AN OPTIMAL GENERALIZED THEORY OF SIGNAL REPRESENTATION

J. Scott Goldstein — MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420 USA

Joseph R. Guerci — SAIC, 4001 North Fairfax Drive, Arlington, VA 22203

Irving S. Reed — Dept. of Electrical Engineering, Univ. of Southern California, Los Angeles, CA 90089 USA

ABSTRACT

A new generalized statistical signal processing framework is introduced for optimal signal representation and compression. Previous work is extended by considering the multiple signal case, where a desired signal is observed only in the presence of other non-white signals. The solution to this multi-signal representation problem yields a generalization of the Karhunen-Lóeve transform and generates a basis selection which is optimal for multiple signals and colored-noise random processes under the minimum mean-square error criterion. The important applications for which this model is valid include detection, prediction, estimation, compression, classification and recognition.

1. INTRODUCTION

This paper is concerned with the representation of discrete time, wide-sense stationary (WSS) signals in the many applications of statistical detection and estimation theory. For the purposes of this work, the efficiency of a signal representation is evaluated based upon its compaction of the useful signal energy as a function of rank. This criterion is equivalent to optimal signal compression.

The multiple-signal problem is considered herein, where a signal of interest is only observed in the presence of at least one other non-white process. Signal processing for multiple signals, under the conditions described above, is described within the general framework of the discrete-time, finite impulse response (FIR) Wiener filter. The Wiener filter is a fundamental component in the solution to virtually every problem which is concerned with optimality for linear filtering, detection, estimation, classification, smoothing and prediction in the framework of statistical signal processing with stationary random processes.

This work was sponsored by the U.S. Air Force under contract F19628-95-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and not necessarily endorsed by the United States Air Force.

The differences between the single and multiple signal representation problems are emphasized in this paper. The fundamental issue in signal representation and compression is the determination of an optimal coordinate system. It is well-known that the eigenvectors associated with the covariance matrix of an N-dimensional WSS signal provide the basis-set for the Karhunen-Lóeve expansion of that signal. The minimax theorem establishes that this set of eigenvectors represent the particular basis for an N-dimensional space that is most efficient in the energy sense for this single process. Optimal representation and compression as a function of rank (or dimension) is then obtained by retention of the principal components [3].

However, there are very few problems where the sole criterion of interest is single-signal representation. This fact is readily verified by considering the popular problems of detection, estimation, prediction, classification, noise-cancellation, spectral estimation. Here, for the problem to be non-trivial, there are a minimum of two non-white signal processes: the signal of interest and a colored-noise process. If one now speaks of signal representation or compression of one process, the solution must take both processes into account in order to determine an optimal basis. In fact, it has been previously established [1]-[2] that the principal-components are no longer the correct enumeration of the eigenvectors for optimal representation. Consequently, the standard Karhunen-Lóeve decomposition no longer provides the solution to optimal basis selection, and a new basis set must be derived which takes into account the presence of the other signals; a generalized Karhunen-Lóeve transformation.

2. MULTIPLE SIGNAL MODELING AND WIENER FILTERING

In a more general setting, the classical problems of statistical signal processing are concerned with joint signal representation and compression. These problems are characterized by a Wiener filter, depicted in

Fig. 1, where there are two processes present. The N-dimensional process $\mathbf{x}_0$ is now considered to be the sum of potentially many processes, while d_0 is a scalar process which is correlated with $\mathbf{x}_0$.

The process d_0, normally termed a desired process, is representative of a signal of interest in some way, and the goal is to estimate d_0 from $\mathbf{x}_0$. For example, in the radar and sonar detection problem the "desired" signal is usually the output of a beamformer or matched-field processor, and the observed data vector $\mathbf{x}_0$ is everything orthogonal to this signal in the data received at a sensor array. In the communications application of multiuser detection and demodulation, the process d_0 may be generated by a known correlation with the signal of interest such as the code of a user in a CDMA wireless network. As a final example, in classification for automatic target recognition the desired signal may be a template image from training data, while the observed data is an image received by the fielded sensor. In general, the mechanism which generates the reference signal is application specific; however nearly every problem in statistical signal processing may be represented using this model.

The process d_0 is a scalar process with variance $\sigma^2_{d_0}$ and $\mathbf{x}_0$ is an observed N-dimensional signal, which itself may be a composite random process, with covariance $\mathbf{R}_{x_0}$. The filter to be defined, $\mathbf{w}$, processes the observed-data to form an estimate of the desired signal $\hat{d}_0 = \mathbf{w}^H\mathbf{x}_0$. The error process ε_0,

$$\varepsilon_0 = d_0 - \mathbf{w}^H\mathbf{x}_0, \tag{1}$$

is the signal which characterizes the performance of the filter, and the optimal Wiener filter minimizes the mean-square value of this error signal. The mean-square value of the error is calculated to be,

$$\mathbf{E}[\ |\varepsilon_0|^2\ ] = \sigma^2_{d_0} - \mathbf{w}^H\mathbf{r}_{x_0d_0} - \mathbf{r}^H_{x_0d_0}\mathbf{w} + \mathbf{w}^H\mathbf{R}_{x_0}\mathbf{w}, \tag{2}$$

where the N-vector $\mathbf{r}_{x_0d_0}$ is the cross-correlation between the processes d_0 and $\mathbf{x}_0$.

The minimum mean-square error (MMSE) optimization criterion is formally stated as follows:

$$\min_{\mathbf{w}}\ \left\{\sigma^2_{d_0} - \mathbf{w}^H\mathbf{r}_{x_0d_0} - \mathbf{r}^H_{x_0d_0}\mathbf{w} + \mathbf{w}^H\mathbf{R}_{x_0}\mathbf{w}\right\}. \tag{3}$$

The well-known solution to (3) is the Wiener filter, which is computed as follows:

$$\mathbf{w} = \mathbf{R}^{-1}_{x_0}\mathbf{r}_{x_0d_0}. \tag{4}$$

3. OPTIMAL SIGNAL REPRESENTATION FOR STATISTICAL SIGNAL PROCESSING

In this section, an optimal rank-ordered and generalized signal represenation is developed based on a multistage Wiener filter decomposition [1]-[2]. The natural optimization criteria at each stage is the minimization of the mean-squared error (MSE). Consider the multistage Wiener filter depicted in Fig. 2. Signal compression is achieved by successive rank-one selections which are chosen to minimize the MSE at each stage, or, equivalently, "whiten" the error residue (innovation). For example, the first rank-one subspace ($\mathbf{h}_1 \in \mathcal{C}^N$) is obtained by choosing the vector in the space spanned by the columns of the covariance matrix $\mathbf{R}_{x_0} \in \mathcal{C}^{N\times N}$ which is maximally correlated with the desired process d_0. The optimal choice is thus $\mathbf{h}_1 = \mathbf{r}_{x_0d_0}/\|\mathbf{r}_{x_0d_0}\|$. This selection results in an error residue $\varepsilon_1 = d_0 - w_1^H d_1$, where $d_1 \triangleq \mathbf{h}_1^H\mathbf{x}_0$, and $w_1 \in \mathcal{C}$ is the optimal Wiener weight for the selected signal d_1. Note that, in general, $\mathbf{h}_1$ will not correspond to the eigenvector with largest associated eigenvalue, as would be the case in a KLT representation. Consequently, a principal components based reduced rank Wiener filter will, in general, produce suboptimal compression [1].

Since the subspace which can contribute to reducing the MSE is likely greater than rank one, additional stages will be required to achieve a desired level of MSE. Thus at the i-th stage, a new rank-one subspace ($\mathbf{h}_i \in \mathcal{C}^{N-i+1}$) is selected which is maximally correlated with the entire optimal Wiener residue ε_i from the previous stages. Thus, referring to Fig. 2, $\mathbf{h}_i = \mathbf{r}_{x_i\varepsilon_i}/\|\mathbf{r}_{x_i\varepsilon_i}\|$, where $\mathbf{r}_{x_i\varepsilon_i} \in \mathcal{C}^{N-i+1}$ is the cross-correlation between $\mathbf{x}_i$ and ε_i. However, a simplification results which eliminates the need to explicitly calculate the optimal Wiener weight vector in order to form $\mathbf{h}_i$, specifically, $\mathbf{r}_{x_i\varepsilon_i} = \sum_{j=0}^{i}\alpha_j\mathbf{r}_{x_id_j} = \alpha_i r_{x_id_i}$, from which the following simplification results, $\mathbf{h}_i = \mathbf{r}_{x_id_i}/\|\mathbf{r}_{x_id_i}\|$.

It is important to note that the above decomposition leads to a tridiagonalization of the joint process covariance matrix [2]. This is in contrast to the standard KLT or principal components based decompositions which impose a unitary constraint, resulting in diagonalization (but with a generally greater MSE). We thus refer to the above multistage Wiener-based signal represenation as a generalized KLT (GKLT).

4. A COMPARATIVE EXAMPLE

Interestingly, major insight into the relative behavior and performance of the various approaches to reduced-rank signal modeling can be gained with relatively simple examples. For instance,consider an example in $\Re^2$, where there is a 2-dimensional observed-data vector $\mathbf{x}_0$

with covariance matrix $\mathbf{R}_{x_0}$,

$$\mathbf{R}_{x_0} = \begin{bmatrix} 10 & 4 \\ 4 & 10 \end{bmatrix},$$

and a desired signal d_0 with a variance $\sigma_{d_0}^2 = 10$. The two processes, d_0 and $\mathbf{x}_0$, are assumed to be zero-mean, jointly stationary, and correlated. The cross-correlation between the two processes is given by the vector $\mathbf{r}_{x_0 d_0}$,

$$\mathbf{r}_{x_0 d_0} = \begin{bmatrix} 9 \\ 1 \end{bmatrix}. \tag{5}$$

The KLT provides the eigendecomposition of the matrix $\mathbf{R}_{x_0}$:

$$\begin{aligned} \mathbf{R}_{x_0} &= \begin{bmatrix} 10 & 4 \\ 4 & 10 \end{bmatrix} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 6 & 0 \\ 0 & 14 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \end{aligned} \tag{6}$$

which demonstrates that one eigenvalue is significantly greater than the other. The KLT takes into account the self-directional preference of the signal and provides the most efficient representation of the autocorrelation energy in the signal.

The question at hand, however, is the determination of the best basis representation for $\mathbf{x}_0$ in terms of estimating d_0. The KLT basis is still thought of as being optimal for detection and estimation by many researchers. To explore whether the KLT and the principal components are still the best basis choice, consider preprocessing the observed process $\mathbf{x}_0$ with a filter composed of the eigenvectors of $\mathbf{R}_{x_0}$. The new process $\mathbf{z}_0 = \mathbf{V}^H\mathbf{x}_0$, has a diagonal covariance matrix given by $\mathbf{\Lambda}$. Also, the cross-correlation between $\mathbf{z}_0$ and d_0 is now given by, $\mathbf{r}_{z_0 d_0} = \mathbf{V}^H\mathbf{r}_{x_0 d_0}$.

The MMSE performance of the eigenvector-basis as a function of rank is now evaluated. The full-rank solution is identical regardless of basis representation since the performance measure is invariant to invertible transformations of the observed signal $\mathbf{x}_0$.

However the results are different in the rank-1 case, where one of the two eigenvectors which compose $\mathbf{V}$ is discarded. The principal-components algorithm states that the eigenvector corresponding with the largest eigenvalue should be retained. Here, the largest eigenvalue magnitude corresponds with the second eigenvector, and the MMSE for this case is given by,

$$\xi_{\text{pc}} = \sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{v}_2 \lambda_2^{-1} \mathbf{v}_2^H \mathbf{r}_{x_0 d_0}. \tag{7}$$

The MMSE performance, converted to decibels, for the full-rank Wiener filter is 0.3951 dB and that for the rank-1 principal-components Wiener filter is 8.0811 dB. Thus, there is a loss of 7.6860 dB in reducing the rank from 2 to 1. Note that the MMSE which results if the smaller eigenvector is retained is 6.6901 dB. Here the MMSE loss is approximately 1.4 dB less than that experienced by the principal-components selection; that is, a performance enhancement is obtained by selecting a different eigenvector than that indicated by the principal-components. This observation led to the development of the cross-spectral metric (CSM) based KLT PC rank reduction method [1]. In the CSM-KLT approach, eigenvectors are ordered based on *both* energy and the amount of cross-correlation with the desired process. The CSM-KLT would have selected the smaller eigenvector in this example.

If, however, the rank one subspace is selected based on the GKLT approach (Section 3), the resulting MSE is 3.9127 dB, an improvement of nearly 3 dB over the CSM-KLT approach, and slightly over 4 dB compared to principal components. Fig. 3 shows the relative performace of the three approaches as the cross correlation vector $\mathbf{r}_{x_0 d_0}$ is rotated through 360°. Note that the GKLT uniformly outperforms the principal components (PC) method except in those rare instances when $\mathbf{r}_{x_0 d_0}$ and $\mathbf{v}_1$ are colinear. Also note that there is additional coincidence with the CSM when $\mathbf{r}_{x_0 d_0}$ and $\mathbf{v}_2$ are colinear.

5. CONCLUSION

A new generalized statistical signal processing framework was introduced for optimal signal representation and compression for the multiple signal case. The solution to this multi-signal representation problem was shown to yield a generalization of the KLT (GKLT) which generates a basis selection which is optimal for multiple signals and colored-noise under the MMSE criterion.

6. REFERENCES

[1] J. S. Goldstein and I. S. Reed. "Reduced rank adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 492-496, February 1997.

[2] J. S. Goldstein, I. S. Reed and L. L. Scharf. "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Transactions on Information Theory*, (to appear) November 1998.

[3] D. W. Tufts, R. Kumaresan, and I. Kirsteins. "Data adaptive signal estimation by singular value decomposition of a data matrix," *Proceedings of the IEEE*, 70(6):684-685, June 1982.

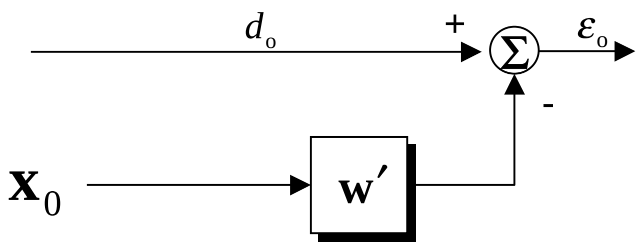

Fig. 1 General Wiener filter structure.

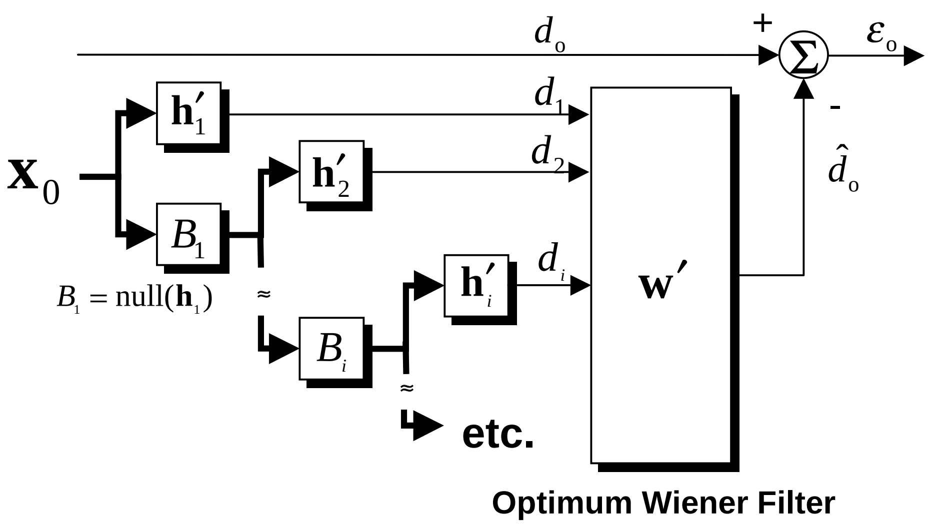

Fig. 2 Multistage Wiener filter structure.

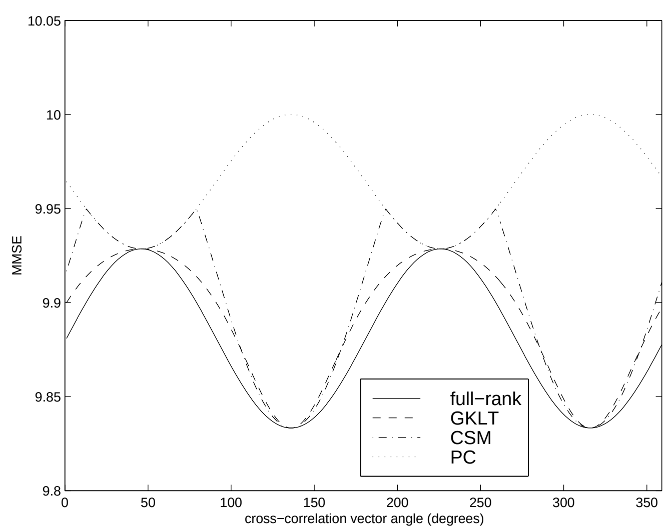

Fig. 3 Relative MMSE performance of the various rank-reduction methods.