Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 49 (2007) 305–316

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

On the optimal linear filtering techniques for noise reduction

Jingdong Chen [a,*], Jacob Benesty [b], Yiteng (Arden) Huang [a]

[a] Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Room 2D-534, Murray Hill, NJ 07974, USA

[b] Université du Québec, INRS-EMT, 800 de la Gauchetière Ouest, Suite 6900, Montréal, Que., Canada H5A 1K6

Received 23 May 2006; received in revised form 4 December 2006; accepted 4 February 2007

Abstract

Noise reduction, which aims at extracting the clean speech from noisy observations, has plenty of applications. It has attracted a considerable amount of research attention over the past several decades. Although many methods have been developed, the most widely used one, by far, is the optimal linear filtering technique, which achieves clean speech estimate by passing the noisy observation through an optimal linear filter/transformation. The representative algorithms of this include Wiener filtering, spectral restoration, subspace method, etc. Many experiments have been carried out, from various points of view, to show that the optimal filtering technique can reduce the level of noise that is present in the speech signal and improve the corresponding signal-to-noise ratio (SNR). However, there is not much theoretical justification so far for the noise reduction and SNR improvement. This paper attempts to provide a theoretical analysis on the performance (including noise reduction, speech distortion, and SNR improvement) of the optimal filtering noise-reduction techniques including the time-domain causal Wiener filter, the subspace method, and the frequency-domain subband Wiener filter. We show that the optimal linear filter, regardless of how we delineate it, can indeed reduce the level of noise (but at a price of attenuating the desired speech signal). Most importantly, we prove that the *a posteriori* SNR (defined after the optimal filtering) is always greater than, or at least equal to the *a priori* SNR, which reveals that the optimal linear filtering technique is indeed able to make noisy speech signals cleaner. We will also discuss the bounds for noise reduction, speech distortion, and SNR improvement.

1. Introduction

Since we live in a natural environment where noise is inevitable and ubiquitous, speech signals can seldom be recorded in pure form and are generally contaminated by acoustic background noise. As a result, the microphone signals have to be "cleaned up" with digital signal processing tools before they are stored, transmitted, or played out.

The cleaning process, which is often referred to as either noise reduction or speech enhancement, can be achieved in many different ways, such as beamforming, adaptive cancellation, temporal filtering, spatial–temporal filtering, etc. The most widely used technique thus far, however, is the single-channel optimal linear filtering approach, which achieves clean speech estimate by passing the noisy observation through an optimal linear filter/transformation. A variety of such algorithms have been developed. They principally fall into one of the following four categories: *Wiener filter*, *spectral restoration*, *subspace method*, and *parametric method*.

Wiener filter: This method restores the desired speech signal by passing the noisy speech through a finite impulse response (FIR) filter whose coefficients are estimated by minimizing the mean square error (MSE) between the clean speech and its estimate (Widrow and Stearns, 1985). The Wiener filter can also be delineated in the frequency domain, resulting in various derivative techniques such as spectral subtraction (Boll, 1979; McAulay and Malpass, 1980; Lim, 1983; Lim and Oppenheim, 1979), parametric Wiener filter (Lim and Oppenheim, 1979; Vary, 1985; Etter and Moschytz, 1994; Chen et al., 2003; Diethorn, 2004), etc.

* Corresponding author. Tel.: +1 908 582 5044; fax: +1 908 582 7308.
E-mail addresses: jingdong@research.bell-labs.com (J. Chen), benesty@emt.inrs.ca (J. Benesty), arden@research.bell-labs.com (Yiteng (Arden) Huang).

0167-6393/$ - see front matter
doi:10.1016/j.specom.2007.02.002

Spectral restoration: In the frequency domain, a speech signal can be factorized into spectral amplitude and phase components. From perceptual point of view, the former is considerably more important than the latter (Lim and Oppenheim, 1979; Vary, 1985; Wang and Lim, 1982). Therefore the spectral-restoration technique recovers only the spectral amplitude (or spectral envelope) of the clean speech from that of the corrupted speech while neglecting the phase corruption (Ephraim and Malah, 1984, 1985; Virag, 1999; Chang and O'Shaughnessy, 1991).

Signal subspace: This method decomposes the vector space of the noisy speech into two orthogonal subspaces using the Karhunen–Loève transform (KLT): one is composed of both speech and noise and the other consists of noise component only. This is possible because it has been proven that the clean speech can be described with a low-rank model. After decomposition, the speech signal is estimated by removing the noise subspace, and cleaning the speech-plus-noise subspace (Ephraim and Van Trees, 1995; Dendrinos et al., 1991; Hansen, 1997; Lev-Ari and Ephraim, 2003; Rezayee and Gazor, 2001; Mittal and Phamdo, 2000; Hu and Loizou, 2003).

Parametric method: It is well known that a speech signal can be modelled as an autoregressive (AR) process. Therefore, noise reduction can be formulated as a parameter estimation problem with its objective to estimate the AR model parameters of the clean speech from the noisy observations (Paliwal and Basu, 1987; Gibson et al., 1991; Gannot et al., 1998).

Although so many optimal filtering algorithms have been developed for noise reduction, there has been remarkably little (if any) theoretical analysis of their performance. The reason may be attributed to the difficulty in quantizing the combinatorial effect between noise reduction and speech distortion. Most existing performance studies have been experimental, including: (1) ranking the mean opinion scores, (2) examining the SNR improvements, (3) inspecting the speech spectrograms, and (4) comparing the noise levels before and after the application of an algorithm. While the results are very helpful for us to understand how the algorithms behave in the specified conditions, the experimental evaluation alone is not enough to justify the algorithms. A more thorough theoretical analysis is important and imperative. Recently, we performed some analysis of the time-domain Wiener filter and proved that, as long as we have an accurate estimate of the statistics of the noisy speech and the noise signal, SNR improvement is guaranteed, no matter whether the noise is white or colored (Chen et al., 2006; Benesty et al., 2005). This paper presents our continued efforts on this topic. The main contribution of this paper is a theoretical analysis on the performance of the optimal [from the minimum-mean-square error (MMSE) sense] filtering techniques including the time-domain causal, the frequency-domain noncausal, and the constrained (subspace) Wiener filters. We show that the optimal filter, regardless of how we delineate it, can indeed reduce the level of noise. Most importantly, we prove that the *a posteriori* SNR is always greater than, or at least equal to the *a priori* SNR, provided that the statistics of the noisy speech and noise signals are accurately estimated. Also discussed are the lower and upper bounds for noise reduction, speech distortion, and SNR improvement.

2. Signal model and problem formulation

The noise-reduction problem considered in this paper is to recover a speech signal of interest $x(n)$ from the noisy observation

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

where $v(n)$ is the unwanted additive noise, which is assumed to be a zero-mean random process (white or colored) and uncorrelated with $x(n)$. This signal model can also be formulated in other forms. For example, in vector/matrix form, it is written as

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

where

$$\mathbf{y}(n) = [\, y(n) \quad y(n-1) \quad \cdots \quad y(n-L+1)\,]^{\mathrm{T}}$$

is a vector consisting of the L most recent samples of the noisy speech signal, superscript T denotes transpose of a vector or a matrix, and $\mathbf{x}(n)$ and $\mathbf{v}(n)$ are defined in a similar way to $\mathbf{y}(n)$. In this case, the noise-reduction problem is formulated as one of estimating $\mathbf{x}(n)$ from the observation $\mathbf{y}(n)$.

If applying the L-point discrete Fourier transform (DFT) to both sides of (2), we have the following relationship in the frequency domain:

$$Y(n, j\omega_k) = X(n, j\omega_k) + V(n, j\omega_k), \tag{3}$$

where

$$Y(n, j\omega_k) = \sum_{l=0}^{L-1} w(l) y(n - L + l + 1) \mathrm{e}^{-j\omega_k l}$$

is the short-time DFT of the noisy speech at time instant n, $\omega_k = 2\pi k/L$, $k = 0, 1, \ldots, L-1$, $w(l)$ is a window function (e.g. Hamming window, Hann window) applied to the frame signal for better spectral estimation, and $X(n,j\omega_k)$ and $V(n,j\omega_k)$ are the short-time DFTs for the clean speech and the noise signal, defined in a similar way to $Y(n,j\omega_k)$. Based on this relationship, the noise-reduction problem can be expressed in the frequency domain as one of estimating $X(n,j\omega_k)$ from $Y(n,j\omega_k)$.

3. Time-domain causal Wiener filter and its performance

The Wiener filter is one of the most fundamental approaches for noise reduction, which can be formulated either in the time or in the frequency domains. In the time-domain Wiener filter, an estimate of the clean speech is obtained by passing the noisy signal $y(n)$ through a temporal filer, i.e.,

$$\hat{x}(n) = \mathbf{h}^{\mathrm{T}}\mathbf{y}(n), \tag{4}$$

where

$$\mathbf{h} = [\, h_0 \quad h_1 \quad \cdots \quad h_{L-1} \,]^{\mathrm{T}}$$

is an FIR filter of length L. The MSE criterion is then written as

$$J_x(\mathbf{h}) \triangleq E\{[x(n) - \hat{x}(n)]^2\} = E\{[x(n) - \mathbf{h}^{\mathrm{T}}\mathbf{y}(n)]^2\}, \tag{5}$$

where $E\{\cdot\}$ denotes mathematical expectation. It is immediately seen that the objective of noise reduction is to find the optimal $\mathbf{h}$ that minimizes $J_x(\mathbf{h})$. Mathematically, the optimal filter can be described as

$$\mathbf{h}_{\mathrm{o}} = \arg\min_{\mathbf{h}} J_x(\mathbf{h}). \tag{6}$$

Differentiating $J_x(\mathbf{h})$ with respect to $\mathbf{h}$ and equating the result to zero, we can find the solution to (6), which is known as the Wiener–Hopf equations

$$\mathbf{R}_y\mathbf{h}_{\mathrm{o}} = \mathbf{r}_{yx}, \tag{7}$$

where

$$\mathbf{R}_y \triangleq E\{\mathbf{y}(n)\mathbf{y}^{\mathrm{T}}(n)\} \tag{8}$$

is the correlation matrix of the observed signal $y(n)$ and

$$\mathbf{r}_{yx} \triangleq E\{\mathbf{y}(n)x(n)\} \tag{9}$$

is the cross-correlation vector between the noisy and clean speech signals. It is seen from (7) that we need to know both $\mathbf{R}_y$ and $\mathbf{r}_{yx}$ in order to compute the Wiener filter $\mathbf{h}_{\mathrm{o}}$. The correlation matrix $\mathbf{R}_y$ can be directly estimated from the observation signal $y(n)$. However, $x(n)$ is unobservable; as a result, an estimation of $\mathbf{r}_{yx}$ may seem difficult to obtain. But since speech and noise are uncorrelated, we have

$$\begin{aligned}\mathbf{r}_{yx} &= E\{\mathbf{y}(n)x(n)\} = E\{\mathbf{y}(n)[y(n) - v(n)]\} \\ &= E[\mathbf{y}(n)y(n)] - E[\mathbf{v}(n)v(n)] = \mathbf{r}_{yy} - \mathbf{r}_{vv}.\end{aligned} \tag{10}$$

Now $\mathbf{r}_{yx}$ depends on two correlation vectors: $\mathbf{r}_{yy}$ and $\mathbf{r}_{vv}$. The vector $\mathbf{r}_{yy}$, which is also the first column of $\mathbf{R}_y$, can be estimated from $y(n)$. The vector $\mathbf{r}_{vv}$ can be estimated during intervals where the speech signal is absent. As a result, the Wiener–Hopf equations given in (7) can be rewritten as

$$\mathbf{R}_y\mathbf{h}_{\mathrm{o}} = \mathbf{r}_{yy} - \mathbf{r}_{vv}. \tag{11}$$

If we assume that the matrix $\mathbf{R}_y$ is full rank, which is the case in most practical situations, the Wiener filter is obtained by solving either (7) or (11), i.e.,

$$\mathbf{h}_{\mathrm{o}} = \mathbf{R}_y^{-1}\mathbf{r}_{yx} = \mathbf{R}_y^{-1}\mathbf{r}_{yy} - \mathbf{R}_y^{-1}\mathbf{r}_{vv}. \tag{12}$$

Now we are ready to check whether the Wiener filter can actually reduce the level of noise as we expected. Substituting $\mathbf{h}_{\mathrm{o}}$ into (4), we can derive the power of the estimated clean speech signal with the optimal Wiener filter,

$$\begin{aligned}E\{\hat{x}_{\mathrm{o}}^2(n)\} &= E\{\mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{y}(n)\mathbf{y}^{\mathrm{T}}(n)\mathbf{h}_{\mathrm{o}}\} \\ &= E\{\mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{h}_{\mathrm{o}} + \mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{h}_{\mathrm{o}}\} \\ &= \mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{R}_x\mathbf{h}_{\mathrm{o}} + \mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_{\mathrm{o}},\end{aligned} \tag{13}$$

where $\mathbf{R}_x$ and $\mathbf{R}_v$ are, respectively, the correlation matrices of the clean speech and the noise signal, defined in a similar way to $\mathbf{R}_y$. It is seen that there are two terms in the right-hand side of (13), where $\mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{R}_x\mathbf{h}_{\mathrm{o}}$ represents the power of the clean speech filtered by the Wiener filter and $\mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_{\mathrm{o}}$ is the residual noise.

In order to check whether the Wiener filter can reduce noise, we evaluate the noise-reduction factor (Chen et al., 2006), which is defined as the ratio between the power of the original noise and that of the residual noise. As seen from (13), the residual noise power is $\mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_{\mathrm{o}}$. The power of the original noise can be computed as

$$\sigma_v^2 = E[v^2(n)] = E[\mathbf{h}_1^{\mathrm{T}}\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{h}_1] = \mathbf{h}_1^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_1, \tag{14}$$

where $\mathbf{h}_1 = [\,1 \quad 0 \quad \cdots \quad 0\,]^{\mathrm{T}}$. Therefore, the noise-reduction factor for the Wiener filter can be written as

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{o}}) = \frac{\mathbf{h}_1^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_1}{\mathbf{h}_{\mathrm{o}}^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_{\mathrm{o}}}. \tag{15}$$

Substituting $\mathbf{h}_{\mathrm{o}} = \mathbf{R}_y^{-1}\mathbf{r}_{yx} = \mathbf{R}_y^{-1}\mathbf{r}_{xx} = \mathbf{R}_y^{-1}\mathbf{R}_x\mathbf{h}_1$ into (15), we obtain

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{o}}) = \frac{\mathbf{h}_1^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_1}{\mathbf{h}_1^{\mathrm{T}}\mathbf{R}_x\mathbf{R}_y^{-1}\mathbf{R}_v\mathbf{R}_y^{-1}\mathbf{R}_x\mathbf{h}_1}, \tag{16}$$

which is a function of all the three correlation matrices $\mathbf{R}_x$, $\mathbf{R}_v$, and $\mathbf{R}_y$. Using generalized eigenvalule decomposition (Fukunaga, 1990), we can decompose the three correlation matrices into the following form:

$$\begin{aligned}\mathbf{R}_x &= \mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}, \\ \mathbf{R}_v &= \mathbf{B}^{\mathrm{T}}\mathbf{B}, \\ \mathbf{R}_y &= \mathbf{B}^{\mathrm{T}}[\mathbf{I} + \mathbf{\Lambda}]\mathbf{B},\end{aligned} \tag{17}$$

where $\mathbf{B}$ is an invertible square matrix, and

$$\mathbf{\Lambda} = \mathrm{diag}[\,\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_L\,] \tag{18}$$

is a diagonal matrix with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_L \geqslant 0$. Substituting (17) into (16), we obtain

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{o}}) = \frac{\sum_{i=1}^{L} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(1+\lambda_i)^2} b_{i1}^2}, \tag{19}$$

where b_{i1}, $i = 1, \ldots, L$, forms the first column of $\mathbf{B}$ and satisfies $\sum_{i=1}^{L} b_{i1}^2 = \sigma_v^2$.

Also with the matrix decomposition in (17), the SNR of the observation signal can be expressed as

$$\mathrm{SNR} = \frac{E[x^2(n)]}{E[v^2(n)]} = \frac{\sigma_x^2}{\sigma_v^2} = \frac{\mathbf{h}_1^{\mathrm{T}}\mathbf{R}_x\mathbf{h}_1}{\mathbf{h}_1^{\mathrm{T}}\mathbf{R}_v\mathbf{h}_1} = \frac{\sum_{i=1}^{L} \lambda_i b_{i1}^2}{\sum_{i=1}^{L} b_{i1}^2}. \tag{20}$$

Using (20), we can rewrite (19) as

$$\xi_{\mathrm{nr}}(\boldsymbol{h}_{\mathrm{o}}) = \frac{1}{\mathrm{SNR}} \cdot \frac{\sum_{i=1}^{L} \lambda_i b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(1+\lambda_i)^2} b_{i1}^2} = \frac{1}{\mathrm{SNR}} \cdot \frac{\sum_{i=1}^{L} \frac{(1+\lambda_i)^2}{(1+\lambda_i)^2} \lambda_i b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(1+\lambda_i)^2} b_{i1}^2} = \frac{1}{\mathrm{SNR}} \cdot \left[\frac{\sum_{i=1}^{L} \frac{\lambda_i+\lambda_i^3}{(1+\lambda_i)^2} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(1+\lambda_i)^2} b_{i1}^2} + 2 \right]. \tag{21}$$

Using the fact that $\lambda_i + \lambda_i^3 \geqslant \lambda_i^3$, we easily deduce from (21) that

$$\xi_{\mathrm{nr}}(\boldsymbol{h}_{\mathrm{o}}) \geqslant \frac{1}{\mathrm{SNR}} \cdot \left[\frac{\sum_{i=1}^{L} \frac{\lambda_i^3}{(1+\lambda_i)^2} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(1+\lambda_i)^2} b_{i1}^2} + 2 \right]. \tag{22}$$

With some algebra, we can prove the following inequality (see Appendix by setting $\mu = 1$ and $q_i = b_{i1}$ or follow the Appendix in Chen et al. (2006))

$$\frac{\sum_{i=1}^{L} \frac{\lambda_i^3}{(1+\lambda_i)^2} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(1+\lambda_i)^2} b_{i1}^2} \geqslant \frac{\sum_{i=1}^{L} \lambda_i b_{i1}^2}{\sum_{i=1}^{L} b_{i1}^2} = \mathrm{SNR}, \tag{23}$$

where equality holds if and only if all the λ_i's corresponding to nonzero b_{i1} are equal, where $i = 1, 2, \ldots, L$. It follows immediately that:

$$\xi_{\mathrm{nr}}(\boldsymbol{h}_{\mathrm{o}}) \geqslant \frac{\mathrm{SNR} + 2}{\mathrm{SNR}}. \tag{24}$$

The right-hand side of (24) is always greater than 1 since SNR is nonnegative. This shows that noise reduction is always feasible with the Wiener filter. It can be checked from (24) that the lower bound of the noise-reduction factor is a monotonically decreasing function of the SNR. It approaches to infinity when SNR comes close to 0 and draws near to 1 as SNR approaches infinity. This indicates that more noise reduction can be achieved with the Wiener filter as the SNR decreases, which is, of course, desirable since as SNR drops, there will be more noise to be eliminated.

Now let us examine the speech distortion. Similarly, we borrow the concept of speech-distortion index from Chen et al. (2006), which is defined as the attenuation in speech power relative to the power of the original clean speech. The power of the clean speech can be computed through $E[x^2(n)] = E[\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{x}(n) \boldsymbol{x}^{\mathrm{T}}(n) \boldsymbol{h}] = \boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{R}_x \boldsymbol{h}_1$. The attenuation of speech power due to the Wiener filter can be computed as $(\boldsymbol{h}_1 - \boldsymbol{h}_{\mathrm{o}})^{\mathrm{T}} \boldsymbol{R}_x (\boldsymbol{h}_1 - \boldsymbol{h}_{\mathrm{o}})$. As a result, the speech-distortion index for the Wiener filter is written as

$$\varphi_{\mathrm{sd}}(\boldsymbol{h}_{\mathrm{o}}) = \frac{(\boldsymbol{h}_1 - \boldsymbol{h}_{\mathrm{o}})^{\mathrm{T}} \boldsymbol{R}_x (\boldsymbol{h}_1 - \boldsymbol{h}_{\mathrm{o}})}{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{R}_x \boldsymbol{h}_1}. \tag{25}$$

Since $\boldsymbol{R}_x$ is positive semi-definite, we obviously have

$$\varphi_{\mathrm{sd}}(\boldsymbol{h}_{\mathrm{o}}) \geqslant 0. \tag{26}$$

In most noisy conditions, $\boldsymbol{h}_1 - \boldsymbol{h}_{\mathrm{o}}$ will neither be equal to the zero vector, nor will it be in the null space of $\boldsymbol{R}_x$. This means we usually have $\varphi_{\mathrm{sd}}(\boldsymbol{h}_{\mathrm{o}}) > 0$. So, speech distortion is unavoidable with the Wiener filter.

The upper bound of the speech-distortion index can be derived using the eigenvalue decomposition given in (17). As a matter of fact, substituting (17) into (25), we obtain

$$\varphi_{\mathrm{sd}}(\boldsymbol{h}_{\mathrm{o}}) = \frac{\sum_{i=1}^{L} \frac{\lambda_i}{(1+\lambda_i)^2} b_{i1}^2}{\sum_{i=1}^{L} \lambda_i b_{i1}^2} \leqslant \frac{\sum_{i=1}^{L} \frac{\lambda_i}{1+2\lambda_i} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i+2\lambda_i^2}{1+2\lambda_i} b_{i1}^2} \leqslant \frac{1}{2 \cdot \mathrm{SNR} + 1}, \tag{27}$$

where we have used the following inequality:

$$\frac{\sum_{i=1}^{L} \frac{\lambda_i^2}{1+2\lambda_i} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i}{1+2\lambda_i} b_{i1}^2} \geqslant \frac{\sum_{i=1}^{L} \lambda_i b_{i1}^2}{\sum_{i=1}^{L} b_{i1}^2} = \mathrm{SNR}. \tag{28}$$

This inequality can be proved by the induction (following the same analysis steps shown in Appendix).

From the previous analysis, we see that while noise reduction is feasible with the Wiener filter, speech attenuation is also unavoidable. In general, the more the noise is reduced, the more the speech is attenuated. A key question is whether the Wiener filter can improve SNR. To answer this question, we give the following proposition.

Proposition 1. *With the Wiener filter given in (7) and (12), the a posteriori SNR (defined after the Wiener filter) is always greater than or at least equal to the a priori SNR.*

Proof. If the noise $v(n)$ is zero, we already see that the Wiener filter has no effect on the speech signal. Now we consider the case where noise is not zero. The *a priori* SNR can be written as

$$\mathrm{SNR} = \frac{E[x^2(n)]}{E[v^2(n)]} = \frac{\sigma_x^2}{\sigma_v^2} = \frac{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{R}_x \boldsymbol{h}_1}{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{R}_v \boldsymbol{h}_1}. \tag{29}$$

Substituting (17) into (29), we can rewrite SNR as

$$\mathrm{SNR} = \frac{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{B} \boldsymbol{h}_1}{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{h}_1} = \frac{\sum_{i=1}^{L} \lambda_i b_{i1}^2}{\sum_{i=1}^{L} b_{i1}^2}. \tag{30}$$

From (13), we can write the *a posteriori* SNR as

$$\mathrm{SNR}_{\mathrm{o}} = \frac{\boldsymbol{h}_{\mathrm{o}}^{\mathrm{T}} \boldsymbol{R}_x \boldsymbol{h}_{\mathrm{o}}}{\boldsymbol{h}_{\mathrm{o}}^{\mathrm{T}} \boldsymbol{R}_v \boldsymbol{h}_{\mathrm{o}}}. \tag{31}$$

Substituting $\boldsymbol{h}_{\mathrm{o}} = \boldsymbol{R}_y^{-1} \boldsymbol{R}_x \boldsymbol{h}_1$ into (31), we obtain

$$\mathrm{SNR}_{\mathrm{o}} = \frac{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{R}_x \boldsymbol{R}_y^{-1} \boldsymbol{R}_x \boldsymbol{R}_y^{-1} \boldsymbol{R}_x \boldsymbol{h}_1}{\boldsymbol{h}_1^{\mathrm{T}} \boldsymbol{R}_x \boldsymbol{R}_y^{-1} \boldsymbol{R}_v \boldsymbol{R}_y^{-1} \boldsymbol{R}_x \boldsymbol{h}_1}. \tag{32}$$

Applying the matrix decomposition given in (17), we deduce that

$$\mathrm{SNR}_{\mathrm{o}} = \frac{\sum_{i=1}^{L} \frac{\lambda_i^3}{(\lambda_i+1)^2} b_{i1}^2}{\sum_{i=1}^{L} \frac{\lambda_i^2}{(\lambda_i+1)^2} b_{i1}^2}. \tag{33}$$

It follows, then, that

$$\frac{\mathrm{SNR_o}}{\mathrm{SNR}} = \frac{\sum_{i=1}^{L} b_{i1}^2 \cdot \sum_{i=1}^{L} \frac{\lambda_i^3}{(\lambda_i+1)^2} b_{i1}^2}{\sum_{i=1}^{L} \lambda_i b_{i1}^2 \cdot \sum_{i=1}^{L} \frac{\lambda_i^2}{(\lambda_i+1)^2} b_{i1}^2}. \tag{34}$$

Following the proof given in Appendix, if setting $\mu = 1$ and $q_i = b_{i1}$, we have

$$\sum_{i=1}^{L} b_{i1}^2 \cdot \sum_{i=1}^{L} \frac{\lambda_i^3}{(\lambda_i+1)^2} b_{i1}^2 \geqslant \sum_{i=1}^{L} \lambda_i b_{i1}^2 \cdot \sum_{i=1}^{L} \frac{\lambda_i^2}{(\lambda_i+1)^2} b_{i1}^2, \tag{35}$$

where equality holds if and only if all the λ_i's corresponding to the nonzero b_{i1}'s (for $i = 1, 2, \ldots, L$) are equal. It follows immediately that:

$$\mathrm{SNR_o} \geqslant \mathrm{SNR}. \tag{36}$$

That completes the proof of the proposition. Therefore, we see that, with the application of the time-domain Wiener filter, we are able to increase SNR of the observed signal. In other words, the Wiener filter can make the noisy speech cleaner. □

4. Subspace method and its performance

Similar to the Wiener filter, the subspace approach is also a linear optimal estimator from the MMSE sense. The difference between the two is that the former is deduced from an unconstrained optimization problem while the latter is derived from a constrained optimization one.

With the signal model given in (2), an estimate of the clean speech can be obtained by applying a linear transformation to the noisy speech vector, i.e.,

$$\hat{\boldsymbol{x}}(n) = \boldsymbol{H}\boldsymbol{y}(n), \tag{37}$$

where $\boldsymbol{H}$ is a matrix of size $L \times L$. The error signal obtained by this estimation is written as

$$\begin{aligned} \boldsymbol{e}(n) &\triangleq \hat{\boldsymbol{x}}(n) - \boldsymbol{x}(n) = \boldsymbol{H}\boldsymbol{y}(n) - \boldsymbol{x}(n) \\ &= (\boldsymbol{H} - \boldsymbol{I})\boldsymbol{x}(n) + \boldsymbol{H}\boldsymbol{v}(n) = \boldsymbol{e}_x(n) + \boldsymbol{e}_v(n), \end{aligned} \tag{38}$$

where

$$\boldsymbol{e}_x(n) = (\boldsymbol{H} - \boldsymbol{I})\boldsymbol{x}(n) \tag{39}$$

and

$$\boldsymbol{e}_v(n) = \boldsymbol{H}\boldsymbol{v}(n) \tag{40}$$

represents, respectively, the speech distortion due to the linear transformation and the residual noise. It is immediately seen that there are three criteria to estimate the optimal $\boldsymbol{H}$

1. Minimizing the energy of $\boldsymbol{e}(n)$.
2. Minimizing the energy of $\boldsymbol{e}_v(n)$ but maintaining the speech distortion less than a certain level.
3. Minimizing the energy of $\boldsymbol{e}_x(n)$ while maintaining the residual noise energy less than a certain threshold.

The first case leads to the Wiener solution, which is similar to the time-domain Wiener filter discussed in Section 3. The second criterion is rarely used in practice because the resulting estimator produces nonstationary residual noise (due to the nonstationarity of speech signals), which is usually intolerable to human perception system. The third case leads to the so-called subspace method.

Mathematically, the optimal linear transformation in the subspace technique can be described as

$$\begin{aligned} &\boldsymbol{H}_\mathrm{o} = \arg\min_{\boldsymbol{H}} \mathrm{tr}\{E[\boldsymbol{e}_x(n)\boldsymbol{e}_x^\mathrm{T}(n)]\} \\ &\text{subject to } \mathrm{tr}\{E[\boldsymbol{e}_v(n)\boldsymbol{e}_v^\mathrm{T}(n)]\} \leqslant L\sigma^2. \end{aligned} \tag{41}$$

If we use a Lagrange multiplier to adjoin the constraint to the cost function, (41) can be rewritten as

$$\boldsymbol{H}_\mathrm{o} = \arg\min_{\boldsymbol{H}} \mathscr{L}(\boldsymbol{H}, \mu), \tag{42}$$

with

$$\begin{aligned} \mathscr{L}(\boldsymbol{H}, \mu) &= \mathrm{tr}\{E[\boldsymbol{e}_x(n)\boldsymbol{e}_x^\mathrm{T}(n)]\} + \mu(\mathrm{tr}\{E[\boldsymbol{e}_v(n)\boldsymbol{e}_v^\mathrm{T}(n)]\} - L\sigma^2) \\ &= \mathrm{tr}[(\boldsymbol{H} - \boldsymbol{I})\boldsymbol{R}_x(\boldsymbol{H} - \boldsymbol{I})^\mathrm{T}] + \mu[\mathrm{tr}(\boldsymbol{H}\boldsymbol{R}_v\boldsymbol{H}^\mathrm{T}) - L\sigma^2], \end{aligned} \tag{43}$$

where $\boldsymbol{R}_x$ and $\boldsymbol{R}_v$ are the correlation matrices of the clean speech and noise respectively, and $\mu > 0$ is the Lagrange multiplier. Using the fact that

$$\frac{\partial}{\partial \boldsymbol{H}} \mathrm{tr}(\boldsymbol{R}_x\boldsymbol{H}) = \frac{\partial}{\partial \boldsymbol{H}} \mathrm{tr}(\boldsymbol{H}\boldsymbol{R}_x) = \boldsymbol{R}_x^\mathrm{T} = \boldsymbol{R}_x, \tag{44}$$

$$\frac{\partial}{\partial \boldsymbol{H}} \mathrm{tr}(\boldsymbol{H}\boldsymbol{R}_x\boldsymbol{H}^\mathrm{T}) = 2\boldsymbol{H}\boldsymbol{R}_x, \tag{45}$$

$$\frac{\partial}{\partial \boldsymbol{H}} \mathrm{tr}(\boldsymbol{H}\boldsymbol{R}_v\boldsymbol{H}^\mathrm{T}) = 2\boldsymbol{H}\boldsymbol{R}_v, \tag{46}$$

we can readily deduce

$$\frac{\partial}{\partial \boldsymbol{H}} \mathscr{L}(\boldsymbol{H}, \mu) = 2\boldsymbol{H}\boldsymbol{R}_x - 2\boldsymbol{R}_x + 2\mu\boldsymbol{H}\boldsymbol{R}_v. \tag{47}$$

Equating the right-hand side of (47) to zero, we obtain the solution to (42)

$$\boldsymbol{H}_\mathrm{o} = \boldsymbol{R}_x[\boldsymbol{R}_x + \mu\boldsymbol{R}_v]^{-1}, \tag{48}$$

where μ satisfies the following equation:

$$\mathrm{tr}\{\boldsymbol{R}_x[\boldsymbol{R}_x + \mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_v[\boldsymbol{R}_x + \mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_x\} = L\sigma^2. \tag{49}$$

To implement the optimal transformation, (48) is often simplified by using either eigenvalue decomposition (Ephraim and Van Trees, 1995; Dendrinos et al., 1991; Hansen, 1997; Lev-Ari and Ephraim, 2003; Rezayee and Gazor, 2001; Mittal and Phamdo, 2000), or generalized eigenvalue decomposition (Hu and Loizou, 2003). In the latter case, substituting (17) into (48), we can rewrite the optimal transformation as

$$\boldsymbol{H}_\mathrm{o} = \boldsymbol{B}^\mathrm{T}\boldsymbol{\Lambda}[\boldsymbol{\Lambda} + \mu\boldsymbol{I}]\boldsymbol{B}^{-T}. \tag{50}$$

Therefore, the estimation of the clean speech can be decoupled into three steps: applying the transformation $\boldsymbol{B}^{-\mathrm{T}}$ to the noisy signal, modifying the noisy signal in the transformed domain by a gain function $\boldsymbol{\Lambda}[\boldsymbol{\Lambda}+\mu\boldsymbol{I}]$, and applying $\boldsymbol{B}^{\mathrm{T}}$ to transform the modified components back to the original domain.

Another interesting interpretation of (50) is to divide the vector space into two subspaces: the speech subspace corresponding to all the λ_i's for $\lambda_i > 0$ and the noise subspace associated with all λ_i's for $\lambda_i = 0$. Suppose that the dimension of the speech subspace is M. We then can rewrite (50) as

$$\boldsymbol{H}_{\mathrm{o}} = \boldsymbol{B}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{\Sigma}_{M\times M} & \boldsymbol{0}_{M\times K} \\ \boldsymbol{0}_{K\times M} & \boldsymbol{0}_{K\times K} \end{bmatrix} \boldsymbol{B}^{-\mathrm{T}}, \tag{51}$$

where $K = L - M$ is the order of the noise subspace and

$$\boldsymbol{\Sigma} = \mathrm{diag}\left[\frac{\lambda_1}{\lambda_1+\mu}, \frac{\lambda_2}{\lambda_2+\mu}, \ldots, \frac{\lambda_M}{\lambda_M+\mu}\right]$$

is a diagonal matrix. We now clearly see that the noise reduction with the subspace method is achieved by nulling the noise subspace and cleaning the speech-plus-noise subspace.

Now we are ready to check whether the optimal linear transformation given in (50) is able to reduce the level of noise. Let us again examine the noise-reduction factor. From the previous analysis, we see that the power of the original noise is $\mathrm{tr}\{\boldsymbol{R}_v\}/L$. The residual noise power can be written as $\mathrm{tr}\{E[\boldsymbol{e}_v(n)\boldsymbol{e}_v^{\mathrm{T}}(n)]\}/L$. Therefore, the noise-reduction factor for the subspace method can be written as

$$\xi_{\mathrm{nr}}(\boldsymbol{H}_{\mathrm{o}}) = \frac{\mathrm{tr}\{\boldsymbol{R}_v\}}{\mathrm{tr}\{E[\boldsymbol{e}_v(n)\boldsymbol{e}_v^{\mathrm{T}}(n)]\}}. \tag{52}$$

Substituting (50) into $\boldsymbol{e}_v(n) = \boldsymbol{H}_{\mathrm{o}}\boldsymbol{v}(n)$, we can obtain

$$\mathrm{tr}\{E[\boldsymbol{e}_v(n)\boldsymbol{e}_v^{\mathrm{T}}(n)]\} = \mathrm{tr}\{\boldsymbol{R}_x[\boldsymbol{R}_x+\mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_v[\boldsymbol{R}_x+\mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_x\}. \tag{53}$$

Following the matrix-decomposition procedure given in (17), we easily deduce that

$$\mathrm{tr}\{\boldsymbol{R}_v\} = \mathrm{tr}\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\} = \mathrm{tr}\{\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\} = \sum_{i=1}^{L}\sum_{j=1}^{L} b_{ij}^2, \tag{54}$$

and

$$\begin{aligned}\mathrm{tr}\{E[\boldsymbol{e}_v(n)\boldsymbol{e}_v^{\mathrm{T}}(n)]\} &= \mathrm{tr}\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Lambda}[\boldsymbol{\Lambda}+\mu\boldsymbol{I}]^{-2}\boldsymbol{\Lambda}\boldsymbol{B}\} \\ &= \mathrm{tr}\{\boldsymbol{\Lambda}[\boldsymbol{\Lambda}+\mu\boldsymbol{I}]^{-2}\boldsymbol{\Lambda}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\} \\ &= \sum_{i=1}^{L}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}\sum_{j=1}^{L} b_{ij}^2. \end{aligned} \tag{55}$$

Therefore, we have

$$\xi_{\mathrm{nr}}(\boldsymbol{H}_{\mathrm{o}}) = \frac{\sum_{i=1}^{L}\sum_{j=1}^{L} b_{ij}^2}{\sum_{i=1}^{L}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}\sum_{j=1}^{L} b_{ij}^2}. \tag{56}$$

Since $\mu > 0$ and $\lambda_i \geqslant 0$ for $i = 1, 2, \ldots, L$, it is obvious that

$$\xi_{\mathrm{nr}}(\boldsymbol{H}_{\mathrm{o}}) > 1. \tag{57}$$

This shows that noise reduction is always feasible with the subspace method.

Since the energy of the speech distortion $\boldsymbol{e}_x(n)$ is always greater than zero, the power of the clean speech is also attenuated when we reduce the noise with the subspace method. We then ask the same question about SNR as we did for the Wiener filter: can the subspace method improve SNR? To answer this question, we give the following proposition.

Proposition 2. *With the linear transformation given in (48), if the noise $v(n)$ is not zero and $\mu > 0$, then the a posteriori SNR with the subspace method is always greater than or at least equal to the a priori SNR.*

Proof. If the noise is zero, we can easily check that the optimal transformation matrix $\boldsymbol{H}_{\mathrm{o}}$ will be the identity matrix so it will not change the input speech. If noise is not zero, according to the signal model given in (2), the *a priori* SNR can be written as

$$\mathrm{SNR} = \frac{\mathrm{tr}\{\boldsymbol{R}_x\}}{\mathrm{tr}\{\boldsymbol{R}_v\}}. \tag{58}$$

After applying the linear transformation $\boldsymbol{H}_{\mathrm{o}}$, the *a posteriori* SNR can be expressed as

$$\mathrm{SNR}_{\mathrm{o}} = \frac{\mathrm{tr}\{E[\boldsymbol{H}_{\mathrm{o}}\boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{T}}(n)\boldsymbol{H}_{\mathrm{o}}^{\mathrm{T}}]\}}{\mathrm{tr}\{E[\boldsymbol{H}_{\mathrm{o}}\boldsymbol{v}(n)\boldsymbol{v}^{\mathrm{T}}(n)\boldsymbol{H}_{\mathrm{o}}^{\mathrm{T}}]\}} = \frac{\mathrm{tr}\{\boldsymbol{H}_{\mathrm{o}}\boldsymbol{R}_x\boldsymbol{H}_{\mathrm{o}}^{\mathrm{T}}\}}{\mathrm{tr}\{\boldsymbol{H}_{\mathrm{o}}\boldsymbol{R}_v\boldsymbol{H}_{\mathrm{o}}^{\mathrm{T}}\}}. \tag{59}$$

It follows then,

$$\begin{aligned}\frac{\mathrm{SNR}_{\mathrm{o}}}{\mathrm{SNR}} &= \frac{\mathrm{tr}\{\boldsymbol{R}_v\}\mathrm{tr}\{\boldsymbol{H}_{\mathrm{o}}\boldsymbol{R}_x\boldsymbol{H}_{\mathrm{o}}^{\mathrm{T}}\}}{\mathrm{tr}\{\boldsymbol{R}_x\}\mathrm{tr}\{\boldsymbol{H}_{\mathrm{o}}\boldsymbol{R}_v\boldsymbol{H}_{\mathrm{o}}^{\mathrm{T}}\}} \\ &= \frac{\mathrm{tr}\{\boldsymbol{R}_v\}\mathrm{tr}\{\boldsymbol{R}_x[\boldsymbol{R}_x+\mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_x[\boldsymbol{R}_x+\mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_x\}}{\mathrm{tr}\{\boldsymbol{R}_x\}\mathrm{tr}\{\boldsymbol{R}_x[\boldsymbol{R}_x+\mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_v[\boldsymbol{R}_x+\mu\boldsymbol{R}_v]^{-1}\boldsymbol{R}_x\}}. \end{aligned} \tag{60}$$

Following the matrix-decomposition procedure given in (17) and using some simple algebra, we can deduce that

$$\begin{aligned}\frac{\mathrm{SNR}_{\mathrm{o}}}{\mathrm{SNR}} &= \frac{\mathrm{tr}\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\}\mathrm{tr}\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-1}\boldsymbol{\Lambda}\boldsymbol{B}\}}{\mathrm{tr}\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{B}\}\mathrm{tr}\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-2}\boldsymbol{\Lambda}\boldsymbol{B}\}} \\ &= \frac{\mathrm{tr}\{\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\}\mathrm{tr}\{\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-1}\boldsymbol{\Lambda}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\}}{\mathrm{tr}\{\boldsymbol{\Lambda}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\}\mathrm{tr}\{\boldsymbol{\Lambda}(\boldsymbol{\Lambda}+\mu\boldsymbol{I})^{-2}\boldsymbol{\Lambda}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\}} \\ &= \frac{\sum_{i=1}^{L}\sum_{j=1}^{L} b_{ij}^2 \cdot \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\lambda_i^3}{(\lambda_i+\mu)^2} b_{ij}^2}{\sum_{i=1}^{L}\sum_{j=1}^{L}\lambda_i b_{ij}^2 \cdot \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\lambda_i^2}{(\lambda_i+\mu)^2} b_{ij}^2}. \end{aligned} \tag{61}$$

Now if following the proof given in Appendix (setting $q_i^2 = \sum_{j=1}^{L} b_{ij}^2$), we can show

$$\sum_{i=1}^{L}\sum_{j=1}^{L} b_{ij}^2 \cdot \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\lambda_i^3}{(\lambda_i+\mu)^2} b_{ij}^2 \geqslant \sum_{i=1}^{L}\sum_{j=1}^{L}\lambda_i b_{ij}^2 \cdot \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\lambda_i^2}{(\lambda_i+\mu)^2} b_{ij}^2. \tag{62}$$

It follows immediately that

$$\mathrm{SNR_o} \geqslant \mathrm{SNR}, \tag{63}$$

with equality if and only if $\lambda_1 = \lambda_2 = \cdots = \lambda_L$. Note that this condition is the same as that obtained previously for the Wiener filter, which should not come as a surprise since both techniques are formulated in a similar way.

From the previous, we see that $\mathrm{SNR_o}$ depends not only on the speech and noise characteristics, but also on the value of μ. With some algebra, it can be shown that

$$\lim_{\mu\to 0}\mathrm{SNR_o} = \lim_{\mu\to 0}\frac{\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}b_{ij}^2}{\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}b_{ij}^2} = \mathrm{SNR}, \tag{64}$$

which is the lower bound of the *a posteriori* SNR. When $\mu \to +\infty$, we see that $\lambda_i + \mu \to \mu$. Therefore,

$$\lim_{\mu\to +\infty}\mathrm{SNR_o} = \frac{\sum_{i=1}^{L}\sum_{j=1}^{L}\lambda_i^3 b_{ij}^2}{\sum_{i=1}^{L}\sum_{j=1}^{L}\lambda_i^2 b_{ij}^2} \leqslant \sum_{i=1}^{L}\lambda_i, \tag{65}$$

which is the upper bound of the *a posteriori* SNR. □

5. Frequency-domain noncausal Wiener filter and its performance

The Wiener filter can also be formulated in the frequency domain. One way to derive such a filter is to transform the time-domain filter into the frequency domain using the so-called overlap-add technique. In this case, the time-domain Wiener filter and its frequency-domain counterpart have exactly the same performance. More often, however, the frequency-domain Wiener filter is formulated by directly estimating the clean speech spectrum from the noisy speech spectrum. The resulting frequency-domain Wiener filter differs in two aspects from the time-domain Wiener filter: first, the former is a causal filter, while the latter can be a noncausal one; second, the former is a full-band technique, while the latter is a subband technique, where each subband filter is independent on the filters of other frequency bands. Therefore, the results achieved from the time-domain full-band Wiener filter may not be applicable to such a frequency-domain subband Wiener filter. It is necessary that we perform some study of the subband Wiener filter.

Given the signal model in (3), the frequency-domain subband Wiener filter is derived by the following criterion (Wiener, 1949):

$$H_{\mathrm{o}}(j\omega_k) = \arg\min_{H(j\omega_k)} E\{|X(j\omega_k) - H(j\omega_k)Y(j\omega_k)|^2\}. \tag{66}$$

It can be easily deduced that,

$$H_{\mathrm{o}}(j\omega_k) = \frac{E[|X(j\omega_k)|^2]}{E[|Y(j\omega_k)|^2]} = \frac{P_x(\omega_k)}{P_y(\omega_k)}, \tag{67}$$

where $P_x(\omega_k) = \frac{1}{L}E[|X(j\omega_k)|^2]$ and $P_y(\omega_k) = \frac{1}{L}E[|Y(j\omega_k)|^2]$ are the power spectral densities (PSDs) of $x(n)$ and $y(n)$ respectively. It can be seen from this expression that the frequency-domain Wiener filter $H_{\mathrm{o}}(j\omega_k)$ is positive and real valued. Therefore, it only modifies the magnitude of the noisy speech spectra, while leaves the phase components unchanged. Since $H_{\mathrm{o}}(j\omega_k)$ is real valued, we shall, from now on, drop the symbol j from its expression, which should not introduce any confusion.

The optimal estimate of the clean speech spectrum, using $H_{\mathrm{o}}(\omega_k)$, is

$$\begin{aligned}\widehat{X}_{\mathrm{o}}(j\omega_k) &= H_{\mathrm{o}}(\omega_k)Y(j\omega_k)\\ &= H_{\mathrm{o}}(\omega_k)X(j\omega_k) + H_{\mathrm{o}}(\omega_k)V(j\omega_k).\end{aligned} \tag{68}$$

Applying the inverse DFT (IDFT) to (68), we can obtain the optimal estimate of the speech samples $\hat{x}_{\mathrm{o}}(n)$.

The power of the estimated clean speech can be evaluated according to the Parseval's relation, i.e.,

$$E[\hat{x}_{\mathrm{o}}^2(n)] = \sum_{k=0}^{L-1}\frac{1}{L}E[|\widehat{X}_{\mathrm{o}}(j\omega_k)|^2] = \sum_{k=0}^{L-1}\frac{P_x^2(\omega_k)}{P_y^2(\omega_k)}P_y(\omega_k). \tag{69}$$

Since speech and noise are uncorrelated, we have

$$P_y(\omega_k) = P_x(\omega_k) + P_v(\omega_k), \tag{70}$$

where $P_v(\omega_k)$ is the PSD of $v(n)$. Substituting (70) into (69), we deduce that

$$E[\hat{x}_{\mathrm{o}}^2(n)] = \sum_{k=0}^{L-1}\frac{P_x^2(\omega_k)}{P_y^2(\omega_k)}P_x(\omega_k) + \sum_{k=0}^{L-1}\frac{P_x^2(\omega_k)}{P_y^2(\omega_k)}P_v(\omega_k), \tag{71}$$

which is the sum of two terms. The first one is the power of the filtered clean speech and the second one is the power of the residual noise.

Now let us examine the noise-reduction performance. The power of the noise in the observation signal can be computed as $\sum_{k=0}^{L-1}P_v(\omega_k)$. From (71), one can see that the power of the residual noise signal is $\sum_{k=0}^{L-1}\frac{P_x^2(\omega_k)}{P_y^2(\omega_k)}P_v(\omega_k)$. So the noise-reduction factor of the frequency-domain Wiener filter is written as

$$\xi_{\mathrm{nr}}[H(\omega_k)] = \frac{\sum_{k=0}^{L-1}P_v(\omega_k)}{\sum_{k=0}^{L-1}\frac{P_x^2(\omega_k)}{P_y^2(\omega_k)}P_v(\omega_k)}. \tag{72}$$

Since $P_x^2(\omega_k) \leqslant P_y^2(\omega_k)$, we easily verify that

$$\xi_{\mathrm{nr}}[H(\omega_k)] \geqslant 1. \tag{73}$$

This indicates that the Wiener filter can reduce the noise level (unless there is no noise at all). Similarly, we can verify that the power of the filtered clean speech is always less than the power of the original clean speech. So the frequency-domain Wiener filter, just like its time-domain counterpart, also reduces noise at a price of attenuating the clean speech. It is key to know, then, whether this Wiener filter can improve SNR. We have the following proposition.

Proposition 3. *With the Wiener filter given in (71), the a posteriori SNR (defined after the Wiener filter) is always greater than or at least equal to the a priori SNR.*

Proof. If there is no noise at all, we see that the Wiener filter has no effect on SNR. Now we consider the generic case where noise is not zero. In the frequency domain, the *a priori* SNR can be expressed as

$$\mathrm{SNR} = \frac{\sum_{k=0}^{L-1} P_x(\omega_k)}{\sum_{k=0}^{L-1} P_v(\omega_k)}. \tag{74}$$

Similarly, we can write, after the Wiener filtering, the *a posteriori* SNR according to (71), as

$$\mathrm{SNR_o} = \frac{\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_x(\omega_k)}{\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_v(\omega_k)}. \tag{75}$$

It follows immediately that

$$\frac{\mathrm{SNR_o}}{\mathrm{SNR}} = \frac{\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_x(\omega_k) \sum_{k=0}^{L-1} P_v(\omega_k)}{\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_v(\omega_k) \sum_{k=0}^{L-1} P_x(\omega_k)}. \tag{76}$$

Now let denote

$$\begin{aligned}\phi(\omega_k) &= \left[\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_x(\omega_k)\right] \cdot \sum_{k=0}^{L-1} P_v(\omega_k) \\ &\quad - \left[\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_v(\omega_k)\right] \cdot \sum_{k=0}^{L-1} P_x(\omega_k).\end{aligned}$$

It follows immediately that

$$\begin{aligned}\phi(\omega_k) &= \sum_{k=0}^{L-1} \sum_{j=0}^{L-1} \frac{P_x^3(\omega_k)}{P_y^2(\omega_k)} P_v(\omega_j) \\ &\quad - \sum_{k=0}^{L-1} \sum_{j=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_v(\omega_k) P_x(\omega_j) \\ &= \sum_{k=0}^{L-1} \sum_{j=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} [P_x(\omega_k) P_v(\omega_j) - P_v(\omega_k) P_x(\omega_j)] \\ &= \sum_{k=0}^{L-1} \sum_{j>k}^{L-1} \left[\frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} - \frac{P_x^2(\omega_j)}{P_y^2(\omega_j)}\right] \\ &\quad \cdot [P_x(\omega_k) P_v(\omega_j) - P_v(\omega_k) P_x(\omega_j)].\end{aligned} \tag{77}$$

Using (70), we can derive $\phi(\omega_k)$ as

$$\begin{aligned}\phi(\omega_k) &= \sum_{k=0}^{L-1} \sum_{j>k}^{L-1} \left[\frac{P_x(\omega_k)}{P_y(\omega_k)} + \frac{P_x(\omega_j)}{P_y(\omega_j)}\right] \\ &\quad \cdot \left[\frac{P_x(\omega_k)}{P_y(\omega_k)} - \frac{P_x(\omega_j)}{P_y(\omega_j)}\right] [P_x(\omega_k) P_v(\omega_j) - P_v(\omega_k) P_x(\omega_j)] \\ &= \sum_{k=0}^{L-1} \sum_{j>k}^{L-1} \frac{1}{P_y(\omega_k) P_y(\omega_j)} \left[\frac{P_x(\omega_k)}{P_y(\omega_k)} + \frac{P_x(\omega_j)}{P_y(\omega_j)}\right] \\ &\quad \cdot [P_x(\omega_k) P_v(\omega_j) - P_v(\omega_k) P_x(\omega_j)]^2.\end{aligned} \tag{78}$$

Since $P_x(\omega_k) \geqslant 0$, $P_v(\omega_k) \geqslant 0$, and $P_y(\omega_k) \geqslant 0$, it is easy to see that the right-hand side of (78) is greater than, or at least equal to 0. Therefore, we have

$$\sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_x(\omega_k) \sum_{k=0}^{L-1} P_v(\omega_k) \geqslant \sum_{k=0}^{L-1} \frac{P_x^2(\omega_k)}{P_y^2(\omega_k)} P_v(\omega_k) \sum_{k=0}^{L-1} P_x(\omega_k),$$

which means that

$$\frac{\mathrm{SNR_o}}{\mathrm{SNR}} \geqslant 1.$$

As a result,

$$\mathrm{SNR_o} \geqslant \mathrm{SNR},$$

where we see from (78) that equality is attained if and only if $\frac{P_x(\omega_0)}{P_v(\omega_0)} = \frac{P_x(\omega_1)}{P_v(\omega_1)} = \cdots = \frac{P_x(\omega_{L-1})}{P_v(\omega_{L-1})}$. In other words, the frequency-domain Wiener filter will be able to increase SNR unless all the subband SNRs are equal, which is understandable. When all the subband SNRs are equal, it means that both speech and noise have the same PSD under the given resolution condition. In this case, the Wiener filter is not able to distinguish noise from speech and, as a result, not able to increase SNR. Otherwise, the subband Wiener filter can improve SNR. That completes the proof. □

6. Experiments

Throughout the text, we have shown by theoretical analysis that the optimal linear filtering techniques including the time-domain Wiener filter, the subspace method, and the frequency-domain Wiener filter, can indeed reduce the level of noise that is present in the speech signal and improve the corresponding SNR, regardless of whether the noise is white or colored. In this section, we will validate the analysis through experiments.

The first experiment is to investigate the SNR behavior of the time-domain Wiener filter. But before doing this, we must implement such a filter, which involves estimation of the correlation matrix $\boldsymbol{R}_y$, and the two correlation vectors $\boldsymbol{r}_{yy}$ and $\boldsymbol{r}_{vv}$. In our experiment, we approximate the mathematical expectation by the exponentially weighted sample average, and the correlation matrix and vectors at time n are estimated as follows:

$$\begin{aligned}\widehat{\boldsymbol{R}}_y(n) &= \sum_{m=0}^{n} \alpha^{n-m} \boldsymbol{y}(m) \boldsymbol{y}^{\mathrm{T}}(m) = \alpha \widehat{\boldsymbol{R}}_y(n-1) + \boldsymbol{y}(n) \boldsymbol{y}^{\mathrm{T}}(n), \\ \hat{\boldsymbol{r}}_{yy}(n) &= \sum_{m=0}^{n} \alpha^{n-m} \boldsymbol{y}(m) y(m) = \alpha \hat{\boldsymbol{r}}_{yy}(n-1) + \boldsymbol{y}(n) y(n), \\ \hat{\boldsymbol{r}}_{vv}(n) &= \sum_{m=0}^{n} \alpha^{n-m} \boldsymbol{v}(m) v(m) = \alpha \hat{\boldsymbol{r}}_{vv}(n-1) + \boldsymbol{v}(n) v(n),\end{aligned} \tag{79}$$

where α $(0 < \alpha < 1)$ is a forgetting factor. In our experiment, α is chosen as a function of the filter length L, i.e., $\alpha = 1 - 1/(9L)$.

Note that in order to compute $\hat{\boldsymbol{r}}_{vv}(n)$, we will need to obtain an estimate of the noise signal $v(n)$. The most popular method for this is to design a voice activity detector (VAD) and estimate noise in the regions where speech is absent. This method works reasonably well in high SNR conditions and when noise is stationary. But when SNR is low or the noise characteristics change over time (even slowly), the noise estimate obtained during the absence of speech may not represent the noise in the presence of speech. To avoid explicit speech/non-speech detection and take into account some non-stationarity of the noise, several other estimation techniques were developed, including the minimum statistics method (Martin, 2001), the quantile method (Hirsch and Ehrlicher, 1995; Stahl et al., 2000), and the sequential estimation (Diethorn, 2004) etc. In this experiment, instead of using the estimated noise, we assume that the noise signal is known *a priori* since our concern is on the SNR behavior of the Wiener filter rather than the implementation of the Wiener filter itself.

The results of this experiment are shown in Figs. 1 and 2, where the clean speech used is a prerecorded speech signal (from a female speaker) that is sampled at 8 kHz and lasts about 3 min. Noise is then added to the signal at an SNR of 10 dB. Three noise signals are used: computer generated White Gaussian noise, a babbling noise signal recorded in a New York Stock Exchange (NYSE) room, and a car noise signal recorded in a Volvo car running at 55 miles/h on a highway with all its windows closed.

We see from Fig. 1 that the *a posteriori* SNR is from 2 to several decibels higher than the *a priori* SNR. This justifies the analysis shown in Section 3. It is also seen from this figure that the *a posteriori* SNR increases monotonically with the filter length L. Therefore, a longer filter should be applied if we expect more noise reduction. However, as the filter length increases, the complexity of the algorithm, the amount of speech distortion, and the estimation variance of the correlation matrix/vectors, they all will increase at the same time. In addition, we see that SNR_o increases with L but with a slower rate for L larger than about 12. Taking all these factors into account, we suggest to set L between 10 and 30 for 8-kHz sampling rate, which is reasonable for most applications.

When L is fixed, but we change SNR from -10 dB to 30 dB, it can be seen that less SNR improvement is achieved for high SNR conditions. This is, of course, understandable. When SNR increases, the speech becomes less noisy, so there is less noise to be reduced. If SNR approaches infinity, the *a posteriori* SNR would be the same as the *a priori* SNR. In all the conditions, we see that the *a posteriori* SNR is always greater than the corresponding *a priori* SNR. This again verifies the analysis given in Section 3.

In the second experiment, we study the SNR behavior of the subspace method. Again, we assume that the noise signal is known *a priori* and we estimate the two correlation matrices $\boldsymbol{R}_x$ and $\boldsymbol{R}_v$ in the same manner as we estimate $\boldsymbol{R}_y$ in (79). The results for this experiments are plotted in Figs. 3 and 4.

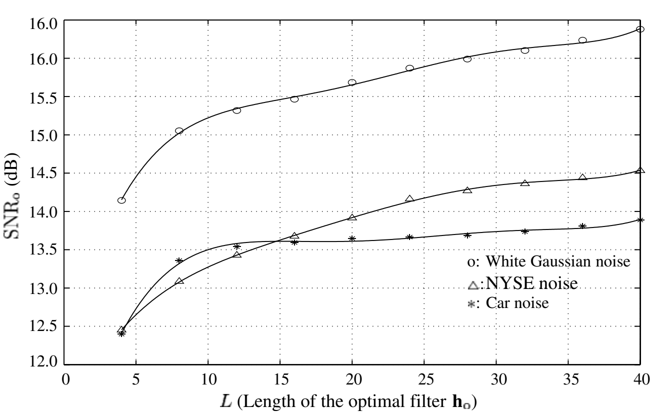

Fig. 1. The *a posteriori* SNR as a function of L: SNR = 10 dB.

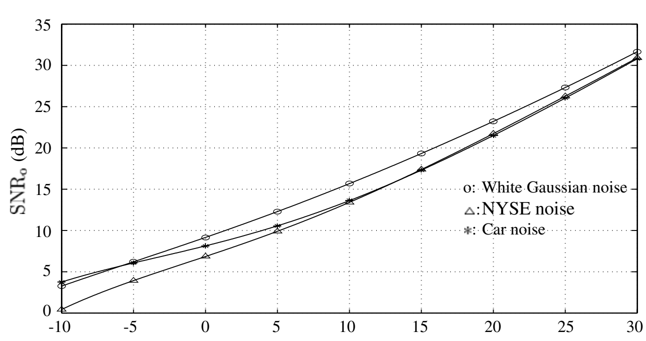

Fig. 2. The *a posteriori* SNR versus the *a priori* SNR: $L = 20$.

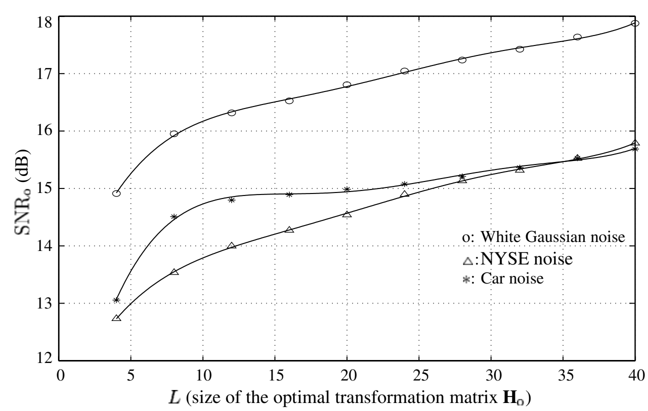

Fig. 3. The *a posteriori* SNR as a function of L: SNR = 10 dB.

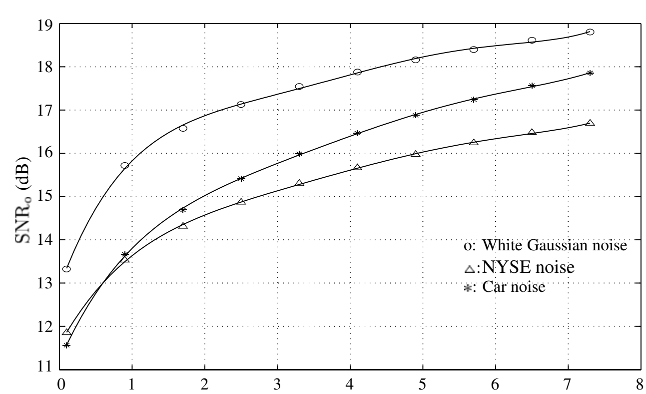

Fig. 4. The *a posteriori* SNR as a function of μ: $L = 20$.

Similar to the previous experiment, one can see from Fig. 3 that when the *a priori* SNR is fixed, the *a posteriori* SNR increases monotonically with the dimension of the optimal transformation $\boldsymbol{H}_o$. Therefore, if more SNR improvement is expected, we should use a transform with a larger dimension L. But when we increase L, we also raise the computational complexity and boost the speech distortion at the same time. Normally, like in the Wiener filter, L should be set between 15 and 30 for 8-kHz sampling rate. It is also seen from Fig. 3 that SNR_o is always higher than SNR, which validates the analysis shown in Section 4.

Another parameter that plays a critical role in the subspace method is the Lagrange multiplier μ. In Section 4, we have shown that if the *a priori* SNR is fixed, the *a posteriori* SNR is a monotonically increasing function of the variable μ. When $\mu \to 0$, the *a posteriori* SNR will approach the *a priori* SNR. When $\mu \to +\infty$, the *a posteriori* SNR will approach its upper bound given in (65). This is justified by the results shown in Fig. 4. In real applications, the selection of μ should depend on system's expectation between noise reduction and speech distortion. If more noise reduction is expected, we should select a large μ. On the other hand, if speech distortion is a big concern, then we should choose a small μ. If we set $\mu = 1$, then the subspace method will be similar to the time-domain Wiener filter.

The third experiment pertains to the frequency-domain Wiener filter. Similar to the previous experiments, we neglect the noise-estimation process and assume the noise signal is known *a priori*. In order to estimate the Wiener filter according to (67), we partition the noisy speech $y(n)$ into overlapping frames. The frame size in this experiment is $L = 64$ (samples), and the overlapping factor is 75%. Each frame is then transformed via a DFT into a block of L spectral samples. Successive frames of spectral samples form a two-dimensional matrix denoted by $Y_t(j\omega_k)$, where subscript t is the frame index and denotes the time dimension. The power spectra of the noisy signal at time t is estimated as

$$\widehat{P}_{y,t}(\omega_k) = \beta \widehat{P}_{y,t-1}(\omega_k) + (1-\beta)|Y_t(j\omega_k)|^2/L, \tag{80}$$

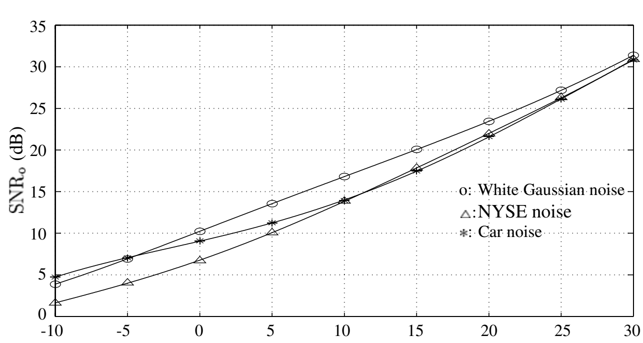

Fig. 5. The *a posteriori* SNR versus the *a priori* SNR: $L = 64$, 75% overlap.

where β is a coefficient to control the time constant of the single-pole recursion. In our experiment, $\beta = 0.9$. The noise power spectra are estimated in a similar manner, and the Wiener filter is computed according to (67). The experimental results are plotted in Fig. 5. We see that the *a posteriori* SNR is always larger than the *a priori* SNR, which justifies the analysis shown in Section 5.

7. Conclusion and discussion

The problem of noise reduction has attracted a considerable amount of research attention over the past several decades. The most widely used approach, thus far, is the optimal linear filtering technique, which achieves clean speech estimate by passing the noisy speech through an optimal linear filter/transformation. While many efforts were made to evaluate how this technique can perform in simulated as well as real noise environments, most existing performance studies have been experimental, which do not provide us with a unified understanding of the performance behavior of the technique. Therefore, more thorough analysis is indispensable.

This paper provided a theoretical analysis on the noise reduction, speech distortion, and SNR behavior of the optimal linear filtering techniques including the time-domain causal Wiener filter, the subspace method, and the frequency-domain noncausal Wiener filter. It showed that the optimal linear filter, regardless of how we delineate it, can indeed reduce the level of noise. Most importantly, it proved that, with the optimal linear filter, the *a posteriori* SNR is always greater than, or at least equal to the *a priori* SNR, which concludes that the optimal linear filtering technique is able to make noisy speech signals cleaner. While the *a posteriori* SNR is lower bounded to the *a priori* SNR, its upper limit depends on the distinction between noise and speech statistics, which was also discussed in the paper. In order to validate the analysis, we carried out a series of experiments using the simulated white Gaussian noise as well as some noise signals recorded in real situations. Theoretical and experimental results agreed very well.

Notice that during the theoretical analysis, we assumed that we knew the second order statistics (covariance matrices, correlation vectors, and power spectral densities) of the noisy speech and the noise signal. In real application, however, they are not known *a priori* and have to be estimated. Any deviation of the estimate from its true value may cause some performance degradation. As how the optimal filtering technique is sensitive to the estimation error of the signal statistics, this is worthy of further research attention.

Appendix

Lemma. *With* λ_i ($i = 1, 2, \ldots, L$ *and* $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_L \geqslant 0$) *being defined in* (17) *and* $\mu > 0$, *we have*

$$\left[\sum_{i=1}^{L}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{L}q_i^2 \geqslant \left[\sum_{i=1}^{L}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{L}\lambda_i q_i^2, \tag{81}$$

where q_i can be any real number.

Proof. This inequality can be proofed by way of induction.

- Basic step: If $L=2$,

$$\left[\sum_{i=1}^{2}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{2}q_i^2 = \frac{\lambda_1^3}{(\lambda_1+\mu)^2}q_1^4+\frac{\lambda_2^3}{(\lambda_2+\mu)^2}q_2^4 + \left[\frac{\lambda_1^3}{(\lambda_1+\mu)^2}+\frac{\lambda_2^3}{(\lambda_2+\mu)^2}\right]q_1^2q_2^2.$$

Since $\lambda_1 \geqslant \lambda_2 \geqslant 0$, it is trivial to show that

$$\frac{\lambda_1^3}{(\lambda_1+\mu)^2}+\frac{\lambda_2^3}{(\lambda_2+\mu)^2} \geqslant \frac{\lambda_1^2\lambda_2}{(\lambda_1+\mu)^2}+\frac{\lambda_1\lambda_2^2}{(\lambda_2+\mu)^2},$$

where "=" holds when $\lambda_1=\lambda_2$. Therefore

$$\begin{aligned}\left[\sum_{i=1}^{2}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{2}q_i^2 &\geqslant \frac{\lambda_1^3}{(\lambda_1+\mu)^2}q_1^4+\frac{\lambda_2^3}{(\lambda_2+\mu)^2}q_2^4 + \left[\frac{\lambda_1^2\lambda_2}{(\lambda_1+\mu)^2}+\frac{\lambda_1\lambda_2^2}{(\lambda_2+\mu)^2}\right]q_1^2q_2^2 \\ &= \left[\sum_{i=1}^{2}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{2}\lambda_i q_i^2,\end{aligned}$$

so the property is true for $L=2$, where "=" holds when $\lambda_1=\lambda_2$ or at least one of q_1 and q_2 is equal to 0.

- Inductive step: Assume that the property is true for $L=n$, i.e.,

$$\left[\sum_{i=1}^{n}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n}q_i^2 \geqslant \left[\sum_{i=1}^{n}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n}\lambda_i q_i^2.$$

We must prove that it is also true for $L=n+1$. As a matter of fact,

$$\begin{aligned}&\left[\sum_{i=1}^{n+1}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n+1}q_i^2 \\ &= \left[\sum_{i=1}^{n}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2+\frac{\lambda_{n+1}^3}{(\lambda_{n+1}+\mu)^2}q_{n+1}^2\right]\cdot\left[\sum_{i=1}^{n}q_i^2+q_{n+1}^2\right] \\ &= \left[\sum_{i=1}^{n}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n}q_i^2+\frac{\lambda_{n+1}^3}{(\lambda_{n+1}+\mu)^2}q_{n+1}^4 \\ &\quad+\sum_{i=1}^{n}\left[\frac{\lambda_i^3}{(\lambda_i+\mu)^2}+\frac{\lambda_{n+1}^3}{(\lambda_{n+1}+\mu)^2}\right]q_i^2q_{n+1}^2.\end{aligned} \tag{82}$$

Using the induction hypothesis, and also the fact that

$$\frac{\lambda_i^3}{(\lambda_i+\mu)^2}+\frac{\lambda_{n+1}^3}{(\lambda_{n+1}+\mu)^2} \geqslant \frac{\lambda_i^2\lambda_{n+1}}{(\lambda_i+\mu)^2}+\frac{\lambda_i\lambda_{n+1}^2}{(\lambda_{n+1}+\mu)^2},$$

we obtain

$$\begin{aligned}\left[\sum_{i=1}^{n+1}\frac{\lambda_i^3}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n+1}q_i^2 &\geqslant \left[\sum_{i=1}^{n}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n}\lambda_i q_i^2 \\ &\quad+\frac{\lambda_{n+1}^3}{(\lambda_{n+1}+\mu)^2}q_{n+1}^4+\sum_{i=1}^{n}\left[\frac{\lambda_i^2\lambda_{n+1}}{(\lambda_i+\mu)^2}+\frac{\lambda_i\lambda_{n+1}^2}{(\lambda_{n+1}+\mu)^2}\right]q_i^2q_{n+1}^2 \\ &= \left[\sum_{i=1}^{n+1}\frac{\lambda_i^2}{(\lambda_i+\mu)^2}q_i^2\right]\cdot\sum_{i=1}^{n+1}\lambda_i q_i^2,\end{aligned} \tag{83}$$

where "=" holds when all the λ_i's corresponding to nonzero q_i are equal, where $i=1,2,\ldots,n+1$. That completes the proof. □

References

Benesty, J., Makino, S., Chen, J. (Eds.), 2005. Speech Enhancement. Springer, Berlin, Germany.

Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. 27 (April), 113–120.

Chang, Y.M., O'Shaughnessy, D., 1991. Speech enhancement based conceptually on auditory evidence. IEEE Trans. Signal Process. 39 (September), 1943–1954.

Chen, J., Huang, Y., Benesty, J., 2003. Filtering techniques for noise reduction and speech enhancement. In: Benesty, J., Huang, Y. (Eds.), Adaptive Signal Processing: Applications to Real-World Problems. Springer, Berlin, Germany, pp. 129–154.

Chen, J., Benesty, J., Huang, Y., Doclo, S., 2006. New insights into the noise reduction Wiener filter. IEEE Trans. Speech Audio Process. 14 (July), 1218–1234.

Dendrinos, M., Bakamidis, S., Garayannis, G., 1991. Speech enhancement from noise: a regenerative approach. Speech Commun. 10 (February), 45–57.

Diethorn, E.J., 2004. Subband noise reduction methods for speech enhancement. In: Huang, Y., Benesty, J. (Eds.), Audio Signal Processing for Next-Generation Multimedia Communication Systems. Kluwer, Boston, MA, USA, pp. 91–115.

Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. 32 (December), 1109–1121.

Ephraim, Y., Malah, D., 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. 33 (April), 443–445.

Ephraim, Y., Van Trees, H.L., 1995. A signal subspace approach for speech enhancement. IEEE Trans. Speech Audio Process. 3 (July), 251–266.

Etter, W., Moschytz, G.S., 1994. Noise reduction by noise-adaptive spectral magnitude expansion. J. Audio Eng. Soc. 42 (May), 341–349.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic, San Diego, CA, USA.

Gannot, S., Burshtein, D., Weinstein, E., 1998. Iterative and sequential Kalman filter-based speech enhancement algorithms. IEEE Trans. Speech Audio Process. 6 (July), 373–385.

Gibson, J.D., Koo, B., Gray, S.D., 1991. Filtering of colored noise for speech enhancement and coding. IEEE Trans. Signal Process. 39 (August), 1732–1742.

Hansen, P.S.K., 1997. Signal subspace methods for speech enhancement. Ph.D. dissertation, Techn. University Denmark, Lyngby, Denmark.

Hirsch, H.G., Ehrlicher, C., 1995. Noise estimation techniques for robust speech recognition. In: Proc. IEEE ICASSP. vol. 1, pp. 153–156.

Hu, Y., Loizou, P.C., 2003. A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans. Speech Audio Process. 11 (July), 334–341.

Lev-Ari, H., Ephraim, Y., 2003. Extension of the signal subspace speech enhancement approach to colored noise. IEEE Trans. Speech Audio Process. 10 (April), 104–106.

Lim, J.S. (Ed.), 1983. Speech Enhancement. Prentice-Hall, Englewood Cliffs, NJ, USA.

Lim, J.S., Oppenheim, A.V., 1979. Enhancement and bandwidth compression of noisy speech. Proc. IEEE 67 (December), 1586–1604.

Martin, R., 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech Audio Process. 9 (July), 504–512.

McAulay, R.J., Malpass, M.L., 1980. Speech enhancement using a soft-decision noise suppression filter. IEEE Trans. Acoust. Speech Signal Process. 28 (April), 137–145.

Mittal, U., Phamdo, N., 2000. Signal/noise KLT based approach for enhancing speech degraded by colored noise. IEEE Trans. Speech Audio Process. 8 (March), 159–167.

Paliwal, K.K., Basu, A., 1987. A speech enhancement method based on Kalman filtering. In: Proc. IEEE ICASSP. pp. 177–180.

Rezayee, A., Gazor, S., 2001. An adpative KLT approach for speech enhancement. IEEE Trans. Speech Audio Process. 9 (February), 87–95.

Stahl, V., Fischer, A., Bippus, R., 2000. Quantile based noise estimation for spectral subtraction and Wiener filtering. In: Proc. IEEE ICASSP. vol. 3, pp. 1875–1878.

Vary, P., 1985. Noise suppression by spectral magnitude estimation – mechanism and theoretical limits. Signal Process. 8 (July), 387–400.

Virag, N., 1999. Single channel speech enhancement based on masking properties of human auditory system. IEEE Trans. Speech Audio Process. 7 (March), 126–137.

Wang, D.L., Lim, J.S., 1982. The unimportance of phase in speech enhancement. IEEE Trans. Acoust. Speech Signal Process. 30 (August), 679–681.

Widrow, B., Stearns, S.D., 1985. Adaptive Signal Processing. Prentice Hall, Englewood Cliffs, NJ, USA.

Wiener, N., 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. John Wiley and Sons, New York, USA.