Optimal Biorthogonal Transform for Colored Noise Suppression With Subband Wiener Filtering

See-May Phoong and Yuan-Pei Lin

***Abstract*—It is well-known that the Karhunen–Loeve transform (KLT) is optimal for white noise suppression. Recently, Akkarakaran and Vaidyanathan showed that for the case of colored noise, if both the noise and signal have a common KLT, then the KLT is optimal for subband noise suppression. In this paper, we will derive the optimal transform (not restricting to the class of unitary transforms) for the noise suppression problem when the signal and noise have arbitrary spectrum.**

***Index Terms*—Denoising, filter bank, noise suppression, subband Wiener filter, transform.**

I. Introduction

IN RECENT years, there has been considerable interest in the application of filter bank (FB) to noise suppression (denoising) (see [1], [2], and references therein). Fig. 1 shows such a FB-based noise reduction scheme. The black boxes in the figure denote the subband denoising operations. There are various subband denoising schemes such as the Wiener filtering, soft thresholding, hard thresholding, input adaptive thresholding, etc. In [2], a practical thresholding scheme that is applied to each subband sample was proposed. It was shown [2] that the proposed thresholding scheme, which thresholds the coefficients to a specific level, provides a quasi optimal min-max estimator of a noisy piecewise-smooth signals.

Unlike [2], our goal is to design the FB-based denoising scheme to minimize the output error variance

$$\mathcal{E}_n = E\{(\mathbf{y}(n) - \mathbf{s}(n))^T(\mathbf{y}(n) - \mathbf{s}(n))\} \tag{1}$$

where $\mathbf{s}(n)$ is the desired signal and $\mathbf{y}(n)$ is the output signal, as shown in Fig. 1. We will consider the case when the subband denoising scheme is Wiener filtering. Recently, Akkarakaran and Vaidyanathan [1] showed that for the white noise case the principle component FB (PCFB) (if exists) is the optimal orthonormal FB that minimizes $\mathcal{E}_n$. Moreover, the optimality of PCFB holds even when any combination of Wiener filter or hard threshold is used in the subband. For the special case of memoryless transform where $\mathbf{T}$ is a constant matrix, PCFB reduces to the well-known KLT. If the additive noise is colored, the only solution known is for the case when the autocorrelation matrices of $\mathbf{s}(n)$ and $\boldsymbol{\nu}(n)$ have a common KLT. For this restrictive case, the common KLT is shown [1] to be the optimal memoryless transform when zeroth-order Wiener filter is applied in the subbands. In this paper, we consider the more general case of arbitrary signal and noise spectrum. We will derive the optimal memoryless transform (not restricting to unitary transform) for colored noise suppression.

Manuscript received August 21, 2001; revised January 17, 2002. The work was supported in part by the National Science Council of Taiwan, R.O.C., under Grants 90-2213-E-002-097 and 90-2213-E-009-108, the Ministry of Education, Taiwan, under Contract 89-E-FA06-2-4, and the Lee and MTI Center for Networking Research. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. John Apostolopoulos.

S.-M. Phoong is with the Department of Electrical Engineering and the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: smp@cc.ee.ntu.edu.tw).

Y.-P. Lin is with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C.

Publisher Item Identifier S 1070-9908(02)06039-X.

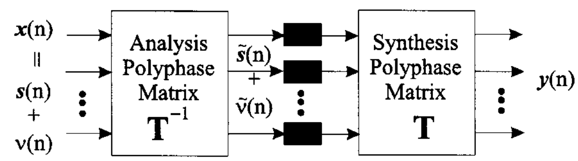

Fig. 1. Filter bank based noise suppression scheme.

II. Optimal Transform

Let the input $\mathbf{x}(n) = \mathbf{s}(n) + \boldsymbol{\nu}(n)$, where $\mathbf{s}(n)$ is the desired signal and $\boldsymbol{\nu}(n)$ is the additive noise. Assume that $\mathbf{s}(n)$ and $\boldsymbol{\nu}(n)$ are real zero-mean WSS uncorrelated vector processes. The $M \times M$ autocorrelation matrices of $\mathbf{x}(n)$, $\mathbf{s}(n)$ and $\boldsymbol{\nu}(n)$ are denoted by $\mathbf{R}_x, \mathbf{R}_s$ and $\mathbf{R}_\nu$ respectively. As $\mathbf{s}(n)$ and $\boldsymbol{\nu}(n)$ are uncorrelated, we have

$$\mathbf{R}_x = \mathbf{R}_s + \mathbf{R}_\nu.$$

Without much loss of generality, we assume that the matrix $\mathbf{R}_x$ is invertible. Hence $\mathbf{R}_x$ is positive definite. Let $\mathbf{R}_x^{1/2}$ be the unique positive definite matrix that satisfies $\mathbf{R}_x^{1/2}\mathbf{R}_x^{1/2} = \mathbf{R}_x$.

We consider only the class of FB with constant polyphase matrices. That is, the matrix $\mathbf{T}$ is a nonsingular constant matrix. Assume that the subband operation is carried out by multiplication with a set of constants k_i. Therefore the subband operations denoted by the black boxes can be written as the diagonal matrix

$$\mathbf{K} = \begin{bmatrix} k_0 & 0 & \cdots & 0 \\ 0 & k_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_{M-1} \end{bmatrix}. \tag{2}$$

Our aim is to find the best transform $\mathbf{T}$ and the optimal k_i such that the output error variance $\mathcal{E}_n$ in (1) is minimized. The optimal solution is given by the following theorem.

Theorem 1: Consider the denoising scheme in Fig. 1. Suppose that the subband operation is taken as (2). Assume that $\mathbf{R}_x$ is nonsingular. Then the output error variance $\mathcal{E}_n$ in (1) is minimized if and only if the following conditions hold:

1) columns of $\mathbf{T}$ are chosen as the eigenvectors of the matrix $\mathbf{R}_s\mathbf{R}_x^{-1}$;
2) scalars k_i are the corresponding eigenvalues.

Moreover the optimal transform can be expressed as $\mathbf{T}_{\rm opt} = \mathbf{R}_x^{1/2}\mathbf{Q}$, where $\mathbf{Q}$ is a unitary matrix that diagonalizes the matrix $(\mathbf{R}_x^{-1/2}\mathbf{R}_s\mathbf{R}_x^{-1/2})$. The minimized output error variance is given by $\mathcal{E}_{\min} = {\rm tr}[\mathbf{R}_s - \mathbf{R}_s\mathbf{R}_x^{-1}\mathbf{R}_s]$.

Proof: Consider Fig. 1. The transfer function from the input to the output $\mathbf{y}(n)$ is the constant matrix

$$\mathbf{P} = \mathbf{TKT}^{-1}.$$

From Wiener theory, we know that the output error variance $\mathcal{E}_n$ is lower bounded by that obtained by the Wiener filter. This lower bound is achieved if and only if the matrix $\mathbf{P}$ is the Wiener filter. For an input with desired signal $\mathbf{s}(n)$ and noise $\boldsymbol{\nu}(n)$, it is known that the Wiener filter is given by

$$\mathbf{P}_{\rm wiener} = \mathbf{R}_s\mathbf{R}_x^{-1}. \tag{3}$$

Observe that if $\mathbf{P}_{\rm wiener}$ is diagonalizable, then we can have $\mathbf{P} = \mathbf{P}_{\rm wiener}$ by choosing the columns of $\mathbf{T}$ be the eigenvectors of $\mathbf{P}_{\rm wiener}$ and k_i to be the eigenvalue of $\mathbf{P}_{\rm wiener}$. To show $\mathbf{P}_{\rm wiener}$ is always diagonalizable, we rewrite $\mathbf{P}_{\rm wiener}$ as

$$\begin{aligned}\mathbf{P}_{\rm wiener} &= \mathbf{R}_x^{1/2}\underbrace{\left[\mathbf{R}_x^{-1/2}\mathbf{R}_s\mathbf{R}_x^{-1/2}\right]}_{\mathbf{A}}\mathbf{R}_x^{-1/2}\\ &= \mathbf{R}_x^{1/2}\mathbf{QDQ}^\dagger\mathbf{R}_x^{-1/2}.\end{aligned}$$

As the matrix $\mathbf{A}$ in this equation is symmetric, there exists a unitary matrix $\mathbf{Q}$ such that $\mathbf{A} = \mathbf{QDQ}^\dagger$ for some diagonal matrix $\mathbf{D}$. Hence, one choice of the optimal transform is $\mathbf{T}_{\rm opt} = \mathbf{R}_x^{1/2}\mathbf{Q}$. From Wiener theory, we know that the minimized the output error variance is given by $\mathcal{E}_{\min} = {\rm tr}[\mathbf{R}_s - \mathbf{R}_s\mathbf{R}_x^{-1}\mathbf{R}_s]$. $\Delta\Delta\Delta$

The optimal transforms $\mathbf{T}_{\rm opt}$ and $\mathbf{K}$ consist of the eigenvectors and eigenvalues of the Wiener filter. From matrix theory, we know that the eigenvalues are unique and the eigenvectors are unique up to a scaling factor. Therefore the optimal k_i are *unique* and the optimal transform $\mathbf{T}_{\rm opt}$ is *unique* up to a nonsingular diagonal matrix. The subband signal is given by $\mathbf{T}_{\rm opt}^{-1}\mathbf{x}(n) = \mathbf{Q}^\dagger\mathbf{R}_x^{-1/2}\mathbf{x}(n)$. The optimal $\mathbf{T}_{\rm opt}^{-1}$ performs two tasks: the matrix $\mathbf{R}_x^{-1/2}$ whitens the input $\mathbf{x}(n)$ while the unitary matrix $\mathbf{Q}^\dagger$ decorrelates the filtered desired signal $\mathbf{R}_x^{-1/2}\mathbf{s}(n)$.

Cases When $\mathbf{R}_s$ or $\mathbf{R}_\nu$ is Nonsingular: Assume that $\mathbf{R}_s$ is nonsingular. We can rewrite the Wiener solution in (3) as

$$\mathbf{P}_{\rm wiener} = \mathbf{R}_s^{1/2}(\mathbf{I} + \mathbf{R}_s^{-1/2}\mathbf{R}_\nu\mathbf{R}_s^{-1/2})^{-1}\mathbf{R}_s^{-1/2}.$$

Let $\mathbf{Q}_0$ be a unitary matrix such that $\mathbf{R}_s^{-1/2}\mathbf{R}_\nu\mathbf{R}_s^{-1/2} = \mathbf{Q}_0\mathbf{D}_0\mathbf{Q}_0^\dagger$ for some diagonal matrix $\mathbf{D}_0$. One can verify that the optimal transform can also be expressed as

$$\mathbf{T}_{\rm opt} = \mathbf{R}_s^{1/2}\mathbf{Q}_0. \tag{4}$$

The optimal $\mathbf{K}$ is given by $\mathbf{K}_{\rm opt} = (\mathbf{I}+\mathbf{D}_0)^{-1}$. It is not difficult to verify that $\mathbf{K}_{\rm opt}$ is the Wiener filter for the subband signal $\tilde{\mathbf{s}}(n)$ plus noise $\tilde{\boldsymbol{\nu}}(n)$. On the other hand, if $\mathbf{R}_\nu$ is nonsingular. Following a similar approach, one can verify that the optimal transform can be expressed as

$$\mathbf{T}_{\rm opt} = \mathbf{R}_\nu^{1/2}\mathbf{Q}_1 \tag{5}$$

where $\mathbf{Q}_1$ is a unitary matrix such that $\mathbf{R}_\nu^{-1/2}\mathbf{R}_s\mathbf{R}_\nu^{-1/2} = \mathbf{Q}_1\mathbf{D}_1\mathbf{Q}_1^\dagger$ for some diagonal matrix $\mathbf{D}_1$. In this case, one can show that the optimal subband operation is the Wiener filter for its input and it is given by $\mathbf{K}_{\rm opt} = \mathbf{D}_1(\mathbf{D}_1 + \mathbf{I})^{-1}$.

Three Interpretations of the Optimal Transforms: From Theorem 1, (4) and (5), we obtain three different expressions for the optimal $\mathbf{T}_{\rm opt}^{-1}$: (1) $\mathbf{Q}^\dagger\mathbf{R}_x^{-1/2}$, where $\mathbf{Q}^\dagger$ diagonalizes $(\mathbf{R}_x^{-1/2}\mathbf{R}_s\mathbf{R}_x^{-1/2})$; the optimal transform is a cascade of an input (signal plus noise) whitener followed by a signal decorrelator. The matrix $\mathbf{R}_x^{-1/2}$ whitens the input $\mathbf{x}(n)$ while $\mathbf{Q}^\dagger$ decorrelates the filtered signal $\mathbf{R}_x^{-1/2}\mathbf{s}(n)$. (2) $\mathbf{T}_{\rm opt}^{-1} = Q_0^\dagger\mathbf{R}_s^{-1/2}$, where $\mathbf{Q}_0^\dagger$ diagonalizes $\mathbf{R}_s^{-1/2}\mathbf{R}_\nu\mathbf{R}_s^{-1/2}$; the optimal transform is a cascade of a signal whitener followed by a noise decorrelator. The matrix $\mathbf{R}_s^{-1/2}$ whitens the signal $\mathbf{s}(n)$ while $\mathbf{Q}_0^\dagger$ decorrelates the filtered noise $\mathbf{R}_s^{-1/2}\boldsymbol{\nu}(n)$. (3) $\mathbf{T}_{\rm opt}^{-1} = Q_1^\dagger\mathbf{R}_\nu^{-1/2}$, where $\mathbf{Q}_1^\dagger$ diagonalizes $\mathbf{R}_\nu^{-1/2}\mathbf{R}_s\mathbf{R}_\nu^{-1/2}$; the optimal transform is a cascade of a noise whitener followed by a signal decorrelator. The matrix $\mathbf{R}_\nu^{-1/2}$ whitens the noise $\boldsymbol{\nu}(n)$ while $\mathbf{Q}_1^\dagger$ decorrelates the filtered signal $\mathbf{R}_\nu^{-1/2}\mathbf{s}(n)$.

Remarks and Discussions:

1) In the case that $\mathbf{R}_s$ and $\mathbf{R}_\nu$ have a common KLT, both $\mathbf{R}_s$ and $\mathbf{R}_\nu$ can be simultaneously diagonalized by the same unitary matrix $\mathbf{Q}$. The optimal transform can simply be chosen as $\mathbf{Q}^\dagger$. Thus, our solution reduces to that given in Theorem 7 of [1].
2) When the noise (or the desired signal) is white, then $\mathbf{R}_\nu = \mathbf{I}$ (or $\mathbf{R}_s = \mathbf{I}$). This becomes a special case of common KLT.
3) The above results can be generalized to the case of unconstrained filter length, where the transform and the subband Wiener filters are allowed to be ideal filters. The entire proof and derivations carry through by simply replacing the correlation matrices by the power spectral matrices.
4) For the case of FIR matrix $\mathbf{T}(z)$, the optimal solution is still an open problem.

III. Simulation

In this section, we compare the performance of the optimal transform and the KLT for colored noise suppression. The dimension of the transform is $M = 8$. The vectors $\mathbf{s}(n)$ and $\boldsymbol{\nu}(n)$ are respectively the blocked versions of scalar uncorrelated WSS processes $s(n)$ and $\nu(n)$. The signal $s(n)$ is an AR(1) process with correlation coefficient $\rho_s^{|k|}$. The noise $\nu(n)$ is an AR(1) process with correlation coefficient of $\rho_\nu^{|k|}$. We compare the output error variances of the following two cases: i) $\mathcal{E}_{\min}$ in Theorem 1 and ii) $\mathcal{E}_{klt}$: the output error variance when $\mathbf{T}$ is the KLT for $s(n)$ and k_i is taken as the zeroth order Wiener filter for its input (this is the optimal transform if the noise were white). Fig. 2 shows the results for $0.7 \le \rho_s \le 0.99$ and and $\rho_\nu = -0.7$. As we can see, $\mathcal{E}_{\min}$ is smaller than $\mathcal{E}_{klt}$ and the gain decreases when ρ_s increases. This is because when ρ_s is

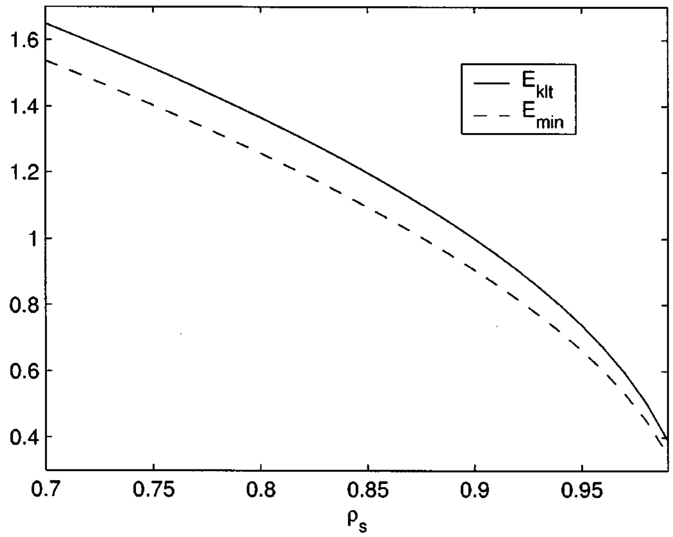

Fig. 2. Output error variances when $0.7 \leq \rho_s \leq 0.99$ and $\rho_\nu = -0.7$.

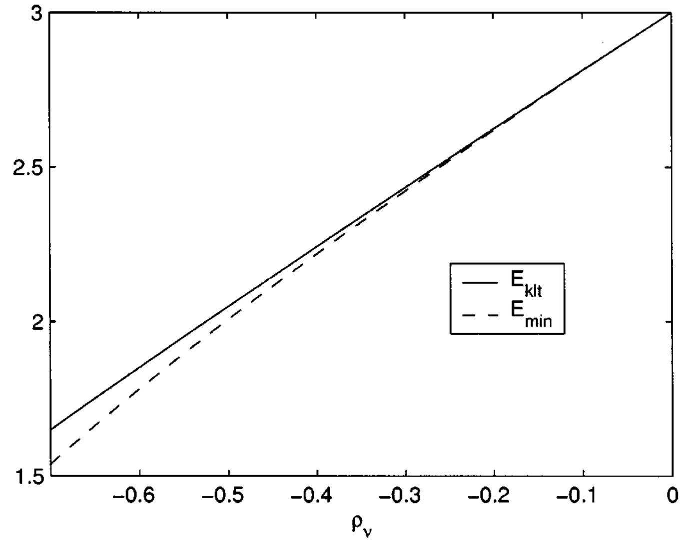

Fig. 3. Output error variances when $\rho_s = 0.7$ and $-0.7 \leq \rho_\nu \leq 0$.

nearly 1, a large portion of the noise reduction can be obtained by exploiting the correlation of the signal alone. Fig. 3 shows the results for $-0.7 \leq \rho_\nu \leq 0$ and $\rho_s = 0.7$. From the figure, we see that as ρ_ν decreases to 0, the two curves converge. When $\rho_\nu = 0$, the noise is white and in this case the optimal transform reduces to the KLT of $s(n)$.

Acknowledgment

The authors would like to thank Dr. S. Akkarakaran and Prof. P. P. Vaidyanathan of California Institute of Technology for their discussion and suggestions that give rise to this simpler derivation of the results and the generalization to the unconstrained case. We would also like to thank the reviewers for their comments that have greatly improved the presentation.

References

[1] S. Akkarakaran and P. P. Vaidyanathan, “Filterbank optimization with convex objectives and the optimality of principle component forms,” *IEEE Trans. Signal Processing*, vol. 49, pp. 100–114, Jan. 2001.

[2] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, pp. 425–455, 1994.