On the Use of Time-Domain Widely Linear Filtering for Binaural Speech Enhancement

Joseph Szurley*, *Student Member, IEEE,* Alexander Bertrand, *Member, IEEE,* and Marc Moonen, *Fellow, IEEE*

***Abstract*—Widely linear (WL) filtering has been shown to improve performance compared to linear filtering due to its ability to incorporate the non-circularity of the signal statistics. However there has been some inconsistency in its application, specifically when constructing complex signals from real signals, which has recently been considered in the context of speech enhancement in binaural or stereo systems. This letter shows that the corresponding WL filtered output contains exactly the same information as the linear filter output while increasing the computational complexity and memory requirements.**

***Index Terms*—Widely linear filtering, binaural speech enhancement**

I. Introduction

Recently there has been a growing interest in applying widely linear (WL) filtering to speech enhancement [1], [2]. The benefit of using a WL filter compared to a linear filter stems from the fact that speech enhancement algorithms often operate in the frequency domain, which yields complex signals with non-circular statistics. With a linear filter, due to circularity assumptions that are imposed, any non-circular second order statistics are neglected which could result in suboptimal solutions. Therefore in order to fully exploit the non-circularity of the second order statistics a WL filter should be used.

In WL filtering a complex signal is augmented with its conjugate and a filter is derived from the corresponding compound signal. This can sometimes improve performance in a mean squared error (MSE) sense but by no more than a factor of 2 [3]. In fact with certain signal models, e.g., double white, it can be shown that the WL filter offers no benefit to the linear filter [4], [5].

In [6], [7], [8], [9], [10], WL filtering has been applied to speech enhancement and echo cancellation in binaural hearing aids or other stereo systems, where two real signals are combined to form a single complex signal which is then used in a WL framework. In this paper we show that while this formulation presents a novel approach, it cannot result in a performance gain, i.e., the corresponding WL filtered output contains exactly the same information as the output from the linear filter. We show this explicitly for the case of multi-channel Wiener filtering (MWF) and time-domain minimum variance distortionless response (MVDR) filtering, but a similar conclusion can be drawn for other types of filters (e.g., those used in [9], [10]). Furthermore, we demonstrate that, while this approach does not improve performance, it increases the computational complexity and memory requirements.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE EF/05/006 'Optimization in Engineering' (OPTEC) and PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office: IUAP P7/19 'Dynamical systems, control and optimization' (DYSCO) 2012-2017, Research Project iMinds, Research Project FWO nr. G.0763.12 'Wireless Acoustic Sensor Networks for Extended Auditory Communication'. Alexander Bertrand is supported by a Postdoctoral Fellowship of the Research Foundation Flanders (FWO). The scientific responsibility is assumed by its authors.

The authors are with the Department of Electrical Engineering ESAT-SCD / iMinds - Future Health Department, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (e-mail: joseph.szurley@esat.kuleuven.be; alexander.bertrand@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

II. Signal model

For the binaural speech enhancement problem considered, we assume that the system contains $2M$ microphones which are assumed to be contained in two different M-microphone arrays or hearing aids. The microphone signals are given as

$$y_{k,m}(t) = x_{k,m}(t) + v_{k,m}(t), \quad k = 0, 1, \quad m = 0 \dots M-1, \tag{1}$$

where $x_{k,m}(t)$ is the speech component and $v_{k,m}(t)$ is the additive noise component. We define the ML-dimensional stacked microphone signal vector, $\mathbf{y}_k \in \mathbb{R}^{ML}$, of each array as

$$\mathbf{y}_k = \left[\mathbf{y}_{k,0}^T, \dots, \mathbf{y}_{k,M-1}^T\right]^T, \quad k = 0, 1 \tag{2}$$

where T indicates the transpose, $\mathbf{y}_{k,m} \in \mathbb{R}^L$ is defined as

$$\mathbf{y}_{k,m} = \begin{bmatrix} y_{k,m}(t) \\ \vdots \\ y_{k,m}(t-L+1) \end{bmatrix}, \quad k = 0, 1 \tag{3}$$

and a 2ML-dimensional signal vector $\mathbf{y} \in \mathbb{R}^{2ML}$ is defined as

$$\mathbf{y} = \left[\mathbf{y}_0^T \ \mathbf{y}_1^T\right]^T \tag{4}$$

where $\mathbf{x}$ and $\mathbf{v}$ are defined similarly.

III. Linear Filtering

In this section, we review two different filtering techniques for speech enhancement which will be compared to their WL counterparts in the following section.

A. Linear Multi-Channel Wiener Filter

The goal of the MWF in speech enhancement is to minimize the mean squared error (MSE) between a desired speech component of a reference microphone signal and a linearly filtered version of the microphone signals. The linear MSE cost function at each array is given as

$$\underset{\mathbf{w}_{k_{\text{MWF}}}}{\text{minimize}} \ \ E\{|d_k - \mathbf{w}_{k_{\text{MWF}}}^T \mathbf{y}|^2\} \tag{5}$$

where d_k is the desired speech component and $E\{.\}$ denotes the expected value. For ease of exposition it is assumed that the first microphone signal of each array, i.e., $d_0 = x_{0,0}$ and $d_1 = x_{1,0}$, acts as the reference microphone.

The solution at each array, takes the form of the MWF [11],

$$\boxed{\hat{\mathbf{w}}_{k_{\text{MWF}}} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k} \tag{6}$$

where $\mathbf{R}_{\mathbf{yy}} = E\{\mathbf{y}\mathbf{y}^H\}$, $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^H\}$, and $\mathbf{e}_k$ is a vector with one entry equal to 1 and 0 otherwise that selects the column of $\mathbf{R}_{\mathbf{xx}}$ that corresponds to the reference microphone. The estimated desired speech component of each array is then given as

$$\begin{aligned}\hat{d}_{k_{\text{MWF}}} &= \hat{\mathbf{w}}_{k_{\text{MWF}}}^T\mathbf{y} \\ &= \mathbf{e}_k^T\mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{y}.\end{aligned} \tag{7}$$

B. Linear Minimum Variance Distortionless Response Filter

It is known that (6) suppresses noise with the adverse effect of distorting the speech. In order to avoid this an MVDR filter can be used, which minimizes the output power while imposing a linear constraint to enforce a distortionless filter response.

It is noted here that a true distortionless response can only be accomplished if the $\mathbf{R}_{\mathbf{xx}}$ has rank-1, which is why MVDR filtering is usually applied in the frequency domain where this rank-1 model holds as in the case of a single target speaker. In [8], MVDR filtering is proposed for time-domain speech enhancement. However, since the rank-1 assumption then generally does not hold, this MVDR approach is not strictly distortionless, in the sense of delivering an undistorted speech signal, despite its name.

The MVDR cost function and linear constraint are given as [6]

$$\begin{aligned}\underset{\mathbf{w}_{k_{\text{MVDR}}}}{\text{minimize}} \quad & \mathbf{w}_{k_{\text{MVDR}}}^T\mathbf{R}_{\mathbf{yy}}\mathbf{w}_{k_{\text{MVDR}}} \\ \text{subject to} \quad & \mathbf{w}_{k_{\text{MVDR}}}^T\mathbf{a}_k = 1\end{aligned} \tag{8}$$

where $\mathbf{a}_k$ is a response vector, which is a scaled version of $\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k$, i.e.,

$$\mathbf{a}_k = \frac{\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k}{\mathbf{e}_k^T\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k}. \tag{9}$$

The solution to (8) is given by [12]

$$\hat{\mathbf{w}}_{k_{\text{MVDR}}} = \frac{\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{a}_k}{\mathbf{a}_k^T\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{a}_k} \tag{10}$$

and using the definition of the response vector (9) the MVDR filter is shown to be equivalent up to a scaling factor, α_k, to the MWF (6), i.e.,

$$\boxed{\hat{\mathbf{w}}_{k_{\text{MVDR}}} = \alpha_k\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k} \tag{11}$$

where

$$\alpha_k = \frac{\mathbf{e}_k^T\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k}{\mathbf{e}_k^T\mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_k}. \tag{12}$$

The estimated desired speech component at each array is then given as

$$\begin{aligned}\hat{d}_{k_{\text{MVDR}}} &= \alpha_k\mathbf{e}_k^T\mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{y} \\ &= \alpha_k\hat{d}_{k_{\text{MWF}}}.\end{aligned} \tag{13}$$

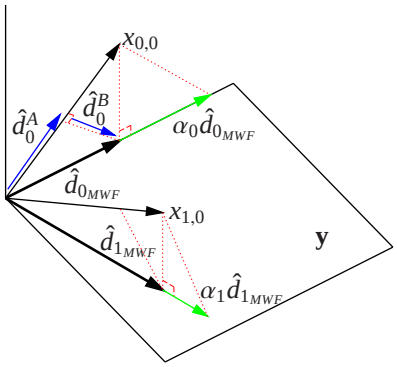

Fig. 1. Graphical representation between the linear MWF and linear MVDR solutions.

The similarity between the MWF and MVDR is shown graphically in Figure 1. The linear MWF projects $x_{0,0}$ and $x_{1,0}$ orthogonally into the $\mathbf{y}$-plane. The MWF estimate for $x_{0,0}$, denoted here again as $\hat{d}_{0_{\text{MWF}}}$, has a component $\hat{d}_0^A$ along $x_{0,0}$ and a component $\hat{d}_0^B$ orthogonal to $x_{0,0}$, where $\hat{d}_{0_{\text{MWF}}} = \hat{d}_0^A + \hat{d}_0^B$. The MVDR is then a stretched version of the MWF solution (scaling with α_0) until the $\hat{d}_0^A$ component lands in $x_{0,0}$. The same process happens for the estimate of $x_{1,0}$.

Since the MWF is known to distort the speech, and since (11) is equivalent to an MWF (up to a fixed scaling), we indeed find that the time-domain MVDR is also not distortionless. Basically, the linear constraint in (9) only ensures that the covariance between the MVDR filter output $\hat{d}_{k_{\text{MVDR}}}$ and the desired signal $x_{k,0}$ is equal to the variance of $x_{k,0}$, i.e., $E\{x_{k,0}\hat{d}_{k_{\text{MVDR}}}\} = E\{x_{k,0}^2\}$ [8].

Despite their theoretical equivalence, an adaptive implementation of (6) or (11) based on block-processing may result in different output signals due to time variations in α_k such that $E\{x_{k,0}\hat{d}_{k_{\text{MVDR}}}\} = E\{x_{k,0}^2\}$ in each block. Finally, it is noted that the MWF is known to preserve the binaural cues of the speech [13]. Since the time-domain MVDR filter corresponds to an MWF with an additional scaling α_k which is different for $k = 0, 1$ this will result in distortion of the binaural cues.

IV. Widely Linear Filtering

The derivation of the linear MWF (6) and MVDR filter (11) also holds for complex-valued signals, but then the transpose operator T should be replaced by a transpose conjugation H in every equation. However, complex signals allow to also exploit the non-circularity of the signals if widely linear filtering techniques are used instead [2], [3], [4].

In [6], [7], [8], [9], a complex signal vector is artificially constructed from the real signals received at both arrays,

$$\bar{\mathbf{y}} = \mathbf{y}_0 + j\mathbf{y}_1 \tag{14}$$

where $j = \sqrt{-1}$, to be able to apply WL filtering techniques. WL filtering then amounts to using the original linear filters on an augmented 2ML-dimensional signal vector, $\tilde{\mathbf{y}} = \mathbb{C}^{2ML}$, defined as

$$\tilde{\mathbf{y}} = \begin{bmatrix}\bar{\mathbf{y}} \\ \bar{\mathbf{y}}^*\end{bmatrix} \tag{15}$$

where $*$ denotes complex conjugation and where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{v}}$ are defined similarly. This augmented signal vector can easily be shown to be a transform of the original signal vector in (4),

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \mathbf{y}. \tag{16}$$

This transformation may then be used to show the equivalence between the estimated desired speech component found with the WL filters and the estimated desired speech components, ($k = 0, 1$), found with the linear filters when applied to real signals.

A. Widely Linear Multi-Channel Wiener Filter

The WL-MWF of (6) is given as

$$\boxed{\tilde{\mathbf{w}}_{\text{MWF}} = \mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0} \tag{17}$$

where $\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = E\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^H\}$ and $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} = E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\}$. The estimated desired speech component using (17) is then given as

$$\tilde{d}_{\text{MWF}} = \mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\tilde{\mathbf{y}} \tag{18}$$

which in (19) is expanded using (16). Simplifying (19) we see that the estimated desired speech component is given by

$$\begin{aligned} \tilde{d}_{\text{MWF}} &= [1|j] \begin{bmatrix} \mathbf{e}_0^T \\ \mathbf{e}_1^T \end{bmatrix} \mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{y} \\ &= [1|j] \begin{bmatrix} \hat{d}_{0_{\text{MWF}}} \\ \hat{d}_{1_{\text{MWF}}} \end{bmatrix}. \end{aligned} \tag{20}$$

The WL-MWF output then fully corresponds to the linear MWF outputs.

B. Widely Linear Minimum Variance Distortionless Response Filter

In [6] a WL response vector, $\tilde{\mathbf{a}}$, is used for the solution to the WL-MVDR filter given as

$$\tilde{\mathbf{a}} = \frac{\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0}. \tag{21}$$

The WL-MVDR of [6] may then be given as

$$\tilde{\mathbf{w}}_{\text{MVDR}} = \frac{\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\tilde{\mathbf{a}}}{\tilde{\mathbf{a}}^H\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\tilde{\mathbf{a}}} \tag{22}$$

and using the definition of the response vector (21) the WL-MVDR filter is shown to be equivalent up to a scaling factor, $\tilde{\alpha}$, to the WL-MWF (17), i.e.,

$$\boxed{\tilde{\mathbf{w}}_{\text{MVDR}} = \tilde{\alpha}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0} \tag{23}$$

where

$$\tilde{\alpha} = \frac{\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0}. \tag{24}$$

The estimated desired speech component using (22) is then given as

$$\tilde{d}_{\text{MVDR}} = \tilde{\alpha}\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\tilde{\mathbf{y}} \tag{25}$$

which uses the same expansion as in (19). Simplifying (25) with (19) we see that the estimated desired speech component is given by

$$\begin{aligned} \tilde{d}_{\text{MVDR}} &= \tilde{\alpha}[1|j] \begin{bmatrix} \mathbf{e}_0^T \\ \mathbf{e}_1^T \end{bmatrix} \mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{y} \\ &= \tilde{\alpha}[1|j] \begin{bmatrix} \hat{d}_{0_{\text{MWF}}} \\ \hat{d}_{1_{\text{MWF}}} \end{bmatrix} \\ &= \tilde{\alpha}[1|j] \begin{bmatrix} \frac{1}{\alpha_0}\hat{d}_{0_{\text{MVDR}}} \\ \frac{1}{\alpha_1}\hat{d}_{1_{\text{MVDR}}} \end{bmatrix}. \end{aligned} \tag{26}$$

The WL-MVDR output then corresponds to the linear MVDR outputs up to a real-valued scaling with $\frac{\tilde{\alpha}}{\alpha_0}$ and $\frac{\tilde{\alpha}}{\alpha_1}$ (or to the linear MWF up to a joint scaling with $\tilde{\alpha}$).

It is noted that

$$\tilde{\alpha} = \frac{\mathbf{e}_0^T\mathbf{R}_{\mathbf{xx}}\mathbf{e}_0 + \mathbf{e}_1^T\mathbf{R}_{\mathbf{xx}}\mathbf{e}_1}{\mathbf{e}_0^T\mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_0 + \mathbf{e}_1^T\mathbf{R}_{\mathbf{xx}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_1} \tag{27}$$

and so $\tilde{\alpha}$ can also be computed from quantities available in the linear filtering approach hence both approaches yield the exact same information.

The WL-MWF gives the same estimates for $x_{0,0}$ and $x_{1,0}$ as the linear MWF. The WL-MVDR is obtained by equally stretching the linear MWF solutions by $\tilde{\alpha}$. If the $\hat{d}_0^A$ component is stretched into something longer than $x_{0,0}$, then the $\hat{d}_1^A$ is stretched into something shorter than $x_{1,0}$, and vice versa because the two stretched components, $\hat{d}_0^A$ and $\hat{d}_1^A$, now jointly satisfy one equation, i.e., $\tilde{\alpha}E\{x_{0,0}\hat{d}_{0_{\text{MWF}}} + x_{1,0}\hat{d}_{1_{\text{MWF}}}\} = E\{x_{0,0}^2 + x_{1,0}^2\}$. The similarity between the linear MWF and WL-MVDR can be shown graphically as in Figure 1 where the vectors representing the WL-MVDR solution would be equal lengths.

V. Equivalence of the MVDR Scaling Factors Under a Rank-1 Model

Originally, the MVDR approach was designed for scenarios with a rank-1 model for $\mathbf{R}_{\mathbf{xx}}$ or $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$. We show that when such a rank-1 model is used for $\mathbf{R}_{\mathbf{xx}}$ or $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$ the scaling factors for the linear MVDR and WL-MVDR are equivalent.

A. Linear MVDR scaling factor

The singular value decomposition (SVD) of the assumed rank-1 $\mathbf{R}_{\mathbf{xx}}$ matrix is given as

$$\mathbf{R}_{\mathbf{xx}} = \mathbf{U}_x\mathbf{\Sigma}_x\mathbf{V}_x^T \tag{28}$$

where $\mathbf{\Sigma}_x = diag(\sigma_x, 0, \ldots, 0)$ and the elements of $\mathbf{U}_x$ and $\mathbf{V}_x$ are given as $u_{x_{i,j}}$ and $v_{x_{i,j}}$ respectively. Using this SVD of the $\mathbf{R}_{\mathbf{xx}}$ matrix the numerator of (12) is shown to be

$$\mathbf{e}_0^T\mathbf{R}_{\mathbf{xx}}\mathbf{e}_0 = \mathbf{e}_0^T\mathbf{U}_x\mathbf{\Sigma}_x\mathbf{V}_x^H\mathbf{e}_0 = \sigma_x u_{x_{1,1}}v_{x_{1,1}}. \tag{29}$$

$$\tilde{d}_{\text{MWF}} = \mathbf{e}_0^T \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \mathbf{R}_{\mathbf{xx}} \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}^H \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}^{-H} \mathbf{R}_{\mathbf{yy}}^{-1} \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \mathbf{y} \tag{19}$$

For the denominator of (12) we express the SVD as

$$\mathbf{R_{xx}}\mathbf{R_{yy}}^{-1}\mathbf{R_{xx}} = \mathbf{U}_x\mathbf{\Sigma}_{xy}\mathbf{V}_x^T \tag{30}$$

where

$$\mathbf{\Sigma}_{xy} = \mathbf{\Sigma}_x\mathbf{V}_x^T\mathbf{R_{yy}}^{-1}\mathbf{U}_x\mathbf{\Sigma}_x \tag{31}$$

which is another diagonal matrix with a single non-zero element, i.e., $diag(\sigma_{xy}, 0, \ldots, 0)$. Therefore

$$\begin{aligned}\mathbf{e}_0^T\mathbf{R_{xx}}\mathbf{R_{yy}}^{-1}\mathbf{R_{xx}}\mathbf{e}_0 &= \mathbf{e}_0^T\mathbf{U}_{xy}\mathbf{\Sigma}_{xy}\mathbf{V}_{xy}^T\mathbf{e}_0 \\ &= \sigma_{xy}u_{x_{1,1}}v_{x_{1,1}}\end{aligned} \tag{32}$$

and the scaling factor α_0 can be shown to be equal to

$$\alpha_0 = \frac{\mathbf{e}_0^T\mathbf{R_{xx}}\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R_{xx}}\mathbf{R_{yy}}^{-1}\mathbf{R_{xx}}\mathbf{e}_0} = \frac{\sigma_x}{\sigma_{xy}} \tag{33}$$

which is also the same for the $k = 1$ array, i.e., $\alpha_0 = \alpha_1$.

B. Widely linear MVDR scaling factor

In the WL case, the numerator of (24) using (16), and the SVD of $\mathbf{R_{xx}}$, (28), is given as

$$\begin{aligned}\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0 &= \mathbf{e}_0^T\begin{bmatrix}\mathbf{I} & j\mathbf{I}\\ \mathbf{I} & -j\mathbf{I}\end{bmatrix}\mathbf{R_{xx}}\begin{bmatrix}\mathbf{I} & j\mathbf{I}\\ \mathbf{I} & -j\mathbf{I}\end{bmatrix}^T\mathbf{e}_0 \\ &= [1|j]\begin{bmatrix}\mathbf{e}_0^T\\ \mathbf{e}_1^T\end{bmatrix}\mathbf{R_{xx}}\begin{bmatrix}\mathbf{e}_0 & \mathbf{e}_1\end{bmatrix}[1|j]^H.\end{aligned} \tag{34}$$

However since $\mathbf{R_{xx}}$ is symmetric, the multiplication of the vectors $[1|j]$ and $[1|j]^H$ cancel out the off-diagonal terms while summing the diagonal terms. Therefore (34) can be represented as

$$\begin{aligned}\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0 &= \mathrm{Tr}\left\{\begin{bmatrix}\mathbf{e}_0^T\\ \mathbf{e}_1^T\end{bmatrix}\mathbf{R_{xx}}\begin{bmatrix}\mathbf{e}_0 & \mathbf{e}_1\end{bmatrix}\right\} \\ &= \mathrm{Tr}\left\{\begin{bmatrix}\mathbf{e}_0^T\\ \mathbf{e}_1^T\end{bmatrix}\mathbf{U}_x\mathbf{\Sigma}_x\mathbf{V}_x^T\begin{bmatrix}\mathbf{e}_0 & \mathbf{e}_1\end{bmatrix}\right\} \\ &= \sigma_x(u_{x_{1,1}}v_{x_{1,1}} + u_{x_{ML+1,1}}v_{x_{ML+1,1}}).\end{aligned} \tag{35}$$

The denominator is expanded in a similar fashion as

$$\begin{aligned}\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0 &= \mathrm{Tr}\left\{\begin{bmatrix}\mathbf{e}_0^T\\ \mathbf{e}_1^T\end{bmatrix}\mathbf{U}_x\mathbf{\Sigma}_{xy}\mathbf{V}_x^T\begin{bmatrix}\mathbf{e}_0 & \mathbf{e}_1\end{bmatrix}\right\} \\ &= \sigma_{xy}(u_{x_{1,1}}v_{x_{1,1}} + u_{x_{ML+1,1}}v_{x_{ML+1,1}}).\end{aligned} \tag{36}$$

The WL-MVDR scaling factor is therefore equivalent to the linear MVDR scaling factor (33),

$$\tilde{\alpha} = \frac{\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{e}_0} = \frac{\sigma_x}{\sigma_{xy}}. \tag{37}$$

VI. Computational Complexity

The WL filtering approach computes a single complex valued filter of length 2ML, while the linear filtering approach computes two real valued filters of length 2ML (one for each desired signal d_k, where $k = 0, 1$). However complex arithmetic is 4 times more expensive than real arithmetic. As a result the WL filtering approach is actually twice as expensive compared to the linear filtering approach with no increase in performance. Furthermore, the two real-valued linear filters can share many of their computations (e.g., the inversion of $\mathbf{R_{yy}}$), which makes the WL filtering approach actually more than twice as expensive compared to the linear filtering approach.

VII. Conclusions

An equivalence was shown between the estimated desired speech components using time-domain linear and widely linear filters in binaural speech enhancement applications when only real signals are used. While the WL filters offer a novel way to represent the received real signals as a single complex signal there is no added benefit in terms of speech enhancement. However by using an artificially constructed complex signal the memory requirement of the system is increased as well as the computational complexity.

References

[1] J. Benesty, J. Chen, and Y.A. Huang, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 469 – 472, May 2010.

[2] J. Benesty, J. Chen, and Y.A. Huang, "On widely linear Wiener and tradeoff filters for noise reduction," *Speech Communication*, vol. 52, no. 5, pp. 427 – 439, 2010.

[3] P.J. Schreier, L.L. Scharf, and C.T. Mullis, "A unified approach to performance comparisons between linear and widely linear processing," in *Proc. IEEE Workshop on Statistical Signal Process.*, Sep. 2003, pp. 114 – 117.

[4] T. Adali, Hualiang Li, and R. Aloysius, "On properties of the widely linear MSE filter and its LMS implementation," in *43rd Annu. Conf. on Inform. Sciences and Systems (CISS '09)*, Mar. 2009, pp. 876 – 881.

[5] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. on Signal Proces.*, vol. 43, no. 8, pp. 2030 – 2033, Aug. 1995.

[6] J. Chen and J. Benesty, "A time-domain widely linear MVDR filter for binaural noise reduction," in *Proc. IEEE Workshop on Applicat. of Signal Proces. to Audio and Acoust. (WASPAA '11)*, Oct. 2011, pp. 105 –108.

[7] J. Benesty and J. Chen, "A multichannel widely linear approach to binaural noise reduction using an array of microphones," in *Proc. IEEE Int.Conf. on Acoust., Speech and Signal Process. (ICASSP '12)*, Mar. 2012, pp. 313 – 316.

[8] J. Benesty, J. Chen, and Y.A. Huang, "Binaural noise reduction in the time domain with a stereo setup," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19, no. 8, pp. 2260–2272, Nov. 2011.

[9] J. Chen and J. Benesty, "On the time-domain widely linear LCMV filter for noise reduction with a stereo system," *IEEE Trans. on Audio, Speech, and Language Process*, vol. 21, no. 7, pp. 1343–1354, 2013.

[10] C. Stanciu, J. Benesty, C. Paleologu, T. Gnsler, and S. Ciochin, "A widely linear model for stereophonic acoustic echo cancellation," *Signal Processing*, vol. 93, no. 2, pp. 511–516, 2013.

[11] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. on Signal Proces.*, vol. 50, no. 9, pp. 2230 – 2244, Sep. 2002.

[12] E.A.P. Habets, J. Benesty, S. Gannot, and I. Cohen, "The MVDR beamformer for speech enhancement," in *Speech Processing in Modern Communication*, Israel Cohen, Jacob Benesty, and Sharon Gannot, Eds., vol. 3 of *Springer Topics in Signal Processing*, pp. 225 – 254. Springer Berlin Heidelberg, 2010.

[13] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.