpt from the Keele database [23], with the spectrogram shown in Fig. 6a. Then, we added white Gaussian noise to the speech signal so the average input SNR was 10 dB, and the maximum SNR, Wiener and MVDR filters were applied to the noisy, speech signal. The spectrogram of the noisy signal is shown in Fig. 6b. To design the filters at each time instance, we used outer product averaged, statistics estimates obtained from the past 400 samples. The length of the filters was $L = 128$, the maximum SNR filter was designed with $P = 50$, and the MVDR filter was designed with both $P = 10$ and $P = 50$. Using this setup, the filters were designed and applied for

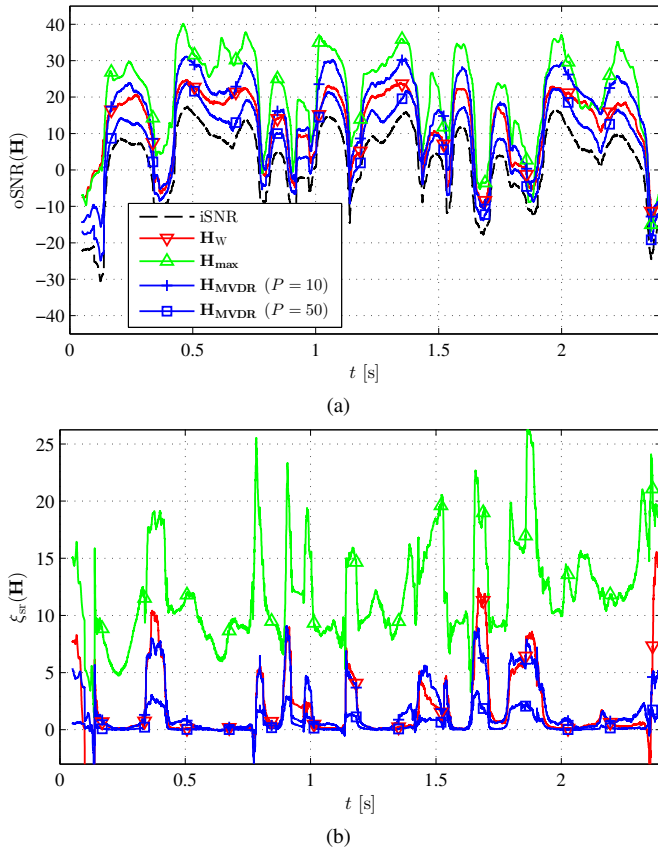


Fig. 7. Plots of the (a) output SNRs and (b) signal reduction factors of the maximum SNR filter ($P = 50$), the Wiener filter, and the MVDR filter ($P = 10$ and $P = 50$) obtained from an experiment with real, female speech in white Gaussian noise at an input SNR of 10 dB.

enhancement, and the resulting spectrograms of the enhanced signals, output SNRs and signal reduction factors are depicted in Figs. 6 and 7. Note that since we get a vector of time-consecutive speech estimates at every time instance, these vectors will be overlapping for one time instance and the following. For one time instance, the final speech estimate is therefore obtained from all vectors containing a speech estimate related to this time instance by averaging those estimates.

From the plots, we first of all observe that all filters improve the SNR. Our informal listening tests also confirmed this. Secondly, the output SNR and signal reduction factor of the MVDR filter depends heavily on the choice of P which is not known in practice. In this experiment, we just used a fixed P , whereas it is known to be time-varying in practice. In most cases, the MVDR filter seems to give a lower signal reduction factor than the Wiener filter, especially so for $P = 50$. The maximum SNR filter yields the highest output SNR but also gives by far the most signal distortion. This was also confirmed by listening. The maximum SNR filter should therefore be regarded as the filter setting a bound on the output SNR rather than a competitor in practical solutions. The above observations are also consistent with the spectrograms of the enhanced signals.

VIII. CONCLUSIONS

In this paper, a new class of optimal filters for speech enhancement has been introduced. These are derived based on the ideas of subspace-based speech enhancement methods so that the observed signal is projected onto the signal subspace after which filtering is performed. By doing this, additional degrees of freedom are achieved in the filter, which means that filters derived this way have the potential to achieve improved output SNRs compared to traditional approaches. In this framework, a number of classical as well as some new filters have been derived. With the new filters, it is possible to trade off signal distortion for better noise reduction. The results confirm that this is indeed the case for both synthetic, periodic signals and real speech signals. In fact, it is possible to seamlessly achieve the maximum output SNR at the cost of speech distortion.

REFERENCES

- [1] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Springer-Verlag, 2009.
- [2] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press, 2007.
- [3] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters – A Theoretical Study*, 1st ed. Springer, 2011, no. VII.
- [4] S. Rangachari and P. Loizou, “A noise estimation algorithm for highly nonstationary environments,” *Speech Communication*, vol. 28, pp. 220–231, 2006.
- [5] I. Cohen, “Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging,” *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, 2003.
- [6] T. Gerkmann and R. C. Hendriks, “Unbiased MMSE-based noise power estimation with low complexity and low tracking delay,” vol. 20, no. 4, pp. 1383–1393, 2012.
- [7] R. C. Hendriks, R. Heusdens, J. Jensen, and U. Kjems, “Low complexity DFT-domain noise PSD tracking using high-resolution periodograms,” *EURASIP J. on Advances in Signal Processing*, vol. 2009, no. 1, p. 15, 2009.
- [8] M. G. Christensen and A. Jakobsson, “Optimal filter designs for separating and enhancing periodic signals,” *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5969–5983, 2010.
- [9] J. R. Jensen, J. Benesty, M. G. Christensen, and S. H. Jensen, “Enhancement of single-channel periodic signals in the time-domain,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 7, pp. 1948–1963, 2012.
- [10] —, “Non-causal time-domain filters for single-channel noise reduction,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1526–1541, 2012.
- [11] Y. Ephraim and H. L. Van Trees, “A signal subspace approach for speech enhancement,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, 1995.
- [12] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, “Reduction of broad-band noise in speech by truncated QSVD,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 6, pp. 439–448, 1995.
- [13] P. Hansen and S. Jensen, “FIR filter representations of reduced-rank noise reduction,” *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1737–1741, 1998.
- [14] 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. on Advances in Signal Processing*, vol. 2007, no. 1, p. 24, 2007.
- [15] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. The John Hopkins University Press, 1996.
- [16] P. S. K. Hansen, “Signal subspace methods for speech enhancement,” Ph.D. dissertation, Techn. Univ. Denmark, Lyngby, Denmark, 1997.
- [17] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [18] R. T. Lacoss, “Data adaptive spectral analysis methods,” *Geophysics*, vol. 36, no. 4, pp. 661–675, Aug. 1971.
- [19] M. Dendrinos, S. Bakamidis, and G. Carayannis, “Speech enhancement from noise: A regenerative approach,” *Speech Commun.*, vol. 10, no. 1, pp. 45–57, 1991.

- [20] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88, no. 4, pp. 972–983, 2008.
- [21] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.
- [22] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Education, Inc., 2005.
- [23] F. Plante, G. F. Meyer, and W. A. Ainsworth, "A pitch extraction reference database," in *Proc. Eurospeech*, 1995, pp. 837–840.