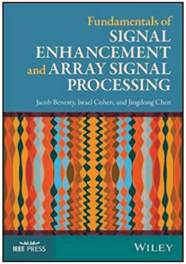
Multichannel Signal Enhancement in the Frequency Domain

J. Benesty, I. Cohen, and J. Chen,
Fundamentals of Signal Enhancement and Array Signal Processing,
Wiley-IEEE Press, 2017.



Outline

Introduction

We study the signal enhancement problem in the frequency domain with multiple sensors.

We explain the signal model and state the problem we wish to solve with the conventional linear filtering technique.

We then derive performance measures and show how to obtain the most well-known optimal linear filters.

Signal Model and Problem Formulation

We consider the conventional signal model in which an array of M sensors with an arbitrary geometry captures a convolved desired source signal in some noise field.

The received signals, at the discrete-time index t, are expressed as

$$\begin{aligned} y_m(t) &= g_m(t) * x(t) + v_m(t) \\ &= x_m(t) + v_m(t),\ m = 1, 2, \ldots, M, \end{aligned} \quad (1)$$

where $g_m(t)$ is the acoustic impulse response from the unknown desired source, $x(t)$, location to the mth sensor, $*$ stands for linear convolution, and $v_m(t)$ is the additive noise at sensor m.

We assume that the signals $x_m(t) = g_m(t) * x(t)$ and $v_m(t)$ are uncorrelated, zero mean, stationary, real, and broadband.

In the frequency domain, at the frequency index f, (1) can be expressed as [1], [2], [3]

$$\begin{aligned} Y_m(f) &= G_m(f)X(f) + V_m(f) \\ &= X_m(f) + V_m(f),\ m = 1, 2, \ldots, M, \end{aligned} \tag{2}$$

where $Y_m(f)$, $G_m(f)$, $X(f)$, $V_m(f)$, and $X_m(f) = G_m(f)X(f)$ are the frequency-domain representations of $y_m(t)$, $g_m(t)$, $x(t)$, $v_m(t)$, and $x_m(t) = g_m(t) * x(t)$, respectively.

Sensor 1 is the reference, so the objective of multichannel noise reduction in the frequency domain is to estimate the desired signal, $X_1(f)$, from the M observations $Y_m(f),\ m = 1, 2, \ldots, M$, in the best possible way.

It is more convenient to write the M frequency-domain sensors' signals in a vector notation:

$$\begin{aligned}\mathbf{y}(f) &= \mathbf{g}(f)X(f) + \mathbf{v}(f) \\ &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}(f)X_1(f) + \mathbf{v}(f),\end{aligned} \tag{3}$$

where

$$\begin{aligned}\mathbf{y}(f) &= \left[\begin{array}{cccc} Y_1(f) & Y_2(f) & \cdots & Y_M(f) \end{array}\right]^T, \\ \mathbf{x}(f) &= \left[\begin{array}{cccc} X_1(f) & X_2(f) & \cdots & X_M(f) \end{array}\right]^T \\ &= X(f)\mathbf{g}(f), \\ \mathbf{g}(f) &= \left[\begin{array}{cccc} G_1(f) & G_2(f) & \cdots & G_M(f) \end{array}\right]^T, \\ \mathbf{v}(f) &= \left[\begin{array}{cccc} V_1(f) & V_2(f) & \cdots & V_M(f) \end{array}\right]^T,\end{aligned}$$

and

$$\mathbf{d}(f) = \begin{bmatrix} 1 & \frac{G_2(f)}{G_1(f)} & \cdots & \frac{G_M(f)}{G_1(f)} \end{bmatrix}^T = \frac{\mathbf{g}(f)}{G_1(f)}. \quad (4)$$

The vector $\mathbf{d}(f)$ can be seen as the steering vector for noise reduction [4] since the acoustic impulse responses ratios from the broadband source to the aperture convey information about the position of the source.

There is another interesting way to write (3). First, it is easy to see that

$$X_m(f) = \gamma^*_{X_1 X_m}(f) X_1(f),\ m = 1, 2, \ldots, M, \quad (5)$$

where

$$\gamma_{X_1X_m}(f) = \frac{E\left[X_1(f)X_m^*(f)\right]}{E\left[\left|X_1(f)\right|^2\right]} = \frac{G_m^*(f)}{G_1^*(f)},\ m = 1, 2, \ldots, M \tag{6}$$

is the partially normalized [with respect to $X_1(f)$] coherence function between $X_1(f)$ and $X_m(f)$.

Using (5), we can rewrite (3) as

$$\mathbf{y}(f) = \boldsymbol{\gamma}_{X_1\mathbf{x}}^*(f)X_1(f) + \mathbf{v}(f), \tag{7}$$

where

$$\boldsymbol{\gamma}_{X_1\mathbf{x}}(f) = \begin{bmatrix} 1 & \gamma_{X_1X_2}(f) & \cdots & \gamma_{X_1X_M}(f) \end{bmatrix}^T = \frac{E\left[X_1(f)\mathbf{x}^*(f)\right]}{E\left[\left|X_1(f)\right|^2\right]} = \mathbf{d}^*(f) \quad (8)$$

is the partially normalized [with respect to $X_1(f)$] coherence vector (of length M) between $X_1(f)$ and $\mathbf{x}(f)$.

In the rest, $\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)$ and $\mathbf{d}(f)$ will be used interchangeably.

By definition, the signal $X_1(f)$ is completely coherent across all sensors [see eq. (5)]; however, $V_1(f)$ is usually partially coherent with the noise components, $V_m(f)$, at the other sensors.

Therefore, any noise term $V_m(f)$ can be easily decomposed into two orthogonal components, i.e.,

$$V_m(f) = \gamma^*_{V_1 V_m}(f) V_1(f) + V'_m(f),\ m = 1, 2, \ldots, M, \tag{9}$$

where $\gamma_{V_1 V_m}(f)$ is the partially normalized [with respect to $V_1(f)$] coherence function between $V_1(f)$ and $V_m(f)$ and

$$E\left[V_1^*(f) V'_m(f)\right] = 0,\ m = 1, 2, \ldots, M. \tag{10}$$

The vector $\mathbf{v}(f)$ can then be written as the sum of two other vectors: one coherent with $V_1(f)$ and the other incoherent with $V_1(f)$, i.e.,

$$\mathbf{v}(f) = \boldsymbol{\gamma}^*_{V_1 \mathbf{v}}(f) V_1(f) + \mathbf{v}'(f), \tag{11}$$

where

$$\boldsymbol{\gamma}_{V_1\mathbf{v}}(f) = \begin{bmatrix} 1 & \gamma_{V_1V_2}(f) & \cdots & \gamma_{V_1V_M}(f) \end{bmatrix}^T \tag{12}$$

is the partially normalized [with respect to $V_1(f)$] coherence vector (of length M) between $V_1(f)$ and $\mathbf{v}(f)$ and

$$\mathbf{v}'(f) = \begin{bmatrix} 0 & V_2'(f) & \cdots & V_M'(f) \end{bmatrix}^T.$$

If $V_1(f)$ is incoherent with $V_m(f)$, where $m \neq 1$, then $\gamma_{V_1V_m}(f) = 0$.

Another convenient way to write the sensors' signals vector is

$$\mathbf{y}(f) = \boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)X_1(f) + \boldsymbol{\gamma}^*_{V_1\mathbf{v}}(f)V_1(f) + \mathbf{v}'(f). \tag{13}$$

We see that $\mathbf{y}(f)$ is the sum of three mutual incoherent components.

Therefore, the correlation matrix of $\mathbf{y}(f)$ is

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}}(f) &= E\left[\mathbf{y}(f)\mathbf{y}^H(f)\right] \\ &= \phi_{X_1}(f)\mathbf{d}(f)\mathbf{d}^H(f) + \mathbf{\Phi}_{\mathbf{v}}(f) \\ &= \phi_{X_1}(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)\boldsymbol{\gamma}^T_{X_1\mathbf{x}}(f) + \phi_{V_1}(f)\boldsymbol{\gamma}^*_{V_1\mathbf{v}}(f)\boldsymbol{\gamma}^T_{V_1\mathbf{v}}(f) + \mathbf{\Phi}_{\mathbf{v}'}(f),\end{aligned} \tag{14}$$

where the superscript ^{H} is the conjugate-transpose operator, $\phi_{X_1}(f) = E\left[|X_1(f)|^2\right]$ and $\phi_{V_1}(f) = E\left[|V_1(f)|^2\right]$ are the variances of $X_1(f)$ and $V_1(f)$, respectively, and $\mathbf{\Phi}_{\mathbf{v}}(f) = E\left[\mathbf{v}(f)\mathbf{v}^H(f)\right]$ and $\mathbf{\Phi}_{\mathbf{v}'}(f) = E\left[\mathbf{v}'(f)\mathbf{v}'^H(f)\right]$ are the correlation matrices of $\mathbf{v}(f)$ and $\mathbf{v}'(f)$, respectively.

The matrix $\mathbf{\Phi}_{\mathbf{y}}(f)$ is the sum of three other matrices: the first two are of rank equal to 1 and the last one (correlation matrix of the incoherent noise) is assumed to be of rank equal to $M-1$.

Linear Filtering

In the frequency domain, conventional multichannel noise reduction is performed by applying a complex weight to the output of each sensor, at frequency f, and summing across the aperture (see Fig. 1):

$$\begin{aligned} Z(f) &= \sum_{m=1}^{M} H_m^*(f) Y_m(f) \\ &= \mathbf{h}^H(f)\mathbf{y}(f), \end{aligned} \tag{15}$$

where $Z(f)$ is the estimate of $X_1(f)$ and

$$\mathbf{h}(f) = \begin{bmatrix} H_1(f) & H_2(f) & \cdots & H_M(f) \end{bmatrix}^T \tag{16}$$

is a filter of length M containing all the complex gains applied to the sensors' outputs at frequency f.

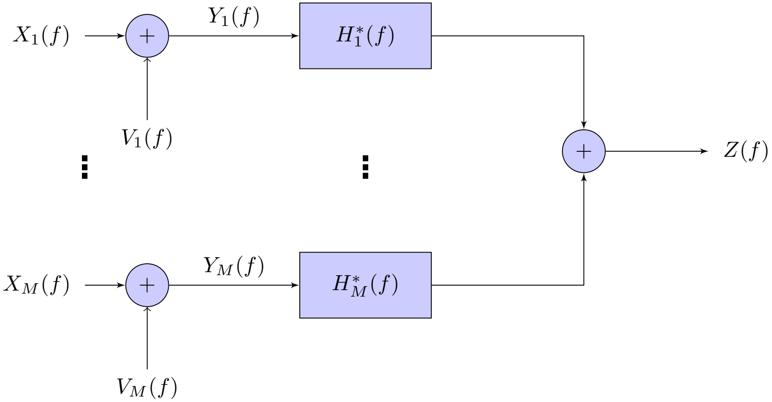


Figure 1: Block diagram of multichannel linear filtering in the frequency domain.

We can express (15) as a function of the steering vector, i.e.,

$$\begin{aligned} Z(f) &= \mathbf{h}^H(f)\left[\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)X_1(f) + \mathbf{v}(f)\right] \\ &= X_{\mathrm{fd}}(f) + V_{\mathrm{rn}}(f), \end{aligned} \tag{17}$$

where

$$X_{\mathrm{fd}}(f) = X_1(f)\mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) \tag{18}$$

is the filtered desired signal and

$$V_{\mathrm{rn}}(f) = \mathbf{h}^H(f)\mathbf{v}(f) \tag{19}$$

is the residual noise. This procedure is called the multichannel signal enhancement problem in the frequency domain.

The two terms on the right-hand side of (17) are incoherent. Hence, the variance of $Z(f)$ is also the sum of two variances:

$$\begin{aligned} \phi_Z(f) &= \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{h}(f) \\ &= \phi_{X_{\text{fd}}}(f) + \phi_{V_{\text{rn}}}(f), \end{aligned} \tag{20}$$

where

$$\phi_{X_{\text{fd}}}(f) = \phi_{X_1}(f)\left|\mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)\right|^2, \tag{21}$$

$$\phi_{V_{\text{rn}}}(f) = \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f). \tag{22}$$

Performance Measures

Signal-to-Noise Ratio

The input SNR gives an idea on the level of the noise as compared to the level of the desired signal at the reference sensor.

The narrowband input SNR is

$$\mathrm{iSNR}(f) = \frac{\phi_{X_1}(f)}{\phi_{V_1}(f)}. \tag{23}$$

The broadband input SNR:

$$\mathrm{iSNR} = \frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{V_1}(f)df}. \tag{24}$$

Notice that

$$\mathrm{iSNR} \neq \int_f \mathrm{iSNR}(f)df. \tag{25}$$

The output SNR quantifies the SNR after performing noise reduction. From (20), we deduce the narrowband output SNR:

$$\begin{aligned} \mathrm{oSNR}\left[\mathbf{h}(f)\right] &= \frac{\phi_{X_{\mathrm{fd}}}(f)}{\phi_{V_{\mathrm{rn}}}(f)} \\ &= \frac{\phi_{X_1}(f)\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2}{\mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)} \end{aligned} \tag{26}$$

and the broadband output SNR:

$$\mathrm{oSNR}\left(\mathbf{h}\right) = \frac{\int_f \phi_{X_1}(f)\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 df}{\int_f \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)df}. \tag{27}$$

It is clear that

$$\mathrm{oSNR}\left(\mathbf{h}\right) \neq \int_f \mathrm{oSNR}\left[\mathbf{h}(f)\right] df. \tag{28}$$

Assume that the matrix $\mathbf{\Phi}_{\mathbf{v}}(f)$ is nonsingular. In this case, for the two vectors $\mathbf{h}(f)$ and $\mathbf{d}(f)$, we have

$$\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 \leq \left[\mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)\right]\left[\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)\right], \tag{29}$$

with equality if and only if $\mathbf{h}(f) \propto \mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)$.

Using the inequality (29) in (26), we deduce an upper bound for the narrowband output SNR:

$$\mathrm{oSNR}\left[\mathbf{h}(f)\right] \leq \phi_{X_1}(f) \times \mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f),\ \forall \mathbf{h}(f). \tag{30}$$

For the particular filter of length M:

$$\mathbf{h}(f) = \mathbf{i}_{\mathrm{i}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T, \tag{31}$$

we have

$$\mathrm{oSNR}\left[\mathbf{i}_{\mathrm{i}}(f)\right] = \mathrm{iSNR}(f), \tag{32}$$

$$\mathrm{oSNR}\left(\mathbf{i}_{\mathrm{i}}\right) = \mathrm{iSNR}. \tag{33}$$

With the identity filter, $\mathbf{i}_{\mathrm{i}}$, the output SNRs cannot be improved and

$$\mathrm{oSNR}\left[\mathbf{i}_{\mathrm{i}}(f)\right] \leq \phi_{X_1}(f) \times \mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f), \tag{34}$$

which implies that

$$\phi_{V_1}(f) \times \mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f) \geq 1. \tag{35}$$

Our objective is then to find the filter, $\mathbf{h}(f)$, within the design constraints, in such a way that $\mathrm{oSNR}\left[\mathbf{h}(f)\right] > \mathrm{iSNR}(f)$.

While the narrowband output SNR is important when we deal with narrowband and broadband signals, the broadband output SNR is even more important when we deal with broadband signals such as speech.

Therefore, we also need to make sure finding $\mathbf{h}(f)$ in such a way that $\mathrm{oSNR}\,(\mathbf{h}) > \mathrm{iSNR}$.

Noise Rejection Factor

The noise reduction factor or noise rejection factor quantifies the amount of noise being rejected by the filter.

This quantity is defined as the ratio of the power of the noise at the reference sensor over the power of the noise remaining at the filter output. Specifically:

- the broadband noise reduction factor,

$$\xi_{\mathrm{n}}(\mathbf{h}) = \frac{\int_f \phi_{V_1}(f)df}{\int_f \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)df} \tag{36}$$

- and the narrowband noise reduction factor,

$$\xi_{\mathrm{n}}[\mathbf{h}(f)] = \frac{\phi_{V_1}(f)}{\mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)}. \tag{37}$$

The broadband noise reduction factor is expected to be lower bounded by 1; otherwise, the filter amplifies the noise received at the senors.

The higher the value of the noise reduction factor, the more noise that is rejected.

Desired-Signal Reduction Factor

In order to quantify the level of this distortion, we define the desired-signal reduction factor or desired-signal cancellation factor as the ratio of the variance of the desired signal at the reference sensor over the variance of the filtered desired signal at the filter output. Specifically:

- the broadband desired-signal reduction factor,

$$\xi_{\mathrm{d}}(\mathbf{h}) = \frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{X_1}(f)\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 df} \quad (38)$$

- and the narrowband desired-signal reduction factor,

$$\xi_{\mathrm{d}}[\mathbf{h}(f)] = \frac{1}{\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2}. \quad (39)$$

Once again, note that

$$\xi_{\mathrm{n}}(\mathbf{h}) \neq \int_f \xi_{\mathrm{n}}[\mathbf{h}(f)]\, df, \tag{40}$$

$$\xi_{\mathrm{d}}(\mathbf{h}) \neq \int_f \xi_{\mathrm{d}}[\mathbf{h}(f)]\, df. \tag{41}$$

Another key observation is that the design of filters that do not cancel the broadband desired signal requires the constraint:

$$\mathbf{h}^H(f)\mathbf{d}(f) = 1. \tag{42}$$

Thus, the desired-signal reduction factor is equal to 1 if there is no cancellation and expected to be greater than 1 when cancellation happens.

Lastly, by making the appropriate substitutions, one can derive the following relationships between the output and input SNRs, noise reduction factor, and desired-signal reduction factor:

$$\frac{\mathrm{oSNR}\left(\mathbf{h}\right)}{\mathrm{iSNR}} = \frac{\xi_{\mathrm{n}}\left(\mathbf{h}\right)}{\xi_{\mathrm{d}}\left(\mathbf{h}\right)}, \qquad (43)$$

$$\frac{\mathrm{oSNR}\left[\mathbf{h}(f)\right]}{\mathrm{iSNR}(f)} = \frac{\xi_{\mathrm{n}}\left[\mathbf{h}(f)\right]}{\xi_{\mathrm{d}}\left[\mathbf{h}(f)\right]}. \qquad (44)$$

Desired-Signal Distortion Index

Another useful way to measure the distortion of the desired signal is via the desired-signal distortion index, which is defined as the MSE between the desired signal and its estimate, normalized by the power of the desired signal. Specifically:

- the broadband desired-signal distortion index,

$$\upsilon_{\mathrm{d}}(\mathbf{h}) = \frac{\int_f \phi_{X_1}(f) \left| \mathbf{h}^H(f) \boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) - 1 \right|^2 df}{\int_f \phi_{X_1}(f) df} \tag{45}$$

- and the narrowband desired-signal distortion index,

$$\begin{aligned} \upsilon_{\mathrm{d}}[\mathbf{h}(f)] &= \frac{E\left[\left| X_{\mathrm{fd}}(f) - X_1(f) \right|^2 \right]}{\phi_{X_1}(f)} \\ &= \left| \mathbf{h}^H(f) \boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) - 1 \right|^2. \end{aligned} \tag{46}$$

It is interesting to point out that the broadband desired-signal distortion index is a linear combination of the narrowband desired-signal distortion indices as the denominator is simply a scaling factor, i.e.,

$$\upsilon_{\mathrm{d}}\left(\mathbf{h}\right) = \frac{\int_f \phi_{X_1}(f)\upsilon_{\mathrm{d}}\left[\mathbf{h}(f)\right] df}{\int_f \phi_{X_1}(f) df}. \quad (47)$$

The distortionless constraint implies that $\upsilon_{\mathrm{d}}\left[\mathbf{h}(f)\right] = 0, \ \forall f$.

Mean-Squared Error

We define the error signal between the estimated and desired signals at frequency f as

$$\begin{aligned}\mathcal{E}(f) &= Z(f) - X_1(f) \\ &= \mathbf{h}^H(f)\mathbf{y}(f) - X_1(f) \\ &= X_{\mathrm{fd}}(f) + V_{\mathrm{rn}}(f) - X_1(f).\end{aligned} \tag{48}$$

This error can also be expressed as

$$\mathcal{E}(f) = \mathcal{E}_{\mathrm{d}}(f) + \mathcal{E}_{\mathrm{n}}(f), \tag{49}$$

where

$$\mathcal{E}_{\mathrm{d}}(f) = \left[\mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) - 1\right] X_1(f) \tag{50}$$

is the desired-signal distortion due to the complex filter and

$$\mathcal{E}_{\mathrm{n}}(f) = \mathbf{h}^H(f)\mathbf{v}(f) \tag{51}$$

represents the residual noise.

The error signals $\mathcal{E}_{\mathrm{d}}(f)$ and $\mathcal{E}_{\mathrm{n}}(f)$ are incoherent.

The narrowband MSE is then

$$\begin{aligned} J[\mathbf{h}(f)] &= E\left[|\mathcal{E}(f)|^2\right] \\ &= \phi_{X_1}(f) + \mathbf{h}^H(f)\boldsymbol{\Phi}_{\mathbf{y}}(f)\mathbf{h}(f) - \phi_{X_1}(f)\mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) \\ &\quad - \phi_{X_1}(f)\boldsymbol{\gamma}^T_{X_1\mathbf{x}}(f)\mathbf{h}(f), \end{aligned} \tag{52}$$

which can be rewritten as

$$\begin{aligned} J[\mathbf{h}(f)] &= E\left[|\mathcal{E}_{\mathrm{d}}(f)|^2\right] + E\left[|\mathcal{E}_{\mathrm{n}}(f)|^2\right] \\ &= J_{\mathrm{d}}[\mathbf{h}(f)] + J_{\mathrm{n}}[\mathbf{h}(f)], \end{aligned} \tag{53}$$

where

$$\begin{aligned} J_{\mathrm{d}}\left[\mathbf{h}(f)\right] &= \phi_{X_1}(f)\left|\mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) - 1\right|^2 \\ &= \phi_{X_1}(f)\upsilon_{\mathrm{d}}\left[\mathbf{h}(f)\right] \end{aligned} \quad (54)$$

and

$$\begin{aligned} J_{\mathrm{n}}\left[\mathbf{h}(f)\right] &= \mathbf{h}^H(f)\boldsymbol{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f) \\ &= \frac{\phi_{V_1}(f)}{\xi_{\mathrm{n}}\left[\mathbf{h}(f)\right]}. \end{aligned} \quad (55)$$

Sometimes, it is also important to examine the MSE from the broadband point of view. We define the broadband MSE as

$$\begin{aligned} J(\mathbf{h}) &= \int_f J[\mathbf{h}(f)]\, df \\ &= \int_f J_\mathrm{d}[\mathbf{h}(f)]\, df + \int_f J_\mathrm{n}[\mathbf{h}(f)]\, df \\ &= J_\mathrm{d}(\mathbf{h}) + J_\mathrm{n}(\mathbf{h}). \end{aligned} \tag{56}$$

It is easy to show the relations between the broadband MSEs and the broadband performance measures:

$$\begin{aligned} \frac{J_\mathrm{d}(\mathbf{h})}{J_\mathrm{n}(\mathbf{h})} &= \mathrm{iSNR} \times \xi_\mathrm{n}(\mathbf{h}) \times \upsilon_\mathrm{d}(\mathbf{h}) \\ &= \mathrm{oSNR}(\mathbf{h}) \times \xi_\mathrm{d}(\mathbf{h}) \times \upsilon_\mathrm{d}(\mathbf{h}). \end{aligned} \tag{57}$$

Optimal Filters

Maximum SNR

Let us rewrite the narrowband output SNR:

$$\mathrm{oSNR}\left[\mathbf{h}(f)\right] = \frac{\phi_{X_1}(f)\mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)\boldsymbol{\gamma}^T_{X_1\mathbf{x}}(f)\mathbf{h}(f)}{\mathbf{h}^H(f)\boldsymbol{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)}. \tag{58}$$

The maximum SNR filter, $\mathbf{h}_{\max}(f)$, is obtained by maximizing the output SNR as given above.

In (58), we recognize the generalized Rayleigh quotient [5].

It is well known that this quotient is maximized with the maximum eigenvector of the matrix $\phi_{X_1}(f)\boldsymbol{\Phi}^{-1}_{\mathbf{v}}(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)\boldsymbol{\gamma}^T_{X_1\mathbf{x}}(f)$.

Let us denote by $\lambda_1(f)$ the maximum eigenvalue corresponding to this maximum eigenvector.

Since the rank of the mentioned matrix is equal to 1, we have

$$\begin{aligned}\lambda_1(f) &= \mathrm{tr}\left[\phi_{X_1}(f)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\right] \\ &= \phi_{X_1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f).\end{aligned} \quad (59)$$

As a result,

$$\begin{aligned}\mathrm{oSNR}\left[\mathbf{h}_{\max}(f)\right] &= \lambda_1(f) \\ &= \mathrm{oSNR}_{\max}(f),\end{aligned} \quad (60)$$

which corresponds to the maximum possible narrowband output SNR.

Obviously, we also have

$$\mathbf{h}_{\max}(f) = \varsigma(f)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f), \tag{61}$$

where $\varsigma(f)$ is an arbitrary frequency-dependent complex number different from zero.

While this factor has no effect on the narrowband output SNR, it has on the broadband output SNR and on the desired-signal distortion.

In fact, all the filters (except for the LCMV) derived in the rest of this section are equivalent up to this complex factor.

These filters also try to find the respective complex factors at each frequency depending on what we optimize.

It is important to understand that while the maximum SNR filter maximizes the narrowband output SNR, it certainly does not maximize the broadband output SNR whose value depends quite a lot on $\varsigma(f)$.

Let us denote by $\mathrm{oSNR}_{\max}^{(m)}(f)$ the maximum narrowband output SNR of an array with m sensors.

By virtue of the inclusion principle [5] for the matrix $\phi_{X_1}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)$, we have

$$\mathrm{oSNR}_{\max}^{(M)}(f) \geq \mathrm{oSNR}_{\max}^{(M-1)}(f) \geq \cdots \geq \mathrm{oSNR}_{\max}^{(2)}(f) \geq \mathrm{oSNR}_{\max}^{(1)}(f) = \mathrm{iSNR}(f). \quad (62)$$

This shows that by increasing the number of sensors, we necessarily increase the narrowband output SNR.

Example 1

Consider a ULA of M sensors. Suppose that a desired signal impinges on the ULA from the direction θ_0, and that an interference impinges on the ULA from the endfire direction ($\theta = 0°$).

Assume that the desired signal is a harmonic pulse of T samples:

$$x(t) = \begin{cases} A \sin\left(2\pi f_0 t + \phi\right), & 0 \leq t \leq T-1 \\ 0, & t < 0,\ t \geq T \end{cases},$$

with fixed, but unknown, amplitude A and frequency f_0, and random phase ϕ, uniformly distributed on the interval from 0 to 2π.

Assume that the interference $u(t)$ is white Gaussian noise, i.e., $u(t) \sim \mathcal{N}\left(0, \sigma_u^2\right)$, uncorrelated with $x(t)$.

In addition, the sensors contain thermal white Gaussian noise, $w_m(t) \sim \mathcal{N}\left(0, \sigma_w^2\right)$, that are mutually uncorrelated.

The desired signal needs to be recovered from the noisy received signals, $y_m(t) = x_m(t) + v_m(t),\ m = 1, \ldots, M$, where $v_m(t) = u_m(t) + w_m(t),\ m = 1, \ldots, M$ are the interference-plus-noise signals.

For simplicity, we choose a sampling interval T_s that satisfies $T_\mathrm{s} = \frac{d}{c}$.

Hence, the desired signal at sensor m is a delayed version of the desired signal at the first sensor:

$$x_m(t) = x_1\left(t - \tau_m\right),$$

where

$$\begin{aligned}\tau_m &= \frac{(m-1)d\cos\theta_0}{cT_\mathrm{s}} \\ &= (m-1)\cos\theta_0,\ m = 1, 2, \ldots, M\end{aligned}$$

is the relative time delay in samples (not necessarily an integer number) between the mth sensor and the first one.

The frequency-domain representation of the desired signal received at the first sensor is given by

$$\begin{aligned} X_1(f) &= \sum_{t=-\infty}^{\infty} x_1(t) e^{\jmath 2\pi f t} \\ &= \frac{A}{2\jmath} e^{\jmath\phi + \jmath\pi(f+f_0)(T-1)} D_T\left[\pi\left(f+f_0\right)\right] + \\ &\quad \frac{A}{2\jmath} e^{-\jmath\phi + \jmath\pi(f-f_0)(T-1)} D_T\left[\pi\left(f-f_0\right)\right], \end{aligned}$$

where

$$D_T(x) = \frac{\sin(Tx)}{\sin(x)}.$$

Therefore, the variance of $X_1(f)$ is

$$\phi_{X_1}(f) = \frac{A^2}{4} D_T^2 \left[\pi \left(f + f_0\right)\right] + \frac{A^2}{4} D_T^2 \left[\pi \left(f - f_0\right)\right].$$

Using the vector notation (3), we have

$$\mathbf{x}(f) = \mathbf{d}(f) X_1(f),$$
$$\mathbf{\Phi}_{\mathbf{x}}(f) = \phi_{X_1}(f)\mathbf{d}(f)\mathbf{d}^H(f),$$

where

$$\mathbf{d}(f) = \begin{bmatrix} 1 & e^{-\jmath 2\pi f \tau_2} & e^{-\jmath 2\pi f \tau_3} & \cdots & e^{-\jmath 2\pi f \tau_M} \end{bmatrix}^T .$$

The interference signal at sensor m is also a delayed version of the interference signal at the first sensor:

$$u_m(t) = u_1\left(t - m + 1\right).$$

The frequency-domain representation of the interference signal received at the first sensor is given by

$$U_1(f) = \sum_{t=0}^{T-1} u_1(t) e^{\jmath 2\pi f t}.$$

Hence, the variance of $U_1(f)$ is $\phi_{U_1}(f) = T\sigma_u^2$.

Using the vector notation (13), we have

$$\begin{aligned}\mathbf{v}(f) &= \boldsymbol{\gamma}_{U_1\mathbf{u}}^{*}(f) U_1(f) + \mathbf{w}(f),\\ \mathbf{\Phi}_{\mathbf{v}}(f) &= \phi_{U_1}(f)\boldsymbol{\gamma}_{U_1\mathbf{u}}^{*}(f)\boldsymbol{\gamma}_{U_1\mathbf{u}}^{T}(f) + T\sigma_w^2 \mathbf{I}_M,\end{aligned}$$

where

$$\boldsymbol{\gamma}_{U_1\mathbf{u}}^{*}(f) = \left[\begin{array}{ccccc} 1 & e^{-\jmath 2\pi f} & e^{-\jmath 2\pi f 2} & \cdots & e^{-\jmath 2\pi f (M-1)} \end{array} \right]^T$$

and $\mathbf{I}_M$ is the $M \times M$ identity matrix.

The narrowband and broadband input SNRs are, respectively,

$$\begin{aligned} \mathrm{iSNR}(f) &= \frac{\phi_{X_1}(f)}{\phi_{V_1}(f)} \\ &= \frac{A^2}{4T\left(\sigma_u^2 + \sigma_w^2\right)} D_T^2\left[\pi\left(f + f_0\right)\right] + \\ &\quad \frac{A^2}{4T\left(\sigma_u^2 + \sigma_w^2\right)} D_T^2\left[\pi\left(f - f_0\right)\right] \end{aligned}$$

and

$$\begin{aligned}\mathrm{iSNR} &= \frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{V_1}(f)df} \\ &= \frac{\sum_t E\left[|x_1(t)|^2\right]}{\sum_t E\left[|v_1(t)|^2\right]} \\ &= \frac{A^2}{2\left(\sigma_u^2 + \sigma_w^2\right)},\end{aligned}$$

where we have used Parseval's identity.

The maximum SNR filter, $\mathbf{h}_{\max}(f)$, is obtained from (61).

Using (60), we can write the narrowband gain in SNR as

$$\begin{aligned}\mathcal{G}\left[\mathbf{h}_{\max}(f)\right] &= \frac{\mathrm{oSNR}\left[\mathbf{h}_{\max}(f)\right]}{\mathrm{iSNR}(f)} \\ &= \mathbf{d}^H(f)\left[\frac{\sigma_u^2}{\sigma_u^2+\sigma_w^2}\boldsymbol{\gamma}^*_{U_1\mathbf{u}}(f)\boldsymbol{\gamma}^T_{U_1\mathbf{u}}(f) + \frac{\sigma_w^2}{\sigma_u^2+\sigma_w^2}\mathbf{I}_M\right]^{-1}\mathbf{d}(f).\end{aligned}$$

To demonstrate the performance of the maximum SNR filter, we choose $\sigma_w^2 = 0.01\sigma_u^2$.

Figure 2 shows the effect of the number of sensors, M, on the narrowband gain in SNR, $\mathcal{G}\left[\mathbf{h}_{\max}(f)\right]$, for different incidence angles of the desired signal and different frequencies.

For a single sensor ($M = 1$), there is no narrowband gain in SNR. As the number of sensors increases, the narrowband gain in SNR increases.

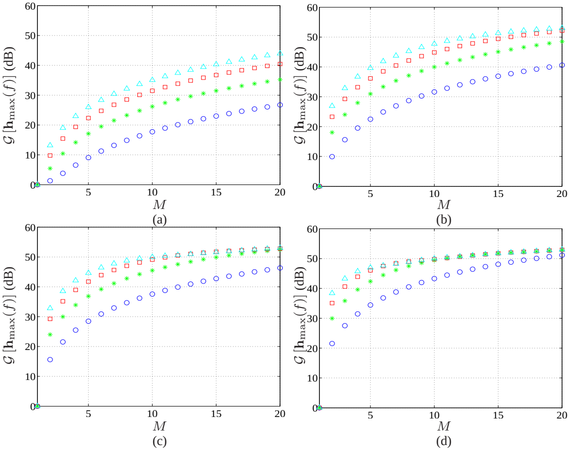


Figure 2: Narrowband gain in SNR of the maximum SNR filter for different incidence angles of the desired signal and different frequencies: $\theta_0 = 30^\circ$ (circles), $\theta_0 = 50^\circ$ (asterisks), $\theta_0 = 70^\circ$ (squares), and $\theta_0 = 90^\circ$ (triangles); (a) $f = 0.01$, (b) $f = 0.05$, (c) $f = 0.1$, and (d) $f = 0.2$.

Wiener

The Wiener filter is found by minimizing the narrowband MSE, $J\left[\mathbf{h}(f)\right]$ [eq. (52)]. We get

$$\mathbf{h}_{\mathrm{W}}(f) = \phi_{X_1}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f). \tag{63}$$

Let

$$\mathbf{\Gamma}_{\mathbf{y}}(f) = \frac{\mathbf{\Phi}_{\mathbf{y}}(f)}{\phi_{Y_1}(f)} \tag{64}$$

be the pseudo-coherence matrix of the observations, we can rewrite (63) as

$$\begin{aligned}\mathbf{h}_{\mathrm{W}}(f) &= \frac{\mathrm{iSNR}(f)}{1+\mathrm{iSNR}(f)}\mathbf{\Gamma}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f) \\ &= H_{\mathrm{W}}(f)\mathbf{\Gamma}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f),\end{aligned} \tag{65}$$

where

$$H_{\mathrm{W}}(f) = \frac{\mathrm{iSNR}(f)}{1 + \mathrm{iSNR}(f)} \tag{66}$$

is the (single-channel) Wiener gain and $\mathbf{\Gamma}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)$ is the spatial information vector.

The decomposition in (65) is very interesting; it shows separately the influence of the spectral and spatial processing on multichannel signal enhancement.

We can express (63) differently, i.e.,

$$\begin{aligned} \mathbf{h}_{\mathrm{W}}(f) &= \mathbf{\Phi}_{\mathbf{y}}^{-1}(f) E\left[\mathbf{x}(f) X_1^*(f)\right] \\ &= \mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{i}_{\mathrm{i}} \\ &= \left[\mathbf{I}_M - \mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{v}}(f)\right]\mathbf{i}_{\mathrm{i}}. \end{aligned} \tag{67}$$

In this form, the Wiener filter relies on the second-order statistics of the observation and noise signals.

We can write the general form of the Wiener filter in another way that will make it easier to compare to other optimal filters. We know that

$$\mathbf{\Phi}_{\mathbf{y}}(f) = \phi_{X_1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^*(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^T(f) + \mathbf{\Phi}_{\mathbf{v}}(f). \tag{68}$$

Determining the inverse of $\mathbf{\Phi}_{\mathbf{y}}(f)$ from the previous expression with the Woodbury's identity, we get

$$\mathbf{\Phi}_{\mathbf{y}}^{-1}(f) = \mathbf{\Phi}_{\mathbf{v}}^{-1}(f) - \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)}{\phi_{X_1}^{-1}(f) + \boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)}. \tag{69}$$

Substituting (69) into (63) gives

$$\mathbf{h}_{\mathrm{W}}(f) = \frac{\phi_{X_1}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)}{1 + \phi_{X_1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)}, \tag{70}$$

that we can rewrite as

$$\begin{aligned} \mathbf{h}_{\mathrm{W}}(f) &= \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\left[\mathbf{\Phi}_{\mathbf{y}}(f) - \mathbf{\Phi}_{\mathbf{v}}(f)\right]}{1 + \mathrm{tr}\left\{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\left[\mathbf{\Phi}_{\mathbf{y}}(f) - \mathbf{\Phi}_{\mathbf{v}}(f)\right]\right\}}\mathbf{i}_{\mathrm{i}} \\ &= \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{1 - M + \mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f)\right]}\mathbf{i}_{\mathrm{i}}. \end{aligned} \tag{71}$$

Comparing (67) with (71), we see that in the former, we invert the correlation matrix of the observations, while in the latter, we invert the correlation matrix of the noise.

It is interesting to see that the two filters $\mathbf{h}_{\mathrm{W}}(f)$ and $\mathbf{h}_{\max}(f)$ differ only by a real-valued factor. Indeed, taking

$$\varsigma(f) = \frac{\phi_{X_1}(f)}{1 + \lambda_1(f)} \tag{72}$$

in (61) (maximum SNR filter), we find (70) (Wiener filter).

We can express $\mathbf{h}_{\mathrm{W}}(f)$ as a function of the narrowband input SNR and the pseudo-coherence matrices, i.e.,

$$\mathbf{h}_{\mathrm{W}}(f) = \frac{\left[1 + \mathrm{iSNR}(f)\right] \boldsymbol{\Gamma}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Gamma}_{\mathbf{y}}(f) - \mathbf{I}_M}{1 - M + \left[1 + \mathrm{iSNR}(f)\right] \mathrm{tr}\left[\boldsymbol{\Gamma}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Gamma}_{\mathbf{y}}(f)\right]} \mathbf{i}_{\mathrm{i}}, \tag{73}$$

where

$$\mathbf{\Gamma}_{\mathbf{v}}(f) = \frac{\mathbf{\Phi}_{\mathbf{v}}(f)}{\phi_{V_1}(f)}. \tag{74}$$

From (70), we deduce that the narrowband output SNR is

$$\begin{aligned} \mathrm{oSNR}\left[\mathbf{h}_{\mathrm{W}}(f)\right] &= \lambda_1(f) \\ &= \mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f)\right] - M \end{aligned} \tag{75}$$

and, obviously,

$$\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{W}}(f)\right] \geq \mathrm{iSNR}(f), \tag{76}$$

since the Wiener filter maximizes the narrowband output SNR.

The desired-signal distortion indices are

$$\upsilon_{\mathrm{d}}\left[\mathbf{h}_{\mathrm{W}}(f)\right]=\frac{1}{\left[1+\lambda_1(f)\right]^2}, \tag{77}$$

$$\upsilon_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right)=\frac{\int_f \phi_{X_1}(f)\left[1+\lambda_1(f)\right]^{-2} df}{\int_f \phi_{X_1}(f)df}. \tag{78}$$

The higher the value of $\lambda_1(f)$ (and/or the number of sensors), the less the desired signal is distorted.

It is also easy to find the noise reduction factors:

$$\xi_{\mathrm{n}}\left[\mathbf{h}_{\mathrm{W}}(f)\right]=\frac{\left[1+\lambda_1(f)\right]^2}{\mathrm{iSNR}(f)\times\lambda_1(f)}, \tag{79}$$

$$\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{W}}\right)=\frac{\int_f \phi_{X_1}(f)\mathrm{iSNR}^{-1}(f)df}{\int_f \phi_{X_1}(f)\lambda_1(f)\left[1+\lambda_1(f)\right]^{-2} df}, \tag{80}$$

and the desired-signal reduction factors:

$$\xi_{\mathrm{d}}\left[\mathbf{h}_{\mathrm{W}}(f)\right] = \frac{\left[1+\lambda_1(f)\right]^2}{\lambda_1^2(f)}, \tag{81}$$

$$\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right) = \frac{\int_f \phi_{X_1}(f) df}{\int_f \phi_{X_1}(f)\lambda_1^2(f)\left[1+\lambda_1(f)\right]^{-2} df}. \tag{82}$$

The broadband output SNR of the Wiener filter is

$$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) = \frac{\displaystyle\int_f \phi_{X_1}(f) \frac{\lambda_1^2(f)}{\left[1+\lambda_1(f)\right]^2} df}{\displaystyle\int_f \phi_{X_1}(f) \frac{\lambda_1(f)}{\left[1+\lambda_1(f)\right]^2} df}. \tag{83}$$

Property

With the frequency-domain multichannel Wiener filter given in (63), the broadband output SNR is always greater than or equal to the broadband input SNR, i.e., $\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) \geq \mathrm{iSNR}$.

Example 2

Returning to Example 1, we now employ the Wiener filter, $\mathbf{h}_{\mathrm{W}}(f)$, given in (63).

To demonstrate the performance of the Wiener filter, we choose $A = 0.5$, $f_0 = 0.1$, $T = 500$, $\theta_0 = 70°$, and $\sigma_w^2 = 0.01\sigma_u^2$.

Figure 3 shows plots of the broadband gain in SNR, $\mathcal{G}\left(\mathbf{h}_{\mathrm{W}}\right)$, the broadband MSE, $J\left(\mathbf{h}_{\mathrm{W}}\right)$, the broadband noise reduction factor, $\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{W}}\right)$, and the broadband desired-signal reduction factor, $\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right)$, as a function of the broadband input SNR, for different numbers of sensors.

For a given broadband input SNR, as the number of sensors increases, the broadband gain in SNR and the broadband noise reduction factor increase, while the broadband MMSE and the broadband desired-signal reduction factor decrease.

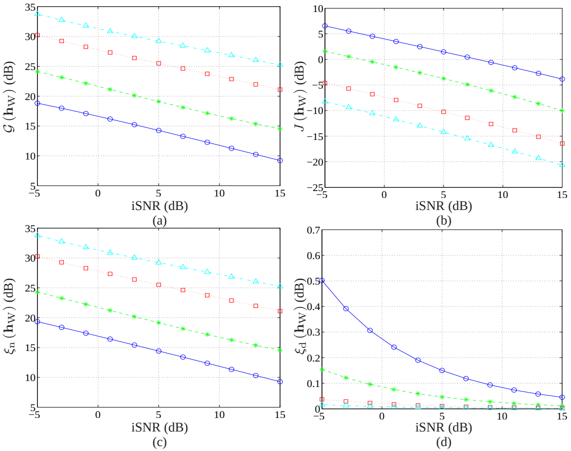


Figure 3: The broadband (a) gain in SNR, (b) MSE, (c) noise reduction factor, and (d) desired-signal reduction factor of the Wiener filter for different numbers of sensors, M: $M = 1$ (solid line with circles), $M = 2$ (dashed line with asterisks), $M = 5$ (dotted line with squares), and $M = 10$ (dash-dot line with triangles).

Figure 4 shows a realization of the frequency-domain noise corrupted signal received at the first sensor, $|Y_1(f)|$, and the error signals $|\mathcal{E}(f)| = |Z(f) - X_1(f)|$ for $\mathrm{iSNR} = -5$ dB and different numbers of sensors.

Figure 5 shows the corresponding time-domain observation signal at the first sensor, $y_1(t)$, and the time-domain estimated signals, $z(t)$.

Obviously, as the number of sensors increases, the Wiener filter better enhances the desired signal.

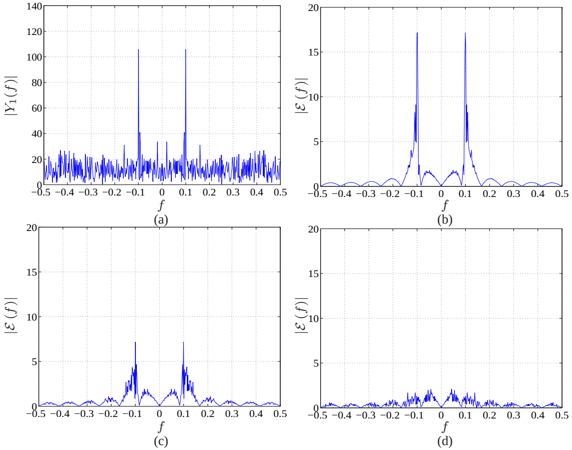


Figure 4: Example of frequency-domain noise corrupted and error signals of the Wiener filter for different numbers of sensors, M: (a) $|Y_1(f)|$ (iSNR $= -5$ dB), and $|\mathcal{E}(f)| = |Z(f) - X_1(f)|$ for (b) $M = 1$ [$\text{oSNR}(\mathbf{h}_\text{W}) = 14.2$ dB], (c) $M = 2$ [$\text{oSNR}(\mathbf{h}_\text{W}) = 19.2$ dB], and (d) $M = 5$ [$\text{oSNR}(\mathbf{h}_\text{W}) = 25.5$ dB].

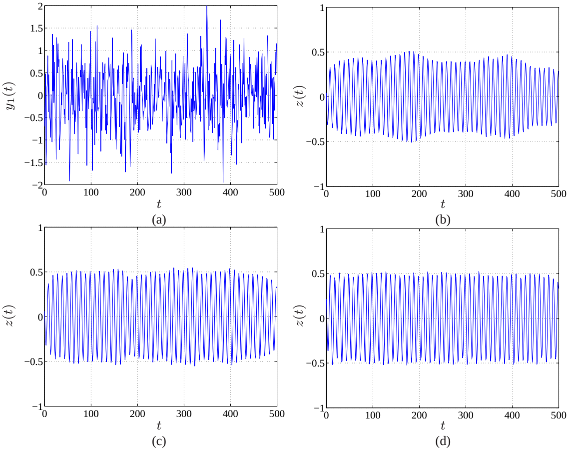


Figure 5: Example of time-domain noise corrupted and Wiener filtered sinusoidal signals for different numbers of sensors, M: (a) $y_1(t)$ ($\mathrm{iSNR} = -5$ dB), and $z(t)$, for (b) $M = 1$ [$\mathrm{oSNR}(\mathbf{h}_\mathrm{W}) = 14.2$ dB], (c) $M = 2$ [$\mathrm{oSNR}(\mathbf{h}_\mathrm{W}) = 19.2$ dB], and (d) $M = 5$ [$\mathrm{oSNR}(\mathbf{h}_\mathrm{W}) = 25.5$ dB].

MVDR

The well-known MVDR filter proposed by Capon [6], [7] is easily derived by minimizing the narrowband MSE of the residual noise, $J_{\mathrm{n}}\left[\mathbf{h}(f)\right]$, with the constraint that the desired signal is not distorted.

Mathematically, this is equivalent to

$$\min_{\mathbf{h}(f)} \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f) \quad \text{subject to} \quad \mathbf{h}^H(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) = 1, \tag{84}$$

for which the solution is

$$\mathbf{h}_{\mathrm{MVDR}}(f) = \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)}{\boldsymbol{\gamma}^T_{X_1\mathbf{x}}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)}. \tag{85}$$

Using the fact that $\mathbf{\Phi}_{\mathbf{x}}(f) = \phi_{X_1}(f)\boldsymbol{\gamma}^{*}_{X_1\mathbf{x}}(f)\boldsymbol{\gamma}^{T}_{X_1\mathbf{x}}(f)$, the explicit dependence of the above filter on the steering vector is eliminated to obtain the following forms:

$$\begin{aligned}
\mathbf{h}_{\mathrm{MVDR}}(f) &= \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{x}}(f)}{\lambda_1(f)}\mathbf{i}_{\mathrm{i}} \\
&= \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{\mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f)\right] - M}\mathbf{i}_{\mathrm{i}} \\
&= \frac{\left[1 + \mathrm{iSNR}(f)\right]\mathbf{\Gamma}_{\mathbf{v}}^{-1}(f)\mathbf{\Gamma}_{\mathbf{y}}(f) - \mathbf{I}_M}{\left[1 + \mathrm{iSNR}(f)\right]\mathrm{tr}\left[\mathbf{\Gamma}_{\mathbf{v}}^{-1}(f)\mathbf{\Gamma}_{\mathbf{y}}(f)\right] - M}\mathbf{i}_{\mathrm{i}}.
\end{aligned} \tag{86}$$

Alternatively, we can also write the MVDR as

$$\mathbf{h}_{\mathrm{MVDR}}(f) = \frac{\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)}{\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)} = \frac{\mathbf{\Gamma}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)}{\boldsymbol{\gamma}_{X_1\mathbf{x}}^{T}(f)\mathbf{\Gamma}_{\mathbf{y}}^{-1}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f)}. \tag{87}$$

Taking

$$\varsigma(f) = \frac{\phi_{X_1}(f)}{\lambda_1(f)} \tag{88}$$

in (61) (maximum SNR filter), we find (85) (MVDR filter), showing how the maximum SNR and MVDR filters are equivalent up to a real-valued factor.

The Wiener and MVDR filters are simply related as follows:

$$\mathbf{h}_{\mathrm{W}}(f) = C_{\mathrm{W}}(f)\mathbf{h}_{\mathrm{MVDR}}(f), \tag{89}$$

where

$$\begin{aligned} C_{\mathrm{W}}(f) &= \mathbf{h}_{\mathrm{W}}^{H}(f)\boldsymbol{\gamma}_{X_1\mathbf{x}}^{*}(f) \\ &= \frac{\lambda_1(f)}{1+\lambda_1(f)} \end{aligned} \tag{90}$$

can be seen as a single-channel frequency-domain Wiener gain.

In fact, any filter of the form:

$$\mathbf{h}(f) = C(f)\mathbf{h}_{\mathrm{MVDR}}(f), \tag{91}$$

where $C(f)$ is a real number, with $0 < C(f) < 1$, removes more noise than the MVDR filter at the price of some desired-signal distortion, which is

$$\xi_{\mathrm{d}}\left[\mathbf{h}(f)\right] = \frac{1}{C^2(f)} \tag{92}$$

or

$$\upsilon_{\mathrm{d}}\left[\mathbf{h}(f)\right] = \left[C(f) - 1\right]^2. \tag{93}$$

It can be verified that we always have

$$\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(f)\right] = \mathrm{oSNR}\left[\mathbf{h}_{\mathrm{W}}(f)\right], \tag{94}$$

$$\upsilon_{\mathrm{d}}\left[\mathbf{h}_{\mathrm{MVDR}}(f)\right] = 0, \tag{95}$$

$$\xi_{\mathrm{d}}\left[\mathbf{h}_{\mathrm{MVDR}}(f)\right] = 1, \tag{96}$$

and

$$\xi_{\mathrm{n}}\left[\mathbf{h}_{\mathrm{MVDR}}(f)\right] \leq \xi_{\mathrm{n}}\left[\mathbf{h}_{\mathrm{W}}(f)\right], \tag{97}$$

$$\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq \xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{W}}\right). \tag{98}$$

The MVDR filter rejects the maximum level of noise allowable without distorting the desired signal at each frequency.

While the narrowband output SNRs of the Wiener and MVDR are strictly equal, their broadband output SNRs are not.

The broadband output SNR of the MVDR is

$$\text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right) = \frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{X_1}(f)\lambda_1^{-1}(f)df} \tag{99}$$

and

$$\text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right) \leq \text{oSNR}\left(\mathbf{h}_{\text{W}}\right). \tag{100}$$

Property

With the frequency-domain MVDR filter given in (85), the broadband output SNR is always greater than or equal to the broadband input SNR, i.e., $\text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right) \geq \text{iSNR}$.

Example 3

Returning to Example 2, we now employ the MVDR filter, $\mathbf{h}_{\mathrm{MVDR}}(f)$, given in (85).

Figure 6 shows plots of the broadband gain in SNR, $\mathcal{G}\left(\mathbf{h}_{\mathrm{MVDR}}\right)$, the broadband MSE, $J\left(\mathbf{h}_{\mathrm{MVDR}}\right)$, the broadband noise reduction factor, $\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{MVDR}}\right)$, and the broadband desired-signal reduction factor, $\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{MVDR}}\right)$, as a function of the broadband input SNR, for different numbers of sensors.

For a given broadband input SNR, as the number of sensors increases, the broadband gain in SNR and the broadband noise reduction factor increase, while the broadband MSE decreases.

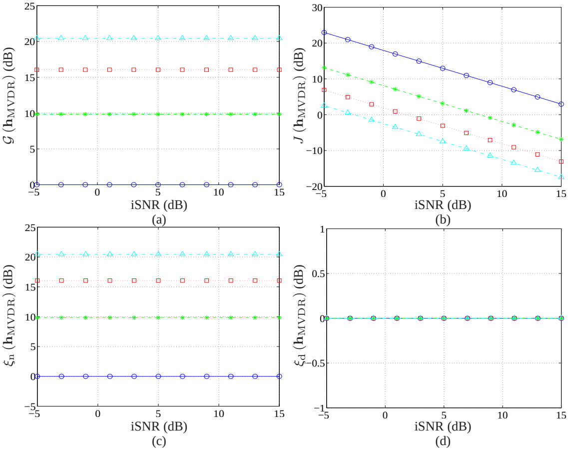


Figure 6: The broadband (a) gain in SNR, (b) MSE, (c) noise reduction factor, and (d) desired-signal reduction factor of the MVDR filter for different numbers of sensors, M: $M = 1$ (solid line with circles), $M = 2$ (dashed line with asterisks), $M = 5$ (dotted line with squares), and $M = 10$ (dash-dot line with triangles).

Tradeoff

In many practical situations, we wish to control the compromise between noise reduction and desired-signal distortion, and one possible way to do this is via the so-called tradeoff filter.

In the tradeoff approach, we minimize the narrowband desired-signal distortion index with the constraint that the narrowband noise reduction factor is equal to a positive value that is greater than 1.

Mathematically, this is equivalent to

$$\min_{\mathbf{h}(f)} J_{\mathrm{d}}\left[\mathbf{h}(f)\right] \quad \text{subject to} \quad J_{\mathrm{n}}\left[\mathbf{h}(f)\right] = \aleph \phi_{V_1}(f), \tag{101}$$

where $0 < \aleph < 1$ to insure that we get some noise reduction.

By using a Lagrange multiplier, $\mu > 0$, to adjoin the constraint to the cost function, we easily deduce the tradeoff filter:

$$\begin{aligned}\mathbf{h}_{\mathrm{T},\mu}(f) &= \phi_{X_1}(f)\left[\mathbf{\Phi}_{\mathbf{x}}(f) + \mu\mathbf{\Phi}_{\mathbf{v}}(f)\right]^{-1}\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) \\ &= \frac{\phi_{X_1}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)}{\mu + \phi_{X_1}(f)\boldsymbol{\gamma}^T_{X_1\mathbf{x}}(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f)} \\ &= \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{\mu - M + \mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f)\right]}\mathbf{i}_{\mathrm{i}},\end{aligned} \tag{102}$$

where the Lagrange multiplier, μ, satisfies

$$J_{\mathrm{n}}\left[\mathbf{h}_{\mathrm{T},\mu}(f)\right] = \aleph\phi_{V_1}(f). \tag{103}$$

However, in practice it is not easy to determine the optimal μ.

Therefore, when this parameter is chosen in a heuristic way, we can see that for

- $\mu = 1$, $\mathbf{h}_{\mathrm{T},1}(f) = \mathbf{h}_{\mathrm{W}}(f)$, which is the Wiener filter;
- $\mu = 0$, $\mathbf{h}_{\mathrm{T},0}(f) = \mathbf{h}_{\mathrm{MVDR}}(f)$, which is the MVDR filter;
- $\mu > 1$, results in a filter with low residual noise at the expense of high desired-signal distortion (as compared to Wiener); and
- $\mu < 1$, results in a filter with high residual noise and low desired-signal distortion (as compared to Wiener).

Note that the MVDR cannot be derived from the first line of (102) since by taking $\mu = 0$, we have to invert a matrix that is not full rank.

It can be observed that the tradeoff, Wiener, and maximum SNR filters are equivalent up to a real-valued number.

As a result, the narrowband output SNR of the tradeoff filter is independent of μ and is identical to the narrowband output SNR of the maximum SNR filter, i.e.,

$$\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{T},\mu}(f)\right] = \mathrm{oSNR}\left[\mathbf{h}_{\max}(f)\right], \ \forall \mu \geq 0. \tag{104}$$

We have

$$v_{\mathrm{d}}\left[\mathbf{h}_{\mathrm{T},\mu}(f)\right] = \left[\frac{\mu}{\mu + \lambda_1(f)}\right]^2, \tag{105}$$

$$\xi_{\mathrm{d}}\left[\mathbf{h}_{\mathrm{T},\mu}(f)\right] = \left[1 + \frac{\mu}{\lambda_1(f)}\right]^2, \tag{106}$$

$$\xi_{\mathrm{n}}\left[\mathbf{h}_{\mathrm{T},\mu}(f)\right] = \frac{\left[\mu + \lambda_1(f)\right]^2}{\mathrm{iSNR}(f) \times \lambda_1(f)}. \tag{107}$$

The tradeoff filter is interesting from several perspectives since it encompasses both the Wiener and MVDR filters.

It is then useful to study the broadband output SNR and the broadband desired-signal distortion index of the tradeoff filter.

It can be verified that the broadband output SNR of the tradeoff filter is

$$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \frac{\displaystyle\int_f \phi_{X_1}(f) \frac{\lambda_1^2(f)}{\left[\mu + \lambda_1(f)\right]^2} df}{\displaystyle\int_f \phi_{X_1}(f) \frac{\lambda_1(f)}{\left[\mu + \lambda_1(f)\right]^2} df}. \tag{108}$$

Property

The broadband output SNR of the tradeoff filter is an increasing function of the parameter μ.

From this property we deduce that the MVDR filter gives the smallest broadband output SNR.

While the broadband output SNR is upper bounded, it is easy to see that the broadband noise reduction factor and broadband desired-signal reduction factor are not.

So when μ goes to infinity, so are $\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{T},\mu}\right)$ and $\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right)$.

The broadband desired-signal distortion index is

$$\upsilon_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) = \frac{\displaystyle\int_f \phi_{X_1}(f)\frac{\mu^2}{\left[\mu+\lambda_1(f)\right]^2}df}{\int_f \phi_{X_1}(f)df}. \tag{109}$$

Property

The broadband desired-signal distortion index of the tradeoff filter is an increasing function of the parameter μ.

It is clear that

$$0 \leq v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) \leq 1, \ \forall \mu \geq 0. \tag{110}$$

Therefore, as μ increases, the broadband output SNR increases at the price of more distortion to the desired signal.

Property

With the frequency-domain tradeoff filter given in (102), the broadband output SNR is always greater than or equal to the broadband input SNR, i.e., $\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{T},\mu}\right) \geq \mathrm{iSNR}, \ \forall \mu \geq 0.$

From the previous results, we deduce that for $\mu \geq 1$,

$$\mathrm{iSNR} \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{T},\mu}\right), \tag{111}$$

$$0 = v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right) \leq v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right), \tag{112}$$

and for $0 \leq \mu \leq 1$,

$$\mathrm{iSNR} \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{T},\mu}\right) \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right), \tag{113}$$

$$0 = v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right) \leq v_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right). \tag{114}$$

Example 4

Returning to Example 2, we now employ the tradeoff filter, $\mathbf{h}_{\mathrm{T},\mu}(f)$, given in (102). We assume $M = 5$ sensors.

Figure 7 shows plots of the broadband gain in SNR, $\mathcal{G}\left(\mathbf{h}_{\mathrm{T},\mu}\right)$, the broadband desired-signal distortion index, $\upsilon_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right)$, the broadband noise reduction factor, $\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{T},\mu}\right)$, and the broadband desired-signal reduction factor, $\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{T},\mu}\right)$, as a function of the broadband input SNR, for several values of μ.

For a given broadband input SNR, the higher is the value of μ, the higher are the broadband gain in SNR and the broadband noise reduction factor, but at the expense of higher broadband desired-signal distortion index and higher broadband desired-signal reduction factor.

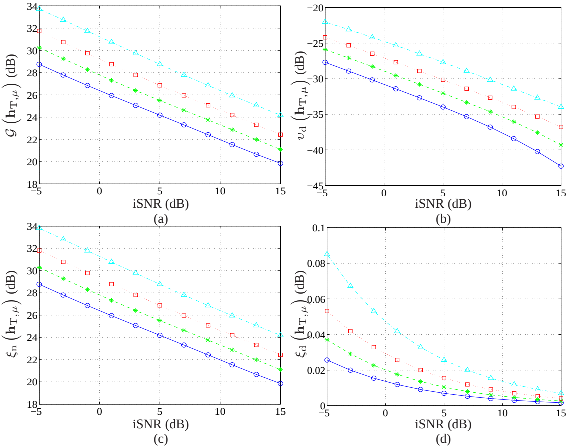


Figure 7: The broadband (a) gain in SNR, (b) desired-signal distortion index, (c) noise reduction factor, and (d) desired-signal reduction factor of the tradeoff filter for several values of μ: $\mu = 0.5$ (solid line with circles), $\mu = 1$ (dashed line with asterisks), $\mu = 2$ (dotted line with squares), and $\mu = 5$ (dash-dot line with triangles).

LCMV

With the linearly constrained minimum variance (LCMV) filter [9], [10], [11], [12], we wish to perfectly recover our desired signal, $X_1(f)$, and completely remove the coherent components, $\boldsymbol{\gamma}^*_{V_1\mathbf{v}}(f)V_1(f)$ [see eq. (11)].

The two constraints can be put together in a matrix form as

$$\mathbf{C}^H_{X_1V_1}(f)\mathbf{h}(f) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{115}$$

where

$$\mathbf{C}_{X_1V_1}(f) = \begin{bmatrix} \boldsymbol{\gamma}^*_{X_1\mathbf{x}}(f) & \boldsymbol{\gamma}^*_{V_1\mathbf{v}}(f) \end{bmatrix} \tag{116}$$

is our constraint matrix of size $M \times 2$.

Then, our optimal filter is obtained by minimizing the energy at the filter output, with the constraints that the coherent noise components are cancelled and the desired signal is preserved, i.e.,

$$\mathbf{h}_{\mathrm{LCMV}}(f) = \arg\min_{\mathbf{h}(f)} \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{h}(f) \quad \text{subject to}$$
$$\mathbf{C}^H_{X_1V_1}(f)\mathbf{h}(f) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{117}$$

The solution to (117) is given by

$$\mathbf{h}_{\mathrm{LCMV}}(f) = \mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}_{X_1V_1}(f)\left[\mathbf{C}^H_{X_1V_1}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}_{X_1V_1}(f)\right]^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{118}$$

We always have

$$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{LCMV}}\right) \leq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MVDR}}\right), \tag{119}$$

$$\upsilon_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{LCMV}}\right) = 0, \tag{120}$$

$$\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{LCMV}}\right) = 1, \tag{121}$$

and

$$\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{LCMV}}\right) \leq \xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{MVDR}}\right) \leq \xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{W}}\right). \tag{122}$$

The LCMV structure can be an interesting solution in practical applications where the coherent noise is more problematic than the incoherent one.

Example 5

Returning to Example 2, we now employ the LCMV filter, $\mathbf{h}_{\mathrm{LCMV}}(f)$, given in (118). We assume $M = 5$ sensors.

Figure 8 shows plots of the broadband gain in SNR, $\mathcal{G}\left(\mathbf{h}_{\mathrm{LCMV}}\right)$, the broadband MSE, $J\left(\mathbf{h}_{\mathrm{LCMV}}\right)$, the broadband noise reduction factor, $\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{LCMV}}\right)$, and the broadband desired-signal reduction factor, $\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{LCMV}}\right)$, as a function of the broadband input SNR, for several values of $\alpha = \sigma_w^2 / \sigma_u^2$.

For a given broadband input SNR, as the ratio between the coherent to incoherent noise increases (α decreases), the LCMV filter yields higher broadband gain in SNR and higher broadband noise reduction factor.

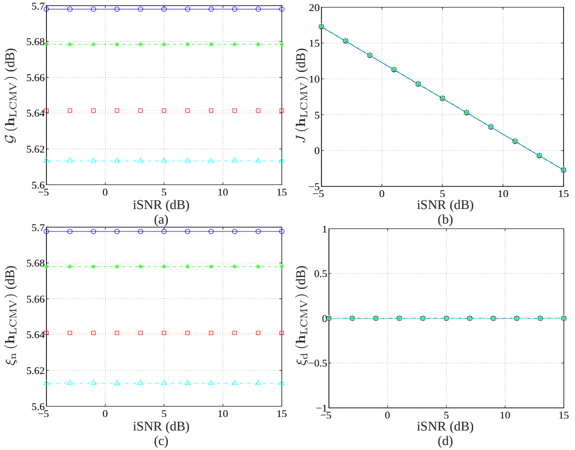


Figure 8: The broadband (a) gain in SNR, (b) MSE, (c) noise reduction factor, and (d) desired-signal reduction factor of the LCMV filter for several values of $\alpha = \sigma_w^2/\sigma_u^2$: $\alpha = 0.03$ (solid line with circles), $\alpha = 0.1$ (dashed line with asterisks), $\alpha = 0.3$ (dotted line with squares), and $\alpha = 1$ (dash-dot line with triangles).

The LCMV filter shown above can, obviously, be extended to any number of linear constraints $M_{\mathrm{c}} \leq M$.

The constraint equation, which includes the distortionless constraint, can be expressed as

$$\mathbf{C}^H(f)\mathbf{h}(f) = \mathbf{i}_{\mathrm{c}}, \tag{123}$$

where

$$\mathbf{C}(f) = \begin{bmatrix} \mathbf{d}(f) & \mathbf{c}_2(f) & \cdots & \mathbf{c}_{M_{\mathrm{c}}}(f) \end{bmatrix} \tag{124}$$

is a matrix of size $M \times M_{\mathrm{c}}$ whose M_{c} columns are linearly independent and $\mathbf{i}_{\mathrm{c}}$ is a vector of length M_{c} whose first component is equal to 1 and the other components are some chosen real numbers to satisfy the constraints on the filter.

Generally, these constraints are null ones where it is desired to completely cancel some interference sources.

Following the same steps as above, we easily find the LCMV filter:

$$\mathbf{h}_{\mathrm{LCMV}}(f) = \mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\left[\mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\right]^{-1}\mathbf{i}_{\mathrm{c}}. \quad (125)$$

For $M_{\mathrm{c}} = 1$, $\mathbf{h}_{\mathrm{LCMV}}(f)$ simplifies to $\mathbf{h}_{\mathrm{MVDR}}(f)$.

In Table 1, we summarize all the optimal filters studied in this section.

Table 1: Optimal linear filters for multichannel signal enhancement in the frequency domain.

Maximum SNR:	$\mathbf{h}_{\max}(f) = \varsigma(f)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f),\ \varsigma(f) \neq 0$
Wiener:	$\mathbf{h}_{\mathrm{W}}(f) = \dfrac{\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{1 - M + \mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Phi}_{\mathbf{y}}(f)\right]}\mathbf{i}_{\mathrm{i}}$
MVDR:	$\mathbf{h}_{\mathrm{MVDR}}(f) = \dfrac{\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{\mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Phi}_{\mathbf{y}}(f)\right] - M}\mathbf{i}_{\mathrm{i}}$
Tradeoff:	$\mathbf{h}_{\mathrm{T},\mu} = \dfrac{\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{\mu - M + \mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(f)\boldsymbol{\Phi}_{\mathbf{y}}(f)\right]}\mathbf{i}_{\mathrm{i}},\ \mu \geq 0$
LCMV:	$\mathbf{h}_{\mathrm{LCMV}}(f) = \boldsymbol{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\left[\mathbf{C}^H(f)\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\right]^{-1}\mathbf{i}_{\mathrm{c}}$

Generalized Sidelobe Canceller Structure

The generalized sidelobe canceller (GSC) structure solves exactly the same problem as the LCMV approach by dividing the filter vector $\mathbf{h}_{\mathrm{LCMV}}(f)$ into two components operating on orthogonal subspaces [13], [14], [15], [16]:

$$\mathbf{h}_{\mathrm{LCMV}}(f) = \mathbf{h}_{\mathrm{MN}}(f) - \mathbf{B}_{\mathbf{C}}(f)\mathbf{w}_{\mathrm{GSC}}(f), \tag{126}$$

where

$$\mathbf{h}_{\mathrm{MN}}(f) = \mathbf{C}(f)\left[\mathbf{C}^H(f)\mathbf{C}(f)\right]^{-1}\mathbf{i}_{\mathrm{c}} \tag{127}$$

is the minimum-norm solution of (123), $\mathbf{B}_{\mathbf{C}}(f)$ is the so-called blocking matrix that spans the nullspace of $\mathbf{C}^H(f)$, i.e.,

$$\mathbf{C}^H(f)\mathbf{B}_{\mathbf{C}}(f) = \mathbf{0}_{M_{\mathrm{c}}\times(M-M_{\mathrm{c}})}, \tag{128}$$

and $\mathbf{w}_{\mathrm{GSC}}(f)$ is a weighting vector derived as explained below.

A block diagram of the GSC is illustrated in Fig. 9.

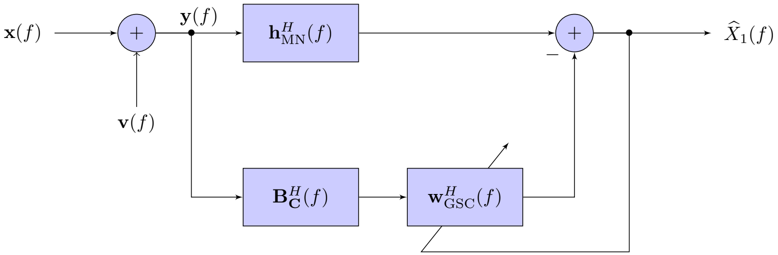


Figure 9: Block diagram of the generalized sidelobe canceller.

The size of $\mathbf{B_C}(f)$ is $M \times (M - M_\mathrm{c})$, where $M - M_\mathrm{c}$ is the dimension of the nullspace of $\mathbf{C}^H(f)$.

Therefore, the length of the vector $\mathbf{w}_\mathrm{GSC}(f)$ is $M - M_\mathrm{c}$.

The blocking matrix is not unique and the most obvious choice is the following:

$$\mathbf{B_C}(f) = \mathbf{P_C}(f) \begin{bmatrix} \mathbf{I}_{M-M_\mathrm{c}} \\ \mathbf{0}_{M_\mathrm{c} \times (M-M_\mathrm{c})} \end{bmatrix}, \quad (129)$$

where

$$\mathbf{P_C}(f) = \mathbf{I}_M - \mathbf{C}(f) \left[\mathbf{C}^H(f)\mathbf{C}(f) \right]^{-1} \mathbf{C}^H(f) \quad (130)$$

is a projection matrix whose rank is equal to $M - M_\mathrm{c}$ and $\mathbf{I}_{M-M_\mathrm{c}}$ is the $(M - M_\mathrm{c}) \times (M - M_\mathrm{c})$ identity matrix.

To obtain the filter $\mathbf{w}_{\mathrm{GSC}}(f)$, the GSC approach is used, which is formulated as the following unconstrained optimization problem:

$$\min_{\mathbf{w}(f)} \left[\mathbf{h}_{\mathrm{MN}}(f) - \mathbf{B}_{\mathbf{C}}(f)\mathbf{w}(f)\right]^H \mathbf{\Phi}_{\mathbf{y}}(f) \left[\mathbf{h}_{\mathrm{MN}}(f) - \mathbf{B}_{\mathbf{C}}(f)\mathbf{w}(f)\right]^H, \quad (131)$$

for which the solution is

$$\mathbf{w}_{\mathrm{GSC}}(f) = \left[\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\right]^{-1} \mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{h}_{\mathrm{MN}}(f). \quad (132)$$

Define the error signal, which is also the estimate of the desired signal, between the outputs of the two filters $\mathbf{h}_{\mathrm{MN}}(f)$ and $\mathbf{B}_{\mathbf{C}}(f)\mathbf{w}(f)$:

$$\widehat{X}_1(f) = \mathbf{h}_{\mathrm{MN}}^H(f)\mathbf{y}(f) - \mathbf{w}^H(f)\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{y}(f). \quad (133)$$

It is easy to see that the minimization of $E\left[\left|\widehat{X}_1(f)\right|^2\right]$ with respect to $\mathbf{w}(f)$ is equivalent to (131).

Now, we need to check if indeed the two filters LCMV and GSC are equivalent, i.e.,

$$\mathbf{i}_{\mathrm{c}}^T \left[\mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\right]^{-1}\mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f) = \mathbf{h}_{\mathrm{MN}}^H(f)\left\{\mathbf{I}_M - \mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\left[\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\right]^{-1}\mathbf{B}_{\mathbf{C}}^H(f)\right\}. \tag{134}$$

For that, we are going to follow the elegant proof given in [17].

The matrix in brackets in the second line of (134) can be rewritten as

$$\mathbf{I}_M - \mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\left[\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\right]^{-1}\mathbf{B}_{\mathbf{C}}^H(f) = \mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\left[\mathbf{I}_M - \mathbf{P}_1(f)\right]\mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f), \tag{135}$$

where

$$\mathbf{P}_1(f) = \mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\mathbf{B}_{\mathbf{C}}(f)\left[\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\right]^{-1}\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{1/2}(f) \quad (136)$$

is a projection operator onto the subspace spanned by the columns of $\mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\mathbf{B}_{\mathbf{C}}(f)$.

We have

$$\begin{aligned}\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{C}(f) &= \mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f)\mathbf{C}(f) \\ &= \mathbf{0}_{(M-M_c)\times M_c}. \quad (137)\end{aligned}$$

This implies that the rows of $\mathbf{B}_{\mathbf{C}}^H(f)$ are orthogonal to the columns of $\mathbf{C}(f)$ and the subspace spanned by the columns of $\mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\mathbf{B}_{\mathbf{C}}(f)$ is orthogonal to the subspace spanned by the columns of $\mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f)\mathbf{C}(f)$.

Since $\mathbf{B}_{\mathbf{C}}(f)$ has a rank equal to $M - M_{\mathrm{c}}$ where M_{c} is the rank of $\mathbf{C}(f)$, then the sum of the dimensions of the two subspaces is M and the subspaces are complementary.

This means that

$$\mathbf{P}_1(f) + \mathbf{P}_2(f) = \mathbf{I}_M, \tag{138}$$

where

$$\mathbf{P}_2(f) = \mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f)\mathbf{C}(f)\left[\mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\right]^{-1}\mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f). \tag{139}$$

When this is substituted and the constraint $\mathbf{i}_{\mathrm{c}}^T = \mathbf{h}_{\mathrm{MN}}^H(f)\mathbf{C}(f)$ is applied, (134) becomes

$$\begin{aligned}
&\mathbf{i}_{\mathrm{c}}^T \left[\mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{C}(f)\right]^{-1} \mathbf{C}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f) = \\
&\quad \mathbf{h}_{\mathrm{MN}}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\mathbf{P}_2(f)\mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f) = \\
&\quad \mathbf{h}_{\mathrm{MN}}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{1/2}(f)\left[\mathbf{I}_M - \mathbf{P}_1(f)\right]\mathbf{\Phi}_{\mathbf{y}}^{-1/2}(f) = \\
&\quad \mathbf{h}_{\mathrm{MN}}^H(f)\left\{\mathbf{I}_M - \mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\left[\mathbf{B}_{\mathbf{C}}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{B}_{\mathbf{C}}(f)\right]^{-1}\mathbf{B}_{\mathbf{C}}^H(f)\right\}.
\end{aligned} \tag{140}$$

Hence, the LCMV and GSC filters are strictly equivalent.

Implementation with the STFT

In this section, we show how to implement the optimal filters in the STFT domain.

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

$$\mathbf{y}_m(t) = \mathbf{x}_m(t) + \mathbf{v}_m(t),\ m = 1, 2, \ldots, M, \tag{141}$$

where

$$\mathbf{y}_m(t) = \begin{bmatrix} y_m(t) & y_m(t-1) & \cdots & y_m(t-L+1) \end{bmatrix}^T \tag{142}$$

is a vector of length L, and $\mathbf{x}_m(t)$ and $\mathbf{v}_m(t)$ are defined in a similar way to $\mathbf{y}_m(t)$ from (142).

A short-time segment of the observation [i.e., $\mathbf{y}_m(t)$], is multiplied with an analysis window of length L:

$$\mathbf{g}_{\mathrm{a}} = \begin{bmatrix} g_{\mathrm{a}}(0) & g_{\mathrm{a}}(1) & \cdots & g_{\mathrm{a}}(L-1) \end{bmatrix}^T \qquad (143)$$

and transformed into the frequency domain by using the discrete Fourier transform (DFT).

Let $\mathbf{W}$ denote the DFT matrix of size $L \times L$, with

$$[\mathbf{W}]_{i,j} = \exp\left(-\frac{\jmath 2\pi ij}{L}\right), \; i, j = 0, \ldots, L-1. \qquad (144)$$

Then, the STFT representation of the observation is defined as [18]

$$\mathbf{Y}_m(t) = \mathbf{W}\mathrm{diag}\left(\mathbf{g}_{\mathrm{a}}\right)\mathbf{y}_m(t), \qquad (145)$$

where

$$\mathbf{Y}_m(t) = \begin{bmatrix} Y_m(t,0) & Y_m(t,1) & \cdots & Y_m(t,L-1) \end{bmatrix}^T . \tag{146}$$

In practice, the STFT representation is decimated in time by a factor R ($1 \leq R \leq L$) [19]:

$$\begin{aligned} \mathbf{Y}_m(rR) &= \mathbf{Y}_m(t)\,|_{t=rR} \qquad (147) \\ &= \begin{bmatrix} Y_m(rR,0) & Y_m(rR,1) & \cdots & Y_m(rR,L-1) \end{bmatrix}^T , \ r \in \mathbb{Z}. \end{aligned}$$

Figure 10 shows the STFT representation of the measured signal at the mth sensor.

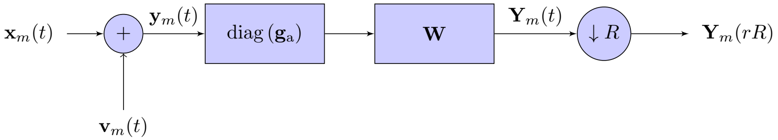


Figure 10: STFT representation of the measured signal at the mth sensor.

Therefore, in the STFT domain, (1) can be written as

$$Y_m(rR, k) = X_m(rR, k) + V_m(rR, k), \tag{148}$$

where $k = 0, \ldots, L-1$ denotes the frequency index, and $X_m(rR, k)$ and $V_m(rR, k)$ are the STFT representations of $x_m(t)$ and $v_m(t)$, respectively.

Assuming that L, the length of the analysis window $\mathbf{g}_\mathrm{a}$, is sufficiently larger than the effective support of the acoustic impulse response $g_m(t)$ [20], we can apply the multiplicative transfer function (MTF) approximation [20] and write the convolved desired signal at the mth sensor as

$$X_m(rR, k) = G_m(k)X(rR, k), \tag{149}$$

where $X(rR, k)$ is the STFT representation of the desired signal, $x(t)$, and $G_m(k)$ is the DFT of $g_m(t)$.

Writing the M STFT representations of the sensors' signals in a vector notation, we have

$$\begin{aligned} \mathbf{y}(rR,k) &= \mathbf{g}(k)X(rR,k) + \mathbf{v}(rR,k) \\ &= \mathbf{x}(rR,k) + \mathbf{v}(rR,k) \\ &= \mathbf{d}(k)X_1(rR,k) + \mathbf{v}(rR,k), \end{aligned} \tag{150}$$

where

$$\begin{aligned} \mathbf{y}(rR,k) &= \begin{bmatrix} Y_1(rR,k) & Y_2(rR,k) & \cdots & Y_M(rR,k) \end{bmatrix}^T, \\ \mathbf{x}(rR,k) &= \begin{bmatrix} X_1(rR,k) & X_2(rR,k) & \cdots & X_M(rR,k) \end{bmatrix}^T \\ &= X(rR,k)\mathbf{g}(k), \\ \mathbf{g}(k) &= \begin{bmatrix} G_1(k) & G_2(k) & \cdots & G_M(k) \end{bmatrix}^T, \\ \mathbf{v}(rR,k) &= \begin{bmatrix} V_1(rR,k) & V_2(rR,k) & \cdots & V_M(rR,k) \end{bmatrix}^T, \end{aligned}$$

and

$$\mathbf{d}(k) = \begin{bmatrix} 1 & \frac{G_2(k)}{G_1(k)} & \cdots & \frac{G_M(k)}{G_1(k)} \end{bmatrix}^T = \frac{\mathbf{g}(k)}{G_1(k)}. \tag{151}$$

The correlation matrix of $\mathbf{y}(rR, k)$ is

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{y}}(rR, k) &= E\left[\mathbf{y}(rR, k)\mathbf{y}^H(rR, k)\right] \\ &= \phi_{X_1}(rR, k)\mathbf{d}(k)\mathbf{d}^H(k) + \mathbf{\Phi}_{\mathbf{v}}(rR, k), \end{aligned} \tag{152}$$

where $\phi_{X_1}(rR, k) = E\left[|X_1(rR, k)|^2\right]$ is the variance of $X_1(rR, k)$ and $\mathbf{\Phi}_{\mathbf{v}}(rR, k) = E\left[\mathbf{v}(rR, k)\mathbf{v}^H(rR, k)\right]$ is the correlation matrix of $\mathbf{v}(f)$.

In the STFT domain, conventional multichannel noise reduction is performed by applying a complex weight to the output of each sensor, at time-frequency bin (rR, k), and summing across the aperture (see Fig. 11):

$$\begin{aligned} Z(rR, k) &= \sum_{m=1}^{M} H_m^*(rR, k) Y_m(rR, k) \\ &= \mathbf{h}^H(rR, k)\mathbf{y}(rR, k), \end{aligned} \tag{153}$$

where $Z(rR, k)$ is the estimate of $X_1(rR, k)$ and

$$\mathbf{h}(rR, k) = \begin{bmatrix} H_1(rR, k) & H_2(rR, k) & \cdots & H_M(rR, k) \end{bmatrix}^T \tag{154}$$

is a filter of length M containing all the complex gains applied to the sensors' outputs at time-frequency bin (rR, k).

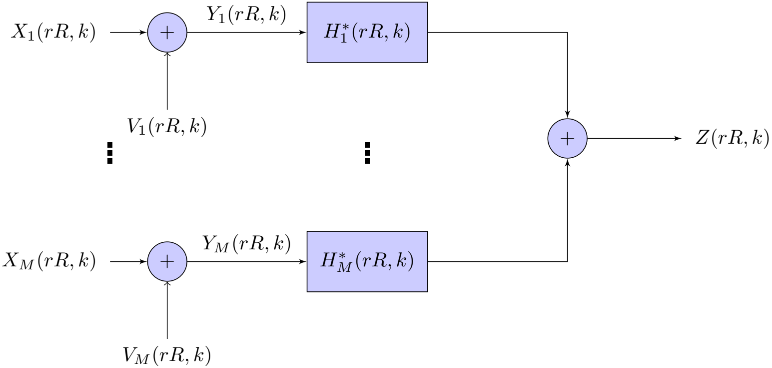


Figure 11: Block diagram of multichannel linear filtering in the STFT domain.

We can express (153) as a function of the steering vector, i.e.,

$$\begin{aligned} Z(rR,k) &= \mathbf{h}^H(rR,k)\left[\mathbf{d}(k)X_1(rR,k) + \mathbf{v}(rR,k)\right] \\ &= X_{\mathrm{fd}}(rR,k) + V_{\mathrm{rn}}(rR,k), \end{aligned} \tag{155}$$

where

$$X_{\mathrm{fd}}(rR,k) = X_1(rR,k)\mathbf{h}^H(rR,k)\mathbf{d}(k) \tag{156}$$

is the filtered desired signal and

$$V_{\mathrm{rn}}(rR,k) = \mathbf{h}^H(rR,k)\mathbf{v}(rR,k) \tag{157}$$

is the residual noise.

This procedure is called multichannel signal enhancement in the STFT domain.

The two terms on the right-hand side of (155) are incoherent.

Hence, the variance of $Z(rR,k)$ is the sum of two variances:

$$\begin{aligned} \phi_Z(rR,k) &= \mathbf{h}^H(rR,k)\mathbf{\Phi}_{\mathbf{y}}(rR,k)\mathbf{h}(rR,k) \\ &= \phi_{X_{\mathrm{fd}}}(rR,k) + \phi_{V_{\mathrm{rn}}}(rR,k), \end{aligned} \tag{158}$$

where

$$\phi_{X_{\mathrm{fd}}}(rR,k) = \phi_{X_1}(rR,k)\left|\mathbf{h}^H(rR,k)\mathbf{d}(k)\right|^2, \tag{159}$$

$$\phi_{V_{\mathrm{rn}}}(rR,k) = \mathbf{h}^H(rR,k)\mathbf{\Phi}_{\mathbf{v}}(rR,k)\mathbf{h}(rR,k). \tag{160}$$

In a similar way to the frequency-domain input SNR, we define the narrowband input SNR as

$$\mathrm{iSNR}(rR,k) = \frac{\phi_{X_1}(rR,k)}{\phi_{V_1}(rR,k)}. \tag{161}$$

The broadband input SNR is obtained by summing over all time-frequency indices the numerator and denominator of $\mathrm{iSNR}(rR, k)$.

We get

$$\mathrm{iSNR} = \frac{\sum_{r,k} \phi_{X_1}(rR, k)}{\sum_{r,k} \phi_{V_1}(rR, k)}. \tag{162}$$

Similarly, the broadband output SNR is

$$\begin{aligned}\mathrm{oSNR}(\mathbf{h}) &= \frac{\sum_{r,k} \phi_{X_{\mathrm{fd}}}(rR, k)}{\sum_{r,k} \phi_{V_{\mathrm{rn}}}(rR, k)} \\ &= \frac{\sum_{r,k} \phi_{X_1}(rR, k) \left|\mathbf{h}^H(rR, k)\mathbf{d}(k)\right|^2}{\sum_{r,k} \mathbf{h}^H(rR, k)\mathbf{\Phi}_{\mathbf{v}}(rR, k)\mathbf{h}(rR, k)},\end{aligned} \tag{163}$$

the broadband noise reduction and desired-signal reduction factors are, respectively,

$$\xi_{\mathrm{n}}(\mathbf{h}) = \frac{\sum_{r,k} \phi_{V_1}(rR,k)}{\sum_{r,k} \mathbf{h}^H(rR,k)\mathbf{\Phi}_{\mathbf{v}}(rR,k)\mathbf{h}(rR,k)} \tag{164}$$

and

$$\xi_{\mathrm{d}}(\mathbf{h}) = \frac{\sum_{r,k} \phi_{X_1}(rR,k)}{\sum_{r,k} \phi_{X_1}(rR,k) \left|\mathbf{h}^H(rR,k)\mathbf{d}(k)\right|^2}, \tag{165}$$

the broadband desired-signal distortion index is

$$\upsilon_{\mathrm{d}}(\mathbf{h}) = \frac{\sum_{r,k} \phi_{X_1}(rR,k) \left|\mathbf{h}^H(rR,k)\mathbf{d}(k) - 1\right|^2}{\sum_{r,k} \phi_{X_1}(rR,k)}, \tag{166}$$

and the broadband MSE is defined as

$$
\begin{aligned}
J(\mathbf{h}) &= \sum_{r,k} J[\mathbf{h}(rR,k)] \\
&= \sum_{r,k} \Big[\phi_{X_1}(rR,k)\left|\mathbf{h}^H(rR,k)\mathbf{d}(k)-1\right|^2 + \\
&\quad \mathbf{h}^H(rR,k)\mathbf{\Phi}_{\mathbf{v}}(rR,k)\mathbf{h}(rR,k)\Big] .
\end{aligned}
\tag{167}
$$

The optimal filters, summarized in Table 1, are employed in the STFT domain by replacing $\mathbf{\Phi}_{\mathbf{y}}(f)$, $\mathbf{\Phi}_{\mathbf{v}}(f)$, and $\mathbf{d}(f)$ with $\mathbf{\Phi}_{\mathbf{y}}(rR,k)$, $\mathbf{\Phi}_{\mathbf{v}}(rR,k)$, and $\mathbf{d}(k)$, respectively.

Example 6

Consider a ULA of M sensors. Suppose that a desired speech signal, $x(t)$, impinges on the ULA from the direction θ_x, and that an interference $u(t)$ impinges on the ULA from the direction θ_u.

Assume that the interference $u(t)$ is white Gaussian noise, i.e., $u(t) \sim \mathcal{N}\left(0, \sigma_u^2\right)$, uncorrelated with $x(t)$.

In addition, the sensors contain thermal white Gaussian noise, $w_m(t) \sim \mathcal{N}\left(0, \sigma_w^2\right)$, that are mutually uncorrelated.

The desired speech signal needs to be recovered from the noisy received signals, $y_m(t) = x_m(t) + v_m(t),\ m = 1, \ldots, M$, where $v_m(t) = u_m(t) + w_m(t),\ m = 1, \ldots, M$ are the interference-plus-noise signals.

Assume that the sampling frequency is 16 kHz, and that the sampling interval T_{s} satisfies $T_{\mathrm{s}} = \frac{d}{c}$. We have

$$x_m(t) = x_1\left(t - \tau_{x,m}\right),$$
$$u_m(t) = u_1\left(t - \tau_{u,m}\right),$$

where

$$\tau_{x,m} = \frac{(m-1)d\cos\theta_x}{cT_{\mathrm{s}}} = (m-1)\cos\theta_x,$$
$$\tau_{u,m} = \frac{(m-1)d\cos\theta_u}{cT_{\mathrm{s}}} = (m-1)\cos\theta_u.$$

In the STFT domain, we obtain

$$\mathbf{x}(rR,k) = X_1(rR,k)\mathbf{d}(k),$$
$$\mathbf{u}(rR,k) = U_1(rR,k)\boldsymbol{\gamma}^*_{U_1\mathbf{u}}(k),$$

where

$$\mathbf{d}(k) = \begin{bmatrix} 1 & e^{-\jmath 2\pi k \tau_{x,2}/L} & e^{-\jmath 2\pi f \tau_{x,3}/L} & \cdots & e^{-\jmath 2\pi f \tau_{x,M}/L} \end{bmatrix}^T,$$

$$\boldsymbol{\gamma}^*_{U_1\mathbf{u}}(k) = \begin{bmatrix} 1 & e^{-\jmath 2\pi k \tau_{u,2}/L} & e^{-\jmath 2\pi f \tau_{u,3}/L} & \cdots & e^{-\jmath 2\pi f \tau_{u,M}/L} \end{bmatrix}^T.$$

To demonstrate noise reduction, we choose $\theta_x = 70^\circ$, $\theta_u = 20^\circ$, $\sigma_w^2 = 0.1\sigma_u^2$, a Hamming window of length $L = 512$ as the analysis window, a decimation factor $R = L/4 = 128$, and Wiener filter:

$$\mathbf{h}_{\mathrm{W}}(rR,k) = \phi_{X_1}(rR,k)\left[\phi_{X_1}(rR,k)\mathbf{d}(k)\mathbf{d}^H(k) + \boldsymbol{\Phi}_{\mathbf{v}}(rR,k)\right]^{-1}\mathbf{d}(k). \tag{168}$$

An estimate for the correlation matrix of $\mathbf{v}(rR, k)$ can be obtained by averaging past cross-spectral power values of the noisy measurement during speech inactivity:

$$\widehat{\mathbf{\Phi}}_{\mathbf{v}}(rR, k) = \begin{cases} \alpha\widehat{\mathbf{\Phi}}_{\mathbf{v}}\left[(r-1)R, k\right] + (1-\alpha)\mathbf{y}(rR, k)\mathbf{y}^H(rR, k), & X(rR, k) = 0 \\ \widehat{\mathbf{\Phi}}_{\mathbf{v}}\left[(r-1)R, k\right], & X(rR, k) \neq 0 \end{cases}, \quad (169)$$

where α ($0 < \alpha < 1$) denotes a smoothing parameter.

This method requires a voice activity detector (VAD), but there are also alternative and more efficient methods that are based on minimum statistics [22], [23].

Finding an estimate for $\phi_{X_1}(rR, k)$ is a much more challenging problem [24], [25].

In this example, for simplicity, we smooth $|Y_1(rR,k)|^2$ in both time and frequency axes and subtract an estimate of the noise, i.e.,

$$\widehat{\phi}_{X_1}(rR,k) = \max\left\{\widehat{\phi}_{Y_1}(rR,k) - \widehat{\phi}_{V_1}(rR,k), 0\right\},$$

where $\widehat{\phi}_{Y_1}(rR,k)$ is obtained as a two-dimensional convolution between $|Y_1(rR,k)|^2$ and a smoothing window $w(rR,k)$.

Here, the smoothing window is a two-dimensional Hamming window of size 3×11, normalized to $\sum_{r,k} w(rR,k) = 1$.

Figure 12 shows the spectrogram and waveform of the clean speech signal received at the first sensor, $x_1(t)$.

Figure 13 shows a realization of the observation signal at the first sensor, $y_1(t)$, and the estimated signals, $z(t)$, for different numbers of sensors, M.

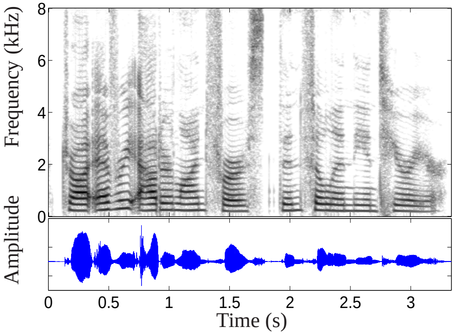


Figure 12: Speech spectrogram and waveform of a clean speech signal received at the first sensor, $x_1(t)$: “Draw every outer line first, then fill in the interior.”

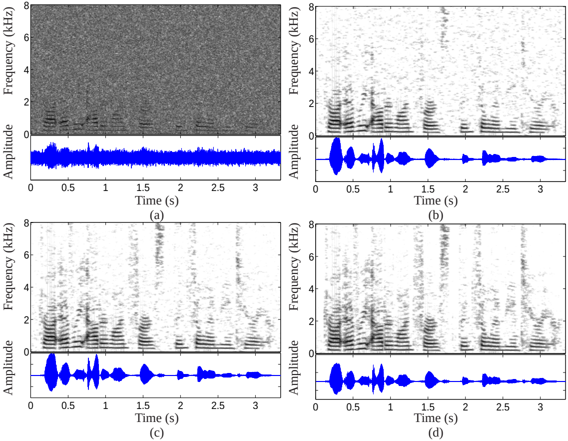


Figure 13: Speech spectrograms and waveforms of (a) noisy speech signal received at the first sensor, $y_1(t)$ ($\mathrm{iSNR} = -5$ dB), and the estimated signal, $z(t)$, for (b) $M = 1$ [$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) = 6.64$ dB], (c) $M = 2$ [$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) = 8.72$ dB], and (d) $M = 5$ [$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) = 13.34$ dB].

Figure 14 shows plots of the broadband gain in SNR, $\mathcal{G}\left(\mathbf{h}_{\mathrm{W}}\right)$, the broadband MSE, $J\left(\mathbf{h}_{\mathrm{W}}\right)$, the broadband noise reduction factor, $\xi_{\mathrm{n}}\left(\mathbf{h}_{\mathrm{W}}\right)$, and the broadband desired-signal reduction factor, $\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right)$, as a function of the broadband input SNR, for different numbers of sensors, M.

Clearly, as the number of sensors increases, the Wiener filter better enhances the desired speech signal in terms of higher SNR and noise reduction, and lower MSE and desired-signal reduction.

Note that more useful algorithms for enhancing noisy speech signals in the STFT domain are presented in [2], [26], [27].

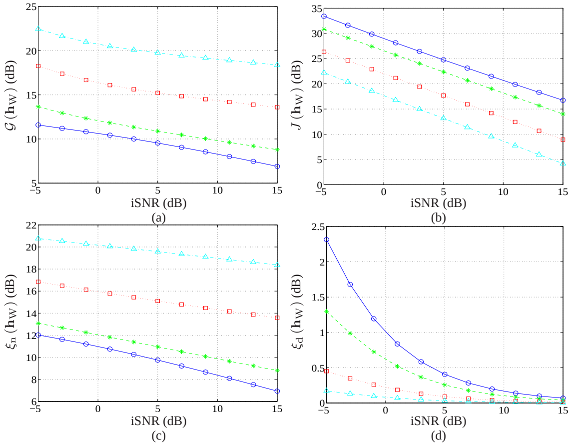


Figure 14: The broadband (a) gain in SNR, (b) MSE, (c) noise reduction factor, and (d) desired-signal reduction factor of the Wiener filter for different numbers of sensors, M: $M = 1$ (solid line with circles), $M = 2$ (dashed line with asterisks), $M = 5$ (dotted line with squares), and $M = 10$ (dash-dot line with triangles).

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

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

[3] J. Benesty, J. Chen, and E. Habets, *Speech Enhancement in the STFT Domain*. Springer Briefs in Electrical and Computer Engineering, 2011.

[4] J. P. Dmochowski and J. Benesty, “Microphone arrays: fundamental concepts,” in *Speech Processing in Modern Communication–Challenges and Perspectives*, I. Cohen, J. Benesty, and S. Gannot, Eds., Berlin, Germany: Springer-Verlag, 2008, Chapter 8, pp. 199–223, 2010.

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

[6] J. Capon, “High resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[7] R. T. Lacoss, “Data adaptive spectral analysis methods,” *Geophysics*, vol. 36, pp. 661–675, Aug. 1971.

[8] M. Souden, J. Benesty, and S. Affes, “On the global output SNR of the parameterized frequency-domain multichannel noise reduction Wiener filter,” *IEEE Signal Process. Lett.*, vol. 17, pp. 425–428, May 2010.

[9] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, “On microhone-array beamforming from a MIMO acoustic signal processing perspective,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, pp. 1053–1065, Mar. 2007.

[10] A. Booker and C. Y. Ong, “Multiple constraint adaptive filtering,” *Geophysics*, vol. 36, pp. 498–509, June 1971.

[11] O. Frost, “An algorithm for linearly constrained adaptive array processing,” *Proc. IEEE*, vol. 60, pp. 926–935, Jan. 1972.

[12] M. Er and A. Cantoni, “Derivative constraints for broad-band element space antenna array processors,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 1378–1393, Dec. 1983.

[13] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, Jan. 1982.

[14] K. M. Buckley, “Broad-band beamforming and the generalized sidelobe canceller,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 1322–1323, Oct. 1986.

[15] K. M. Buckley and L. J. Griffiths, “An adaptive generalized sidelobe canceller with derivative constraints,” *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 311–319, Mar. 1986.

[16] S. Werner, J. A. Apolinário, Jr., and M. L. R. de Campos, “On the equivalence of RLS implementations of LCMV and GSC processors,” *IEEE Signal Process. Lett.*, vol. 10, pp. 356–359, Dec. 2003.

[17] B. R. Breed and J. Strauss, “A short proof of the equivalence of LCMV and GSC beamforming,” *IEEE Signal Process. Lett.*, vol. 9, pp. 168–169, June 2002.

[18] J. Wexler and S. Raz, “Discrete Gabor expansions,” *Speech Process.*, vol. 21, pp. 207–220, Nov. 1990.

[19] S. Qian and D. Chen, “Discrete Gabor transform,” *IEEE Trans. Signal Process.*, vol. 41, pp. 2429–2438, July 1993.

[20] Y. Avargel and I. Cohen, “On multiplicative transfer function approximation in the short-time Fourier transform domain,” *IEEE Signal Process. Lett.*, vol. 14, pp. 337–340, May 2007.

[21] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1983.

[22] R. Martin, “Noise power spectral density estimation based on optimal smoothing and minimum statistics,” *IEEE Trans. Speech, Audio Process.*, vol. 9, pp. 504–512, July 2001.

[23] I. Cohen, “Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging,” *IEEE Trans. Speech, Audio Process.*, vol. 11, pp. 466–475, Sept. 2003.

[24] I. Cohen, “Relaxed statistical model for speech enhancement and a priori SNR estimation,” *IEEE Trans. Speech, Audio Process.*, vol. 13, pp. 870–881, Sept. 2005.

[25] I. Cohen, “Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models,” *Signal Process.*, vol. 86, pp. 698–709, Apr. 2006.

[26] I. Cohen and B. Berdugo, “Speech enhancement for non-stationary noise environments,” *Signal Process.*, vol. 81, pp. 2403–2418, Nov. 2001.

[27] I. Cohen and S. Gannot, "Spectral enhancement methods," in J. Benesty, M. M. Sondhi, and Y. Huang (Eds.), *Springer Handbook of Speech Processing*, Springer-Verlag, 2008, Part H, Chapter 44, pp. 873–901.